



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

# **Regressão Quantílica com Distribuições Assimétricas**

por

**Alan da Silva**

Brasília, 11 de janeiro de 2021

# **Regressão Quantílica com Distribuições Assimétricas**

**por**

**Alan da Silva**

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

Orientador: Prof. Dr. Helton Saulo Bezerra dos Santos

Brasília, 11 de janeiro de 2021

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como requisito para a obtenção do grau de Mestre em Estatística.

Texto aprovado por:

Prof. Dr. Helton Saulo Bezerra dos Santos  
Orientador, EST/UnB

Prof. Dr. José Augusto Fiorucci  
Membro interno, EST/UnB

Prof. Dr. Jeremias da Silva Leão  
Membro externo, DE/UFAM

Prof. Dr. Eduardo Yoshio Nakano  
Suplente, EST/UnB

*Sofrer, é só uma vez; vencer, é para toda a eternidade.*

(Søren Kierkegaard)

Para meus amores Dayane, Nilda e Neves.

# Agradecimentos

A Deus por ter me dado forças para superar todas as adversidades que apareceram ao longo do mestrado.

A toda a minha família e principalmente a minha mãe por todo o apoio incondicional nos momentos difíceis ao longo desses dois anos de mestrado. Um agradecimento especial aos meus familiares de Brasília que me deram todo o suporte desde a minha chegada até os dias atuais na capital federal.

A minha noiva Dayane, por todo o seu amor, carinho e compreensão por todos os momentos que tive que me ausentar para me dedicar aos estudos.

A todos professores do PPGEST/UnB por todos os conhecimentos transmitidos nas disciplinas cursadas ao longo do mestrado. Um agradecimento especial, ao meu orientador, o professor Helton Saulo, por todo apoio, confiança e dedicação na orientação para realização desta dissertação.

A todos os meus colegas de mestrado, em especial, aos meus amigos e companheiros de estudo, Adolfo, Débora e Thays.

A toda equipe da ASMEQ/IPEA pelo ambiente favorável para o desenvolvimento desse trabalho.

Aos professores José Augusto Fiorucci (Est/UnB) e Jeremias da Silva Leão (DE/UFAM) por terem aceito o convite de participar da banca e também pelas valiosas contribuições e sugestões que acrescentaram ainda mais a este trabalho.

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

# Resumo

Modelos de regressão baseados na família de distribuições log-simétricas e na família de distribuições mistura-escala Birnbaum-Saunders são particularmente úteis quando a variável resposta é estritamente positiva e assimétrica. Neste trabalho, propomos duas classes de modelos de regressão quantílica que são apresentados em dois artigos, o primeiro deles baseado nas distribuições log-simétricas e o segundo baseado nas distribuições mistura-escala Birnbaum-Saunders. Em ambos os casos, é feita uma reparametrização das distribuições inserindo um parâmetro de quantil. Em cada artigo, dois estudos de simulação foram realizados utilizando o software R, sendo o primeiro deles para análise do desempenho dos estimadores de máxima verossimilhança, dos critérios de informação AIC, BIC e AICc, e dos resíduos Cox-Snell e quantil aleatorizado. O segundo avaliou o tamanho/poder dos testes Wald, razão de verossimilhança, score e gradiente. Os dois estudos de simulação foram conduzidos considerando diferentes quantis de interesse e tamanhos amostrais. Por fim, uma aplicação utilizando dados reais é apresentada para cada modelo proposto.

**Palavras-chave:** Distribuições log-simétricas; Distribuição mistura-escala Birnbaum-Saunders; Regressão quantílica; Simulação de Monte Carlo; Teste de hipóteses.

# Abstract

Regression models based on the family of log-symmetric distributions and the family of scale-mixture Birnbaum-Saunders distributions are particularly useful when the response variable is strictly positive and asymmetric. In this work, we propose two classes of quantile regression models that are presented in two articles, the first based on the log-symmetric distributions and the second based on the scale-mixture Birnbaum-Saunders distributions. In both cases, a reparameterization of the distribution is introduced by inserting a quantile parameter. For each article, two simulation studies are carried out using the R software. The first one analyzes the performance of the maximum likelihood estimators, the information criteria AIC, BIC and AICc, and the generalized Cox-Snell and random quantile residuals. The second one evaluates the performance of the size and power of the Wald, likelihood ratio, score and gradient tests. The two simulation studies are conducted considering different quantiles of interest and sample sizes. Finally, an application using real data is presented for each proposed model.

**Keywords:** Log-symmetric distributions; Scale-mixture Birnbaum-Saunders distributions; Quantile regression; Monte Carlo simulation; Hypothesis testing.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Modelos de regressão log-simétricos quantílicos</b>	<b>4</b>
2.1	Introdução . . . . .	5
2.2	Preliminares . . . . .	7
2.2.1	A distribuição log-simétrica clássica . . . . .	7
2.2.2	Uma distribuição log-simétrica baseada no quantil . . . . .	8
2.3	Modelo de regressão log-simétrico quantílico . . . . .	10
2.3.1	Testes de hipóteses . . . . .	12
2.3.2	Critérios de informação . . . . .	13
2.3.3	Resíduos . . . . .	14
2.4	Estudos de simulação de Monte Carlo . . . . .	15
2.4.1	Estudo de Monte Carlo 1: desempenho dos estimadores, critérios de informação e resíduos . . . . .	15
2.4.2	Estudo de Monte Carlo 2: tamanho e poder dos testes . . . . .	24
2.5	Exemplo ilustrativo: Bilheteria de filmes . . . . .	35
2.6	Considerações finais . . . . .	40
<b>3</b>	<b>Modelos de regressão quantílica mistura-escala Birnbaum-Saunders</b>	<b>42</b>
3.1	Introdução . . . . .	43
3.2	Preliminares . . . . .	45
3.2.1	Distribuições SBS . . . . .	45
3.2.2	Distribuições QSBS . . . . .	46
3.2.3	Casos especiais das distribuições QSBS . . . . .	48

3.2.4	Análise da forma das distribuições QSBS . . . . .	51
3.3	Um modelo de regressão quantílica QSBS . . . . .	52
3.3.1	Aproximação para os erros padrão . . . . .	54
3.3.2	Critérios de seleção de modelos . . . . .	55
3.3.3	Resíduos . . . . .	56
3.3.4	Testes de hipóteses . . . . .	57
3.4	Estudos de simulação . . . . .	58
3.4.1	Estudo de simulação 1 . . . . .	59
3.4.2	Estudo de simulação 2 . . . . .	68
3.5	Exemplo ilustrativo: Seguro de acidentes pessoais . . . . .	76
3.6	Conclusões . . . . .	80
<b>4</b>	<b>Considerações Finais</b>	<b>82</b>
	<b>Bibliografia</b>	<b>83</b>

# Lista de Tabelas

2.1	Gerador de densidade $g(u)$ para algumas distribuições log-simétricas. . . . .	8
2.2	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-NO. . . . .	18
2.3	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log- $t$ ( $\vartheta = 3$ ). . . . .	18
2.4	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-PE ( $\vartheta = 0.3$ ). . . . .	18
2.5	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-HP ( $\vartheta = 2$ ). . . . .	19
2.6	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-SL ( $\vartheta = 4$ ). . . . .	19
2.7	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-CN ( $\vartheta_1 = 0.1, \vartheta_1 = 0.2$ ). . . . .	19
2.8	Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica EBS ( $\vartheta = 0.5$ ). . . . .	20
2.9	Viés, EQM e PC dos parâmetros do modelo de regressão com erros Log-ST(0.5,3). . . . .	20
2.10	Taxas de acertos segundo os critérios de informação para os modelos de regressão log-simétricos quantílicos. . . . .	21
2.11	Resultados para os resíduos GCS considerando diferentes modelos de regressão log-simétricos quantílicos. . . . .	23
2.12	Resultados para os resíduos RQ considerando diferentes modelos de regressão log-simétricos quantílicos. . . . .	24

2.13	Tamanho dos testes para o modelo de regressão log-NO quantílico, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	27
2.14	Tamanho dos testes para o modelo de regressão log- $t$ ( $\vartheta = 3$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	28
2.15	Tamanho dos testes para o modelo de regressão log-PE ( $\vartheta = 0.3$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	29
2.16	Tamanho dos testes para o modelo de regressão log-HP ( $\vartheta = 2$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	30
2.17	Tamanho dos testes para o modelo de regressão log-SL ( $\vartheta = 4$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	31
2.18	Tamanho dos testes para o modelo de regressão log-CN ( $\vartheta_1 = 0.1, \vartheta_2 = 0.2$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	32
2.19	Tamanho dos testes para o modelo de regressão EBS ( $\vartheta = 0.5$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	33
2.20	Tamanho dos testes para o modelo de regressão EBS- $t$ ( $\vartheta_1 = 0.5, \vartheta_2 = 3$ ) quantílica, sob $H_0 : \beta_1 = \dots = \beta_t = 0$ . . . . .	34
2.21	Sumário de estatísticas descritivas para a bilheteria, em milhões de US\$. . . . .	35
2.22	Média dos critérios AIC, BIC, AICc e RMSE dos modelos de regressão quantílica log-simétrica considerando os quantis $q = 0.01, 0.02, \dots, 0.98, 0.99$ . . . . .	37
2.23	Estimativas e erros-padrões (entre parênteses) para o modelo de regressão quantílica com erros log-HP(1). . . . .	38
2.24	Estatísticas observadas $S_W, S_{LR}, S_R$ e $S_T$ e o seu respectivo p-valor para o modelo de regressão quantílica log-simétrica. . . . .	39
3.1	Viés, EQM e CP para o modelo de regressão quantílica CN-BS, com $\nu = 0.1, \delta = 0.3$ . . . . .	62
3.2	Viés, EQM e CP para o modelo de regressão quantílica SL-BS, com $\nu = 4$ . . . . .	63
3.3	Viés, EQM e CP para o modelo de regressão quantílica $t_\nu$ -BS, com $\nu = 11$ . . . . .	64
3.4	Sumário de estatísticas para os resíduos GCS considerando o modelo de regressão quantílica SBS. . . . .	65

3.5	Sumário de estatísticas para os resíduos RQ considerando o modelo de regressão quantílica SBS. . . . .	66
3.6	Taxa de acerto considerando os critérios de informação AIC, BIC, AICc, HIC e o REQM para o modelo de regressão quantílica. . . . .	67
3.7	Taxa de rejeição sob a hipótese nula para testar $H_0 : \beta_3 = 0$ , para o modelo de regressão CN-BS, com $\nu = 0.1, \delta = 0.3$ . . . . .	70
3.8	Taxa de rejeição sob a hipótese nula para testar $H_0 : \beta_3 = 0$ , para o modelo de regressão SL-BS, com $\nu = 4$ . . . . .	71
3.9	Taxa de rejeição sob a hipótese nula para testar $H_0 : \beta_3 = 0$ , para o modelo de regressão $t_\nu$ -BS, com $\nu = 11$ . . . . .	72
3.10	Taxa de rejeição sob a hipótese nula para testar $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , para o modelo de regressão CN-BS, com $\nu = 0.1, \delta = 0.3$ . . . . .	73
3.11	Taxa de rejeição sob a hipótese nula para testar $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , para o modelo de regressão SL-BS, com $\nu = 4$ . . . . .	74
3.12	Taxa de rejeição sob a hipótese nula para testar $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , para o modelo de regressão $t_\nu$ -BS, com $\nu = 11$ . . . . .	75
3.13	Sumário de estatísticas descritivas para a variável montante. . . . .	77
3.14	Média dos critérios AIC, BIC, AICc, HIC e REQM dos modelos de regressão quantílica SBS considerando os quantis $q = 0.01, 0.02, \dots, 0.98, 0.99$ . . . . .	78
3.15	Estimativas e erro padrão (em parênteses) dos parâmetros do modelo de regressão quantílica CN-BS considerando os quantis $q = 0.25, 0.5, 0.75$ . . . . .	78
3.16	Estatísticas observadas e $p$ -valor (em parênteses) para testar $H_0$ considerando os parâmetros do modelo de regressão quantílica CN-BS para os quantis $q = 0.25, 0.5, 0.75$ . . . . .	80

# Lista de Figuras

2.1	Curva de poder dos testes para o modelo de regressão log-NO quantílico. . . .	27
2.2	Curva de poder dos testes para o modelo de regressão log- $t$ ( $\vartheta = 3$ ) quantílico.	28
2.3	Curva de poder dos testes para o modelo de regressão log-PE ( $\vartheta = 0.3$ ) quantílico.	29
2.4	Curva de poder dos testes para o modelo de regressão log-HP ( $\vartheta = 2$ ) quantílico.	30
2.5	Curva de poder dos testes para o modelo de regressão log-SL ( $\vartheta = 4$ ) quantílico.	31
2.6	Curva de poder dos testes para o modelo de regressão log-CN ( $\vartheta_1 = 0.1, \vartheta_2 = 0.2$ ) quantílico. . . . .	32
2.7	Curva de poder dos testes para o modelo de regressão EBS ( $\vartheta = 0.5$ ) quantílico.	33
2.8	Curva de poder dos testes para o modelo de regressão EBS- $t$ ( $\vartheta_1 = 0.5, \vartheta = 3$ ) quantílico. . . . .	34
2.9	Gráfico das estimativas dos parâmetros do modelo log-simétrico quantílico, com erros log-HP(1), para os quantis $q = 0.01, 0.02, \dots, 0.98, 0.99$ . . . . .	37
2.10	Gráfico de envelope dos resíduos GCS e RQ para o modelo ajustado considerando os quantis $q = 0.25, 0.5, 0.75$ . . . . .	40
3.1	FDP de alguns membros da família de distribuições QSBS, considerando diferentes valores de $\alpha$ e $Q = 1$ fixado, para os valores de $q = 0.25, 0.5, 0.75$ , (nas linhas) e para os modelos $t_\nu$ -BS, SL-BS e CN-BS (nas colunas), com $\nu = 2, 2, (0.2, 0.2)$ , respectivamente. . . . .	51
3.2	Poder do teste das estatísticas $S_W, S_{LR}, S_R$ e $S_T$ para os modelos CN-BS, SL-BS e $t_\nu$ -BS considerando o nível de significância de 1%, $n = 100$ e $q = 0.5$ . . . . .	76
3.3	Gráfico de frequência por estimativa de kernel (a), boxplot padrão (b) e boxplot ajustado (c) para a variável montante, em dólares australianos. . . . .	77

3.4	Estimativas de máxima verossimilhança dos parâmetros do modelo de regressão quantílica CN-BS para os quantis $q = 0.01, 0.02, \dots, 0.99$ . . . . .	79
3.5	Gráficos QQ dos resíduos GCS ((a), (b) e (c)) e RQ ((d), (e) e (f)) do modelo de regressão quantílica CN-BS para os quantis $q = 0.25, 0.5, \dots, 0.75$ . . . . .	80

# Capítulo 1

## Introdução

Dados estritamente positivos com assimetria à direita vem se tornando cada vez mais comum em diversas áreas do conhecimento. Alguns exemplos podem ser encontrados em trabalhos práticos, podendo ser encontrados em diferentes contextos, como na modelagem de tempos de vida (Lawless, 2011; Lee e Wang, 2003), concentração de poluentes no ar (Leiva et al., 2008a; Leiva et al., 2010; Marchant et al., 2013), entre outros. Devido a crescente demanda por modelos que forneçam um bom ajuste para esse tipo de dado, diversos propostas de modelos probabilísticos vem aparecendo na literatura, alguns desses e outros modelos podem ser vistos em ver Johnson, Kotz e Balakrishnan (1994), Johnson, Kotz e Balakrishnan (1995), por exemplo.

Buscando, principalmente, uma maior flexibilidade de ajuste de dados com diferentes características, diversas famílias de distribuições de probabilidade vêm sendo propostas. Geralmente, essas famílias gozam de diversas propriedades em comum a todos os membros da família. Além disso, em muitos casos, apresentam mais parâmetros que a distribuições usuais, uma vez que, para certos valores específicos, a família de distribuições assume uma distribuição em particular. Algumas das classes de distribuições que apresentam essas características são as famílias de distribuições simétricas (Fang e Anderson, 1990), log-simétricas (Vanegas e Paula, 2016b), mistura-escala de normais (Andrews e Mallows, 1974), e mistura-escala Birnbaum-Saunders (Balakrishnan et al., 2009). Em particular, podemos destacar duas das classes de distribuições citadas que são utilizadas na modelagem de dados positivos assimétricos à direita, a saber, as distribuições log-simétricas as quais são obtidas por meio de uma transformação exponencial em uma variável aleatória que pertence a família de distribuições simétricas, e as distribui-

ções mistura-escala Birnbaum-Saunders as quais são obtidas por meio de uma relação entre a distribuição Birnbaum-Saunders e a classe de distribuições mistura-escala de normais.

Os modelos de regressão quantílica constituem uma alternativa frente aos modelos de regressão para a média condicional. Os modelos de regressão quantílica foram estudados inicialmente por Koenker e Bassett Jr (1978). Além da vantagem de estimar uma medida de tendência central (mediana) mais robusta em relação aos valores atípicos, os *outliers*, o modelo de regressão quantílica fornece a possibilidade de obter um ajuste mais global, explorando o efeito das covariáveis nos quantis mais extremos, apresentando assim uma ideia geral a respeito da distribuição condicional da variável resposta. Desde então, os modelos de regressão quantílica vem se tornando uma técnica cada vez mais atrativa, ganhando cada vez mais importância contando com inúmeros trabalhos publicados; ver Koenker e Hallock (2001), Yu e Moyeed (2001), Yu, Lu e Stander (2003), Koenker (2004), entre outros.

É possível encontrar alguns modelos de regressão quantílica para dados estritamente positivos na literatura. Mais recentemente, podemos destacar o modelo de regressão quantílica baseado na distribuição Birnbaum-Saunders clássica (Birnbaum e Saunders, 1969a), o qual foi proposto por Sánchez et al. (2020a), com uma estrutura similar a de modelos lineares generalizados, além de contar com um exemplo prático para dados de economia; outros trabalhos usando esse modelo para dados espaciais foram estudados em Sánchez et al. (2020b) e Leiva et al. (2020). Apesar desse modelo apresentar uma boa alternativa para casos em que a resposta é assimétrica, em situações em que a distribuição da variável resposta apresente características como bimodalidade ou caudas pesadas, o ajuste de modelos de regressão quantílica mais usuais não apresentam um ajuste satisfatório, necessitando de modelos mais robustos para tais situações.

Neste trabalho, propomos dois modelos de regressão quantílica para dados positivos com assimetria à direita. Mais precisamente, o primeiro modelo de regressão quantílica proposto baseia-se na família de distribuições log-simétricas, enquanto o segundo baseia-se na família de distribuições mistura-escala Birnbaum-Saunders (SBS). Os dois modelos são apresentados em dois artigos autocontidos. Estudos de simulação são realizados para analisar o desempenho das estimativas de máxima verossimilhança, critérios de informação, resíduos e testes de hipóteses.

O restante desse trabalho está organizado da seguinte forma. No Capítulo 2 é apresentado o modelo de regressão log-simétrico quantílico, o qual é baseado distribuição log-simétrica

reparametrizada pelo quantil. No Capítulo 3 é apresentado o modelo de regressão quantílica mistura-escala Birnbaum-Saunders, a qual é baseado em uma reparametrização pelo quantil da distribuição mistura-escala Birnbaum-Saunders. Ressalta-se que os Capítulos 2 e 3 são apresentados no formato de artigo científico. No Capítulo 4 são apresentadas as considerações finais e possibilidades de trabalhos futuros. Trabalhos considerando esses problemas estão em andamento e espera-se publicação dos respectivos resultados em artigos futuros.

## Capítulo 2

# Modelos de regressão log-simétricos quantílicos

**Resumo:** Modelos de regressão baseados na família de distribuições log-simétricas são particularmente úteis quando a resposta é estritamente positiva e assimétrica ou bimodal. Neste artigo, propomos uma classe de modelos de regressão log-simétrica quantílica baseado em uma reparametrização das distribuições log-simétrica que têm o quantil como um dos seus parâmetros. Dois estudos de simulação foram realizados utilizando o software R, sendo o primeiro deles para análise do desempenho dos estimadores de máxima verossimilhança, dos critérios de informação AIC, BIC e AICc, e dos resíduos Cox-Snell e quantil aleatorizado. O segundo avaliou o tamanho/poder dos testes Wald, razão de verossimilhança, score e gradiente. Os dois estudos de simulação foram conduzidos considerando diferentes quantis de interesse e tamanhos amostrais. Por fim, uma aplicação utilizando dados referentes a bilheteria de filmes é apresentada.

**Palavras-chave:** Distribuições log-simétricas; Regressão quantílica; Simulação de Monte Carlo; Teste de hipóteses.

## 2.1 Introdução

Modelos de regressão para dados com resposta contínua estritamente positiva são cada vez mais frequentes em estudos em diferentes áreas do conhecimento, como é o caso de estudos na área de análise de sobrevivência, confiabilidade, economia, entre outros; ver Lawless (2011), Saulo e Leão (2017), Ventura et al. (2019), Sánchez et al. (2020a), Sánchez et al. (2020b) e Leiva et al. (2020). Em diversos trabalhos, é possível encontrar diferentes modelos probabilísticos que podem ser utilizados na modelagem de dados contínuos estritamente positivos. Entretanto, algumas distribuições de probabilidade se destacam para essa finalidade, podemos citar como exemplos as distribuições: log-normal, gama, exponencial, Weibull, Birnbaum-Saunders, além de muitas outras distribuições que são descritas detalhadamente em Johnson, Kotz e Balakrishnan (1994) e Johnson, Kotz e Balakrishnan (1995).

Além disso, na literatura para distribuições de probabilidade é possível encontrar diversas propostas de famílias de distribuições de probabilidade, essas classes de modelos probabilísticos geralmente contemplam diversas outras distribuições de probabilidade como casos particulares, como se pode ver nas famílias exponencial e locação e escala; ver McCullagh e Nelder (1983). A principal utilidade dessas famílias de distribuições está na flexibilidade para a modelagem de dados com diferentes características, isso geralmente ocorre devido a versatilidade proporcionada pelo número maior de parâmetros no modelo, proporcionando assim um ajuste de melhor qualidade do modelo estatístico aos dados.

A família de distribuições log-simétricas, apresentada em Vanegas e Paula (2016b), compõem uma classe de distribuições de probabilidade que geralmente são usadas na modelagem de dados assimétricos, estritamente positivas e bimodais, além de flexibilidade nas caudas. Essa família tem como base em sua formulação a classe de distribuições simétricas (Fang e Anderson, 1990), ou seja, ela é obtida por meio de uma transformação exponencial em um variável aleatória que segue uma distribuição simétrica. Assim, se  $Y$  é uma variável aleatória simétrica, então  $T = \exp(Y)$  segue uma distribuição log-simétrica. Como é característico de famílias de distribuições, a grande vantagem da distribuição log-simétrica está em um ajuste mais flexível a dados contínuos estritamente positivos.

Os modelos de regressão compõem uma das mais importantes áreas da estatística. A capacidade de estimar ou prever uma quantidade de interesse por meio da relação existente entre

diferentes variáveis aliada a utilidade prática dos modelos de regressão, visto em sua multidisciplinaridade, fazem com que essa técnica tenha um amplo alcance em diversas áreas do conhecimento. Na literatura para modelos de regressão, podemos encontrar diversas propostas para diferentes tipos de dados além do uso de distintas reparametrizações, tendo como principal objeto a estimação de diferentes quantidades, como é o caso da média ou mediada, que são os casos mais comuns. Dentro da literatura de regressão, os modelos de regressão quantílica (Hao e Naiman, 2007) desempenham um papel importante. Tal nome se justifica pela estimação de algum quantil de interesse. A ideia por trás do uso de modelos de regressão quantílica está na sua robustez, principalmente na presença de valores atípicos, os *outliers*, além da sua flexibilidade devida a escolha do quantil de interesse, podendo ser possível a estimação da mediana condicional. Na regressão quantílica, há uma caracterização mais rica dos efeitos das variáveis explicativas (covariáveis) sobre a variável dependente (resposta), que pode depender do espectro da última.

Modelos de regressão com base nas distribuições log-simétricas foram propostos por Vanegas e Paula (2015), Vanegas e Paula (2016a) e Vanegas e Paula (2017), em que os autores apresentam a modelagem simultânea da mediana e da assimetria com e sem a presença de censura. Outros trabalhos abordando aplicações dos modelos log-simétricos, principalmente na área de economia, podem ser vistos nos trabalhos de Saulo e Leão (2017) e Ventura et al. (2019).

O principal objetivo desse artigo é propor uma nova classe de modelos de regressão quantílica baseado na classe das distribuições log-simétricas. É proposto uma reparametrização das distribuições log-simétricas através da inserção do quantil como um dos parâmetros. Buscando avaliar o modelo proposto em diferentes aspectos, dois estudos de simulação de Monte Carlo utilizando o software R são realizados. O primeiro estudo avalia o desempenho dos estimadores de máxima verossimilhança, dos resíduos Generalizados Cox-Snell (GCS) (Cox e Snell, 1968) e quantil aleatorizado (RQ) (Dunn e Smyth, 1996), e dos critérios de informação de Akaike (AIC), (Akaike, 1974), Bayesiano (BIC), (Schwarz, 1978) e o Akaike corrigido (AICc) (Bozdogan, 1987). O segundo estudo de simulação analisa o desempenho dos testes Wald, score, razão de verossimilhança e gradiente. Por fim, um conjunto de dados reais referentes à bilheteria de filmes é utilizado para ilustrar a metodologia proposta.

O restante desse capítulo é organizado da seguinte forma. Na Seção 2.2, aspectos prelimina-

res a respeito da família de distribuições log-simétricas, bem como a distribuição log-simétrica reparametrizada pelo quantil proposta, são apresentados. Na Seção 2.3, o modelo de regressão quantílica log-simétrica é introduzido, após isso, serão apresentados: a função score, a informação de Fisher, testes de hipóteses baseados nos estimadores de máxima verossimilhança, critérios de seleção de modelos e os resíduos GCS e RQ. Na Seção 2.4 os resultados obtidos dos dois estudos de simulação de Monte Carlo para avaliar o desempenho dos estimadores de máxima verossimilhança, critérios de informação, resíduos e testes de hipótese, são apresentados. Na Seção 2.5, uma aplicação utilizando um conjunto de dados reais é feita, e por fim, na Seção 2.6, são apresentadas as conclusões finais sobre o modelo proposto.

## 2.2 Preliminares

Nesta seção serão apresentados alguns aspectos da família de distribuições log-simétricas, bem como uma reparametrização com base no quantil da distribuição, além de algumas propriedades do modelo reparametrizado.

### 2.2.1 A distribuição log-simétrica clássica

Uma variável aleatória  $Y$ , estritamente contínua com suporte no intervalo  $(0, \infty)$  segue uma distribuição log-simétrica com parâmetro de escala  $\lambda > 0$  e parâmetro de forma  $\phi > 0$ , denotada por  $Y \sim LS(\lambda, \phi, g)$ , se a função densidade de probabilidade (FDP) e função de distribuição acumulada (FDA) são dadas, respectivamente, por

$$f_Y(y; \lambda, \phi) = \frac{\xi_{nc}}{\sqrt{\phi}y} g\left(\frac{1}{\phi} [\log(y) - \log(\lambda)]^2\right), \quad y > 0, \quad (2.1)$$

e

$$F_Y(y; \lambda, \phi) = G\left(\frac{1}{\phi} [\log(y) - \log(\lambda)]^2\right), \quad y > 0, \quad (2.2)$$

em que  $g(\cdot)$  é o kernel gerador de densidade possivelmente associado a um parâmetro adicional  $\vartheta$  (ou um vetor de parâmetros  $\boldsymbol{\vartheta}$ ),  $\xi_{nc}$  é uma constante normalizadora e  $G(\omega) = \xi_{nc} \int_{-\infty}^{\omega} g(z^2) dz$ ,  $\omega \in \mathbb{R}$ ; ver Vanegas e Paula (2016b). Os parâmetros  $\lambda > 0$  e  $\phi > 0$  representam, respectivamente, a mediana e a assimetria (ou dispersão relativa) da distribuição de  $Y$ . Note que  $V = \log(Y)$  é uma variável aleatória que segue uma distribuição simétrica com parâmetro de locação  $\mu =$

$\log(\lambda) \in \mathbb{R}$ , parâmetro de escala  $\phi > 0$  e kernel gerador de densidade  $g(\cdot)$ , denotado por  $V \sim S(\mu, \phi, g)$ , e com FDP dada por

$$f_V(v; \mu, \phi) = \frac{\xi_{nc}}{\sqrt{\phi}} g\left(\frac{1}{\phi} [v - \mu]^2\right), \quad v \in \mathbb{R}. \quad (2.3)$$

Note também que  $G$  em (2.2) é a FDA de uma variável aleatória simétrica  $V \sim S(0, 1, g)$ . O  $100q$ -ésimo quantil de  $Y \sim \text{LS}(\lambda, \phi, g)$  é dado por

$$Q_Y(q; \lambda, \phi) = \lambda \exp(\sqrt{\phi} z_q), \quad (2.4)$$

em que  $z_q = G^{-1}(q)$  é o  $100q$ -ésimo quantil de  $V \sim S(\mu, \phi, g)$ . Na Tabela 2.1 temos a função  $g$  para algumas distribuições log-simétricas. Nessa Tabela, a constante  $l$ ,  $\text{IGF}(x, l) = \int_0^1 \exp(-xt)t^{l-1}dt$ ,  $l > 0$ ,  $x \geq 0$ , e a função gama incompleta. Mais detalhes sobre a classe das distribuições log-simétricas, bem como de outras distribuições log-simétricas podem ser vistas em Vanegas e Paula (2016b).

**Tabela 2.1:** Gerador de densidade  $g(u)$  para algumas distribuições log-simétricas.

Distribuição	$g(u)$
Log-normal( $\lambda, \phi$ )	$\exp(-\frac{1}{2}u)$
Log-Student- $t$ ( $\lambda, \phi, \vartheta$ )	$(1 + \frac{u}{\vartheta})^{-\frac{\vartheta+1}{2}}$ , $\vartheta > 0$
Log-exponencial-potência( $\lambda, \phi, \vartheta$ )	$\exp(-\frac{1}{2}u^{\frac{1}{1+\vartheta}})$ , $-1 < \vartheta \leq 1$
Log-hiperbólica( $\lambda, \phi, \vartheta$ )	$\exp(-\vartheta\sqrt{1+u})$ , $\vartheta > 0$
Log-slash( $\lambda, \phi, \vartheta$ )	$\text{IGF}(\vartheta + \frac{1}{2}, \frac{u}{2})$ , $\vartheta > 0$
Log-normal-contaminada( $\lambda, \phi, \vartheta = (\vartheta_1, \vartheta_2)^\top$ )	$\sqrt{\vartheta_2} \exp(-\frac{1}{2}\vartheta_2 u) + \frac{(1-\vartheta_1)}{\vartheta_1} \exp(-\frac{1}{2}u)$ , $0 < \vartheta_1, \vartheta_2 < 1$
Birnbaum-Saunders estendida( $\lambda, \phi, \vartheta$ )	$\cosh(u^{1/2}) \exp(-\frac{2}{\vartheta^2} \sinh^2(u^{1/2}))$ , $\vartheta > 0$
Birnbaum-Saunders- $t$ estendida( $\lambda, \phi, \vartheta = (\vartheta_1, \vartheta_2)^\top$ )	$\cosh(u^{1/2}) (\vartheta_2 \vartheta_1^2 + 4 \sinh^2(u^{1/2}))^{-\frac{\vartheta_2+1}{2}}$ , $\vartheta_1, \vartheta_2 > 0$

### 2.2.2 Uma distribuição log-simétrica baseada no quantil

Considere um número fixo  $q \in (0, 1)$  e considere a transformação um-a-um  $(\lambda, \phi) \mapsto (Q, \phi)$ , em que  $Q$  é o  $100q$ -ésimo quantil de  $Y$ , dado na Equação (2.4). Então, definindo a reparametrização da distribuição log-simétrica baseada em  $Q$ , fazendo  $\lambda = Q / \exp(\sqrt{\phi} z_p)$ ,

obtemos a FDP e a FDA de  $Y$ , respectivamente, como

$$F_Y(y; Q, \phi) = G\left(\frac{1}{\phi} [\log(y) - \log(Q)]^2\right) = G\left(\frac{1}{\phi} [\log(y) - \log(Q) + \sqrt{\phi} z_p]^2\right), \quad y > 0, \quad (2.5)$$

e

$$f_Y(y; Q, \phi) = \frac{\xi_{nc}}{\sqrt{\phi} y} g\left(\frac{1}{\phi} [\log(y) - \log(Q) + \sqrt{\phi} z_p]^2\right), \quad y > 0, \quad (2.6)$$

em que  $g(\cdot)$  e  $G(\cdot)$  são como em (2.1) e (2.2), respectivamente. Se  $Y$  segue uma distribuição log-simétrica reparametrizada pelo quantil, então denotamos  $Y \sim \text{QLS}(Q, \phi, g)$ , de tal forma que  $V = \log(Y)$  segue uma distribuição simétrica com FDP

$$f_V(v; \Psi, \phi) = \frac{\xi_{nc}}{\sqrt{\phi}} g\left(\frac{1}{\phi} [v - \Psi + \sqrt{\phi} z_p]^2\right), \quad v \in \mathbb{R}. \quad (2.7)$$

em que  $\Psi = \log(Q)$ . A seguinte notação é utilizada  $V \sim \text{QS}(\Psi, \phi, g)$ . A seguir temos algumas propriedades da distribuição log-simétrica reparametrizada pelo quantil.

**Proposição 1.** Seja  $Y \sim \text{QLS}(Q, \phi, g)$ . Então,  $cY \sim \text{QLS}(cQ, \phi, g)$ , com  $c > 0$ .

*Demonstração.* Se  $Y \sim \text{QLS}(Q, \phi, g)$  e  $c > 0$ , então

$$\begin{aligned} P(cY \leq y) &= P\left(Y \leq \frac{y}{c}\right) = G\left(\frac{1}{\phi} \left[\log\left(\frac{y}{c}\right) - \log(Q) + \sqrt{\phi} z_p\right]^2\right) \\ &= G\left(\frac{1}{\phi} \left[\log(y) - \log(cQ) + \sqrt{\phi} z_p\right]^2\right). \end{aligned}$$

□

**Proposição 2.** Seja  $Y \sim \text{QLS}(Q, \phi, g)$ . Então,  $Y^c \sim \text{QLS}(Q^c, c^2\phi, g)$ , com  $c > 0$ .

*Demonstração.* Se  $Y \sim \text{QLS}(Q, \phi, g)$  e  $c \neq 0$ , Então,

$$\begin{aligned} P(Y^c \leq y) &= P\left(Y \leq y^{1/c}\right) = G\left(\frac{1}{\phi} \left[\frac{1}{c} \log(y) - \log(Q) + \sqrt{\phi} z_p\right]^2\right) \\ &= G\left(\frac{1}{c^2\phi} \left[\log(y) - \log(Q^c) + \sqrt{c^2\phi} z_p\right]^2\right). \end{aligned}$$

□

### 2.3 Modelo de regressão log-simétrico quantílico

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes, de modo que  $Y_i \sim \text{QLS}(Q_i, \phi_i, g)$ , e  $\mathbf{y} = (y_1, \dots, y_n)^\top$  sejam suas observações correspondentes. Então, definimos

$$Y_i = Q_i \epsilon_i^{\sqrt{\phi_i}}, \quad \epsilon_i \sim \text{QLS}(1, 1, g), \quad (2.8)$$

que implica

$$\begin{aligned} V_i = \log(Y_i) &= \log(Q_i) + \sqrt{\phi_i} \log(\epsilon_i) \\ &= \Psi_i + \sqrt{\phi_i} \epsilon_i \quad i = 1, \dots, n, \end{aligned} \quad (2.9)$$

em que  $V_i \sim \text{QS}(\Psi_i, \phi_i, g)$ ,  $\epsilon_i = \log(\epsilon_i) \sim \text{QS}(0, 1, g)$ ,  $\log(Q_i) = \Psi_i = \mathbf{x}_i^\top \boldsymbol{\beta}$  e  $\log(\phi_i) = \mathbf{w}_i^\top \boldsymbol{\tau}$ , tal que  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^\top$  e  $\boldsymbol{\tau} = (\tau_0, \dots, \tau_l)^\top$  denotam os vetores de parâmetros desconhecidos a serem estimados,  $\mathbf{x}_i^\top = (1, x_{i1}, \dots, x_{ik})^\top$  e  $\mathbf{w}_i^\top = (1, w_{i1}, \dots, w_{il})^\top$  denotam as  $k$  e  $l$  covariáveis que modelam a quantil  $Q_i$  e assimetria (ou dispersão relativa)  $\phi_i$ , respectivamente.

A estimação dos parâmetros do modelo de regressão é feita utilizando o método da máxima verossimilhança. Denotando o vetor de parâmetros do modelo  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ , então o logaritmo da função verossimilhança (sem a constante) para o modelo de regressão quantílica é dado por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \log(g(z_i^2)) - \frac{1}{2} \log(\phi_i) \right\}, \quad (2.10)$$

onde  $z_i = [\log(y_i) - \log(Q_i) + \sqrt{\phi_i} z_p] / \sqrt{\phi_i}$ , para  $i = 1, \dots, n$ , com  $z_p$  como na Equação (2.4). As estimativas de máxima verossimilhança são obtidas maximizando a expressão definida em (2.10). Isso é feito igualando o vetor score  $\dot{\ell}(\boldsymbol{\theta})$ , no qual está contido as primeiras derivadas de  $\dot{\ell}(\boldsymbol{\theta})$ , a zero, o que nos fornece as equações de verossimilhança. O vetor score  $\dot{\ell}(\boldsymbol{\theta})$  pode ser dividido em dois outros componentes, sendo eles

$$\dot{\ell}(\boldsymbol{\theta}) = (\dot{\ell}_\eta(\boldsymbol{\theta}), \dot{\ell}_\phi(\boldsymbol{\theta}))^\top, \quad (2.11)$$

em que o vetor score para o parâmetro  $\boldsymbol{\eta}$  é dado por  $\dot{\ell}_\eta(\boldsymbol{\theta}) = (\dot{\ell}_{\beta_0}(\boldsymbol{\theta}), \dots, \dot{\ell}_{\beta_k}(\boldsymbol{\theta}))^\top$ , ou seja, pelo vetor das primeiras derivadas de  $\ell(\boldsymbol{\theta})$  em relação ao vetor de parâmetros  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)$ ,

em que

$$\dot{\ell}_{\beta_j}(\boldsymbol{\theta}) = \sum_{i=1}^n v(z_i) z_i \frac{x_{ij}}{\sqrt{\phi_i}}, \quad j = 0, \dots, k,$$

com  $v(z_i) = -2g'(z_i^2)/g(z_i^2)$  denotando os pesos atribuídos a função  $g(\cdot)$ , em que  $g'(u)$  é a primeira derivada de  $g(u)$  em relação a  $u$ . Por outro lado, o vetor score  $\dot{\ell}_\phi(\boldsymbol{\theta}) = (\dot{\ell}_{\tau_0}(\boldsymbol{\theta}), \dots, \dot{\ell}_{\tau_l}(\boldsymbol{\theta}))^\top$  é dado pelas primeiras derivadas de  $\ell(\boldsymbol{\theta})$  em relação ao vetor de parâmetros  $\boldsymbol{\tau} = (\tau_0, \dots, \tau_l)$ , ou seja,

$$\dot{\ell}_{\tau_j}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^n \frac{w_{ij}}{\phi_i} [v(z_i) z_i (z_i - z_p) - 1], \quad j = 0, \dots, l.$$

Entretanto, a solução do sistema de equações de estimação dos parâmetros não podem ser obtidas de forma analítica, o que faz necessário o uso de métodos iterativos, como o método de Newton-Raphson ou scoring de Fisher. Mais detalhes sobre os métodos de otimização citados podem ser vistos em Nocedal e Wright (2006) e Lange (2010).

Uma vez obtidos os estimadores de máxima verossimilhança, a inferência para  $\boldsymbol{\theta}$  do modelo de regressão log-simétrico quantílico pode ser baseado na distribuição assintótica de  $\hat{\boldsymbol{\theta}}$ . Para  $n$  suficientemente grande e sob as condições usuais de regularidade (Efron e Hinkley, 1978), o estimador de máxima verossimilhança  $\hat{\boldsymbol{\theta}}$  converge em distribuição para uma distribuição normal multivariada, ou seja,

$$\sqrt{n} [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] \xrightarrow{D} \mathbf{N}_{2+k+l}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})^{-1}), \quad (2.12)$$

com  $n \rightarrow \infty$ , em que  $\xrightarrow{D}$  significa “convergência em distribuição”,  $\mathbf{0}$  é um vetor de zeros  $(2 + k + l) \times 1$  e  $\mathcal{I}(\boldsymbol{\theta})$  corresponde a matriz de informação de Fisher esperada, a qual é obtida calculando  $\mathcal{I}(\boldsymbol{\theta}) = \mathbb{E}[-\partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top]$ . Note que a diagonal da matriz de informação de Fisher esperada é dada por  $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}) = \text{diag} \{ \mathcal{I}_{\beta\beta}(\boldsymbol{\theta}), \mathcal{I}_{\tau\tau}(\boldsymbol{\theta}) \}$ , em que

$$\mathcal{I}_{\beta_j\beta_j}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{x_{ij}^2}{\phi_i} d_g(\zeta), \quad j = 0, \dots, k,$$

e

$$\mathcal{I}_{\tau_j\tau_j}(\boldsymbol{\theta}) = \frac{1}{4} \sum_{i=1}^n \left( \frac{w_{ij}}{\phi_i} \right)^2 f_g(\zeta), \quad j = 0, \dots, l,$$

em que  $d_g(\zeta) = \mathbb{E}[v^2(Z)Z^2]$ ,  $Z = [\log(Y) - \log(Q) + \sqrt{\phi} z_p] / \sqrt{\phi} \sim \mathbf{S}(0, 1, g)$ , e  $f_g(\zeta) = \mathbb{E}[(v(Z)Z(Z - z_p) - 1)^2]$ .

Note que em (2.10) o possível parâmetro extra ( $\vartheta$ ) é assumido fixo. A motivação para

deixá-lo fixo vem dos trabalhos Lucas (1997a) e Kano, Berkane e Bentler (1993). No primeiro trabalho é mostrado que robustez contra outliers sob o modelo Student- $t$  acontece apenas se os graus de liberdade são fixos ao invés de estimados pelo método da máxima verossimilhança. O segundo trabalho reporta dificuldades de estimação do parâmetro extra da exponencial-potência; ver Vanegas e Paula (2016b). Desse modo, dois passos são realizados para estimar o parâmetro extra: Passo 1) considere um grid de valores para  $\vartheta$ ,  $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ , e compute as estimativas de máxima verossimilhança de  $\theta$  para cada valor de  $\vartheta_i$ ,  $i = 1, \dots, k$ . Compute também o valor da função de log-verossimilhança; Passo 2) as estimativas finais de  $\vartheta$  de  $\theta$  são aquelas que maximizam a função de log-verossimilhança.

### 2.3.1 Testes de hipóteses

Suponha que queremos testar a hipótese  $H_0 : \theta = \theta_0$ , contra a hipótese alternativa  $H_1 : \theta \neq \theta_0$  em que  $\theta = (\beta^\top, \tau^\top)^\top$  é um vetor de parâmetros desconhecidos. Na literatura sobre testes de hipóteses para modelos de regressão, três testes assintóticos clássicos são amplamente utilizados para testar estas hipóteses, sendo estes os testes Wald, score e razão de verossimilhança, propostos respectivamente, por Wald (1947), Rao (1948) e Wilks (1938). As estatísticas Wald, score e razão de verossimilhança, denotadas por  $S_W$ ,  $S_R$  e  $S_{LR}$  são dadas, respectivamente, por:

$$\begin{aligned} S_W &= (\hat{\theta} - \theta_0)^\top \mathcal{J}(\hat{\theta})(\hat{\theta} - \theta_0), \\ S_R &= \dot{\ell}(\theta_0)^\top \mathcal{I}(\theta_0)^{-1} \dot{\ell}(\theta_0), \\ S_{LR} &= -2 \left[ \ell(\theta_0) - \ell(\hat{\theta}) \right], \end{aligned}$$

em que  $\mathcal{J}(\theta) = - \{ \partial^2 \ell(\theta) / \partial \theta \partial \theta^\top \}$  é a informação de Fisher observada. Rao (2005) sugere uma modificação para o teste  $S_R$  substituindo  $\mathcal{I}(\theta_0)$  por  $\mathcal{J}(\theta_0)$ , o que simplifica o cálculo da estatística score principalmente sob condições em que o cálculo de  $\mathcal{I}(\theta)$  é muito complicado.

Um outro teste com propriedades bem similares aos apresentados anteriormente foi desenvolvido por Terrell (2002), o chamado teste gradiente. Ele fundamenta-se em uma modificação da estatística  $S_R$ . Terrell (2002) mostrou que estatística gradiente, denotada por  $S_T$ , possui propriedades similares às três estatísticas clássicas. A grande vantagem da estatística gradiente está na facilidade de calculá-la, uma vez que assim como a estatística  $S_{LR}$ , não há necessidade de se

obter a matriz de informação de Fisher. A estatística  $S_T$  é dada por

$$S_T = \dot{\ell}(\boldsymbol{\theta}_0)^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Vale ressaltar que, sob as condições usuais, as estatísticas  $S_W$ ,  $S_{LR}$ ,  $S_R$  e  $S_T$  são assintoticamente equivalentes, isto é, sob  $H_0$ , ambas seguem uma distribuição aproximadamente  $\chi_t^2$ , em que  $t$  é o número de parâmetros fixados na hipótese  $H_0$ . Para os testes em questão, rejeitamos a hipótese  $H_0$  se o valor observado da estatística de teste for maior que o valor crítico  $C_{t,1-\alpha}$ , em que  $C_{t,1-\alpha}$  é o quantil de ordem  $1 - \alpha$  de uma distribuição  $\chi_t^2$ .

### 2.3.2 Critérios de informação

O problema de seleção de modelos em regressão é de grande relevância. Nesse sentido, os critérios de informação se apresentam como uma metodologia bem efetiva e simples para a seleção do modelo ótimo. Os critérios de informação de Akaike (AIC), bayesiano (BIC) e Akaike corrigido (AICc) serão considerados para seleção dos modelos.

O critério AIC (Akaike, 1974) parte da suposição da existência de um “modelo verdadeiro”, desconhecido, de tal forma que, para um grupo de modelos candidatos, o melhor modelo é aquele que apresenta a menor divergência em relação ao “modelo verdadeiro”. Tomando a função de log-verossimilhança maximizada como estimativa da divergência, penalizada pelo número de parâmetros  $p$  indexados ao modelo, o critério AIC é dado por

$$\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2p, \quad (2.13)$$

em que  $\ell(\hat{\boldsymbol{\theta}})$  é a função de log-verossimilhança maximizada e  $p$  representa o número de parâmetros do vetor  $\hat{\boldsymbol{\theta}}$ . Sakamoto, Ishiguro e Kitagawa (1986) e Sugiura (1978) afirmam que o AIC pode apresentar desempenho ruim diante de um número elevado de parâmetros, o que, por sua vez, pode induzir a uma conclusão equivocada em relação ao melhor modelo. Buscando corrigir o viés causado em virtude da circunstância apresentada, Bozdogan (1987) propôs uma correção ao AIC, conhecido como critério AICc, o qual é dado por

$$\text{AICc} = \text{AIC} + \frac{2p(p+1)}{n-p-1}, \quad (2.14)$$

em que  $n$  é o tamanho da amostra.

O critério BIC foi proposto por Schwarz (1978) e assume o pressuposto da existência de um ‘modelo verdadeiro’, o critério bayesiano é definido como a estatística que maximiza a probabilidade de escolha do modelo verdadeiro, dentre os modelos postulados. Além de penalizar o número de parâmetros, o critério BIC também leva em consideração o tamanho da amostra e é dado por

$$\text{BIC} = -2\ell(\hat{\theta}) + p \log(n).$$

Vale ressaltar que, para os três critérios de informação apresentados, o melhor modelo, dentre os modelos candidatos, é aquele que apresenta o menor valor do critério de informação.

### 2.3.3 Resíduos

Os resíduos compõem um importante ferramenta na análise de diagnóstico de modelos de regressão, sendo úteis na detecção de indícios da inadequabilidade de ajuste, *outliers*, entre outros problemas. Dois resíduos que são bastante utilizados são os resíduos Cox-Snell generalizado (GCS) e o resíduo quantil aleatorizado (RQ). Os resíduos GCS (Cox e Snell, 1968) são geralmente utilizados para avaliar a qualidade do ajuste global do modelo, sendo amplamente utilizados para dados com resposta assimétrica, principalmente na avaliação da qualidade do ajuste do modelos de análise de sobrevivência. Os resíduos GCS são dados por

$$\hat{r}_i^{\text{GCS}} = -\log(1 - F_Y(y_i; \hat{Q}_i, \hat{\phi}_i)), \quad i = 1, 2, \dots, n, \quad (2.15)$$

em que  $F_Y(y_i; \hat{Q}_i, \hat{\phi}_i)$  é dada pela Equação (2.5). Segundo Lee e Wang (2003), se o modelo está corretamente especificado, os resíduos GCS seguem uma distribuição exponencial padrão.

Os resíduos RQ (Dunn e Smyth, 1996) são usualmente utilizados em modelos aditivos generalizados para localização, escala e forma. Esses resíduos são dados por

$$\hat{r}_i^{\text{QR}} = \Phi^{-1}(F_Y(y_i; \hat{Q}_i, \hat{\phi}_i)), \quad i = 1, 2, \dots, n, \quad (2.16)$$

em que  $\Phi(\cdot)^{-1}$  é a função inversa da FDA da distribuição normal padrão. Os resíduos RQ seguem uma distribuição normal padrão se o modelo postulado está corretamente especificado.

Para ambos os resíduos, o pressuposto da distribuição pode ser verificado por meio de técnicas gráficas, testes de hipóteses e estatísticas descritivas.

## 2.4 Estudos de simulação de Monte Carlo

Nesta seção iremos apresentar os resultados de dois estudos de simulação de Monte Carlo para o modelo de regressão log-simétrico quantílico. Na primeira simulação são avaliados o desempenho dos estimares de máxima verossimilhança, dos critérios de informação e dos resíduos. Na segunda simulação, é avaliada o desempenho dos testes Wald ( $S_W$ ), score ( $S_R$ ), razão de verossimilhança ( $S_{LR}$ ) e gradiente ( $S_T$ ). Em ambas as simulações, o seguinte cenário é considerado: tamanhos de amostra:  $n = 50, 100, 200$ , quantis  $q = 0.25, 0.5, 0.75, 5,000$  réplicas de Monte Carlo, com as seguintes distribuições log-simétricas: log-normal (log-NO), log-Student- $t$  (log- $t$ ), log-exponencial-potência (log-PE), log-hiperbólica (log-HP), log-slash (log-SL), log-normal-contaminada (log-CN), Birnbaum-Saunders estendida (EBS), e Birnbaum-Saunders- $t$  estendida (EBS- $t$ ). Os valores dos parâmetros extras para as distribuições consideradas foram:  $\vartheta = 3$  (log- $t$ ),  $\vartheta = 0.3$  (log-PE),  $\vartheta = 2$  (log-HP),  $\vartheta = 4$  (log-SL),  $\vartheta_1 = 0.1, \vartheta_2 = 0.2$  (log-CN),  $\vartheta = 0.5$  (EBS), e  $\vartheta_1 = 0.5, \vartheta_2 = 3$  (EBS- $t$ ).

### 2.4.1 Estudo de Monte Carlo 1: desempenho dos estimadores, critérios de informação e resíduos

O modelo utilizado para geração das amostras é dado por

$$Y_i = (\beta_0 + \beta_1 x_i) \epsilon_i^{\sqrt{\tau_0 + \tau_1 w_i}}, \quad i = 1, 2, \dots, n,$$

com os valores verdadeiros dos parâmetros fixados em  $\beta_0 = 1.5$ ,  $\beta_1 = 0.5$ ,  $\tau_0 = 1$ ,  $\tau_1 = 0.5$ , e  $\epsilon_i \sim \text{QLS}(1, 1, g)$ . Os valores das covariáveis  $x_i, w_i$  são geradas das distribuições Bernoulli(0.5) e Uniforme(0,1), respectivamente. Para avaliar o desempenho dos estimadores de máxima verossimilhança, as estimativas do viés e do erro quadrático médio (EQM) são computadas via simulação de Monte Carlo, e são dadas, respectivamente, por

$$\text{Viés}(\hat{\varphi}) = \frac{1}{M} \sum_{r=1}^M (\hat{\varphi}^{(r)} - \varphi), \quad \text{EQM}(\hat{\varphi}) = \frac{1}{M} \sum_{r=1}^M (\hat{\varphi}^{(r)} - \varphi)^2,$$

em que  $\varphi$  e  $\widehat{\varphi}^{(r)}$  representam o valor verdadeiro do parâmetro e sua respectiva  $r$ -ésima estimativa, respectivamente, enquanto que  $M$  representa o número de réplicas de Monte Carlo utilizadas. São computadas também nessa simulação as probabilidades de cobertura (PC) do intervalo de confiança (IC) de 95% baseado na normalidade assintótica (2.12), as quais são obtidas fazendo

$$\text{PC}(\widehat{\varphi}) = \frac{1}{M} \sum_{r=1}^M I(\varphi \in [\widehat{\varphi}_L^{(r)}, \widehat{\varphi}_U^{(r)}]),$$

em que  $I(\cdot)$  é uma função indicadora tal que  $\varphi$  pertence ao  $r$ -ésimo intervalo  $[\widehat{\varphi}_L^{(r)}, \widehat{\varphi}_U^{(r)}]$ , com  $\widehat{\varphi}_L^{(r)}$  e  $\widehat{\varphi}_U^{(r)}$  sendo os  $r$ -ésimos limites inferior e superior estimados do IC de 95%.

Além da análise da performance dos estimadores de máxima verossimilhança, os critérios de informação para seleção de modelos AIC, BIC e AICc também foram avaliados para o modelo de regressão quantílica proposto. A ideia é avaliar quanto bem esses critérios escolhem a distribuição log-simétrica correta. Por exemplo, se geramos amostras do modelo de regressão log-normal quantílico, devemos computar, para cada réplica de Monte Carlo, o AIC, BIC e AICc para os modelos log-normal e todas as outras distribuições. A média então desses valores é computada e o modelo com os menores valores para os critérios é o escolhido (nesse caso, espera-se que seja o log-normal). Por fim, uma análise do desempenho dos resíduos GCS e RQ é realizada, ou seja, avalia-se a distribuição empírica dos resíduos. Nesse caso, para cada réplica de Monte Carlo, os resíduos GCS e RQ associados a um modelo de regressão log-simétrico quantílico são computados bem como suas correspondentes medidas descritivas, tais como média, desvio-padrão (SD), coeficiente de assimetria (CS) e coeficiente de curtose (CK). Então, são computadas as médias de cada medida descritiva. Espera-se que tais médias sejam próximas aos valores teóricos. Os passos desse primeiro estudo de simulação de Monte Carlo é descrito no Algoritmo 1.

---

**Algoritmo 1** Passos para o estudo de Monte Carlo 1.

---

- 1: Escolher o modelo de regressão log-simétrico quantílico baseado em algum gerador de densidade  $g(\cdot)$  da Tabela 2.1 e fixar os parâmetros do modelo.
  - 2: Gerar 5000 amostras baseadas no modelo escolhido.
  - 3: Estimar os parâmetros do modelo pelo método da máxima verossimilhança para cada amostra.
  - 4: Estimar o viés, o EQM e a PC com base nas 5000 réplicas.
  - 5: Computar, para cada réplica, os resíduos GCS e RQ e as respectivas estatísticas descritivas: media, mediana, SD, CS e CK. Computar as médias das estatísticas descritivas com base nas 5000 réplicas.
  - 6: Computar, para cada réplica, os valores dos critérios de informação AIC, BIC e AICc para o modelo escolhido e para diferentes modelos da Tabela 2.1. Computar a percentagem que acertadamente os critérios de informação AIC, BIC e AICc, escolhem o modelo correto, ou seja, o modelo baseado no processo gerador dos dados.
- 

As Tabelas 2.2-2.9 apresentam os resultados das simulações de Monte Carlo baseados nas distribuições log-NO, log- $t$ , log-PE, log-HP, log-SL, log-CN, EBS e EBS- $t$ . Os seguintes resultados para as estimativas de máxima verossimilhança são apresentados: viés, EQM e PC. Uma análise das Tabelas 2.2-2.9 permite concluir que como esperado, em geral, a medida que o tamanho da amostra aumenta, o viés e o EQM diminuem, para todas as distribuições consideradas. Ainda, ressalta-se que a PC se aproxima do nível de confiança nominal de 95% a medida que aumentamos o tamanho da amostra. Por exemplo, para amostras de tamanho  $n = 50$ , a PC é de aproximadamente 93%, para  $n = 200$ , a PC é de aproximadamente 95%. Com os resultados apresentados nessas tabelas, pode-se observar também que o EQM do estimador de  $\tau_1$  é sempre maior para o modelo de regressão com  $q = 0.5$  que em relação aos quantis  $q = 0.25$  e  $q = 0.75$ , para todas as distribuições analisadas. Outro aspecto que deve ser ressaltado é que os EQMs dos estimadores de  $\beta_1$  e  $\tau_1$  são sempre maiores do que os estimadores associados aos outros coeficientes do modelo.

**Tabela 2.2:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-NO.

$q$		$n = 50$			$n = 100$			$n = 200$		
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC
0.25	$\beta_0$	0.04494	0.21187	0.9378	0.02548	0.07570	0.9408	0.00639	0.03417	0.9490
	$\beta_1$	-0.00517	0.31041	0.9366	-0.01253	0.14261	0.9400	0.00313	0.07205	0.9446
	$\tau_0$	-0.08922	0.19829	0.9376	-0.04883	0.07609	0.9418	-0.02215	0.03902	0.9502
	$\tau_1$	0.01857	0.47784	0.9430	0.01845	0.20340	0.9422	0.00481	0.10433	0.9516
0.5	$\beta_0$	0.00802	0.19609	0.9330	0.00623	0.06855	0.9398	-0.00258	0.03038	0.9498
	$\beta_1$	-0.00565	0.31264	0.9358	-0.01274	0.14346	0.9384	0.00266	0.07201	0.9480
	$\tau_0$	-0.09039	0.23368	0.9332	-0.04718	0.08827	0.9386	-0.02093	0.04420	0.9522
	$\tau_1$	0.02074	0.59772	0.9374	0.01492	0.25472	0.9432	0.00232	0.12644	0.9540
0.75	$\beta_0$	-0.03438	0.13609	0.9366	-0.01331	0.07714	0.9396	-0.01170	0.03487	0.9424
	$\beta_1$	0.00155	0.31954	0.9336	-0.01247	0.14316	0.9398	0.00253	0.07186	0.9488
	$\tau_0$	-0.10746	0.19714	0.9360	-0.04470	0.07654	0.9384	-0.01996	0.03862	0.9512
	$\tau_1$	0.05292	0.47134	0.9426	0.00976	0.20408	0.9424	0.00033	0.10397	0.9496

**Tabela 2.3:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log- $t$  ( $\vartheta = 3$ ).

$q$		$n = 50$			$n = 100$			$n = 200$		
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC
0.25	$\beta_0$	0.05723	0.35426	0.9226	0.02137	0.13608	0.9390	0.01011	0.06332	0.9438
	$\beta_1$	-0.01033	0.48476	0.9340	0.00568	0.21711	0.9390	-0.00399	0.10860	0.9496
	$\tau_0$	-0.09946	0.36506	0.9332	-0.05036	0.14223	0.9354	-0.01374	0.07637	0.9452
	$\tau_1$	0.03662	0.84097	0.9414	0.01999	0.36923	0.9360	-0.00971	0.19696	0.9446
0.5	$\beta_0$	-0.00009	0.33458	0.9278	-0.00367	0.11682	0.9366	-0.00144	0.05375	0.9476
	$\beta_1$	0.00682	0.51008	0.9288	0.00620	0.21809	0.9392	-0.00339	0.10864	0.9478
	$\tau_0$	-0.08664	0.43086	0.9302	-0.05151	0.17611	0.9370	-0.01108	0.09616	0.9408
	$\tau_1$	0.02129	1.16259	0.9358	0.02734	0.51431	0.9414	-0.01276	0.27128	0.9450
0.75	$\beta_0$	-0.04706	0.37799	0.9250	-0.02775	0.13861	0.9364	-0.01245	0.06492	0.9430
	$\beta_1$	0.00673	0.50690	0.9298	0.00665	0.21735	0.9396	-0.00313	0.10819	0.9498
	$\tau_0$	-0.09495	0.34629	0.9320	-0.05336	0.13944	0.9388	-0.01596	0.07606	0.9476
	$\tau_1$	0.02797	0.83592	0.9356	0.02692	0.35996	0.9440	-0.00512	0.19132	0.9482

**Tabela 2.4:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-PE ( $\vartheta = 0.3$ ).

$q$		$n = 50$			$n = 100$			$n = 200$		
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC
0.25	$\beta_0$	0.03967	0.37253	0.9258	0.02905	0.13351	0.9380	0.01162	0.07233	0.9388
	$\beta_1$	0.00865	0.53300	0.9272	0.00046	0.22886	0.9434	0.00072	0.11441	0.9438
	$\tau_0$	-0.09984	0.23753	0.9344	-0.04151	0.10144	0.9440	-0.02117	0.04064	0.9500
	$\tau_1$	0.03592	0.58700	0.9376	0.00376	0.27762	0.9430	0.00477	0.11494	0.9472
0.5	$\beta_0$	-0.00349	0.20361	0.9296	0.00462	0.11724	0.9416	0.00002	0.06459	0.9392
	$\beta_1$	-0.00374	0.45932	0.9378	0.00065	0.22971	0.9406	0.00077	0.11494	0.9430
	$\tau_0$	-0.08881	0.29855	0.9330	-0.04069	0.11968	0.9464	-0.02006	0.04712	0.9472
	$\tau_1$	0.01916	0.96609	0.9314	0.00210	0.35393	0.9464	0.00228	0.14724	0.9422
0.75	$\beta_0$	-0.05048	0.23064	0.9310	-0.01718	0.12940	0.9330	-0.00948	0.06530	0.9428
	$\beta_1$	-0.00121	0.45677	0.9412	-0.00203	0.22307	0.9438	-0.00268	0.11625	0.9378
	$\tau_0$	-0.07954	0.19601	0.9342	-0.03317	0.09536	0.9364	-0.01860	0.04493	0.9480
	$\tau_1$	0.00446	0.54319	0.9334	-0.01392	0.30320	0.9406	-0.00088	0.13713	0.9478

**Tabela 2.5:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-HP ( $\vartheta = 2$ ).

$q$	$n = 50$			$n = 100$			$n = 200$			
	Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC	
0.25	$\beta_0$	0.04341	0.12272	0.9244	0.01370	0.06625	0.9460	0.00484	0.01944	0.9510
	$\beta_1$	-0.01192	0.25651	0.9228	0.00355	0.11393	0.9446	0.00100	0.03268	0.9546
	$\tau_0$	-0.09152	0.26537	0.9386	-0.04850	0.09725	0.9414	-0.01914	0.04941	0.9456
	$\tau_1$	0.02235	0.78818	0.9368	0.01685	0.27689	0.9466	0.00534	0.14225	0.9464
0.5	$\beta_0$	0.00800	0.11653	0.9250	-0.00464	0.03579	0.9358	0.00132	0.02724	0.9534
	$\beta_1$	-0.00928	0.24365	0.9308	0.00624	0.07025	0.9374	-0.00139	0.05645	0.9488
	$\tau_0$	-0.09105	0.23209	0.9356	-0.04459	0.11275	0.9422	-0.02084	0.05383	0.9480
	$\tau_1$	0.02449	0.60907	0.9388	0.01181	0.31179	0.9470	0.00564	0.16596	0.9520
0.75	$\beta_0$	-0.02912	0.14252	0.9330	-0.01710	0.04095	0.9362	-0.00667	0.03108	0.9514
	$\beta_1$	-0.00218	0.22892	0.9400	0.00600	0.07011	0.9384	-0.00112	0.05636	0.9492
	$\tau_0$	-0.08729	0.20085	0.9366	-0.04657	0.09803	0.9408	-0.01984	0.04618	0.9464
	$\tau_1$	0.01938	0.56486	0.9388	0.01505	0.24663	0.9444	0.00288	0.13103	0.9528

**Tabela 2.6:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-SL ( $\vartheta = 4$ ).

$q$	$n = 50$			$n = 100$			$n = 200$			
	Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC	
0.25	$\beta_0$	0.04513	0.18733	0.9360	0.02660	0.09976	0.9406	0.00591	0.05687	0.9448
	$\beta_1$	-0.01673	0.38210	0.9398	-0.01562	0.18943	0.9394	0.00751	0.10493	0.9444
	$\tau_0$	-0.08512	0.17147	0.9306	-0.04243	0.08096	0.9410	-0.01660	0.04662	0.9408
	$\tau_1$	0.01686	0.46471	0.9350	0.01067	0.23064	0.9442	0.00140	0.12478	0.9434
0.5	$\beta_0$	0.00024	0.15159	0.9354	-0.00169	0.11140	0.9456	-0.00082	0.04713	0.9510
	$\beta_1$	-0.00578	0.39991	0.9350	0.00150	0.20521	0.9466	0.00134	0.09196	0.9462
	$\tau_0$	-0.09047	0.19744	0.9228	-0.04150	0.11226	0.9404	-0.01601	0.04652	0.9490
	$\tau_1$	0.02633	0.57069	0.9280	0.01002	0.32505	0.9428	-0.00529	0.15561	0.9528
0.75	$\beta_0$	-0.04315	0.18198	0.9320	-0.01470	0.11162	0.9434	-0.00857	0.05156	0.9452
	$\beta_1$	0.00021	0.40340	0.9322	-0.01007	0.18708	0.9460	-0.00455	0.09277	0.9464
	$\tau_0$	-0.09574	0.18798	0.9344	-0.03966	0.08969	0.9430	-0.02439	0.04361	0.9416
	$\tau_1$	0.03542	0.46394	0.9414	0.00483	0.25233	0.9408	0.01075	0.11783	0.9452

**Tabela 2.7:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica log-CN ( $\vartheta_1 = 0.1, \vartheta_1 = 0.2$ ).

$q$	$n = 50$			$n = 100$			$n = 200$			
	Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC	
0.25	$\beta_0$	0.03428	0.29328	0.9298	0.01734	0.10322	0.9394	0.00958	0.04582	0.9476
	$\beta_1$	-0.00047	0.40343	0.9352	0.00098	0.17773	0.9370	-0.00548	0.08839	0.9462
	$\tau_0$	-0.08920	0.19994	0.9344	-0.04330	0.09862	0.9404	-0.01747	0.05315	0.9486
	$\tau_1$	0.03943	0.52512	0.9388	0.01604	0.26394	0.9446	0.00265	0.14103	0.9508
0.5	$\beta_0$	0.00152	0.14127	0.9348	-0.00149	0.08951	0.9416	0.00158	0.03971	0.9514
	$\beta_1$	-0.00866	0.37150	0.9310	0.00227	0.17737	0.9384	-0.00565	0.08850	0.9454
	$\tau_0$	-0.07597	0.26109	0.9344	-0.04236	0.11553	0.9362	-0.01608	0.06270	0.9518
	$\tau_1$	0.01602	0.81035	0.9380	0.01602	0.34129	0.9388	0.00080	0.18008	0.9462
0.75	$\beta_0$	-0.03343	0.17552	0.9368	-0.02022	0.10162	0.9374	-0.00631	0.04599	0.9492
	$\beta_1$	-0.00868	0.36991	0.9374	0.00286	0.17763	0.9372	-0.00573	0.08824	0.9458
	$\tau_0$	-0.07384	0.24572	0.9282	-0.04242	0.09721	0.9408	-0.01600	0.05257	0.9516
	$\tau_1$	0.00432	0.66144	0.9344	0.01407	0.26288	0.9384	-0.00014	0.13954	0.9476

**Tabela 2.8:** Viés, EQM e PC baseados nos dados simulados do modelo de regressão quantílica EBS ( $\vartheta = 0.5$ ).

$q$		$n = 50$			$n = 100$			$n = 200$		
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC
0.25	$\beta_0$	0.01121	0.00881	0.9350	0.00491	0.00430	0.9414	0.00248	0.00226	0.9418
	$\beta_1$	-0.00343	0.01558	0.9380	-0.00037	0.00810	0.9488	0.00062	0.00413	0.9422
	$\tau_0$	-0.09553	0.14441	0.9276	-0.05220	0.07498	0.9378	-0.02107	0.03415	0.9466
	$\tau_1$	0.01635	0.45149	0.9368	0.02227	0.20758	0.9460	0.00437	0.10093	0.9470
0.5	$\beta_0$	-0.00116	0.00698	0.9342	-0.00022	0.00383	0.9404	0.00015	0.00210	0.9444
	$\beta_1$	0.00448	0.01744	0.9310	0.00009	0.00805	0.9452	0.00098	0.00414	0.9446
	$\tau_0$	-0.09376	0.19738	0.9332	-0.05998	0.08609	0.9440	-0.02061	0.03796	0.9482
	$\tau_1$	0.01882	0.73957	0.9316	0.03481	0.24082	0.9472	0.00267	0.12472	0.9474
0.75	$\beta_0$	-0.01071	0.01067	0.9334	-0.00507	0.00457	0.9374	-0.00294	0.00236	0.9418
	$\beta_1$	0.00300	0.01839	0.9264	0.00085	0.00849	0.9400	0.00081	0.00420	0.9428
	$\tau_0$	-0.09893	0.23746	0.9292	-0.04283	0.08525	0.9330	-0.02446	0.02988	0.9416
	$\tau_1$	0.02173	0.60652	0.9346	0.01077	0.22118	0.9382	0.00422	0.08813	0.9404

**Tabela 2.9:** Viés, EQM e PC dos parâmetros do modelo de regressão com erros Log-ST(0.5,3).

$q$		$n = 50$			$n = 100$			$n = 200$		
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC
0.25	$\beta_0$	0.01005	0.01464	0.9318	0.00464	0.00826	0.9382	0.00212	0.00375	0.9474
	$\beta_1$	0.00289	0.02661	0.9328	0.00164	0.01367	0.9390	0.00096	0.00662	0.9470
	$\tau_0$	-0.09564	0.23770	0.9374	-0.04201	0.13874	0.9462	-0.01412	0.06065	0.9464
	$\tau_1$	0.03524	0.72409	0.9428	0.00279	0.41689	0.9496	-0.00590	0.16126	0.9524
0.5	$\beta_0$	-0.00034	0.01362	0.9308	0.00017	0.00773	0.9420	0.00100	0.00334	0.9476
	$\beta_1$	-0.00056	0.02825	0.9296	-0.00122	0.01328	0.9426	-0.00125	0.00654	0.9446
	$\tau_0$	-0.08863	0.37166	0.9340	-0.03873	0.15450	0.9360	-0.02320	0.07412	0.9516
	$\tau_1$	0.01687	1.01367	0.9300	0.00324	0.44647	0.9404	0.00502	0.21397	0.9514
0.75	$\beta_0$	-0.01242	0.01326	0.9346	-0.00528	0.00779	0.9412	-0.00219	0.00388	0.9408
	$\beta_1$	0.00672	0.02939	0.9320	0.00133	0.01348	0.9406	-0.00126	0.00669	0.9422
	$\tau_0$	-0.09278	0.32558	0.9374	-0.03276	0.13614	0.9410	-0.01992	0.05917	0.9468
	$\tau_1$	0.03374	0.91462	0.9380	-0.00390	0.33330	0.9426	-0.00047	0.16003	0.9482

Na Tabela 2.10 são apresentadas os resultados de simulação para os critérios de informação AIC, BIC e AICc. Essa tabela reporta as taxas de acerto segundo esses critérios, ou seja, a porcentagem de vezes em que acertadamente os critérios escolheram o modelo correto (distribuição correta). Dessa tabela, observa-se que as taxas de acerto tendam a aumentar com o aumento do tamanho da amostra  $n$ , como esperado, e que essas taxas são maiores para os modelos log-NO e EBS.

**Tabela 2.10:** Taxas de acertos segundo os critérios de informação para os modelos de regressão log-simétricos quantílicos.

$q$	Distribuição	$n = 50$			$n = 100$			$n = 200$		
		AIC	BIC	AICc	AIC	BIC	AICc	AIC	BIC	AICc
0.25	Log-NO	0.7058	0.7058	0.7058	0.7108	0.7108	0.7108	0.7320	0.7320	0.7320
	Log- $t(3)$	0.3464	0.3464	0.3464	0.5226	0.5226	0.5226	0.6484	0.6484	0.6484
	Log-PE(0.3)	0.2526	0.2526	0.2526	0.3402	0.3402	0.3402	0.4550	0.4550	0.4550
	Log-HP(2)	0.0630	0.0630	0.0630	0.1200	0.1200	0.1200	0.2272	0.2272	0.2272
	Log-SL(4)	0.0348	0.0348	0.0348	0.0934	0.0934	0.0934	0.2024	0.2024	0.2024
	Log-CN(0.1,0.2)	0.2300	0.2300	0.2300	0.3542	0.3542	0.3542	0.5100	0.5100	0.5100
	EBS(0.5)	0.7386	0.7386	0.7386	0.7206	0.7206	0.7206	0.7230	0.7230	0.7230
	EBS- $t(0.5, 3)$	0.2068	0.2068	0.2068	0.3242	0.3242	0.3242	0.4790	0.4790	0.4790
0.5	Log-NO	0.7208	0.7208	0.7208	0.7242	0.7242	0.7242	0.7366	0.7366	0.7366
	Log- $t(3)$	0.4008	0.4008	0.4008	0.5112	0.5112	0.5112	0.6400	0.6400	0.6400
	Log-PE(0.3)	0.2474	0.2474	0.2474	0.3514	0.3514	0.3514	0.4430	0.4430	0.4430
	Log-HP(2)	0.0618	0.0618	0.0618	0.1206	0.1206	0.1206	0.2210	0.2210	0.2210
	Log-SL(4)	0.0308	0.0308	0.0308	0.0928	0.0928	0.0928	0.2038	0.2038	0.2038
	Log-CN(0.1,0.2)	0.2280	0.2280	0.2280	0.3562	0.3562	0.3562	0.5084	0.5084	0.5084
	EBS(0.5)	0.7462	0.7462	0.7462	0.7298	0.7298	0.7298	0.7280	0.7280	0.7280
	EBS- $t(0.5, 3)$	0.1990	0.1990	0.1990	0.3248	0.3248	0.3248	0.4774	0.4774	0.4774
0.75	Log-NO	0.7096	0.7096	0.7096	0.7122	0.7122	0.7122	0.7306	0.7306	0.7306
	Log- $t(3)$	0.4176	0.4176	0.4176	0.5212	0.5212	0.5212	0.6452	0.6452	0.6452
	Log-PE(0.3)	0.2426	0.2426	0.2426	0.3538	0.3538	0.3538	0.4486	0.4486	0.4486
	Log-HP(2)	0.0648	0.0648	0.0648	0.1252	0.1252	0.1252	0.2256	0.2256	0.2256
	Log-SL(4)	0.0356	0.0356	0.0356	0.0938	0.0938	0.0938	0.2044	0.2044	0.2044
	Log-CN(0.1,0.2)	0.2310	0.2310	0.2310	0.3574	0.3574	0.3574	0.5050	0.5050	0.5050
	EBS(0.5)	0.7376	0.7376	0.7376	0.7294	0.7294	0.7294	0.7266	0.7266	0.7266
	EBS- $t(0.5, 3)$	0.2050	0.2050	0.2050	0.3318	0.3318	0.3318	0.4750	0.4750	0.4750

Na simulação dos resíduos, buscamos avaliar o comportamento dos resíduos GCS e RQ do modelo de regressão quantílica. Mais precisamente, o objetivo aqui é verificar se os resíduos GCS e RQ se comportam de acordo com suas distribuições teóricas, sob os valores de  $q = 0.25, 0.5, 0.75$ , bem como se há mudanças quando variamos os tamanhos amostrais  $n = 50, 100, 200$ . Vale ressaltar também que os valores verdadeiros dos parâmetros do modelo foram mantidos.

As Tabelas 2.11-2.12 apresentam os resultados da simulação para os resíduos GCS e RQ, respectivamente. Nesse sentido, as tabelas apresentam a média, mediana, SD, CS e CK, cujos valores se espera que sejam 1, 0.69, 1, 2 e 6, respectivamente, para o resíduo GCS, e 0, 0, 1, 0 e 0, respectivamente, para o resíduo RQ. Pela Tabela 2.11 podemos observar que as médias, medianas e SDs ficam próximas a 1, 0.69, 1, respectivamente, ou seja, do valores referência de uma distribuição exponencial padrão, para todos os cenários considerados. Por outro lado, nota-se que os valores de CS e CK tendem, me geral, para seus valores de referência a medida que a amostra aumenta, sendo que os valores do CK tendem a ficar um pouco abaixo do valor de referência. Na Tabela 2.12 os valores da média, mediana e do SD ficam bem próximos de 0,

0, 1, ou seja, dos valores teóricos da distribuição normal padrão. Nota-se que a medida que o tamanho da amostra cresce, os valores de CS e CK tendem, em geral, a ficar mais próximo de 0, como esperado.

**Tabela 2.11:** Resultados para os resíduos GCS considerando diferentes modelos de regressão log-simétricos quantílicos.

		$n = 50$							
$q$	Estatística	Log-NO	Log- $t$	Log-PE	Log-HP	Log-SL	Log-CN	EBS	EBS- $t$
0.25	Média	0.99936	0.99404	0.99644	0.99611	0.99758	1.07521	1.00123	0.99435
	Mediana	0.70226	0.69642	0.69922	0.69828	0.70149	0.63628	0.70325	0.69687
	SD	0.98180	0.98300	0.98114	0.98371	0.97898	1.26828	0.97875	0.98342
	CS	1.52807	1.56819	1.53842	1.55628	1.51692	1.40280	1.48478	1.56940
	CK	2.55102	2.75474	2.58906	2.68969	2.48426	1.70423	2.31197	2.76390
0.5	Média	1.00067	0.99685	0.99964	0.99897	0.99900	1.05825	1.00188	0.99961
	Mediana	0.69720	0.69518	0.69591	0.69761	0.69643	0.59712	0.69827	0.69473
	SD	0.99083	0.98565	0.99004	0.98957	0.98663	1.30431	0.99038	0.99158
	CS	1.55564	1.55944	1.55711	1.56082	1.54231	1.83512	1.53950	1.57030
	CK	2.65167	2.69073	2.66652	2.68922	2.59905	3.94936	2.57457	2.73967
0.75	Média	1.00208	1.00216	1.00253	1.00273	1.00069	0.67645	1.00189	1.00243
	Mediana	0.69355	0.69318	0.69286	0.69338	0.69325	0.24333	0.69429	0.69370
	SD	1.00088	0.99531	0.99963	1.00020	0.99712	1.03971	0.99888	0.99773
	CS	1.59439	1.57219	1.58905	1.59299	1.58222	2.08323	1.58600	1.58787
	CK	2.82522	2.71662	2.79731	2.84244	2.77479	4.70738	2.82211	2.81158
		$n = 100$							
0.25	Média	0.99987	0.99697	0.99876	0.99849	0.99910	1.28750	1.00067	0.99611
	Mediana	0.69714	0.69533	0.69572	0.69557	0.69651	0.78760	0.69647	0.69531
	SD	0.99084	0.99236	0.99127	0.99113	0.99066	1.45686	0.98949	0.98831
	CS	1.72120	1.75991	1.73983	1.73707	1.73105	1.57111	1.69862	1.74351
	CK	3.73464	4.00959	3.85506	3.84957	3.83319	2.76641	3.60417	3.90985
0.5	Média	1.00059	0.99882	1.00014	0.99915	0.99954	1.03951	1.00105	0.99884
	Mediana	0.69532	0.69495	0.69383	0.69476	0.69447	0.52992	0.69412	0.69401
	SD	0.99504	0.99385	0.99500	0.99332	0.99527	1.32493	0.99442	0.99269
	CS	1.73446	1.75317	1.74640	1.74562	1.74992	1.84787	1.72423	1.74386
	CK	3.79463	3.96013	3.88064	3.91642	3.92758	3.90925	3.75120	3.87944
0.75	Média	1.00128	1.00145	1.00136	1.00056	1.00074	0.81053	1.00109	1.00069
	Mediana	0.69339	0.69407	0.69192	0.69296	0.69286	0.32290	0.69245	0.69363
	SD	1.00080	0.99851	1.00049	0.99889	0.99978	1.18421	1.00061	0.99672
	CS	1.76260	1.75754	1.77238	1.76733	1.75966	2.24027	1.76164	1.74963
	CK	3.95368	3.95836	4.03496	4.03327	3.95530	6.07540	3.96037	3.90162
		$n = 200$							
0.25	Média	1.00018	0.99908	0.99960	0.99931	0.9992	1.28156	1.00047	0.99891
	Mediana	0.69487	0.69412	0.69428	0.69466	0.69592	0.79174	0.69639	0.69376
	SD	0.99608	0.99675	0.99596	0.99618	0.99468	1.42813	0.99300	0.99496
	CS	1.84419	1.86845	1.85341	1.86476	1.85288	1.57580	1.82405	1.85192
	CK	4.61888	4.83472	4.69717	4.83238	4.73117	2.73989	4.49670	4.71112
0.5	Média	1.00044	0.99994	1.00031	0.99955	1.00009	1.02630	1.00051	0.99887
	Mediana	0.69378	0.69361	0.69327	0.69510	0.69403	0.52399	0.69500	0.69414
	SD	0.99827	0.99760	0.99773	0.99684	0.99801	1.29699	0.99688	0.99518
	CS	1.85139	1.86526	1.85437	1.86244	1.85947	1.87475	1.84482	1.85365
	CK	4.65578	4.80970	4.68798	4.78175	4.75484	4.06773	4.61053	4.71592
0.75	Media	1.00060	1.00113	1.00091	1.00036	1.00060	0.79130	1.00037	1.00022
	Mediana	0.69238	0.69302	0.69282	0.69426	0.69341	0.30988	0.69382	0.69362
	SD	1.00109	0.99994	1.00088	0.99948	1.00057	1.15141	0.99890	0.99836
	CS	1.86760	1.86810	1.87219	1.87235	1.87140	2.27928	1.85969	1.86558
	CK	4.76094	4.81148	4.81338	4.83858	4.83023	6.37006	4.72970	4.79141

**Tabela 2.12:** Resultados para os resíduos RQ considerando diferentes modelos de regressão log-simétricos quantílicos.

		$n = 50$							
$q$	Estatística	Log-NO	Log- $t$	Log-PE	Log-HP	Log-SL	Log-CN	EBS	EBS- $t$
0.25	Média	-0.00038	-0.00448	-0.00285	-0.00268	-0.00067	0.18892	0.00080	-0.00444
	Mediana	0.00653	0.00143	0.00479	0.00311	0.00579	0.21746	0.00711	0.00186
	SD	1.01013	1.00519	1.00885	1.00740	1.00780	1.16427	1.01208	1.00652
	CS	-0.02458	-0.01694	-0.02598	-0.01574	-0.02771	-0.08877	-0.03768	-0.01799
	CK	-0.31632	-0.29297	-0.29107	-0.29093	-0.30447	-0.58239	-0.33039	-0.28323
0.5	Média	-0.00014	-0.00099	-0.00031	0.00008	-0.00011	-0.00200	0.00020	0.00108
	Mediana	0.00019	-0.00042	0.00067	0.00219	-0.00048	-0.00075	0.00098	0.00102
	SD	1.00998	1.00367	1.00861	1.00662	1.00756	1.23045	1.01191	1.00508
	CS	0.00680	-0.00422	0.00182	0.00465	-0.00058	0.00357	-0.00134	0.00789
	CK	-0.33186	-0.30871	-0.30544	-0.30779	-0.32195	-0.20582	-0.32464	-0.30257
0.75	Média	0.00035	0.00339	0.00197	0.00309	0.00064	-0.17968	-0.00062	0.00345
	Mediana	-0.00439	-0.00268	-0.00306	-0.00301	-0.00457	-0.20295	-0.00423	-0.00217
	SD	1.01012	1.00532	1.00874	1.00713	1.00779	1.17251	1.01181	1.00606
	CS	0.03333	0.00835	0.02420	0.02689	0.02616	0.08275	0.02336	0.01545
	CK	-0.31318	-0.29367	-0.29421	-0.29202	-0.30382	-0.50849	-0.30819	-0.28281
		$n = 100$							
0.25	Média	-0.00005	-0.00250	-0.00087	-0.00064	-0.00012	0.23444	0.00024	-0.00291
	Mediana	0.00262	0.00147	0.00193	0.00145	0.00210	0.26566	0.00159	0.00141
	SD	1.00501	1.00265	1.00447	1.00323	1.00392	1.25715	1.00610	1.00284
	CS	-0.01025	-0.00522	-0.01251	-0.00663	-0.01102	-0.10634	-0.01571	-0.01513
	CK	-0.16536	-0.15415	-0.14267	-0.15708	-0.15116	-0.39488	-0.17582	-0.14859
0.5	Média	0.00016	-0.00017	0.00021	-0.00024	0.00003	-0.00104	0.00012	-0.00021
	Mediana	0.00031	0.00092	-0.00045	0.00027	-0.00054	-0.00022	-0.00142	-0.00030
	SD	1.00498	1.00170	1.00435	1.00344	1.00357	1.26945	1.00602	1.00241
	CS	0.00452	0.00124	0.00027	-0.00037	0.00317	-0.00300	0.00026	-0.00107
	CK	-0.17381	-0.16176	-0.15378	-0.15974	-0.14998	-0.42956	-0.17600	-0.16302
0.75	Média	0.00031	0.00199	0.00106	0.00076	0.00056	-0.23798	-0.00049	0.00134
	Mediana	-0.00207	-0.00015	-0.00283	-0.00195	-0.00257	-0.26895	-0.00353	-0.00074
	SD	1.00510	1.00256	1.00443	1.00362	1.00398	1.26510	1.00597	1.00284
	CS	0.01961	0.00747	0.01308	0.01236	0.01712	0.10668	0.01872	0.00490
	CK	-0.16354	-0.15548	-0.14104	-0.14852	-0.15370	-0.36732	-0.16617	-0.15229
		$n = 200$							
0.25	Média	0.00016	-0.00058	-0.00024	-0.00035	-0.00030	0.23705	0.00052	-0.00103
	Mediana	0.00098	0.00060	0.00079	0.00107	0.00243	0.26940	0.00278	0.00103
	SD	1.00243	1.00123	1.00221	1.00190	1.00176	1.24485	1.00297	1.00141
	CS	-0.00194	-0.00108	-0.00487	-0.00403	-0.00663	-0.11239	-0.01314	-0.00346
	CK	-0.08664	-0.07952	-0.07249	-0.07082	-0.07804	-0.40279	-0.08985	-0.08044
0.5	Média	0.00009	0.00045	0.00032	-0.00013	0.00016	-0.00058	0.00006	-0.00057
	Mediana	-0.00038	-0.00006	-0.00049	0.00168	0.00003	0.00009	0.00104	0.00008
	SD	1.00249	1.00087	1.00216	1.00159	1.00183	1.25481	1.00301	1.00127
	CS	0.00605	0.00235	0.00152	0.00091	0.00558	-0.00009	-0.00049	-0.00437
	CK	-0.09175	-0.08312	-0.07903	-0.08080	-0.08750	-0.43546	-0.08946	-0.08011
0.75	Média	-0.00001	0.00143	0.00070	0.00048	0.00044	-0.23952	-0.00027	0.00051
	Mediana	-0.00214	-0.00077	-0.00102	0.00063	-0.00074	-0.27161	-0.00046	-0.00004
	SD	1.00249	1.00121	1.00221	1.00173	1.00194	1.24974	1.00294	1.00142
	CS	0.01398	0.00592	0.00897	0.00642	0.01168	0.11432	0.00587	0.00316
	CK	-0.08627	-0.07922	-0.07190	-0.07527	-0.08225	-0.38922	-0.08485	-0.07795

## 2.4.2 Estudo de Monte Carlo 2: tamanho e poder dos testes

No segundo estudo de simulação de Monte Carlo foi realizada uma análise do desempenho dos testes de hipóteses  $S_W$ ,  $S_{LR}$ ,  $S_R$  e  $S_T$  para os modelos de regressão log-simétricos quan-

tílicos, sendo que estes testes são analisados sob duas perspectivas: tamanho (taxa de rejeição nula) e poder (taxa de rejeição não-nula). Os níveis de significância nominais utilizados foram  $\alpha = 0.01, 0.05, 0.1$ , os quantis e tamanhos amostrais utilizados foram os mesmos do estudo de simulação anterior. O modelo utilizado para geração das amostras é dado por

$$Y_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \epsilon_i^{\sqrt{\phi}}, \quad 1, \dots, n,$$

em que  $\epsilon_i \sim \text{QLS}(1, 1, g)$ , as variáveis independentes  $x_i$  são simuladas de acordo com uma distribuição Bernoulli(0.5), os coeficientes não fixados sob  $H_0$  são  $\beta_l = 1, \forall \beta_l \neq 0$  e  $\phi = 3$ , isto é, os coeficientes  $\beta_l$  não fixados sob  $H_0$  são iguais a 1. Um processo descritivo para obtenção das estimativas de Monte Carlo para o tamanho/poder do teste segue de acordo com o Algoritmo 2.

---

**Algoritmo 2** Etapas para obtenção do tamanho e poder do teste.

---

- 1: Fixar o número de parâmetros  $t$  sob  $H_0$  e o nível de significância nominal  $\alpha$ .
  - 2: Fixar os parâmetros e gerar 5000 amostras de tamanho  $n$  baseada na  $g(\cdot)$  escolhida da Tabela 2.1, de acordo com o modelo postulado em  $H_0$  para o tamanho do teste, ou de acordo com o modelo postulado em  $H_1$  para o poder do teste.
  - 3: Estimar os parâmetros do modelo pelo método da máxima verossimilhança para cada amostra.
  - 4: Calcular as estatísticas de teste  $S_W, S_{LR}, S_R$  e  $S_T$  para cada amostra.
  - 5: Calcular o valor crítico  $C_{t,1-\alpha}$  com base no valor de  $\alpha$  e no número de parâmetros  $t$  fixado sob  $H_0$ .
  - 6: Obter o tamanho/poder do teste calculando a proporção de réplicas da estatística de teste ( $S_W, S_{LR}, S_R$  e  $S_T$ ) maior que o valor crítico  $C_{t,1-\alpha}$ .
- 

Nas Tabelas 2.13-2.20 são apresentadas as taxas de rejeição nula para diferentes modelos de regressão log-simétricos quantílicos. O interesse aqui é testar  $H_0 : \beta_1 = \dots = \beta_t = 0$ , para  $t = 1, 3$ , ou seja, a proporção de vezes que cada estatística ( $S_W, S_{LR}, S_R$  e  $S_T$ ) é maior que o quantil  $1 - \alpha$  da distribuição de referência  $\chi_t^2$ . A escolha de analisar o comportamento do teste fixando diferentes quantidades de parâmetros foi com o objetivo de observar se existem diferenças na desempenho do teste sob esses cenários. De fato, podemos notar uma diferença nas taxas de rejeição nula dos testes. Por exemplo, para  $t = 1$ , as taxas de rejeição nula se aproximam mais dos níveis nominais adotados se comparados quando testamos parâmetros simultaneamente, com  $t = 3$ . Vale ressaltar também que a medida que aumentamos o tamanho da amostra, as taxas de rejeição nula se aproximam dos correspondentes níveis nominais. Em

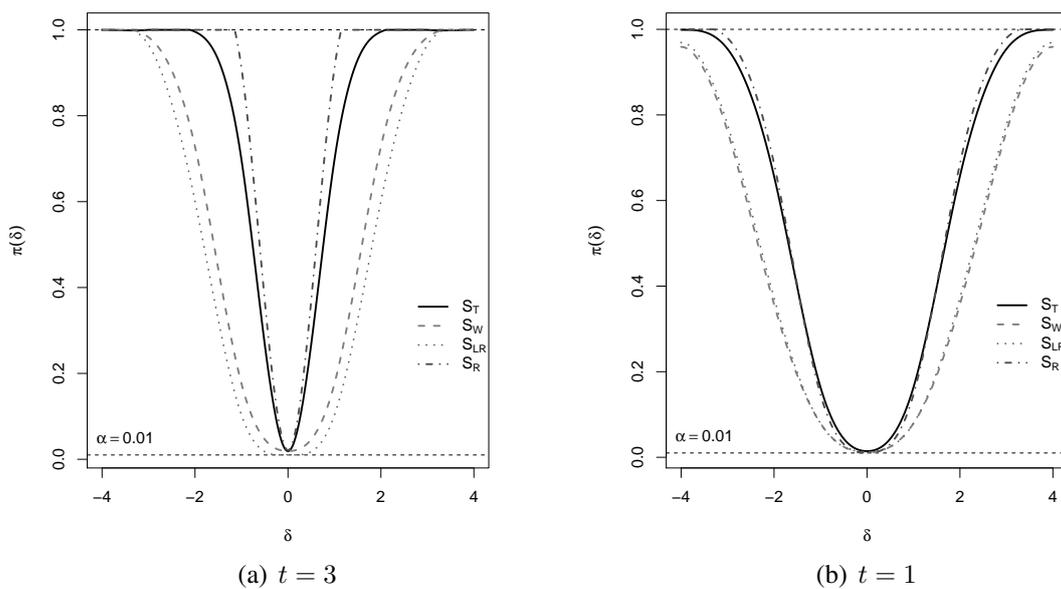
geral, as estatísticas  $S_W$  e  $S_{LR}$ , em relação às demais estatísticas, apresentam taxas de rejeição mais próximas do nível nominal fixado a medida que aumentamos a amostra.

As Figuras 2.1-2.8 apresentam as curvas de poder correspondentes aos quatro testes. Foram computados as taxas de rejeição não-nula considerando a hipótese  $H_0 : \beta_1 = \dots = \beta_t$  contra  $H_1 : \beta_1 = \dots = \beta_t = \delta, |\delta| = 0, 1, \dots, 4$ , em que a função poder é denotada por  $\pi(\delta)$ . Ressalta-se que o poder foi calculado considerando o tamanho amostral de  $n = 100$  e  $\alpha = 0.01$ , pois resultados similares foram observados para outras configurações. Com exceção dos modelos com distribuições EBS ( $\vartheta = 0.5$ ) e EBS- $t$  ( $\vartheta_1 = 0.5, \vartheta_2 = 3$ ), o poder do teste para testar  $H_0$  com  $t = 3$  parâmetros, se comparado com  $t = 1$ , é mais elevado, o que indica que os testes estudados são mais eficientes quando testamos múltiplos parâmetros. No geral, vemos que o poder do teste é maior para a estatística  $S_R$ , seguida pela estatística  $S_T$ , para testar  $t = 3$  parâmetros simultaneamente. Para  $t = 1$ , o poder do teste usando as estatísticas  $S_R$  e  $S_T$  praticamente coincidem. Para o modelo de regressão quantílica com distribuições EBS ( $\vartheta = 0.5$ ) e EBS- $t$  ( $\vartheta_1 = 0.5, \vartheta_2 = 3$ ), as curvas de poder dos quatro testes se sobrepuseram, exceção para a curva de poder do teste Wald (um pouco acima das demais) com  $t = 1$  para o modelo EBS- $t$  ( $\vartheta_1 = 0.5, \vartheta_2 = 3$ ). Em geral, todos os poderes dos testes se aproxima de 1 conforme  $|\delta|$  aumenta, como esperado.

**Tabela 2.13:** Tamanho dos testes para o modelo de regressão log-NO quantílico, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

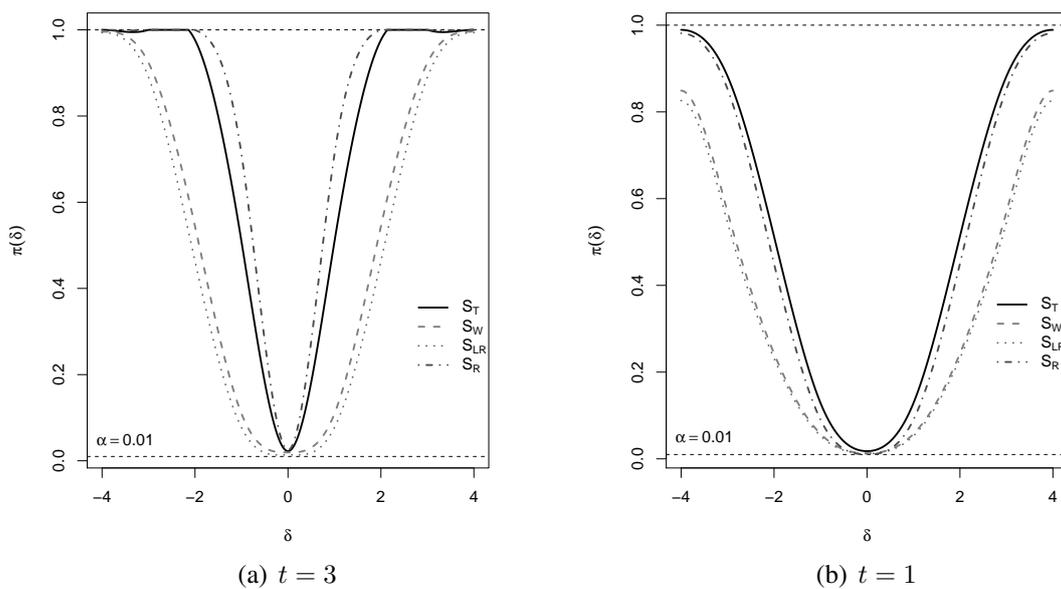
$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
3	0.25	$S_W$	0.0276	0.0864	0.1272	0.0136	0.0586	0.1078	0.0150	0.0588	0.1056
		$S_{LR}$	0.0132	0.0644	0.1162	0.0118	0.0538	0.1134	0.0068	0.0530	0.1046
		$S_R$	0.0238	0.0716	0.1172	0.0244	0.0698	0.1158	0.0240	0.0722	0.1204
		$S_T$	0.0186	0.0704	0.1562	0.0210	0.0674	0.1224	0.0174	0.0712	0.1186
3	0.5	$S_W$	0.0294	0.0848	0.1438	0.0146	0.0596	0.1100	0.0154	0.0594	0.1090
		$S_{LR}$	0.0170	0.0732	0.1124	0.0130	0.0538	0.1114	0.0110	0.0504	0.1064
		$S_R$	0.0248	0.0716	0.1172	0.0244	0.0698	0.1158	0.0240	0.0722	0.1204
		$S_T$	0.0186	0.0680	0.1276	0.0210	0.0674	0.1224	0.0174	0.0712	0.1186
3	0.75	$S_W$	0.0296	0.0836	0.1412	0.0140	0.0590	0.1098	0.0158	0.0584	0.1074
		$S_{LR}$	0.0146	0.0668	0.1184	0.0130	0.0548	0.1148	0.0116	0.0544	0.1040
		$S_R$	0.0248	0.0716	0.1172	0.0244	0.0698	0.1158	0.0240	0.0722	0.1162
		$S_T$	0.0186	0.0704	0.1276	0.0212	0.0674	0.1224	0.0174	0.0694	0.1186
1	0.25	$S_W$	0.0172	0.0818	0.1390	0.0136	0.0588	0.1154	0.0104	0.0554	0.1026
		$S_{LR}$	0.0420	0.1296	0.1908	0.0254	0.0906	0.1494	0.0160	0.0482	0.0788
		$S_R$	0.0102	0.0470	0.1024	0.0102	0.0506	0.0954	0.0104	0.0486	0.0980
		$S_T$	0.0140	0.0722	0.1216	0.0196	0.0650	0.1200	0.0178	0.0670	0.1280
1	0.5	$S_W$	0.0226	0.0824	0.1382	0.0138	0.0606	0.1158	0.0098	0.0552	0.1048
		$S_{LR}$	0.0406	0.1290	0.1984	0.0264	0.1086	0.1344	0.0140	0.0516	0.0944
		$S_R$	0.0080	0.0470	0.1024	0.0102	0.0506	0.0954	0.0104	0.0486	0.0980
		$S_T$	0.0190	0.0722	0.1216	0.0196	0.0650	0.1200	0.0170	0.0670	0.1282
1	0.75	$S_W$	0.0230	0.0830	0.1398	0.0138	0.0606	0.1158	0.0100	0.0556	0.1052
		$S_{LR}$	0.0388	0.1308	0.1944	0.0250	0.0894	0.1482	0.0156	0.0448	0.0766
		$S_R$	0.0080	0.0470	0.1024	0.0102	0.0506	0.0954	0.0104	0.0486	0.0980
		$S_T$	0.0206	0.0722	0.1216	0.0196	0.0650	0.1200	0.0170	0.0670	0.1282

**Figura 2.1:** Curva de poder dos testes para o modelo de regressão log-NO quantílico.



**Tabela 2.14:** Tamanho dos testes para o modelo de regressão log- $t$  ( $\vartheta = 3$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

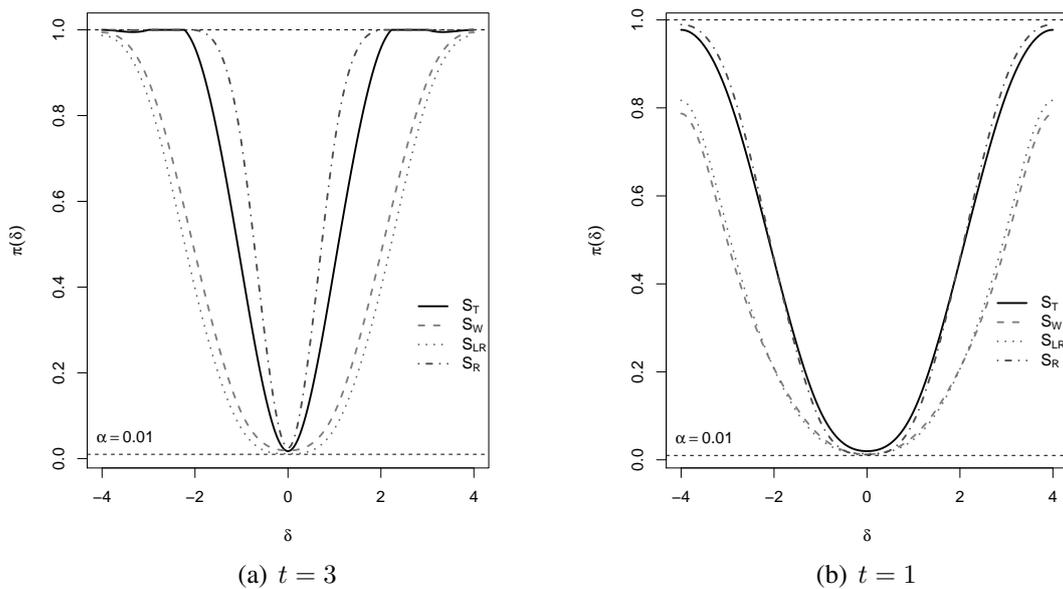
$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
3	0.25	$S_W$	0.0270	0.0920	0.1486	0.0160	0.0696	0.1186	0.0152	0.0562	0.1098
		$S_{LR}$	0.0162	0.0644	0.1280	0.0114	0.0544	0.1108	0.0122	0.0534	0.1082
		$S_R$	0.0244	0.0714	0.1178	0.0226	0.0654	0.1078	0.0232	0.0664	0.1132
		$S_T$	0.0204	0.0782	0.1422	0.0188	0.0718	0.1236	0.0192	0.0696	0.1242
	0.5	$S_W$	0.0274	0.0920	0.1486	0.0156	0.0696	0.1186	0.0150	0.0562	0.1098
		$S_{LR}$	0.0146	0.0728	0.1248	0.0134	0.0584	0.1132	0.0124	0.0506	0.1092
		$S_R$	0.0244	0.0714	0.1178	0.0232	0.0654	0.1078	0.0232	0.0664	0.1132
		$S_T$	0.0214	0.0782	0.1300	0.0188	0.0718	0.1236	0.0192	0.0696	0.1242
	0.75	$S_W$	0.0290	0.0912	0.1494	0.0148	0.0676	0.1182	0.0152	0.0582	0.1094
		$S_{LR}$	0.0138	0.0754	0.1212	0.0122	0.0562	0.1110	0.0096	0.0570	0.1014
		$S_R$	0.0244	0.0714	0.1178	0.0226	0.0654	0.1078	0.0232	0.0664	0.1132
		$S_T$	0.0214	0.0782	0.1424	0.0188	0.0718	0.1236	0.0192	0.0696	0.1242
1	0.25	$S_W$	0.0176	0.0658	0.1230	0.0158	0.0702	0.1216	0.0154	0.0652	0.1130
		$S_{LR}$	0.0456	0.1486	0.2258	0.0318	0.1172	0.1826	0.0242	0.0776	0.1232
		$S_R$	0.0088	0.0496	0.1006	0.0098	0.0452	0.0928	0.0098	0.0468	0.0968
		$S_T$	0.0232	0.0684	0.1208	0.0178	0.0690	0.1208	0.0184	0.0642	0.1192
	0.5	$S_W$	0.017	0.0700	0.1224	0.0142	0.0646	0.1216	0.0152	0.0692	0.1152
		$S_{LR}$	0.0432	0.1486	0.2258	0.0362	0.1172	0.1826	0.0214	0.0776	0.123
		$S_R$	0.0088	0.0496	0.1006	0.0098	0.0452	0.0928	0.0098	0.0468	0.0968
		$S_T$	0.0224	0.0752	0.1332	0.0178	0.0690	0.1208	0.0184	0.0642	0.1192
	0.75	$S_W$	0.0178	0.0702	0.1242	0.0152	0.0638	0.1250	0.0154	0.0684	0.1158
		$S_{LR}$	0.0566	0.1486	0.2258	0.0348	0.1172	0.1826	0.0232	0.0776	0.1230
		$S_R$	0.0088	0.0496	0.1006	0.0098	0.0452	0.0928	0.0098	0.0468	0.0968
		$S_T$	0.0224	0.0752	0.1332	0.0178	0.0690	0.1208	0.0184	0.0640	0.1192

**Figura 2.2:** Curva de poder dos testes para o modelo de regressão log- $t$  ( $\vartheta = 3$ ) quantílico.

**Tabela 2.15:** Tamanho dos testes para o modelo de regressão log-PE ( $\vartheta = 0.3$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

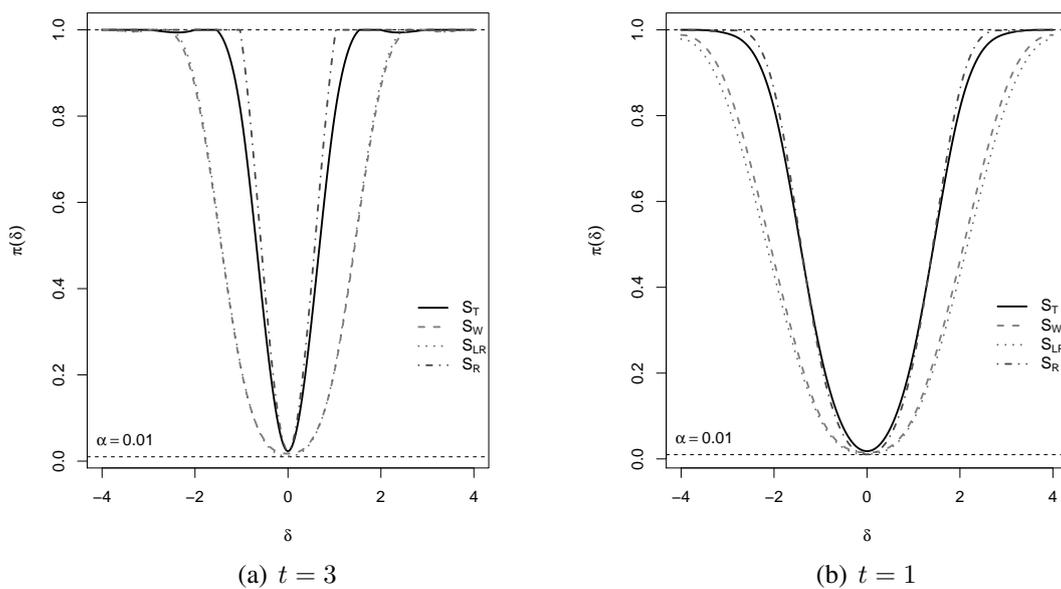
$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
	0.25	$S_W$	0.0284	0.0906	0.1508	0.0180	0.0684	0.1180	0.0148	0.0532	0.1066
		$S_{LR}$	0.0114	0.0630	0.1154	0.0142	0.0610	0.1158	0.0110	0.0544	0.1006
		$S_R$	0.0218	0.0644	0.1078	0.0276	0.0714	0.1134	0.0222	0.0646	0.1122
		$S_T$	0.0218	0.0772	0.1270	0.0218	0.0706	0.1246	0.0170	0.0628	0.1174
3	0.5	$S_W$	0.0276	0.0894	0.1548	0.0204	0.0688	0.1196	0.0158	0.0540	0.1060
		$S_{LR}$	0.0130	0.0610	0.1184	0.0130	0.0552	0.1138	0.0104	0.0564	0.1052
		$S_R$	0.0228	0.0644	0.1078	0.0276	0.0714	0.1134	0.0222	0.0646	0.1122
		$S_T$	0.0192	0.0704	0.1264	0.0218	0.0706	0.1244	0.0170	0.0630	0.1172
	0.75	$S_W$	0.0262	0.0900	0.1534	0.0198	0.0694	0.1286	0.0150	0.0538	0.1176
		$S_{LR}$	0.0144	0.0652	0.1210	0.0108	0.0542	0.1170	0.0108	0.0552	0.1048
		$S_R$	0.0228	0.0644	0.1078	0.0276	0.0714	0.1134	0.0222	0.0646	0.1122
		$S_T$	0.0192	0.0736	0.1264	0.0218	0.0706	0.1244	0.0170	0.0628	0.1168
	0.25	$S_W$	0.0166	0.0658	0.1196	0.0154	0.0536	0.1050	0.0108	0.0528	0.1042
		$S_{LR}$	0.0436	0.1580	0.2318	0.0338	0.1062	0.1716	0.0198	0.0720	0.1124
		$S_R$	0.0092	0.0508	0.0984	0.0092	0.0522	0.1014	0.0094	0.0468	0.0936
		$S_T$	0.0206	0.0784	0.1336	0.0154	0.0690	0.1262	0.0160	0.0584	0.1122
1	0.5	$S_W$	0.0204	0.0680	0.1236	0.0160	0.0564	0.1032	0.0100	0.0548	0.1054
		$S_{LR}$	0.0460	0.1580	0.2318	0.0338	0.1062	0.1716	0.0198	0.0720	0.1124
		$S_R$	0.0068	0.0508	0.0984	0.0092	0.0522	0.1014	0.0094	0.0468	0.0936
		$S_T$	0.0194	0.0784	0.1332	0.0184	0.0690	0.1262	0.0160	0.0584	0.1122
	0.75	$S_W$	0.0204	0.0674	0.1252	0.0162	0.0564	0.1056	0.0112	0.0556	0.1072
		$S_{LR}$	0.0506	0.1580	0.2318	0.0338	0.1062	0.1716	0.0198	0.0720	0.1124
		$S_R$	0.0068	0.0508	0.0984	0.0092	0.0522	0.1014	0.0094	0.0468	0.0936
		$S_T$	0.0206	0.0784	0.1332	0.0184	0.0690	0.1262	0.0160	0.0584	0.1122

**Figura 2.3:** Curva de poder dos testes para o modelo de regressão log-PE ( $\vartheta = 0.3$ ) quantílico.



**Tabela 2.16:** Tamanho dos testes para o modelo de regressão log-HP ( $\vartheta = 2$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

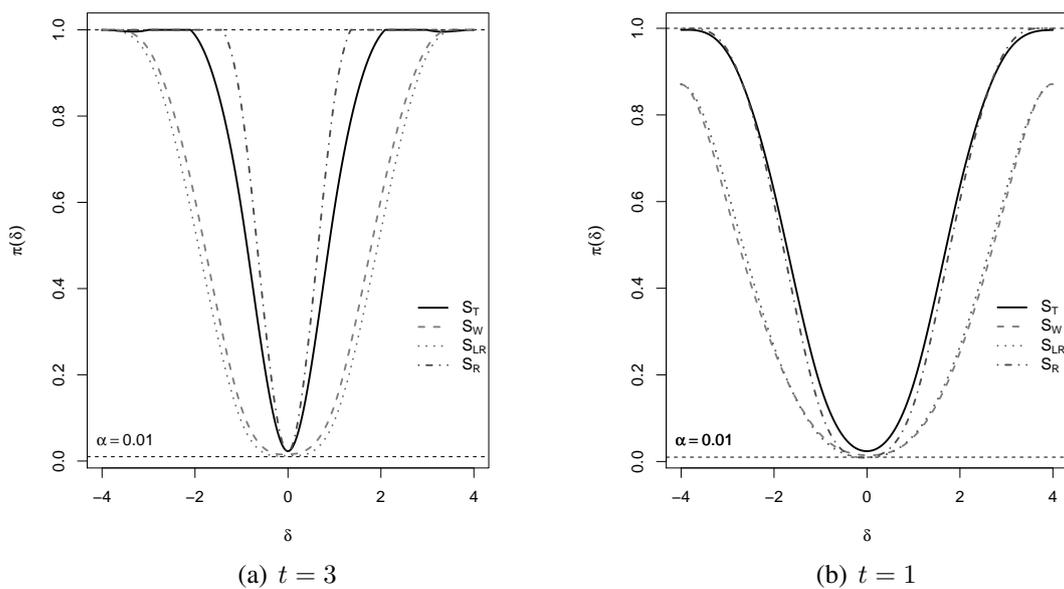
$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
3	0.25	$S_W$	0.0228	0.0784	0.1322	0.0138	0.0676	0.1218	0.0174	0.0602	0.1084
		$S_{LR}$	0.0130	0.0630	0.1122	0.0144	0.0596	0.1130	0.0106	0.0542	0.1028
		$S_R$	0.0280	0.0712	0.1176	0.0240	0.0698	0.1110	0.0252	0.0722	0.1124
		$S_T$	0.0196	0.0896	0.1432	0.0196	0.0688	0.1200	0.0208	0.0698	0.1294
	0.5	$S_W$	0.0264	0.0782	0.1318	0.0138	0.0676	0.1212	0.0168	0.0594	0.1100
		$S_{LR}$	0.0150	0.0630	0.1122	0.0144	0.0596	0.1130	0.0106	0.0542	0.1028
		$S_R$	0.0240	0.0698	0.1110	0.0252	0.0722	0.1124	0.0252	0.0722	0.1124
		$S_T$	0.0266	0.0896	0.1432	0.0196	0.0688	0.1200	0.0208	0.0698	0.1238
	0.75	$S_W$	0.0208	0.0828	0.1320	0.0190	0.0688	0.1254	0.0136	0.0586	0.1002
		$S_{LR}$	0.0150	0.0630	0.1122	0.0144	0.0596	0.1130	0.0106	0.0542	0.1028
		$S_R$	0.0214	0.0712	0.1176	0.0240	0.0698	0.1110	0.0252	0.0722	0.1124
		$S_T$	0.0196	0.0896	0.1432	0.0196	0.0688	0.1202	0.0194	0.0780	0.1238
1	0.25	$S_W$	0.0166	0.0616	0.1172	0.0142	0.0564	0.1096	0.0110	0.0490	0.0990
		$S_{LR}$	0.0372	0.1196	0.2064	0.0220	0.0782	0.1368	0.0104	0.0378	0.0664
		$S_R$	0.0114	0.0492	0.0998	0.0100	0.0536	0.1008	0.0108	0.0476	0.0970
		$S_T$	0.0156	0.0862	0.1424	0.0184	0.0674	0.1212	0.0176	0.0676	0.1320
	0.5	$S_W$	0.0170	0.0628	0.1178	0.0148	0.0576	0.1102	0.0104	0.0486	0.099
		$S_{LR}$	0.0356	0.1208	0.1902	0.0226	0.0848	0.1300	0.0140	0.0428	0.0704
		$S_R$	0.0106	0.0492	0.0998	0.0100	0.0536	0.1008	0.0108	0.0476	0.0970
		$S_T$	0.0238	0.0860	0.1424	0.0184	0.0674	0.1212	0.0176	0.0676	0.1320
	0.75	$S_W$	0.0168	0.0634	0.1176	0.0146	0.0580	0.1116	0.0112	0.0492	0.0994
		$S_{LR}$	0.0394	0.1018	0.1806	0.0276	0.0748	0.1260	0.0126	0.0372	0.0658
		$S_R$	0.0106	0.0492	0.0998	0.0100	0.0536	0.1008	0.0108	0.0476	0.0970
		$S_T$	0.0238	0.0862	0.1424	0.0184	0.0674	0.1212	0.0176	0.0676	0.1320

**Figura 2.4:** Curva de poder dos testes para o modelo de regressão log-HP ( $\vartheta = 2$ ) quantílico.

**Tabela 2.17:** Tamanho dos testes para o modelo de regressão log-SL ( $\vartheta = 4$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
3	0.25	$S_W$	0.0274	0.0764	0.1482	0.0166	0.0644	0.1224	0.0134	0.0554	0.1026
		$S_{LR}$	0.0136	0.0620	0.1276	0.0112	0.0590	0.1138	0.0122	0.0534	0.1094
		$S_R$	0.0196	0.0666	0.1134	0.0262	0.0702	0.1186	0.0240	0.0672	0.1142
		$S_T$	0.0226	0.0814	0.1396	0.0258	0.0750	0.1330	0.0218	0.0702	0.1282
	0.5	$S_W$	0.0260	0.0766	0.1414	0.0168	0.0638	0.1188	0.0130	0.0630	0.1146
		$S_{LR}$	0.0134	0.0622	0.1218	0.0096	0.0534	0.1082	0.0102	0.0496	0.0992
		$S_R$	0.0196	0.0666	0.1134	0.0262	0.0702	0.1186	0.0240	0.0672	0.1142
		$S_T$	0.0264	0.0814	0.1396	0.0258	0.0750	0.1330	0.0178	0.0702	0.1282
	0.75	$S_W$	0.0264	0.0822	0.1376	0.0174	0.0640	0.1180	0.0126	0.0520	0.1130
		$S_{LR}$	0.0154	0.0650	0.1192	0.0134	0.0550	0.1036	0.0090	0.0536	0.1056
		$S_R$	0.0196	0.0712	0.1134	0.0262	0.0702	0.1186	0.0240	0.0672	0.1142
		$S_T$	0.0264	0.0814	0.1396	0.0258	0.0750	0.1330	0.0218	0.0742	0.1282
1	0.25	$S_W$	0.0158	0.0660	0.1188	0.0108	0.0630	0.1122	0.0124	0.0502	0.0998
		$S_{LR}$	0.0478	0.1416	0.2150	0.0336	0.1198	0.1548	0.0198	0.0606	0.1112
		$S_R$	0.0096	0.0530	0.1026	0.0112	0.0526	0.1030	0.0090	0.0446	0.0966
		$S_T$	0.0240	0.0686	0.1288	0.0188	0.0722	0.1248	0.0156	0.0686	0.1216
	0.5	$S_W$	0.0162	0.0660	0.1218	0.0106	0.0608	0.1124	0.0116	0.0498	0.1008
		$S_{LR}$	0.0474	0.1376	0.1980	0.0338	0.1058	0.1644	0.0238	0.0658	0.1038
		$S_R$	0.0090	0.0530	0.1026	0.0112	0.0526	0.1030	0.0090	0.0446	0.0966
		$S_T$	0.0196	0.0716	0.1244	0.0188	0.0720	0.1248	0.0156	0.0686	0.1218
	0.75	$S_W$	0.0150	0.0682	0.1230	0.0110	0.0616	0.1132	0.0124	0.0500	0.0996
		$S_{LR}$	0.0396	0.1378	0.2114	0.0366	0.0904	0.1586	0.0186	0.0716	0.1000
		$S_R$	0.0090	0.0530	0.1026	0.0112	0.0526	0.1030	0.0090	0.0446	0.0966
		$S_T$	0.0196	0.0718	0.1242	0.0188	0.0720	0.1248	0.0156	0.0684	0.1218

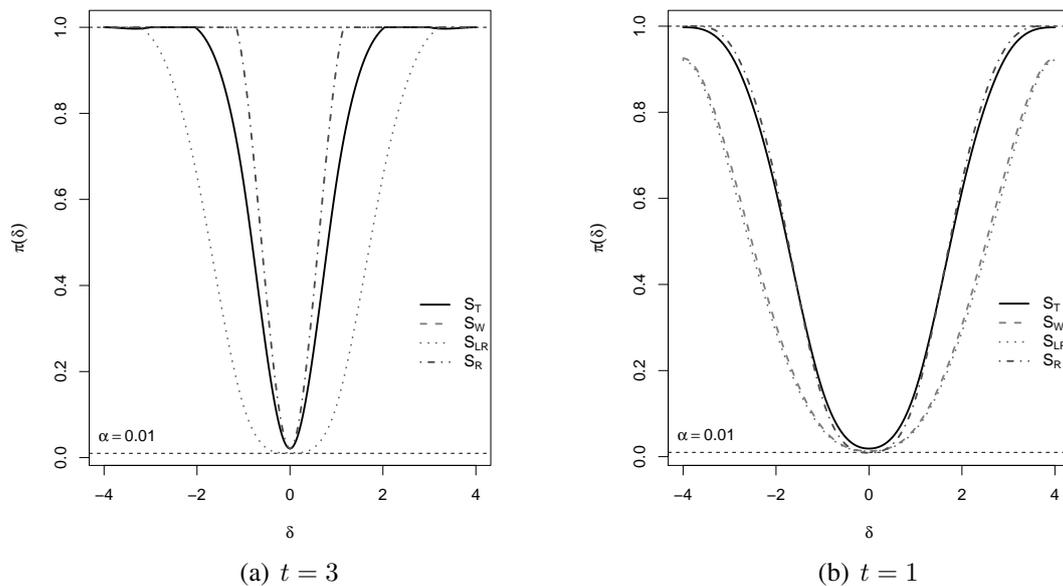
**Figura 2.5:** Curva de poder dos testes para o modelo de regressão log-SL ( $\vartheta = 4$ ) quantílico.



**Tabela 2.18:** Tamanho dos testes para o modelo de regressão log-CN ( $\vartheta_1 = 0.1, \vartheta_2 = 0.2$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

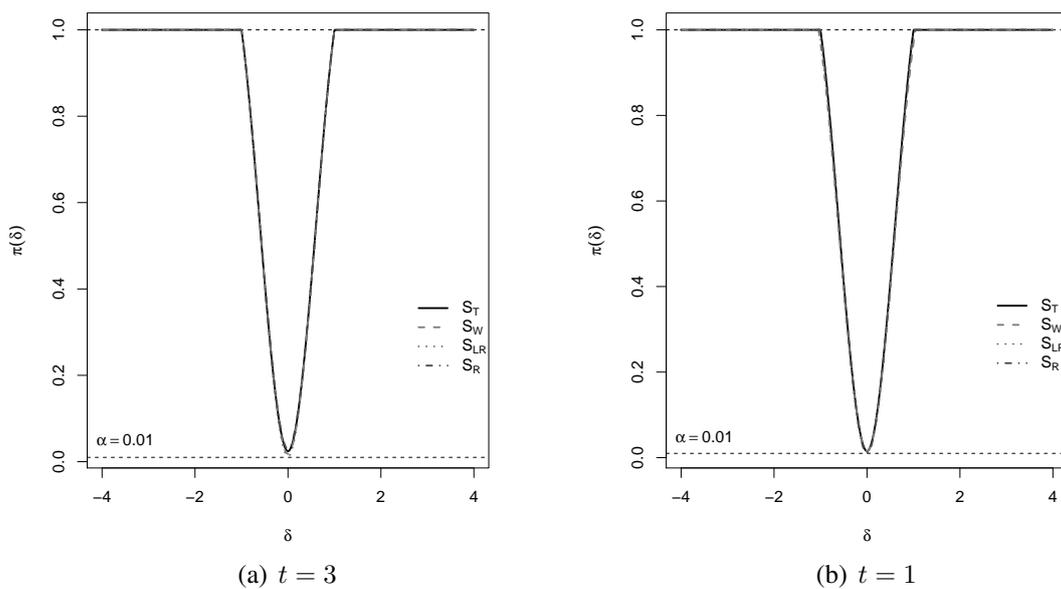
$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
	0.25	$S_W$	0.0312	0.0906	0.1432	0.0164	0.0674	0.1206	0.0136	0.0628	0.1120
		$S_{LR}$	0.0144	0.0706	0.1170	0.0122	0.0492	0.1032	0.0102	0.0488	0.1008
		$S_R$	0.0274	0.0724	0.1132	0.0218	0.0672	0.1154	0.0288	0.0802	0.1298
		$S_T$	0.0118	0.0830	0.1380	0.0162	0.0618	0.1178	0.0260	0.0766	0.1372
3	0.5	$S_W$	0.0264	0.0894	0.1436	0.0158	0.0682	0.1208	0.0134	0.0606	0.1118
		$S_{LR}$	0.0146	0.0646	0.1220	0.0112	0.0562	0.1134	0.0122	0.0546	0.0998
		$S_R$	0.0274	0.0724	0.1166	0.0218	0.0672	0.1154	0.0288	0.0802	0.1298
		$S_T$	0.0206	0.0830	0.1380	0.0162	0.0618	0.1178	0.0258	0.0766	0.1372
	0.75	$S_W$	0.0308	0.0894	0.1458	0.0160	0.0666	0.1200	0.0132	0.0622	0.1114
		$S_{LR}$	0.0146	0.0644	0.1222	0.0122	0.0550	0.1152	0.0128	0.0506	0.1076
		$S_R$	0.0274	0.0616	0.1166	0.0218	0.0672	0.1154	0.0288	0.0802	0.1298
		$S_T$	0.0240	0.0832	0.1378	0.0162	0.0618	0.1178	0.0258	0.0766	0.1372
	0.25	$S_W$	0.0146	0.0620	0.1166	0.0124	0.0610	0.1102	0.0102	0.0546	0.1038
		$S_{LR}$	0.0498	0.1298	0.1982	0.0240	0.0946	0.1480	0.0210	0.0592	0.0988
		$S_R$	0.0100	0.0496	0.1028	0.0090	0.0496	0.0988	0.0104	0.0572	0.1166
		$S_T$	0.0224	0.0748	0.1372	0.0154	0.0682	0.1208	0.0186	0.0756	0.1350
1	0.5	$S_W$	0.0166	0.0614	0.1156	0.0128	0.0618	0.1108	0.0108	0.0546	0.1054
		$S_{LR}$	0.0366	0.1298	0.1982	0.0240	0.0946	0.1480	0.0210	0.0592	0.0988
		$S_R$	0.0104	0.0496	0.1028	0.0090	0.0496	0.0988	0.0104	0.0572	0.1166
		$S_T$	0.0194	0.0748	0.1372	0.0154	0.0682	0.1210	0.0186	0.0756	0.1352
	0.75	$S_W$	0.0170	0.0632	0.1172	0.0132	0.0634	0.1106	0.0104	0.0550	0.1064
		$S_{LR}$	0.0366	0.1298	0.1982	0.0240	0.0946	0.1480	0.0210	0.0592	0.0988
		$S_R$	0.0104	0.0496	0.1028	0.0090	0.0496	0.0988	0.0104	0.0572	0.1166
		$S_T$	0.0190	0.0748	0.1372	0.0154	0.0682	0.1210	0.0186	0.0754	0.1352

**Figura 2.6:** Curva de poder dos testes para o modelo de regressão log-CN ( $\vartheta_1 = 0.1, \vartheta_2 = 0.2$ ) quantílico.



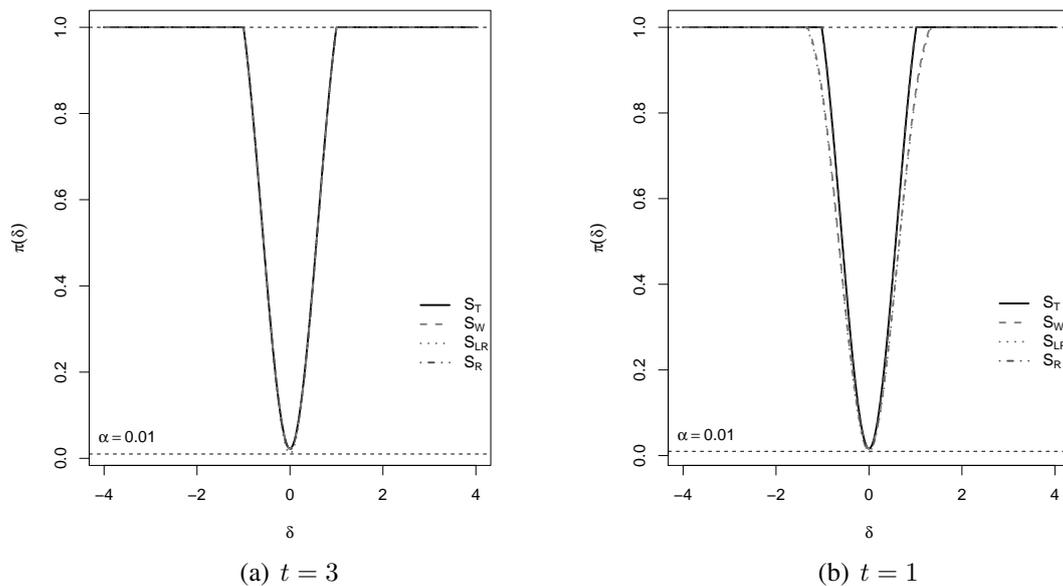
**Tabela 2.19:** Tamanho dos testes para o modelo de regressão EBS ( $\vartheta = 0.5$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
	0.25	$S_W$	0.0222	0.0756	0.1384	0.0150	0.0630	0.1180	0.0126	0.0576	0.1060
		$S_{LR}$	0.0118	0.0612	0.1194	0.0114	0.0576	0.1110	0.0110	0.0532	0.1012
		$S_R$	0.0272	0.0728	0.1208	0.0266	0.0720	0.1140	0.0252	0.0694	0.1202
		$S_T$	0.0192	0.0684	0.1246	0.0222	0.0696	0.1226	0.0190	0.0720	0.1270
3	0.5	$S_W$	0.0222	0.0756	0.1386	0.0150	0.0632	0.1176	0.0126	0.0576	0.106
		$S_{LR}$	0.0162	0.0612	0.1194	0.0114	0.0576	0.1110	0.0110	0.0532	0.1012
		$S_R$	0.0262	0.0728	0.1208	0.0266	0.0720	0.1140	0.0252	0.0694	0.1202
		$S_T$	0.0206	0.0684	0.1246	0.0222	0.0696	0.1226	0.0190	0.0720	0.1272
	0.75	$S_W$	0.0222	0.0756	0.1386	0.0150	0.0632	0.1178	0.0126	0.0576	0.1062
		$S_{LR}$	0.0142	0.0612	0.1194	0.0114	0.0576	0.1110	0.0110	0.0532	0.1012
		$S_R$	0.0262	0.0728	0.1208	0.0266	0.0720	0.1140	0.0252	0.0694	0.1202
		$S_T$	0.0206	0.0684	0.1246	0.0222	0.0696	0.1226	0.0190	0.0720	0.1272
	0.25	$S_W$	0.0242	0.0232	0.0538	0.0132	0.0404	0.0846	0.0068	0.0448	0.0926
		$S_{LR}$	0.0114	0.0602	0.1096	0.0092	0.0526	0.0970	0.0080	0.0520	0.0978
		$S_R$	0.0106	0.0474	0.0948	0.0118	0.0536	0.1014	0.0118	0.0522	0.0988
		$S_T$	0.0306	0.0740	0.1216	0.0256	0.0818	0.1386	0.0162	0.0718	0.1292
1	0.5	$S_W$	0.0046	0.0232	0.0536	0.0134	0.0644	0.0848	0.0036	0.0450	0.0926
		$S_{LR}$	0.0120	0.0604	0.1096	0.0112	0.0524	0.0980	0.0082	0.0516	0.0988
		$S_R$	0.0116	0.0474	0.0948	0.0118	0.0536	0.1014	0.0118	0.0522	0.0988
		$S_T$	0.0206	0.0740	0.1216	0.0256	0.0818	0.1386	0.0162	0.0720	0.1292
	0.75	$S_W$	0.0046	0.0232	0.0536	0.0134	0.0644	0.1216	0.0036	0.0446	0.0928
		$S_{LR}$	0.0150	0.0608	0.1108	0.0110	0.0532	0.0958	0.0080	0.0516	0.1000
		$S_R$	0.0116	0.0474	0.0948	0.0118	0.0536	0.1014	0.0118	0.0522	0.0988
		$S_T$	0.0206	0.0740	0.1216	0.0256	0.0818	0.1386	0.0162	0.0722	0.1292

**Figura 2.7:** Curva de poder dos testes para o modelo de regressão EBS ( $\vartheta = 0.5$ ) quantílico.

**Tabela 2.20:** Tamanho dos testes para o modelo de regressão EBS- $t$  ( $\vartheta_1 = 0.5, \vartheta_2 = 3$ ) quantílica, sob  $H_0 : \beta_1 = \dots = \beta_t = 0$ .

$t$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
3	0.25	$S_W$	0.0256	0.0806	0.1382	0.0226	0.0718	0.1292	0.0114	0.0578	0.1102
		$S_{LR}$	0.0142	0.0712	0.1288	0.0122	0.0586	0.1108	0.0100	0.0556	0.0992
		$S_R$	0.0264	0.0708	0.1146	0.0262	0.0624	0.1050	0.0296	0.0746	0.1178
		$S_T$	0.0194	0.0642	0.1200	0.0234	0.0680	0.1262	0.0204	0.0670	0.1192
	0.5	$S_W$	0.0234	0.0806	0.1380	0.0228	0.0718	0.1288	0.0110	0.0582	0.1106
		$S_{LR}$	0.0138	0.0674	0.1144	0.0126	0.0560	0.1064	0.0104	0.0534	0.0946
		$S_R$	0.0264	0.0708	0.1146	0.0262	0.0624	0.1050	0.0296	0.0746	0.1178
		$S_T$	0.0142	0.0644	0.1200	0.0234	0.0680	0.1262	0.0202	0.0670	0.1192
	0.75	$S_W$	0.0236	0.0806	0.1382	0.0228	0.0718	0.1292	0.0112	0.0580	0.1108
		$S_{LR}$	0.0144	0.0608	0.1144	0.0142	0.0556	0.1100	0.0128	0.0510	0.1040
		$S_R$	0.0264	0.0708	0.1146	0.0262	0.0624	0.1050	0.0296	0.0746	0.1178
		$S_T$	0.0142	0.0642	0.1200	0.0234	0.0680	0.1244	0.0202	0.0670	0.1192
1	0.25	$S_W$	0.0124	0.0530	0.1060	0.0192	0.0434	0.0906	0.0074	0.0380	0.0816
		$S_{LR}$	0.0148	0.0624	0.1170	0.0122	0.0570	0.1088	0.0108	0.0554	0.1010
		$S_R$	0.0120	0.0496	0.1026	0.0110	0.0470	0.0958	0.0110	0.0498	0.0980
		$S_T$	0.0244	0.0742	0.1274	0.0270	0.0788	0.1294	0.0174	0.0698	0.1234
	0.5	$S_W$	0.0146	0.0566	0.1062	0.0192	0.0434	0.0906	0.0072	0.0380	0.0814
		$S_{LR}$	0.0130	0.0624	0.1170	0.0122	0.0572	0.1088	0.0108	0.0554	0.1010
		$S_R$	0.0120	0.0496	0.1026	0.0104	0.0470	0.0958	0.0110	0.0498	0.0980
		$S_T$	0.0186	0.0742	0.1274	0.0272	0.0788	0.1294	0.0174	0.0670	0.1234
	0.75	$S_W$	0.0150	0.0568	0.1066	0.0192	0.0432	0.0910	0.0074	0.0378	0.0814
		$S_{LR}$	0.0130	0.0624	0.1170	0.0122	0.0570	0.1088	0.0108	0.0554	0.1010
		$S_R$	0.0120	0.0496	0.1026	0.0104	0.0470	0.0958	0.0110	0.0498	0.0980
		$S_T$	0.0186	0.0742	0.1272	0.0272	0.0788	0.1294	0.0174	0.0698	0.1236

**Figura 2.8:** Curva de poder dos testes para o modelo de regressão EBS- $t$  ( $\vartheta_1 = 0.5, \vartheta = 3$ ) quantílico.

## 2.5 Exemplo ilustrativo: Bilheteria de filmes

Nessa seção, os modelos de regressão log-simétricos quantílicos propostos são ilustrados usando uma base de dados referentes a filmes exibidos entre 1929 e 2016. Essa base foi montada por Ventura et al. (2019) e pode ser obtida por meio de *web scraping* no *Internet Movie Database* (IMDb), hospedado em <https://www.imdb.com/>. A base de dados conta com 250 filmes catalogados, entretanto, devido ao fato de que alguns filmes antigos não produzidos nos EUA não constarem com suas respectivas informações financeiras, a análise do conjunto de dados procedeu com 155 filmes. As seguintes variáveis são reportadas:  $Y$ , bilheteria, em milhões de US\$;  $x_1$ , quantidade de votos, em milhões;  $x_2$ , orçamento, em milhares de US\$; e  $x_3$ , classificação, segundo os usuários. Aqui  $Y$  é a resposta e  $x_1, x_2, x_3$  são as respectivas covariáveis.

A Tabela 2.21 apresenta um sumário de estatísticas descritivas para as observações da resposta, bilheteria (em milhões de US\$). Essa tabela contém o tamanho da amostra, média, mediana, mínimo, máximo, SD, CS, CK e o coeficiente de Variação (CV). Pela Tabela 2.21 podemos observar que, em média, a bilheteria do filmes contidos no catálogo é de 127.915 milhões de US\$, com SD de 160.9 milhões de US\$, que evidencia uma grande variabilidade na receita dos filmes, juntamente com o coeficiente de variação de 126.29%. Além do elevado CV da bilheteria, o CS indica que esses dados apresentam uma assimetria à direita, levando a utilização da mediana como uma medida de tendência central mais apropriada, a qual foi 60.99 US\$.

**Tabela 2.21:** Sumário de estatísticas descritivas para a bilheteria, em milhões de US\$.

$n$	Min.	Mediana	Média	Max.	SD	CS	CK	CV
155	0.1149	60.9947	127.3915	936.6274	160.8849	1.9370	4.4831	126.2916

Ventura et al. (2019) analisaram esses dados de bilheteria utilizando modelos de regressão log-simétricos (Vanegas e Paula, 2016a), o que segundo os autores é justificado sobretudo pela natureza assimétrica dos dados. No entanto, como explicado anteriormente, na regressão quantílica há uma caracterização mais rica dos efeitos das covariáveis sobre a resposta. Desse modo, iremos ajustar o modelo proposto de regressão log-simétrico quantílico como alternativa mais geral ao modelo de regressão log-simétrico de Vanegas e Paula (2016a), utilizando os

quantis  $q = 0.25, 0.50, 0.75$ . Temos, então

$$Y_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \varepsilon_i^{\sqrt{\tau_0 + \tau_1 x_{1i} + \tau_2 x_{2i} + \tau_3 x_{3i}}}, \quad i = 1, 2, \dots, 155,$$

em que  $\varepsilon_i \sim \text{QLS}(1, 1, g)$ .

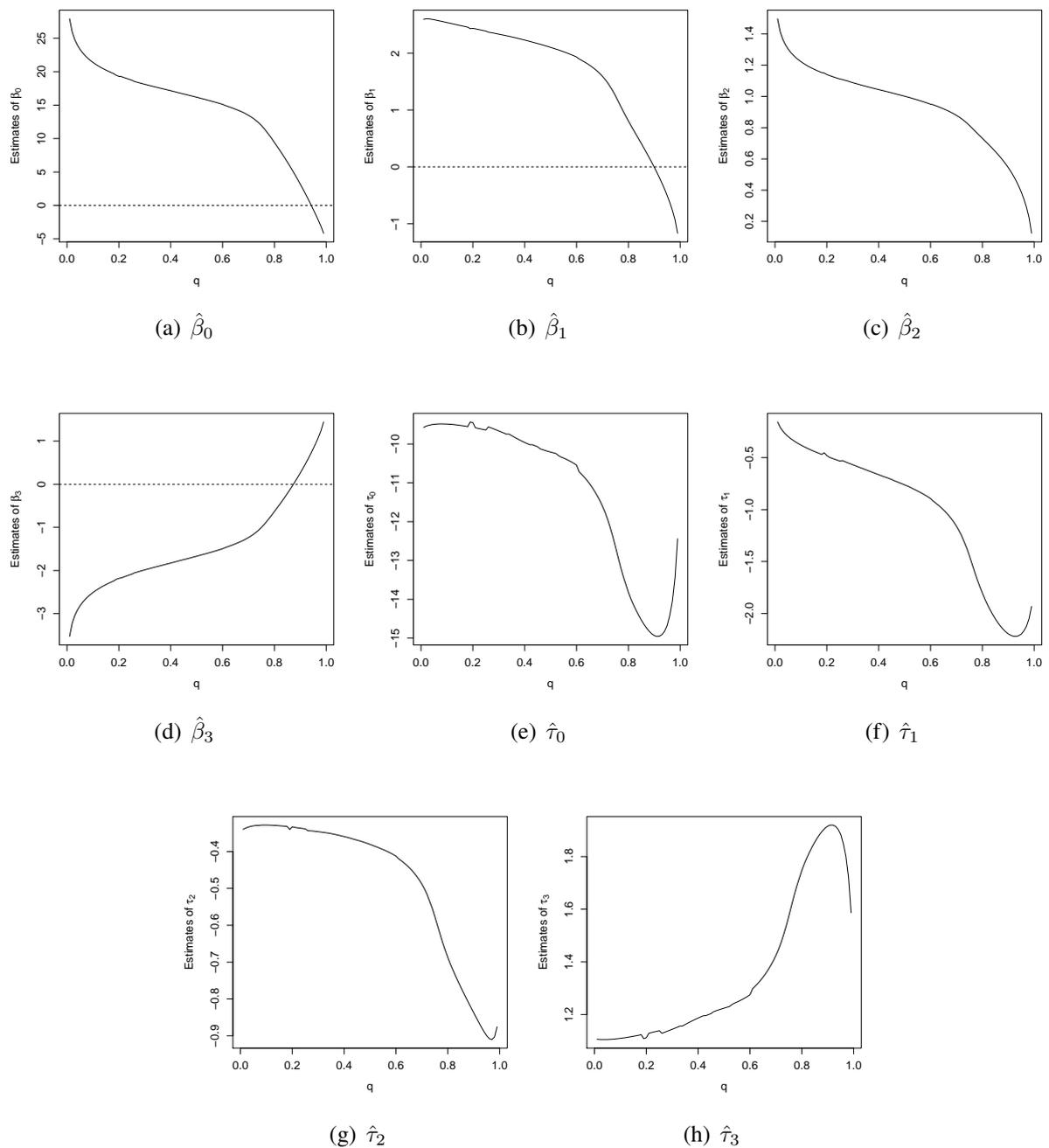
Inicialmente, temos que buscar um modelo, baseado em uma distribuição da Tabela 2.1, com melhor ajuste para os dados. Nesse sentido, dois critérios são utilizados: critérios de informação AIC, BIC e AICc; e a raiz do erro quadrático médio (REQM) da predição. Buscando uma análise do ajuste fornecido pelo modelo de regressão como um todo, por esse motivo o ajuste prestado pelo modelo foi analisado considerando os quantis  $q = 0.01, 0.02, \dots, 0.98, 0.99$ . Além disso, foi considerado um grid de valores para  $\vartheta$  de acordo com a distribuição usada na Tabela 2.1, a ideia aqui é obter o valor do parâmetro extra que maximiza a função de verossimilhança dos dados. Em virtude do número elevado de modelos ajustados para cada quantil, o valor ótimo do parâmetro extra mudou em alguns casos conforme o  $q$  fixado. A Tabela 2.22 apresenta os valores das médias dos critérios de informação e dos valores do REQM da predição, respectivamente. Pela Tabela 2.22, podemos observar que o modelo com uma distribuição log-CN apresenta o menor valor para os critérios de informação. Entretanto, para as distribuições log- $t$ , log-PE, log-HP e log-SL, os valores dos critérios são muito próximos aos da log-CN. Na Tabela 2.22, podemos perceber que o REQM para os modelos baseados nas distribuições log-NO, EBS e log-HP apresentam os menores valores, indicando que são os modelos mais precisos dentre os analisados, em termos de predição. Dentre os modelos que se destacam com menores REQM, o modelo com distribuição log-NO apresenta o menor REQM. Tomando como critério de escolha os critérios de seleção de modelos e o REQM, o melhor modelo é o modelo baseado na distribuição log-HP. Especialmente no caso da distribuição log-HP, o valor ótimo do parâmetro extra foi  $\vartheta = 1$  para todos os quantis  $q = 0.01, 0.02, \dots, 0.98, 0.99$ .

A Figura 2.9 apresenta o comportamento das estimativas para o modelo de regressão quantílica com a distribuição log-HP(1) conforme o valor de  $q$  fixado. Na Figura 2.9 podemos observar que as estimativas  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$  decrescem e  $\hat{\beta}_3$  aumenta à medida que o valor de  $q$  aumenta. Enquanto que as estimativas de  $\hat{\tau}_0, \hat{\tau}_1, \hat{\tau}_2$  começam decrescendo e a partir de  $q = 0.9$  crescem e  $\hat{\tau}_3$  aumenta e após  $q = 0.9$  decresce.

**Tabela 2.22:** Média dos critérios AIC, BIC, AICc e RMSE dos modelos de regressão quantílica log-simétrica considerando os quantis  $q = 0.01, 0.02, \dots, 0.98, 0.99$ .

Critério	Log-NO	Log- $t$	Log-PE	Log-HP	Log-SL	Log-CN	EBS	EBS- $t$
AIC	514.0027	499.4266	500.5683	501.1267	499.3116	497.9390	514.1874	499.4132
BIC	538.3501	523.7740	524.9157	525.4741	523.6590	522.2864	538.5348	523.7606
AICc	514.9890	500.4129	501.5546	502.1130	500.2979	498.9253	515.1737	500.3995
REQM	191.7217	225.3356	212.2050	198.2404	239.7783	230.8775	192.0232	233.8590

**Figura 2.9:** Gráfico das estimativas dos parâmetros do modelo log-simétrico quantílico, com erros log-HP(1), para os quantis  $q = 0.01, 0.02, \dots, 0.98, 0.99$ .



A Tabela 2.23 apresenta as estimativas dos parâmetros para o modelo baseado na log-HP, considerando os quantis  $q = 0.25, 0.5, 0.75$ . Observa-se que as estimativas de máxima verossimilhança dos parâmetros mudam de acordo com o valor de  $q$  fixado. A estimativa de  $\hat{\vartheta}$  foi igual a 1 para todos os valores de  $q$ . As estimativas de  $\beta_0, \beta_1, \beta_3$  (em valor absoluto) crescem a medida que o valor de  $q$  aumenta, enquanto que de  $\beta_2$  se mantém constante independente do valor de  $q$  fixado. Para os coeficientes que modelam a assimetria (ou dispersão relativa), a estimativa de  $\tau_0$  (em valor absoluto) fica constante com a mudança do valor de  $q$ . Pode-se interpretar os coeficientes estimados em termos de seu efeito na resposta  $Y_i$  (bilheteria, em milhões de US\$); ver Weisberg (2014, p.82) para uma interpretação semelhante. Por exemplo, um aumento no número de votos em 1 milhão, aumenta o 25° percentil ( $q = 0.25$ ) da bilheteria em  $(\exp(2.3891) - 1) \times 100\% = 990, 37\%$ , enquanto o aumento no percentil 75° ( $q = 0.75$ ) é de  $(\exp(1, 2510) - 1) \times 100\% = 249, 39\%$ . Portanto, o efeito do número de votos na bilheteria é menor para filmes com bilheteria maior (quantis maiores). Observa-se também que o orçamento se comporta de forma semelhante ao número de votos, ou seja, o efeito na bilheteria é menor para filmes com orçamentos maiores. Por fim, nota-se que o efeito da classificação na bilheteria é negativo, em que filmes com bilheteria menor (quantis menores) têm efeitos negativos maiores.

**Tabela 2.23:** Estimativas e erros-padrões (entre parênteses) para o modelo de regressão quantílica com erros log-HP(1).

$q$	Estimativa (Erro padrão)				
	Intercept ( $\hat{\beta}_0/\hat{\tau}_0$ )	Votes ( $\hat{\beta}_1/\hat{\tau}_1$ )	Budget ( $\hat{\beta}_2/\hat{\tau}_2$ )	Rating ( $\hat{\beta}_3/\hat{\tau}_3$ )	
0.25	$\hat{\beta}$	18.7352 (4.4784)	2.3891 (0.4168)	1.1115 (0.1716)	-2.0848 (0.5508)
	$\hat{\tau}$	-9.6409 (8.5539)	-0.5325 (0.7200)	-0.3385 (0.2954)	1.1387 (1.0594)
0.5	$\hat{\beta}$	16.1816 (3.9019)	2.1025 (0.4026)	1.0013 (0.1499)	-1.6689 (0.4814)
	$\hat{\tau}$	-10.2079 (3.8779)	-0.7619 (0.5187)	-0.3806 (0.2923)	1.2243 (0.4798)
0.75	$\hat{\beta}$	11.9536 (6.5034)	1.2510 (0.8749)	0.8142 (0.1644)	-1.0023 (0.8438)
	$\hat{\tau}$	-12.6406 (6.3288)	-1.4446 (0.9279)	-0.5768 (0.3268)	1.5746 (0.8140)

Na Tabela 2.24 temos os valores observados das estatísticas e o correspondente p-valor para testar:  $H_0 : \beta_1 = 0$ ,  $H_0 : \beta_2 = 0$ ,  $H_0 : \beta_3 = 0$  e  $H_0 : \tau_1 = 0$ ,  $H_0 : \tau_2 = 0$ ,  $H_0 : \tau_3 = 0$  dos testes  $S_W$ ,  $S_{LR}$ ,  $S_R$  e  $S_T$ , baseado no modelo de regressão log-simétrico quantílico com distribuição log-HP(1). Considerando o nível de significância de 5%, não rejeitamos as hipóteses  $H_0 : \beta_1 = 0$  para as estatísticas  $S_W$  e  $S_R$ , para  $q = 0.75$ . Para  $H_0 : \beta_2 = 0$ , não rejeitamos a hipótese nula para nenhum dos testes. Testando  $H_0 : \beta_3 = 0$ , ao nível de 5%, não rejeitamos a hipótese nula para  $S_R$ , com  $q = 0.25, 0.5, 0.75$  e também para  $S_W$  e  $S_T$ , com  $q = 0.75$ . Além

disso, para os coeficientes que modelam a assimetria, isto é,  $\tau_1, \tau_2, \tau_3$  não rejeitamos  $H_0 : \tau_1 = 0$  apenas para o teste gradiente e  $H_0 : \tau_3 = 0$  para os testes  $S_T$  e  $S_{LR}$ , com  $q = 0.75$ , ao nível de 5%. Um aspecto relevante que pode ser notado, é que para testar os parâmetros  $\beta_1, \beta_2, \beta_3$  o valor observado das estatísticas diminuem a medida que  $q$  aumenta. Removidos os coeficientes não significantes, o modelo reduzido é da seguinte forma

$$Y_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \varepsilon_i^{\sqrt{\tau_0}}, \quad i = 1, 2, \dots, 155. \quad (2.17)$$

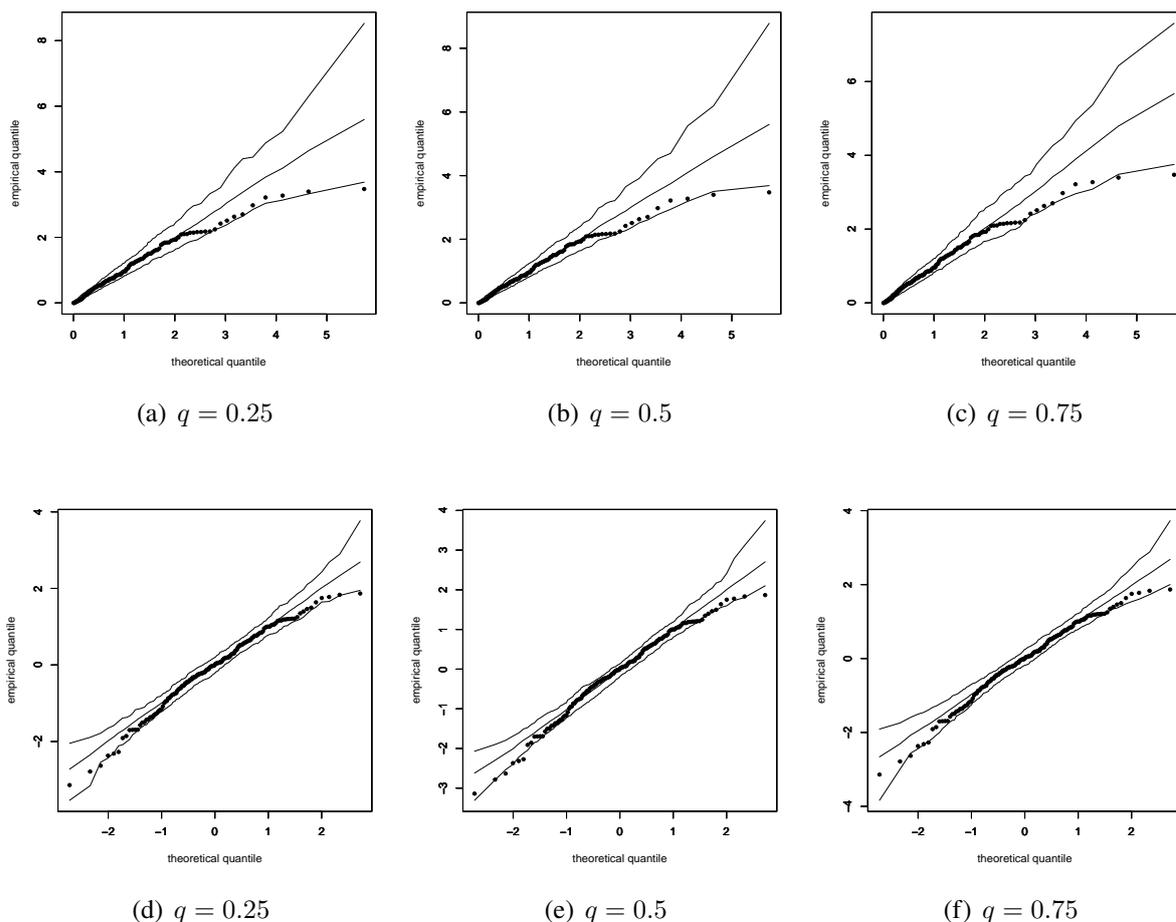
**Tabela 2.24:** Estatísticas observadas  $S_W, S_{LR}, S_R$  e  $S_T$  e o seu respectivo p-valor para o modelo de regressão quantílica log-simétrica.

$q$	$H_0 : \beta_1 = 0$		$H_0 : \beta_2 = 0$		$H_0 : \beta_3 = 0$		
	Observado	p-valor	Observado	p-valor	Observado	p-valor	
0.25	$S_W$	32.8550	< 0.0001	41.9580	< 0.0001	14.3270	0.0002
	$S_{LR}$	34.8390	< 0.0001	26.1230	< 0.0001	14.9210	0.0001
	$S_R$	4.0034	0.0454	7.8686	0.0050	0.0061	0.9378
	$S_T$	31.6740	< 0.0001	19.9820	< 0.0001	13.7980	0.0002
0.5	$S_W$	27.2720	< 0.0001	44.6410	< 0.0001	12.0190	0.0005
	$S_{LR}$	26.6200	< 0.0001	30.3810	< 0.0001	13.2230	0.0003
	$S_R$	1.6362	0.2008	10.6870	0.0011	0.0080	0.9289
	$S_T$	27.7960	< 0.0001	23.8190	< 0.0001	12.8890	0.0003
0.75	$S_W$	2.0448	0.1527	24.5320	< 0.0001	1.4109	0.2349
	$S_{LR}$	5.6346	0.0176	25.2240	< 0.0001	3.0571	0.0804
	$S_R$	0.1559	0.6929	4.8372	0.0279	0.0009	0.9757
	$S_T$	6.5448	0.0105	21.4550	< 0.0001	3.0775	0.0794
$q$	$H_0 : \tau_1 = 0$		$H_0 : \tau_2 = 0$		$H_0 : \tau_3 = 0$		
	Observado	p-valor	Observado	p-valor	Observado	p-valor	
0.25	$S_W$	0.5470	0.4595	1.3128	0.2519	1.1553	0.2824
	$S_{LR}$	0.8809	0.3480	1.3403	0.2470	2.6462	0.1038
	$S_R$	0.1296	0.7189	0.5341	0.4649	0.0014	0.9700
	$S_T$	0.8795	0.3483	1.3290	0.2490	2.6321	0.1047
0.5	$S_W$	2.1577	0.1419	1.6959	0.1928	6.5109	0.0107
	$S_{LR}$	1.5583	0.2119	1.6458	0.1995	2.9014	0.0885
	$S_R$	0.2368	0.6265	0.7194	0.3963	0.0018	0.9663
	$S_T$	1.5671	0.2106	1.6174	0.2034	2.9183	0.0876
0.75	$S_W$	2.4239	0.1195	3.1166	0.0775	3.7418	0.0531
	$S_{LR}$	3.5582	0.0593	3.0260	0.0819	4.1351	0.0420
	$S_R$	0.2980	0.5851	0.7495	0.3866	0.0017	0.9673
	$S_T$	3.8473	0.0498	2.9802	0.0843	4.2370	0.0396

Na Figura 2.10 estão plotados os gráficos QQ com envelope simulado dos resíduos GCS e RQ para baseado no modelo de regressão log-simétrico quantílico com distribuição log-HP

reduzido (2.17), considerando os diferentes valores do quantil  $q$ . Em geral, observamos que a maioria dos pontos está dentro das bandas, indicando que modelo baseado na distribuição log-HP apresenta um bom ajuste.

**Figura 2.10:** Gráfico de envelope dos resíduos GCS e RQ para o modelo ajustado considerando os quantis  $q = 0.25, 0.5, 0.75$ .



## 2.6 Considerações finais

Nesse trabalho introduzimos uma nova classe de modelos de regressão quantílica baseada em uma reparametrização proposta das distribuições log-simétricas. O modelo proposto se apresenta como uma alternativa bastante flexível na modelagem de dados assimétricos positivos além de ser mais informativo, uma vez que detalha de forma mais aprofundada os efeitos das covariáveis sobre a resposta, que pode depender do espectro da última. Dois estudos de simulação foram realizados, o primeiro para avaliar o desempenho dos estimativas de máxima verossimilhança, dos critérios de informação AIC, BIC e AICc, e dos resíduos Cox-Snell

generalizado e quantil aleatorizado, é o segundo estudo para avaliar o desempenho dos testes Wald ( $S_W$ ), score ( $S_R$ ), razão de verossimilhança ( $S_{LR}$ ) e gradiente ( $S_T$ ). Os critérios de seleção de modelos se apresentam como uma boa alternativa para a seleção do melhor modelo. Além disso, os resíduos Cox-Snell generalizado e quantil aleatorizado se apresentaram como uma alternativa simples e eficiente na análise da qualidade do ajuste prestado pelo modelo de regressão log-simétrico quantílico, a restrição fica sob o caso em que utilizamos a distribuição log-normal-contaminada, isso porque, os resíduos não se comportaram de acordo com a sua respectiva distribuição teórica, no estudo de simulação de Monte Carlo. Em relação aos testes de hipóteses, foi observado no estudo de simulação que, na maioria das situações, as estatísticas  $S_W$  e  $S_{LR}$ , em relação às demais estatísticas, apresentam taxas de rejeição mais próximas do nível nominal quando testamos múltiplos parâmetros simultaneamente. No teste de apenas um parâmetro, as estatísticas  $S_W$  e  $S_R$  apresentam taxas de rejeição mais próximas do nível nominal. Por outro lado, vemos que o poder do teste (exceto os modelos Birnbaum-Saunders estendida e Birnbaum-Saunders- $t$  estendida) é maior para a estatística  $S_R$ , seguida pela estatística  $S_T$ , para testar  $t = 3$  parâmetros simultaneamente, já para  $t = 1$  parâmetro, o poder do teste usando as estatísticas  $S_R$  e  $S_T$  praticamente coincidem. Por fim, uma aplicação a dados reais ilustrou a utilidade do modelo proposto, mostrando uma caracterização mais rica dos efeitos das covariáveis sobre a resposta.

## Capítulo 3

# Modelos de regressão quantílica mistura-escala Birnbaum-Saunders

**Resumo:** Modelos de regressão Birnbaum-Saunders são utilizados na modelagem de dados positivos e assimétricos à direita. Neste artigo propomos um novo modelo de regressão quantílica baseado na distribuição mistura-escala Birnbaum-Saunders reparametrizada pelo quantil, em que as estimativas de máxima verossimilhança dos parâmetros são obtidas utilizando o algoritmo EM. Dois estudos de simulação de Monte Carlo foram realizados utilizando o software R. O primeiro estudo tem como objetivo analisar a performance dos estimadores de máxima verossimilhança, os critérios de informação AIC, AICc, BIC, HIC, a raiz do erro quadrático médio, os resíduos Cox-Snell e quantil aleatorizado. No segundo estudo de simulação foram avaliados os testes de hipóteses em duas perspectivas: o tamanho e o poder dos testes Wald, razão de verossimilhança, score e gradiente. Os dois estudos de simulação foram conduzidos considerando diferentes quantis de interesse e tamanhos amostrais. Por fim, é realizada uma aplicação à dados de seguros pessoais a fim de ilustrar a metodologia proposta.

**Palavras-chave:** Distribuição Birnbaum-Saunders; Algoritmo EM; Testes de hipóteses; Simulação de Monte Carlo; Regressão quantílica.

### 3.1 Introdução

A classe das distribuições mistura-escala de normais (SMN) compõem uma família de distribuições simétricas estudada inicialmente por Andrews e Mallows (1974). Essas distribuições vem ganhando bastante notoriedade com a publicação de inúmeros trabalhos, como pode ser visto em Efron e Olshen (1978), West (1987), Lange e Sinsheimer (1993), Gneiting (1997), Taylor e Verbyla (2004), Walker e Gutiérrez-Peña (2007) e Lachos e Labra (2007). Essa família de distribuições conta com uma gama de distribuições como casos especiais, as quais tem como principal característica o ajuste flexível para caudas pesadas, ajuste esse obtido por meio de uma estimação robusta dos parâmetros da distribuição; ver Lange, Little e Taylor (1989) e Lucas (1997b). Apesar das vantagens apresentadas anteriormente, em diversas situações em que o conjunto de dados a ser analisado apresenta uma distribuição assimétrica e geralmente com caudas pesadas, é necessário o uso de distribuições flexíveis com boas propriedades para se ajustar a tais dados.

Um dos modelos mais utilizados atualmente na modelagem de dados estritamente positivos é a distribuição Birnbaum-Saunders (BS), proposta por Birnbaum e Saunders (1969a). Esse modelo foi derivado com base em algumas suposições sobre o processo de fadiga de materiais. O fato da distribuição BS apresentar assimétrica à direita com suporte positivo, além de ter uma relação direta com a distribuição normal fazem com que a mesma apresente propriedades bem interessantes para a modelagem de dados positivos assimétricos. Devido a tais propriedades, a distribuição BS vem ganhando cada vez mais importância com diversos trabalhos publicados, como pode ser visto em Birnbaum e Saunders (1969b), Leiva et al. (2008a), Leiva et al. (2008b), Leiva, Sanhueza e Angulo (2009), Cordeiro e Lemonte (2011), Ferreira, Gomes e Leiva (2012), entre outros.

Com base na relação existente entre a distribuição BS e a distribuição normal, Balakrishnan et al. (2009) propuseram uma nova família de distribuições com base na classe das distribuições SMN, denominada distribuição mistura-escala Birnbaum-Saunders (SBS). Balakrishnan et al. (2009) derivaram uma série de propriedades, a descrição do processo de obtenção dos estimadores de máxima verossimilhança via algoritmo EM (Dempster, Laird e Rubin, 1977) e três aplicações utilizando dados reais. Posteriormente, Lachos et al. (2017) propuseram um modelo de regressão bayesiano com base na distribuição log-SBS para dados censurados.

Os modelos de regressão quantílica surgiram como uma proposta mais ampla na análise da relação entre a variável resposta e as respectivas covariáveis de um modelo de regressão. Esses modelos foram estudados inicialmente por Koenker e Bassett Jr (1978) e desde então vem se tornando cada vez mais utilizadas na análise de regressão; ver Koenker e Hallock (2001), Yu e Moyeed (2001) e Koenker (2004). Os modelos de regressão quantílica se apresentam como uma alternativa mais robusta em relação aos modelos de regressão para a média, pelo fato do processo de estimação dos parâmetros do modelo de regressão quantílica não estarem sujeito a influência dos *outliers* como no modelo de regressão linear, bem como a perspectiva de uma análise mais ampla da relação entre a variável resposta e as covariáveis que compõem o modelo de regressão, provida pelo ajuste de mais de um modelo de regressão para cada quantil de interesse.

Neste artigo, propomos um novo modelo de regressão quantílica para dados estritamente positivos com base nas distribuições SBS reparametrizadas pelo quantil (QSBS). Essa reparametrização é proposta nesse artigo e é feita inserindo um parâmetro de quantil através de uma transformação um-a-um, similarmente ao trabalho de Sánchez et al. (2020a) que introduziu um parâmetro de quantil na distribuição BS. As estimativas de máxima verossimilhança dos parâmetros do modelo são obtidas utilizando o algoritmo EM. Dois estudos de simulação são realizados, o primeiro tem como objetivo analisar os estimadores de máxima verossimilhança para o modelo de regressão proposto, além da análise dos resíduos generalizados Cox-Snell (Cox e Snell, 1968) e quantil aleatorizado (Dunn e Smyth, 1996), bem como a taxa de acerto dos critérios de informação AIC, BIC, AICc, HIC e da raiz do erro quadrático médio (REQM). O segundo estudo de simulação foca na análise dos testes de hipóteses Wald, Score, razão de verossimilhança e gradiente em duas perspectivas: na análise do tamanho e do poder dos testes. Por fim, um exemplo ilustrativo é apresentado mostrando a metodologia proposta.

O restante desse capítulo é organizado como segue. Na Seção 3.2, as distribuições SBS e as correspondentes versões reparametrizadas pelo quantil QSBS propostas, são apresentadas. Nessa seção, são apresentadas ainda algumas propriedades das distribuições QSBS bem como alguns casos especiais. Na Seção 3.3, o modelo de regressão quantílica QSBS é introduzido. Detalhes de estimação via algoritmo EM, obtenção de erros-padrões, critérios de seleção, resíduos, e testes de hipóteses, são também abordados na Seção 3.3. Na Seção 3.4, os resultados obtidos de dois estudos de simulação de Monte Carlo para avaliar o desempenho dos estimadores de

máxima verossimilhança, critérios de informação, REQM, resíduos e testes de hipótese, são apresentados. Na Seção 3.5, uma aplicação utilizando um conjunto de dados reais é feita, e por fim, na Seção 3.6, são apresentadas as conclusões finais sobre o modelo proposto.

## 3.2 Preliminares

Essa seção apresenta as distribuições SBS bem como as distribuições QSBS propostas nesse artigo. Algumas propriedades e casos especiais das distribuições QSBS são também apresentados.

### 3.2.1 Distribuições SBS

Seja  $Y$  uma variável aleatória (v.a) cuja sua representação estocástica pode ser escrita da seguinte forma:

$$Y = \mu + \sqrt{g(U)}X, \quad (3.1)$$

onde  $\mu$  é um parâmetro de locação,  $X \sim N(0, \sigma^2)$ ,  $U$  é uma v.a independente de  $X$ , com função de distribuição acumulada (FDA) dada por  $H(\cdot)$ , a qual é indexada por um escalar ou vetor de parâmetros  $\nu$  e  $g(\cdot)$  uma função positiva. No caso em que  $g(U) = 1/U$ , a distribuição de  $Y$ , na Equação (3.1), se reduz a distribuição normal/independente, apresentada em Lange e Sinsheimer (1993). De forma análoga, quando  $g(U) = U$  em (3.1), a distribuição de  $Y$  se reduz a distribuição SMN, estudada por Fernández e Steel (1999).

A função densidade de probabilidade (FDP) de  $Y$ , uma v.a que segue uma distribuição SMN com parâmetros de locação e escala,  $\mu \in \mathcal{R}$  e  $\sigma^2 > 0$ , respectivamente, é dada por:

$$\phi_{\text{SMN}}(y) = \int_0^\infty \phi(y; \mu, g(u)\sigma^2) dH(u), \quad (3.2)$$

onde  $\phi(y; \mu, g(u)\sigma^2)$  é a FDP de uma distribuição normal com média  $\mu$  e variância  $g(u)\sigma^2$  e  $H(u)$  é a FDA de  $U$ . Denotamos  $Y \sim \text{SMN}(\mu, \sigma^2; H)$  se a FDP de  $Y$  é da forma de (3.2). Para o caso em que  $\mu = 0$  e  $\sigma^2 = 1$ , denotamos apenas por  $Y \sim \text{SMN}(H)$ .

De acordo com Balakrishnan et al. (2009), uma v.a  $T$  segue uma distribuição mistura-escala Birnbaum-Saunders (SBS), se  $T$  tem representação estocástica dada por

$$T = \frac{\beta}{4} \left[ \alpha \sqrt{g(U)}Z + \sqrt{(\alpha \sqrt{g(U)}Z)^2 + 4} \right]^2, \quad (3.3)$$

onde  $Y = \sqrt{g(U)}Z \sim \text{SMN}(H)$ , tal que  $Z \sim \text{N}(0, 1)$  e  $\alpha > 0$  e  $\beta > 0$  são, respectivamente, os parâmetros de forma e escala de uma v.a com distribuição Birnbaum-Saunders (BS). No caso em que  $Y \sim \text{N}(0, 1)$ , então  $T$  se resume apenas a uma distribuição BS clássica. Denotamos  $T \sim \text{SBS}(\alpha, \beta; H)$ . A FDP de  $T \sim \text{SBS}(\alpha, \beta; H)$  é dada por:

$$f_T(t) = \phi_{\text{SMN}}(a(t))A(t), \quad t > 0, \quad (3.4)$$

em que  $\phi_{\text{SMN}}(\cdot)$  é a FDP dada pela Equação (3.2) com  $\mu = 0$  e  $\sigma^2 = 1$ ,  $a(t) = [\sqrt{t/\beta} - \sqrt{\beta/t}]/\alpha$ , e  $A(t) = t^{-3/2}[t + \beta]/[2\alpha\beta^{1/2}]$  é a derivada de  $a(t)$  em relação a  $t$ . A FDA de  $T$  é dada seguinte forma

$$F_T(t) = \Phi_{\text{SMN}}(a(t)), \quad (3.5)$$

onde  $\Phi_{\text{SMN}}(\cdot)$  é a FDA de uma distribuição  $\text{SMN}(H)$ . Dado  $q \in (0, 1)$ , podemos obter o  $q \times 100$ -ésimo quantil de  $T$  com base na representação estocástica (3.3) da seguinte forma

$$Q = t_q = \frac{\beta}{4} \left[ \alpha y_q + \sqrt{(\alpha y_q)^2 + 4} \right]^2, \quad (3.6)$$

onde  $y_q$  é o  $q \times 100$ -ésimo quantil de uma distribuição  $Y \sim \text{SMN}(H)$ .

### 3.2.2 Distribuições QSBS

Seja  $q \in (0, 1)$  um número fixado e a transformação um-a-um

$$(\alpha, \beta; H) \mapsto (\alpha, Q; H),$$

onde  $Q$  é definido na Equação (3.6), obtemos a representação estocástica com a reparametrização pelo quantil

$$T = \frac{Q}{\gamma_\alpha^2} \left[ \alpha \sqrt{g(U)}Z + \sqrt{(\alpha \sqrt{g(U)}Z)^2 + 4} \right]^2, \quad (3.7)$$

onde  $\gamma_\alpha = \alpha y_q + \sqrt{(\alpha y_q)^2 + 4}$ ,  $Z \sim \text{N}(0, 1)$ , com  $g(U) = 1/U$  ou  $U$ , tal que a distribuição  $g(U)$  tem uma FDP conhecida. A representação estocástica (3.7) pode ser utilizada para a geração de números aleatórios, obtenção dos momentos da distribuição e na implementação do algoritmo EM para obtenção dos estimadores de máxima verossimilhança.

Então, pela representação estocástica (3.7), definimos a reparametrização da distribuição SBS baseada no quantil  $Q$ , dessa forma, podemos reescrever as funções densidade de probabilidade e função de distribuição acumulada, como segue

$$F_T(t) = \Phi_{\text{SMN}}(a(t)), \quad t > 0, \quad (3.8)$$

$$f_T(t) = \phi_{\text{SMN}}(a(t))A(t), \quad t > 0, \quad (3.9)$$

onde  $a(t) = \sqrt{4Q/t}(\gamma_\alpha^2 t/4Q - 1)/[\alpha\gamma_\alpha]$ ,  $A(t) = (\gamma_\alpha^2/2 + 2Q/t)/[\alpha\gamma_\alpha\sqrt{4Qt}]$  e  $\gamma_\alpha = \alpha y_q + \sqrt{(\alpha y_q)^2 + 4}$ . Denotamos  $T \sim \text{QSBS}(\alpha, Q; H)$  se a FDP e a FDA de  $T$ , são dadas, respectivamente, pelas Equações (3.8) e (3.9). Algumas propriedades da distribuição QSBS são apresentadas a seguir.

**Proposição 3.** Se  $cT \sim \text{QSBS}(\alpha, cQ; H)$ , com  $c > 0$ .

*Demonstração.* Se  $T \sim \text{QSBS}(\alpha, Q; H)$  e  $c > 0$ . Então,

$$\begin{aligned} P(cT \leq t) = P(T \leq t/c) &= \Phi_{\text{SMN}} \left( \sqrt{\frac{4Q}{t/c}} \left( \frac{\gamma_\alpha^2 t/c}{4Q} - 1 \right) \frac{1}{\alpha\gamma_\alpha} \right) \\ &= \Phi_{\text{SMN}} \left( \sqrt{\frac{4(cQ)}{t}} \left( \frac{\gamma_\alpha^2 t}{4(cQ)} - 1 \right) \frac{1}{\alpha\gamma_\alpha} \right). \end{aligned}$$

□

**Proposição 4.** Se  $1/T \sim \text{QSBS}(\alpha, 1/Q; H)$ .

*Demonstração.* Se  $T \sim \text{QSBS}(\alpha, Q; H)$ . Então,

$$\begin{aligned} P(1/T \leq t) = P(T \leq 1/t) &= \Phi_{\text{SMN}} \left( \sqrt{\frac{4Q}{(1/t)}} \left( \frac{\gamma_\alpha^2 (1/t)}{4Q} - 1 \right) \frac{1}{\alpha\gamma_\alpha} \right) \\ &= \Phi_{\text{SMN}} \left( \sqrt{\frac{(1/4Q)}{t}} \left( \frac{\gamma_\alpha^2 t}{(1/4Q)} - 1 \right) \frac{1}{\alpha\gamma_\alpha} \right). \end{aligned}$$

□

**Proposição 5.** Seja  $T \sim \text{QSBS}(\alpha, Q; H)$  e  $g(U)$  uma v.a. como na Equação (3.1) com todos os

momentos finitos de todas as ordens, então o  $k$ -ésimo momento de  $T$  é dado por

$$E[T^k] = \left(\frac{4Q}{\gamma_\alpha^2}\right)^k \sum_{i=0}^k \binom{2k}{2i} \sum_{j=0}^i \binom{i}{j} \omega_{k+j-i} \left(\frac{\alpha}{2}\right)^{2(k+j-i)},$$

onde  $\omega_r = E[\{g(U)\}^r]$  denota o  $r$ -ésimo momento da distribuição de  $g(U)$ .

*Demonstração.* Segue pela representação estocástica em (3.7) e pela aplicação do teorema binomial.  $\square$

Então, segue que, pela Proposição 5, se  $T \sim \text{QSBS}(\alpha, Q; H)$  a esperança, variância, e os coeficientes de variação (CV), assimetria (CS), curtose (SK), são dadas, respectivamente, por

$$\begin{aligned} E[T] &= \frac{2Q}{\gamma_\alpha^2} (2 + \omega_1 \alpha^2), \quad \text{Var}[T] = \frac{4Q^2 \alpha^2}{\gamma_\alpha^4} [\omega_1 + (2\omega_2 - \omega_1^2) \alpha^2], \\ \text{CV}[T] &= \frac{\alpha \sqrt{4\omega_1 + (2\omega_2 - \omega_1^2) \alpha^2}}{2 + \omega_1 \alpha^2}, \\ \text{CS}[T] &= \frac{4\alpha [3(\omega_2 - \omega_1^2) + (2\omega_3 - 3\omega_1 \omega_2 + \omega_1^3) \alpha^2 / 2]}{[4\omega_1 + (2\omega_2 - \omega_1^2) \alpha^2]^{3/2}}, \\ \text{SK}[T] &= \frac{16\omega_2 + (32\omega_3 - 48\omega_1 \omega_2 + 24\omega_1^3) \alpha^2 + (8\omega_4 - 16\omega_1 \omega_3 + 12\omega_1^2 \omega_2 - \omega_1^4) \alpha^4}{[4\omega_1 + (2\omega_2 - \omega_1^2) \alpha^2]^2}, \end{aligned}$$

onde  $\omega_r = E[\{g(U)\}^r]$  denota o  $r$ -ésimo momento da distribuição de  $g(U)$ .

### 3.2.3 Casos especiais das distribuições QSBS

Nesta seção iremos apresentar alguns casos particulares da família de distribuições QSBS. Dentre alguns desses casos estão as distribuições normal contaminada, slash e t-student. Onde estes são obtidos pela representação estocástica em (3.7). A seguir são apresentadas alguns casos da distribuição QSBS, além da distribuição condicional  $U|T = t$ , bem como a esperança condicional para cada caso.

### Distribuição normal contaminada Birnbaum-Saunders

Considere o caso em que  $T \sim \text{QSBS}(\alpha, Q; H)$ , com  $H$  sendo a FDA de uma v.a discreta  $U$ , com FDP da seguinte forma

$$h_U(u) = \nu \mathbb{I}_{\{u=\delta\}}(u) + [1 - \nu] \mathbb{I}_{\{u=1\}}(u), \quad 0 < \nu < 1, \quad 0 < \delta < 1,$$

onde  $\mathbb{I}_{\{A\}}(\cdot)$  é uma função indicadora do conjunto  $A$ . Segue que, por (3.2) e (3.8), temos que a FDP de  $T$  é dada por

$$f_T(t) = \left[ \nu \sqrt{\delta} \phi(\sqrt{\delta} a(t)) + [1 - \nu] \phi(a(t)) \right] \frac{1}{\alpha \gamma_\alpha \sqrt{4Qt}} \left( \frac{\gamma_\alpha^2}{2} + \frac{2Q}{t} \right), \quad t > 0,$$

onde  $\phi(\cdot)$  é a FDP de uma distribuição normal padrão. Denotamos  $T \sim \text{CN-BS}(\alpha, Q, \boldsymbol{\nu})$ , com  $\boldsymbol{\nu} = (\nu, \delta)$ . A FDP da distribuição condicional  $U|T = t$  é dada por

$$h_{U|T}(u|t) = \nu p(t, u) \mathbb{I}_{\{u=\delta\}}(u) + [1 - \nu] p(t, u) \mathbb{I}_{\{u=1\}}(u),$$

de tal forma que obtemos  $p(t, u)$  calculando

$$p(t, u) = \frac{\sqrt{u} \exp\left(-\frac{ua(t)^2}{2}\right)}{\nu \sqrt{\delta} \exp\left(-\frac{\delta a(t)^2}{2}\right) + [1 - \nu] \exp\left(-\frac{a(t)^2}{2}\right)},$$

de onde obtemos a esperança condicional de  $U|T = t$

$$\mathbb{E}[U|T = t] = \frac{1 - \nu + \nu \delta^{3/2} \exp\left(\frac{[1-\delta]a(t)^2}{2}\right)}{1 - \nu + \nu \sqrt{\delta} \exp\left(\frac{[1-\delta]a(t)^2}{2}\right)}. \quad (3.10)$$

### Distribuição Slash Birnbaum-Saunders

Considere o caso em que  $T \sim \text{QSBS}(\alpha, Q; H)$ , onde  $H$  é a FDA de uma v.a  $U \sim \text{Beta}(\nu, 1)$ , com FDP dada por

$$h_U(u) = \nu u^{\nu-1}, \quad 0 < u < 1, \quad \nu > 0. \quad (3.11)$$

Donde segue que, por (3.2), (3.8) e (3.11), a FDP de  $T$  é dada por

$$f_T(t) = \left[ \nu \int_0^1 u^{\nu-1} \phi(a(t); 0, u^{-1}) du \right] \frac{1}{\alpha \gamma_\alpha \sqrt{4Qt}} \left( \frac{\gamma_\alpha^2}{2} + \frac{2Q}{t} \right), \quad t > 0,$$

neste caso, denotamos  $T \sim \text{SL-BS}(\alpha, Q, \nu)$ . Além disso, a distribuição condicional  $U|T = t \sim \text{Gama}(1/2 + \nu, a(t)^2/2)$  truncada no intervalo  $[0,1]$ . Dessa forma, obtemos a esperança condicional

$$E[U|T = t] = \left( \frac{1 + 2\nu}{a(t)^2} \right) \frac{P_1\left(\frac{3}{2} + \nu, \frac{a(t)^2}{2}\right)}{P_1\left(\frac{1}{2} + \nu, \frac{a(t)^2}{2}\right)}, \quad (3.12)$$

onde  $P_x(a, b)$  é a FDA de uma distribuição Gama, com parâmetros  $a$  e  $b$  avaliada em  $x$ , com parametrização de acordo com a FDP em (3.13).

### Distribuição t-Student Birnbaum-Saunders

Considere o caso em que  $T \sim \text{QSBS}(\alpha, Q; H)$ , em que  $H$  é FDA de uma v.a  $U \sim \text{Gama}(\nu/2, \nu/2)$ , com FDP dada por:

$$h_U(u) = \frac{(\nu/2)^{\nu/2} u^{\nu/2-1}}{\Gamma(\nu/2)} \exp\left(-\frac{\nu u}{2}\right), \quad u > 0, \nu > 0. \quad (3.13)$$

Segue que, por (3.2), (3.8) e (3.13), temos que a FDP de  $T$  é da seguinte forma

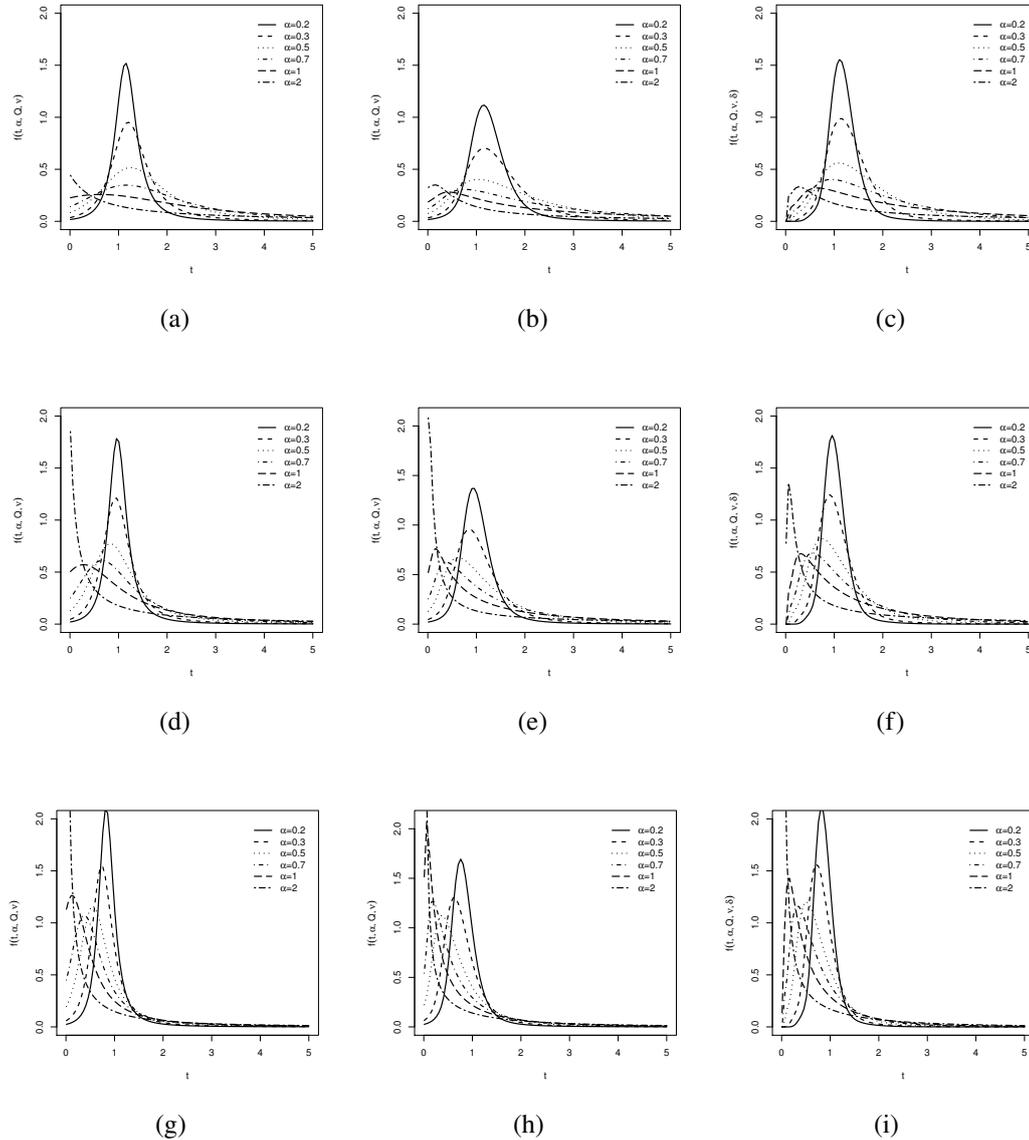
$$f_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi} \sqrt{\nu} \Gamma(\frac{\nu}{2})} \left[ 1 + \frac{4Q}{\nu t \alpha^2 \gamma_\alpha^2} \left( \frac{\gamma_\alpha^2 t}{4Q} - 1 \right)^2 \right]^{-\frac{\nu+1}{2}} \frac{1}{\alpha \gamma_\alpha \sqrt{4Qt}} \left( \frac{\gamma_\alpha^2}{2} + \frac{2Q}{t} \right), \quad t > 0, \quad (3.14)$$

denotamos  $T \sim t_\nu\text{-BS}(\alpha, Q, \nu)$ , se  $T$  tem FDP como em (3.14). Neste caso, a distribuição condicional  $U|T = t \sim \text{Gama}([\nu + 1]/2, [\nu + a(t)^2]/2)$ , com esperança dada por

$$E[U|T = t] = \frac{\nu + 1}{\nu + a(t)^2}. \quad (3.15)$$

### 3.2.4 Análise da forma das distribuições QSBS

**Figura 3.1:** FDP de alguns membros da família de distribuições QSBS, considerando diferentes valores de  $\alpha$  e  $Q = 1$  fixado, para os valores de  $q = 0.25, 0.5, 0.75$ , (nas linhas) e para os modelos  $t_\nu$ -BS, SL-BS e CN-BS (nas colunas), com  $\nu = 2, 2, (0.2, 0.2)$ , respectivamente.



Na Figura 3.1 estão alguns gráficos da função de densidade de probabilidade dos casos especiais da distribuição QSBS mencionados aqui, o vetor de parâmetros. Para verificar de forma mais precisa o comportamento da distribuição, utilizamos diferentes valores para o parâmetro de forma  $\alpha$  e também os quantis  $q = 0.25, 0.5, 0.75$ , além disso, o parâmetro  $Q$  foi fixado em 1. Podemos observar que à medida que aumentamos  $q$ , a curva da distribuição se torna mais achatada, outro detalhe importante é a influência do parâmetro de forma na cauda da distribuição,

quanto maior o valor de  $\alpha$ , mais pesada é a cauda da distribuição.

### 3.3 Um modelo de regressão quantílica QSBS

Seja  $T_1, T_2, \dots, T_n$  uma amostra aleatória, onde  $T_i \sim \text{QSBS}(\alpha, Q_i; H)$ , para  $i = 1, 2, \dots, n$ , e as suas respectivas observações  $\mathbf{t} = (t_1, t_2, \dots, t_n)^\top$ . Então, definimos o modelo com componente sistemático dada por

$$h(Q_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i = 1, 2, \dots, n, \quad (3.16)$$

tal que, a função de ligação  $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  é estritamente monótona, positiva e pelo menos duas vezes diferenciável,  $Q_i = h^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ , com  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ , o vetor contendo  $(p + 1) < n$  coeficientes desconhecidos e  $\mathbf{x}_i^\top = (1, x_{i1}, \dots, x_{ip})$  suas respectivas covariáveis. De maneira análoga utilizada por Balakrishnan et al. (2009), podemos obter as estimativas de máxima verossimilhança do vetor de parâmetros  $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})^\top$  que indexa o modelo de regressão quantílica utilizando o algoritmo EM, Dempster, Laird e Rubin (1977), onde  $\boldsymbol{\nu}$  é o vetor de parâmetros que indexa a distribuição de  $H$ .

Sejam  $\ell_c(\boldsymbol{\theta}; \mathbf{t}_c)$  e  $\mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}; \mathbf{t}_c); \mathbf{t}, \hat{\boldsymbol{\theta}}]$  que denotam a função de log-verossimilhança completa e o seu valor esperado condicionado aos dados observados, respectivamente. Dessa forma, podemos descrever o processo iterativo utilizando o algoritmo EM da seguinte maneira. Cada iteração do algoritmo EM envolve dois passos, sendo eles, o passo expectativa (passo-E) e o passo maximização (passo-M), no qual definimos:

**Passo-E.** Calcule  $\mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}^{(r)})$ ,  $r = 1, 2, \dots$

**Passo-M.** Encontre  $\boldsymbol{\theta}^{(r+1)}$  tal que  $\mathcal{Q}(\boldsymbol{\theta}^{(r+1)}; \hat{\boldsymbol{\theta}}^{(r)}) = \max_{\boldsymbol{\theta} \in \Theta} \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}^{(r)})$ ,  $r = 1, 2, \dots$

De maneira análoga a propriedade apresentada em Balakrishnan et al. (2009), é fácil mostrar que, se tivermos  $T \sim \text{QSBS}(\alpha_u, Q, H)$ , com  $\alpha_u = \sqrt{g(u)}\alpha$ , então  $T|U = u \sim \text{QBS}(\sqrt{g(u)}\alpha, Q)$ , em que QBS denota a distribuição BS reparametrizada pelo quantil proposta por Sánchez et al. (2020a). Dessa forma podemos definir a seguinte representação hierárquica

como

$$T_i | (U_i = u_i) \stackrel{\text{ind}}{\sim} \text{QBS}(\sqrt{g(u_i)}\alpha, Q_i), \quad (3.17)$$

$$U_i \stackrel{\text{ind}}{\sim} h_U(u_i), \quad i = 1, 2, \dots, n. \quad (3.18)$$

Vale ressaltar que o parâmetro  $\nu$  que indexa a FDP de  $h_U(\cdot)$  é considerado conhecido. Uma forma de obter um valor ótimo para  $\nu$  pode ser feito utilizando o critério de informação Schwarz, como pode ser visto em Spiegelhalter et al. (2002). Então, definimos a representação hierárquica com base em (3.17) e (3.18), temos que a função de log-verossimilhança completa com base em  $\mathbf{t}_c = [\mathbf{t}^\top, \mathbf{u}^\top]^\top$  é dada por

$$\begin{aligned} \ell_c(\boldsymbol{\theta}; \mathbf{t}_c) \propto & -n \log(\alpha\gamma_\alpha) - \frac{1}{2} \sum_{i=1}^n \log(Q_i) + \sum_{i=1}^n \log\left(\frac{\gamma_\alpha^2}{2} + \frac{2Q_i}{t_i}\right) \\ & - \frac{2}{(\alpha\gamma_\alpha)^2} \sum_{i=1}^n \frac{1}{g(u_i)} \frac{Q_i}{t_i} \left(\frac{\gamma_\alpha^2 t_i}{4Q_i} - 1\right)^2. \end{aligned}$$

Tomando a esperança condicional da função de log-verossimilhança completa e substituindo  $\hat{u}_i = E[1/g(U_i); t_i, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}]$ ,  $i = 1, 2, \dots, n$ , a qual obtemos de acordo com a distribuição condicional de  $g(U)$ , por meio de (3.10), (3.12) ou (3.15). Então, temos

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}) \propto & -n \log(\alpha\gamma_\alpha) - \frac{1}{2} \sum_{i=1}^n \log(Q_i) + \sum_{i=1}^n \log(\gamma_\alpha^2 t_i + 4Q_i) \\ & - \frac{2}{(\alpha\gamma_\alpha)^2} \sum_{i=1}^n \hat{u}_i \frac{Q_i}{t_i} \left(\frac{\gamma_\alpha^2 t_i}{4Q_i} - 1\right)^2. \end{aligned} \quad (3.19)$$

A partir disso, as estimativas de máxima verossimilhança dos parâmetros do modelo de regressão são obtidas realizando da seguinte forma:

**Passo-E.** Dado  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , calcular  $\hat{u}_i$ , para  $i = 1, 2, \dots, n$ .

**Passo-M.** Atualizar  $\hat{\boldsymbol{\theta}}$  maximizando  $\mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})$  com respeito a  $\boldsymbol{\theta}$ . A qual podemos maximizar derivando a Equação (3.19) em relação a  $\boldsymbol{\theta} = (\alpha, \beta_0, \dots, \beta_p)$ , igualado a zero e resolvendo o sistema de equações. As derivadas parciais de (3.19) em relação a  $\boldsymbol{\theta}$  são dadas, respectivamente,

por

$$\begin{aligned} \frac{\partial \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})}{\partial \alpha} &= -n \left( \frac{\gamma_\alpha + \alpha \gamma'_\alpha}{\alpha \gamma_\alpha} \right) + 2\gamma_\alpha \gamma'_\alpha \sum_{i=1}^n \frac{t_i}{\gamma_\alpha^2 t_i + 4Q_i} + \frac{1}{\alpha^3} \sum_{i=1}^n \hat{u}_i \left[ \frac{\gamma_\alpha^2 t_i}{4Q_i} + \frac{4Q_i}{\gamma_\alpha^2 t_i} - 2 \right] \\ &\quad - \frac{1}{\alpha^2} \sum_{i=1}^n \hat{u}_i \left[ \frac{\gamma_\alpha \gamma'_\alpha t_i}{4Q_i} - \frac{4Q_i \gamma'_\alpha}{\gamma_\alpha^3 t_i} \right], \quad (3.20) \\ \frac{\partial \mathcal{Q}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})}{\partial \beta_j} &= \sum_{i=1}^n \left\{ \left[ -\frac{1}{2Q_i} + \frac{4}{\gamma_\alpha^2 t_i + 4Q_i} - \frac{\hat{u}_i}{\alpha^2} \left( -\frac{\gamma_\alpha^2 t_i}{8Q_i^2} + \frac{2}{\gamma_\alpha^2 t_i} \right) \right] \left( \frac{\partial Q_i}{\partial \eta_i} \right) \left( \frac{\partial \eta_i}{\partial \beta_j} \right) \right\}, \quad (3.21) \end{aligned}$$

com  $j = 0, 1, \dots, p$  e  $\gamma'_\alpha = y_q + \alpha y_q^2 / \sqrt{(\alpha y_q)^2 + 4}$  sendo a primeira derivada de  $\gamma_\alpha$  em relação a  $\alpha$ . Entretanto, não é possível resolver o sistema de equações de forma analítica, fazendo-se necessário a utilização de algum método tipo-Newton no passo-M; ver Jamshidian e Jennrich (1997) e McLachlan e Krishnan (2007). A convergência do algoritmo EM pode ser estabelecida utilizando a seguinte regra de parada, se  $|\mathcal{Q}(\boldsymbol{\theta}^{(r+1)}; \hat{\boldsymbol{\theta}}^{(r)}) - \mathcal{Q}(\boldsymbol{\theta}^{(r)}; \hat{\boldsymbol{\theta}}^{(r)})| < \varepsilon$ , onde  $\varepsilon$  é o valor de tolerância; ver McLachlan e Krishnan (2007).

### 3.3.1 Aproximação para os erros padrão

Uma desvantagem do algoritmo EM em relação aos métodos tipo-Newton é que no primeiro, não podemos obter as estimativas dos erros padrão dos estimadores por meio da matriz de informação de Fisher de forma direta. Existem diversas abordagens propostas para a obtenção dos erros padrão dos estimadores de máxima verossimilhança, como Baker (1992), Oakes (1999) e Louis (1982), entre outras propostas.

A ideia de Louis (1982) se baseia no princípio da falta de informação, o qual foi desenvolvido pelo autor. De acordo com o princípio da falta de informação de Louis (1982), a função score de uma log-verossimilhança dos dados incompletos se relaciona com a esperança condicional de uma log-verossimilhança dos dados completos da seguinte forma: seja  $S_o(\mathbf{y}; \boldsymbol{\theta}) = E[\partial \ell_c(\boldsymbol{\theta}; \mathbf{t}) / \partial \boldsymbol{\theta}]$ , onde  $S_o(\mathbf{y}; \boldsymbol{\theta}) = \partial \ell_o(\boldsymbol{\theta}; \mathbf{t}) / \partial \boldsymbol{\theta}$  e  $S_c(\mathbf{y}; \boldsymbol{\theta}) = \partial \ell_c(\boldsymbol{\theta}; \mathbf{t}) / \partial \boldsymbol{\theta}$ , são as funções scores dos dados incompletos e completos, respectivamente. Meilijson (1989) apresenta uma definição da matriz de informação empírica como segue

$$\mathbf{I}_e(\boldsymbol{\theta}; \mathbf{t}) = \sum_{i=1}^n s(t_i; \boldsymbol{\theta}) s^\top(t_i; \boldsymbol{\theta}) - \frac{1}{n} S(\mathbf{t}; \boldsymbol{\theta}) S^\top(\mathbf{t}; \boldsymbol{\theta}),$$

tal que  $S(\mathbf{t}; \boldsymbol{\theta}) = \sum_{i=1}^n s(t_i, \boldsymbol{\theta})$  e  $s(t_i, \boldsymbol{\theta})$  é denominada como a função score empírica da  $i$ -ésima observação,  $i = 1, 2, \dots, n$ . Substituindo  $\boldsymbol{\theta}$  por suas respectivas estimativas de máxima verossimilhança  $\hat{\boldsymbol{\theta}}$ , temos que  $S(\mathbf{t}; \hat{\boldsymbol{\theta}}) = 0$ , então a matriz de informação empírica se resume apenas

$$\mathbf{I}_e(\hat{\boldsymbol{\theta}}; \mathbf{t}) = \sum_{i=1}^n s(t_i; \hat{\boldsymbol{\theta}}) s^\top(t_i; \hat{\boldsymbol{\theta}}),$$

em que  $s(t_i; \hat{\boldsymbol{\theta}})^{(r)}$  é a função score empírica obtida na  $r$ -ésima iteração utilizando o algoritmo EM, a qual é dada por

$$s(t_i; \hat{\boldsymbol{\theta}})^{(r)} = \mathbb{E}[s(t_i, u_i^{(r)}; \boldsymbol{\theta}^{(r)}) | t_i],$$

onde  $u_i^{(r)}$  é a variável latente obtida na  $r$ -ésima iteração com distribuição condicional  $h_{U|T}(u_i | t_i)$  e as derivadas parciais da função de log-verossimilhança completa em relação a  $\alpha$  e  $\beta$  são dadas, respectivamente, por (3.20) e (3.21).

Pelo método de Louis (1982), uma aproximação para a matriz de informação observada na  $r$ -ésima iteração pode ser obtida como  $\mathbf{I}_e(\hat{\boldsymbol{\theta}}; \mathbf{t})^{(r)} = \sum_{i=1}^n s(t_i; \hat{\boldsymbol{\theta}})^{(r)} s^\top(t_i; \hat{\boldsymbol{\theta}})^{(r)}$ . Dessa forma, uma aproximação da matriz de covariância é obtida invertendo a matriz de informação de Fisher empírica  $\mathbf{I}_e(\hat{\boldsymbol{\theta}}; \mathbf{t})^{(r)}$  após a convergência.

### 3.3.2 Critérios de seleção de modelos

Os critérios de informação são bastante utilizados na seleção do modelo de regressão que forneça o melhor ajuste a um conjunto de dados. Existem diversos critérios de informação utilizados na seleção de modelos. Mais especificamente, quando usamos o algoritmo EM para obter as estimativas de máxima verossimilhança, substituímos a função de log-verossimilhança pelo valor esperado da função de log-verossimilhança. Além disso, os critérios de informação podem ser utilizados na estimação dos parâmetros extras associados a um modelo representado por uma estrutura hierárquica; ver Spiegelhalter et al. (2002).

O critério de informação proposto por Akaike (1974) (AIC) é um dos critérios de informação mais utilizados. Este se baseia na existência de um “modelo verdadeiro” e sob essa suposição, dentre uma lista de modelos postulados, o melhor modelo que apresenta a menor divergência

em relação ao “modelo verdadeiro”, de tal forma que o AIC é obtido da seguinte forma:

$$\text{AIC} = -2Q(\hat{\theta}; \hat{\theta}) + 2p,$$

onde  $Q(\hat{\theta}, \hat{\theta})$  é a função de verossimilhança completa maximizada e  $p$  representa o número de parâmetros que indexa o modelo de regressão.

Nos trabalhos de Sugiura (1978) e Sakamoto, Ishiguro e Kitagawa (1986), foi verificado que o AIC é viesado em situações que o modelo apresenta um elevado número de parâmetros. Buscando resolver esse problema Bozdogan (1987) propôs uma correção ao AIC, o qual é definido da seguinte forma:

$$\text{AICc} = \text{AIC} + \frac{2p(p+1)}{n-p-1},$$

onde  $n$  é o tamanho da amostra. O Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978), parte do pressuposto da existência de um “modelo verdadeiro” donde o BIC é definido como a estatística que maximiza a probabilidade do modelo postulado ser o melhor modelo dentre todos os modelos candidatos. O BIC é definido sendo

$$\text{BIC} = -2Q(\hat{\theta}; \hat{\theta}) + p \log(n).$$

Uma alternativa aos critérios AIC e BIC é o critério de informação de Hannan-Quinn (HIC). Este foi proposto por Hannan e Quinn (1979) inicialmente na utilizados na seleção de modelos de séries temporais. O HIC é definido como

$$\text{HIC} = -2Q(\hat{\theta}; \hat{\theta}) + 2p \log(\log(n)).$$

Para todos os critérios de informação apresentados, o melhor modelo é aquele que apresenta o menor critério de informação, isto é, que apresenta menor divergência em relação ao “modelo verdadeiro”.

### 3.3.3 Resíduos

Os resíduos são uma importante ferramenta na análise da qualidade de ajuste apresentada por um modelo postulado. Alguns desses resíduos são bastante utilizados principalmente na análise

do ajuste global provido pelo modelo, como é o caso dos resíduos Generalizados Cox-Snell (GCS) e Quantil Aleatorizado (RQ). Os dois resíduos citados anteriormente são muito utilizados na avaliação da qualidade de ajuste de alguns modelos de regressão Birnbaum-Saunders; ver Dasilva et al. (2020).

Os resíduos de Generalizados Cox-Snell são utilizados na avaliação da qualidade do ajuste do modelo de regressão principalmente para modelos de análise de sobrevivência Lee e Wang (2003). Estes resíduos foram estudados inicialmente por Cox e Snell (1968). Podemos obter os resíduos GCS por meio da função de risco acumulado sob os parâmetros estimados do modelo da seguinte forma

$$\hat{r}_i^{\text{GCS}} = -\log[1 - \Phi_{\text{SMN}}(a(t))], \quad i = 1, 2, \dots, n,$$

onde  $\Phi_{\text{SMN}}(a(t))$  é a função de distribuição acumulada de uma v.a.  $T \sim \text{QSBS}(\hat{\alpha}, \hat{\beta}, \nu)$ .

O outro resíduo utilizado na avaliação da qualidade do ajuste provida pelo modelo, são os resíduos Dunn e Smyth (1996). Esses resíduos comumente são utilizados na avaliação de modelos aditivos generalizados para locação, escala e forma. Podemos obter os resíduos RQ por meio dos quantis de uma distribuição normal padrão da seguinte forma

$$\hat{r}_i^{\text{RQ}} = \Phi^{-1}[\Phi_{\text{SMN}}(a(t))], \quad i = 1, 2, \dots, n,$$

onde  $\Phi(\cdot)^{-1}$  é a função de distribuição acumulada inversa. Dizemos que o modelo postulado é adequado quando os resíduos GCS e RQ seguem uma distribuição exponencial padrão e normal padrão, respectivamente. Para ambos os resíduos, o pressuposto da distribuição pode ser verificado por meio de técnicas gráficas, testes de hipóteses e estatísticas descritivas.

### 3.3.4 Testes de hipóteses

Diversos testes de hipóteses são utilizados na análise da significância das covariáveis para o modelo, dentre esses testes, há três testes de hipóteses que se destacam para essa finalidade, sendo eles os testes Wald (1947), Score (Rao, 1948) e razão de verossimilhança (Wilks, 1938). Os três testes mencionados anteriormente se baseiam nos estimadores de máxima verossimilhança, o que permite a utilização de uma série de propriedades assintóticas dos estimadores. Um outro teste

que apresenta propriedades bem semelhantes aos três testes citados é o teste gradiente proposto por Terrell (2002), a estatística do teste de gradiente foi obtida por Terrel por meio de uma modificação na estatística do teste Score. As estatísticas wald, Score, razão de verossimilhança e gradiente para testar  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  são dadas, respectivamente, por:

$$\begin{aligned} S_W &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathcal{J}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \\ S_R &= S(\hat{\boldsymbol{\theta}}_0)^\top \mathcal{J}(\hat{\boldsymbol{\theta}})^{-1} S(\hat{\boldsymbol{\theta}}_0), \\ S_{LR} &= -2[\ell(\hat{\boldsymbol{\theta}}_0) - \ell(\hat{\boldsymbol{\theta}})], \\ S_T &= S(\hat{\boldsymbol{\theta}}_0)^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \end{aligned}$$

onde  $\ell(\hat{\boldsymbol{\theta}}_0)$ ,  $\ell(\hat{\boldsymbol{\theta}})$ ,  $S(\hat{\boldsymbol{\theta}}_0)$  e  $\mathcal{J}(\hat{\boldsymbol{\theta}})$  representam a função log-verossimilhança esperada maximizada sob  $H_0$  e  $H_1$ , a função score sob  $H_0$  e a matriz de informação de observada de Fisher, respectivamente. Em particular, neste caso em que utilizamos o algoritmo EM para obter as estimativas de máxima verossimilhança, a função de log-verossimilhança na estatística  $S_{LR}$  é substituída pela esperança da função de log-verossimilhança, nas estatísticas  $S_W$ ,  $S_R$  e  $S_T$  a função score e a matriz de informação de Fisher são aproximadas utilizando as suas respectivas versões empíricas apresentadas na Seção 3.3.1.

Sob  $H_0$ , as estatísticas  $S_W$ ,  $S_R$ ,  $S_{LR}$ ,  $S_T$  seguem uma distribuição  $\chi_\kappa^2$ , com  $n \rightarrow \infty$  e  $\kappa$  sendo o número de parâmetros fixados na hipótese nula. Vale ressaltar que rejeitamos  $H_0$  se o valor observado da estatística de teste for maior que  $\chi_{\kappa, 1-\alpha}^2$ , onde  $\chi_{\kappa, 1-\alpha}^2$  é o quantil de ordem  $1 - \alpha$  de uma distribuição Qui-quadrado com  $\kappa$  graus de liberdade.

### 3.4 Estudos de simulação

Nesta Seção serão apresentados os resultados obtidos nos estudos de simulação de Monte Carlo. As simulações foram realizadas utilizando o software R, o qual é de domínio público. Os estudos foram realizados utilizando 5.000 réplicas de Monte Carlo. Além disso, foram considerados os tamanhos amostrais  $n = 50, 100, 200$  e os quantis  $q = 0.25, 0.5, 0.75$ , com função de ligação logarítmica.

### 3.4.1 Estudo de simulação 1

Considere o modelo de regressão quantílica:

$$\log(Q_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \quad i = 1, \dots, n,$$

foram fixados os seguintes valores  $\beta_0 = 2.5$ ,  $\beta_1 = 3$  e  $\beta_2 = 0.9$ , com as covariáveis simuladas de acordo com uma distribuição Uniforme(0,1). Além disso, foram considerados os seguintes valores do parâmetro de forma  $\alpha = 0.2, 0.5, 1.0$  e os quantis  $q = 0.25, 0.5, 0.75$ . Os parâmetros extras, indexados as distribuições CN-BS, SL-BS e  $t_\nu$ -BS, são, respectivamente,  $\nu = (0.1, 0.3)$ ,  $\nu = 4$  e  $\nu = 11$ . Nesse estudo de simulação de Monte Carlo foram avaliadas as performances dos estimadores de máxima verossimilhança usando o algoritmo EM sob diferentes critérios como o Viés, o Erro Quadrático Médio (EQM) e Probabilidade de Cobertura (CP). As estimativas de Monte Carlo para o Viés e o EQM dos estimadores de máxima verossimilhança são dados, respectivamente, por

$$\text{Viés}(\hat{\varphi}) = \frac{1}{M} \sum_{r=1}^M (\hat{\varphi}^{(r)} - \varphi), \quad \text{EQM}(\hat{\varphi}) = \frac{1}{M} \sum_{r=1}^M (\hat{\varphi}^{(r)} - \varphi)^2,$$

onde  $M$  representa o número de replicas utilizadas no estudo de Monte Carlo,  $\varphi$  o parâmetro verdadeiro e  $\hat{\varphi}^{(r)}$  sua  $r$ -ésima estimativa de máxima verossimilhança. Por sua vez, a probabilidade de cobertura (CP) do intervalo de confiança (95%) assintótico, é dada por

$$\text{CP}(\hat{\varphi}) = \frac{1}{M} \sum_{r=1}^M I(\varphi \in [\hat{\varphi}_L^{(r)}, \hat{\varphi}_U^{(r)}]),$$

em que  $I(\cdot)$  é uma função indicadora, tal que  $\varphi$  pertence ao  $r$ -ésimo intervalo assintótico  $[\hat{\varphi}_L^{(r)}, \hat{\varphi}_U^{(r)}]$ , com  $\hat{\varphi}_L^{(r)}$  e  $\hat{\varphi}_U^{(r)}$  os limites inferior e superior, respectivamente.

Além da análise da performance dos estimadores de máxima verossimilhança dos estimadores, esse estudo de simulação de Monte Carlo conta com mais duas análises: uma análise da taxa de acertos com qual os critérios de informação AIC, BIC AICc, HIC e a raiz do erro quadrático médio (REQM) da predição selecionam o modelo correto, isto é, o modelo sob o processo gerador dos dados, a outra análise busca avaliar a suposição da distribuição teórica dos resíduos GCS e RQ por meio de estatísticas descritivas considerando o modelo baseado no processo

gerador dos dados. As etapas realizadas para a obtenção dos resultados do estudo de simulação 1 são descritas no Algoritmo 3.

---

**Algoritmo 3** Passos para o estudo de Monte Carlo 1.

---

- 1: Escolher o modelo de regressão quantílica mistura-escala Birnbaum-Saunders baseado em uma variável latente  $U$  com densidade  $h_U(u)$  das Equações (3.10), (3.11) ou (3.13) e fixar os parâmetros  $\nu$  do modelo.
  - 2: Gerar 5000 amostras baseadas no modelo escolhido.
  - 3: Estimar os parâmetros do modelo pelo método da máxima verossimilhança para cada amostra.
  - 4: Estimar o viés, o EQM e a PC com base nas 5000 réplicas.
  - 5: Computar, para cada réplica, os resíduos GCS e RQ e as respectivas estatísticas descritivas: media, mediana, SD, CS e CK. Computar as médias das estatísticas descritivas com base nas 5000 réplicas.
  - 6: Computar, para cada réplica, os valores dos critérios de informação AIC, BIC, AICc, HIC e REQM da predição para o modelo escolhido e para diferentes modelos considerando  $h_U(u)$  com densidade dada por (3.10), (3.11) ou (3.13). Computar a percentagem que acertadamente os critérios de informação AIC, BIC, AICc, HIC e o REQM escolhem o modelo correto, ou seja, o modelo baseado no processo gerador dos dados.
- 

Nas Tabelas 3.1-3.3 encontram-se o viés, erro quadrático médio (EQM) e a probabilidade de cobertura (CP) para as estimativas de máxima verossimilhança dos parâmetros do modelo de regressão quantílica considerando as distribuições CN-BS, SL-BS e  $t_\nu$ -BS. Pelas Tabelas 3.1-3.3, podemos observar que o Viés e o EQM diminuem a medida que aumentamos o tamanho da amostra, ainda sobre o Viés e o EQM, é possível observar que estes aumentam para todos os parâmetros, a medida que aumentamos o valor do parâmetro  $\alpha$  fixado a priori, em relação a probabilidade de cobertura (CP) foi possível observar que a medida que aumentamos o tamanho amostral, a probabilidade de cobertura do parâmetro tende a se aproximar do nível fixado de 95%, isso fica mais evidente quando observamos o coeficiente  $\beta_0$ , por exemplo.

Nas Tabelas 3.4-3.5 temos um sumário de estatísticas descritivas, contando com a média (MN), mediana (MD) desvio padrão (SD), coeficiente de assimetria (SK) e curtose (CK) para os resíduos GCS e RQ, respectivamente. Pela Tabela 3.4, os modelos para as distribuições CN-BS e  $t_\nu$ -BS, a média e a mediana, são aproximadamente, 1 e 0.69, além disso, o SD dos resíduos GCS é aproximadamente 1, para os dois modelos mencionados, indicando assim que os resíduos GCS se comportam de acordo com uma distribuição exponencial padrão, para o modelo de regressão associado a distribuição SL-BS, os resíduos GCS não se comportam de acordo com uma distribuição exponencial padrão. Pela Tabela 3.5, os modelos baseados nas

distribuições CN-BS e  $t_\nu$ -BS, os resíduos RQ se comportam de acordo com uma distribuição normal padrão, isto é, apresentando média e SD, aproximadamente 0 e 1, respectivamente. De forma análoga aos resíduos GCS, os resíduos RQ para o modelo baseado na distribuição SL-BS não se comportam de acordo com uma distribuição normal padrão.

Na Tabela 3.6 estão dispostas as taxas de acerto para o modelo baseado no processo gerador dos dados considerando os critérios de informação AIC, BIC, AICc, HIC e também a raiz do erro quadrático médio (REQM). Pela Tabela 3.6 podemos observar que a taxa de acertos sob os critérios utilizados não muda de acordo que mudamos o valor do parâmetro de forma  $\alpha$ . Por outro lado, considerando apenas o REQM, este muda de acordo com o valor do quantil de interesse, para os modelos CN-BS e  $t_\nu$ -BS, a taxa de acertos considerando o REQM como critério de escolha do modelo é maior para os quantis  $q = 0.25, 0.75$ , enquanto que para o modelo SL-BS, a taxa de acertos sob o REQM como critério foi maior para  $q = 0.5$  e menor para os quantis de 0.25 e 0.75.

**Tabela 3.1:** Viés, EQM e CP para o modelo de regressão quantílica CN-BS, com  $\nu = 0.1$ ,  $\delta = 0.3$ .

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC	
0.2	0.25	$\hat{\alpha}$	-0.0064	0.0006	0.9440	-0.0030	0.0003	0.9436	-0.0015	0.0001	0.9522
		$\hat{\beta}_0$	0.0053	0.0050	0.9210	0.0028	0.0029	0.9334	0.0007	0.0017	0.9440
		$\hat{\beta}_1$	0.0020	0.0129	0.9360	-0.0013	0.0060	0.9406	0.0007	0.0028	0.9480
		$\hat{\beta}_2$	-0.0021	0.0092	0.9440	0.0004	0.0060	0.9436	0.0001	0.0027	0.9476
	0.5	$\hat{\alpha}$	-0.0063	0.0005	0.9474	-0.0030	0.0003	0.9436	-0.0015	0.0001	0.9522
		$\hat{\beta}_0$	0.0002	0.0087	0.9184	0.0007	0.0028	0.9334	-0.0004	0.0017	0.9440
		$\hat{\beta}_1$	0.0010	0.0133	0.9406	-0.0013	0.0060	0.9416	0.0007	0.0028	0.9466
		$\hat{\beta}_2$	-0.0005	0.0101	0.9476	0.0004	0.0060	0.9436	0.0001	0.0027	0.9476
	0.75	$\hat{\alpha}$	-0.0063	0.0005	0.9474	-0.0030	0.0003	0.9436	-0.0015	0.0001	0.9524
		$\hat{\beta}_0$	-0.0043	0.0090	0.9184	-0.0015	0.0029	0.9334	-0.0014	0.0018	0.9440
		$\hat{\beta}_1$	0.0010	0.0133	0.9404	-0.0013	0.0060	0.9408	0.0007	0.0028	0.9458
		$\hat{\beta}_2$	-0.0005	0.0101	0.9474	0.0004	0.0060	0.9436	0.0001	0.0027	0.9476
0.5	0.25	$\hat{\alpha}$	-0.0162	0.0034	0.9480	-0.0078	0.0017	0.9456	-0.0039	0.0009	0.9520
		$\hat{\beta}_0$	0.0119	0.0536	0.9182	0.0070	0.0171	0.9320	0.0018	0.0104	0.9432
		$\hat{\beta}_1$	0.0022	0.0797	0.9424	-0.0031	0.0360	0.9416	0.0017	0.0169	0.9474
		$\hat{\beta}_2$	-0.0014	0.0607	0.9506	0.0010	0.0361	0.9450	0.0005	0.0165	0.9468
	0.5	$\hat{\alpha}$	-0.0162	0.0034	0.9480	-0.0078	0.0017	0.9456	-0.0039	0.0009	0.9520
		$\hat{\beta}_0$	0.0005	0.0519	0.9182	0.0015	0.0164	0.9320	-0.0010	0.0102	0.9432
		$\hat{\beta}_1$	0.0022	0.0797	0.9434	-0.0031	0.0360	0.9448	0.0017	0.0169	0.9472
		$\hat{\beta}_2$	-0.0014	0.0607	0.9506	0.0010	0.0361	0.9450	0.0005	0.0165	0.9468
	0.75	$\hat{\alpha}$	-0.0162	0.0034	0.9480	-0.0078	0.0017	0.9456	-0.0039	0.0009	0.9520
		$\hat{\beta}_0$	-0.0109	0.0536	0.9182	-0.0039	0.0174	0.9320	-0.0037	0.0109	0.9432
		$\hat{\beta}_1$	0.0022	0.0797	0.9442	-0.0031	0.0360	0.9440	0.0017	0.0169	0.9444
		$\hat{\beta}_2$	-0.0014	0.0607	0.9506	0.0010	0.0361	0.9450	0.0050	0.0165	0.9468
1	0.25	$\hat{\alpha}$	-0.0341	0.0137	0.9530	-0.0163	0.0067	0.9514	-0.0082	0.0035	0.9518
		$\hat{\beta}_0$	0.0245	0.1833	0.9174	0.0138	0.0584	0.9330	0.0036	0.0357	0.9426
		$\hat{\beta}_1$	0.0037	0.2732	0.9498	-0.0055	0.1231	0.9438	0.0030	0.0582	0.9512
		$\hat{\beta}_2$	-0.0028	0.2095	0.9568	0.0020	0.1233	0.9514	0.0018	0.0573	0.9492
	0.5	$\hat{\alpha}$	-0.0341	0.0137	0.9530	-0.0163	0.0067	0.9514	-0.0082	0.0035	0.9518
		$\hat{\beta}_0$	0.0013	0.1772	0.9174	0.0026	0.0558	0.9330	-0.0020	0.0352	0.9424
		$\hat{\beta}_1$	0.0037	0.2732	0.9496	-0.0055	0.1231	0.9506	0.0030	0.0582	0.9490
		$\hat{\beta}_2$	-0.0028	0.2095	0.9568	0.0020	0.1233	0.9514	0.0018	0.0573	0.9492
	0.75	$\hat{\alpha}$	-0.0341	0.0137	0.9530	-0.0163	0.0067	0.9514	-0.0082	0.0035	0.9518
		$\hat{\beta}_0$	-0.0220	0.1834	0.9174	-0.0086	0.0593	0.9330	-0.0076	0.0379	0.9424
		$\hat{\beta}_1$	0.0037	0.2732	0.9496	-0.0055	0.1231	0.9486	0.0030	0.0582	0.9462
		$\hat{\beta}_2$	-0.0028	0.2095	0.9568	0.0020	0.1233	0.9514	0.0018	0.0573	0.9492

**Tabela 3.2:** Viés, EQM e CP para o modelo de regressão quantílica SL-BS, com  $\nu = 4$ .

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC	
0.2	0.25	$\widehat{\alpha}$	-0.0063	0.0005	0.9536	-0.0035	0.0002	0.9502	-0.0016	0.0001	0.9460
		$\widehat{\beta}_0$	0.0035	0.0067	0.9258	0.0041	0.0051	0.9338	0.0007	0.0018	0.9452
		$\widehat{\beta}_1$	0.0010	0.0105	0.9478	-0.0014	0.0067	0.9472	-0.0003	0.0033	0.9448
		$\widehat{\beta}_2$	0.0008	0.0131	0.9516	-0.0015	0.0061	0.9462	0.0008	0.0031	0.9504
	0.5	$\widehat{\alpha}$	-0.0072	0.0005	0.9542	-0.0033	0.0002	0.9498	-0.0019	0.0001	0.9530
		$\widehat{\beta}_0$	-0.0013	0.0067	0.9198	-0.0006	0.0037	0.9382	-0.0003	0.0023	0.9422
		$\widehat{\beta}_1$	0.0035	0.0118	0.9456	0.0016	0.0067	0.9500	-0.0001	0.0037	0.9506
		$\widehat{\beta}_2$	-0.0001	0.0117	0.9468	0.0000	0.0068	0.9516	0.0008	0.0033	0.9472
	0.75	$\widehat{\alpha}$	-0.0071	0.0005	0.9470	-0.0036	0.0002	0.9522	-0.0017	0.0001	0.9540
		$\widehat{\beta}_0$	-0.0041	0.0107	0.9212	-0.0038	0.0033	0.9316	-0.0013	0.0018	0.9418
		$\widehat{\beta}_1$	-0.0017	0.0176	0.9496	0.0003	0.0064	0.9470	-0.0002	0.0031	0.9522
		$\widehat{\beta}_2$	-0.0020	0.0140	0.9498	0.0005	0.0067	0.9486	-0.0006	0.0031	0.9554
0.5	0.25	$\widehat{\alpha}$	-0.0180	0.0031	0.9486	-0.0096	0.0015	0.9514	-0.0044	0.0007	0.9504
		$\widehat{\beta}_0$	0.0158	0.0421	0.9244	0.0074	0.0228	0.9346	0.0039	0.0108	0.9392
		$\widehat{\beta}_1$	-0.0068	0.0647	0.9404	0.0012	0.0327	0.9464	-0.0007	0.0178	0.9506
		$\widehat{\beta}_2$	0.0034	0.0734	0.9568	0.0001	0.0382	0.9494	0.0012	0.0193	0.9484
	0.5	$\widehat{\alpha}$	-0.0183	0.0031	0.9538	-0.0083	0.0015	0.9508	-0.0043	0.0007	0.9552
		$\widehat{\beta}_0$	-0.0013	0.0459	0.9170	0.0004	0.0282	0.9356	0.0004	0.0119	0.9354
		$\widehat{\beta}_1$	-0.0006	0.0813	0.9506	-0.0009	0.0383	0.9530	-0.0004	0.0208	0.9520
		$\widehat{\beta}_2$	-0.0015	0.0649	0.9492	0.0002	0.0427	0.9548	-0.0001	0.0180	0.9446
	0.75	$\widehat{\alpha}$	-0.0182	0.0030	0.9460	-0.0087	0.0014	0.9508	-0.0049	0.0007	0.9516
		$\widehat{\beta}_0$	-0.0122	0.0671	0.9264	-0.0075	0.0264	0.9378	-0.0048	0.0108	0.9432
		$\widehat{\beta}_1$	0.0009	0.0893	0.9442	0.0006	0.0456	0.9514	-0.0015	0.0164	0.9522
		$\widehat{\beta}_2$	-0.0047	0.0767	0.9504	0.0030	0.0353	0.9466	0.0029	0.0190	0.9484
1	0.25	$\widehat{\alpha}$	-0.0414	0.0127	0.9616	-0.0206	0.0061	0.9580	-0.0103	0.0029	0.9530
		$\widehat{\beta}_0$	0.0278	0.1608	0.9196	0.0120	0.0731	0.9284	0.0081	0.0361	0.9442
		$\widehat{\beta}_1$	0.0031	0.2456	0.9554	0.0035	0.1498	0.9566	-0.0028	0.0620	0.9524
		$\widehat{\beta}_2$	0.0041	0.2423	0.9606	0.0039	0.1439	0.9580	0.0017	0.0641	0.9536
	0.5	$\widehat{\alpha}$	-0.0372	0.0129	0.9680	-0.0177	0.0057	0.9580	-0.0086	0.0029	0.9536
		$\widehat{\beta}_0$	-0.0057	0.1361	0.9200	-0.0002	0.0736	0.9388	0.0025	0.0385	0.9394
		$\widehat{\beta}_1$	-0.0088	0.2754	0.9582	-0.0013	0.1190	0.9532	-0.0064	0.0595	0.9524
		$\widehat{\beta}_2$	0.0192	0.2526	0.9618	0.0037	0.1142	0.9554	0.0013	0.0621	0.9528
	0.75	$\widehat{\alpha}$	-0.0398	0.0130	0.9634	-0.0170	0.0059	0.9544	-0.0091	0.0029	0.9528
		$\widehat{\beta}_0$	-0.0289	0.1368	0.9158	-0.0088	0.0708	0.9372	-0.0096	0.0337	0.9446
		$\widehat{\beta}_1$	-0.0062	0.2156	0.9482	-0.0076	0.1237	0.9532	-0.0001	0.0592	0.9492
		$\widehat{\beta}_2$	0.0062	0.2486	0.9608	-0.0019	0.1315	0.9524	0.0053	0.0575	0.9542

**Tabela 3.3:** Viés, EQM e CP para o modelo de regressão quantílica  $t_\nu$ -BS, com  $\nu = 11$ .

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		Viés	EQM	PC	Viés	EQM	PC	Viés	EQM	PC	
0.2	0.25	$\hat{\alpha}$	-0.0064	0.0006	0.9474	-0.0030	0.0003	0.9416	-0.0015	0.0001	0.9522
		$\hat{\beta}_0$	0.0060	0.0112	0.9192	0.0028	0.0028	0.9318	0.0007	0.0017	0.9432
		$\hat{\beta}_1$	-0.0009	0.0144	0.9376	-0.0013	0.0060	0.9404	0.0007	0.0028	0.9486
		$\hat{\beta}_2$	-0.0009	0.0122	0.9426	0.0003	0.0060	0.9426	0.0001	0.0027	0.9468
	0.5	$\hat{\alpha}$	-0.0063	0.0006	0.9468	-0.0030	0.0003	0.9416	-0.0015	0.0001	0.9520
		$\hat{\beta}_0$	0.0003	0.0086	0.9186	0.0007	0.0027	0.9318	-0.0003	0.0017	0.9432
		$\hat{\beta}_1$	0.0009	0.0132	0.9402	-0.0013	0.0060	0.9412	0.0007	0.0028	0.9470
		$\hat{\beta}_2$	-0.0005	0.0101	0.9446	0.0003	0.0060	0.9428	0.0001	0.0027	0.9468
	0.75	$\hat{\alpha}$	-0.0063	0.0006	0.9468	-0.0030	0.0003	0.9416	-0.0015	0.0001	0.9520
		$\hat{\beta}_0$	-0.0041	0.0089	0.9186	-0.0014	0.0029	0.9318	-0.0014	0.0018	0.9432
		$\hat{\beta}_1$	0.0009	0.0132	0.9410	-0.0013	0.0060	0.9402	0.0007	0.0028	0.9446
		$\hat{\beta}_2$	-0.0005	0.0101	0.9446	0.0003	0.0060	0.9428	0.0001	0.0027	0.9468
0.5	0.25	$\hat{\alpha}$	-0.0163	0.0034	0.9462	-0.0078	0.0017	0.9442	-0.0039	0.0009	0.9538
		$\hat{\beta}_0$	0.0121	0.0533	0.9194	0.0072	0.0171	0.9328	0.0018	0.0104	0.9440
		$\hat{\beta}_1$	0.0022	0.0791	0.9436	-0.0030	0.0359	0.9422	0.0016	0.0169	0.9480
		$\hat{\beta}_2$	-0.0012	0.0605	0.9494	0.0007	0.0361	0.9448	0.0004	0.0165	0.9472
	0.5	$\hat{\alpha}$	-0.0164	0.0034	0.9462	-0.0078	0.0017	0.9442	-0.0039	0.0009	0.9538
		$\hat{\beta}_0$	0.0008	0.0516	0.9194	0.0018	0.0164	0.9328	-0.0009	0.0102	0.9438
		$\hat{\beta}_1$	0.0022	0.0791	0.9456	-0.0030	0.0359	0.9438	0.0016	0.0169	0.9480
		$\hat{\beta}_2$	-0.0012	0.0605	0.9494	0.0007	0.0361	0.9448	0.0004	0.0165	0.9474
	0.75	$\hat{\alpha}$	-0.0163	0.0034	0.9462	-0.0078	0.0017	0.9440	-0.0039	0.0009	0.9538
		$\hat{\beta}_0$	-0.0106	0.0532	0.9194	-0.0036	0.0173	0.9328	-0.0035	0.0109	0.9440
		$\hat{\beta}_1$	0.0022	0.0791	0.9444	-0.0030	0.0359	0.9426	0.0016	0.0169	0.9446
		$\hat{\beta}_2$	-0.0012	0.0605	0.9494	0.0007	0.0361	0.9448	0.0004	0.0165	0.9472
1	0.25	$\hat{\alpha}$	-0.0350	0.0139	0.9526	-0.0167	0.0069	0.9518	-0.0084	0.0035	0.9526
		$\hat{\beta}_0$	0.0250	0.1827	0.9190	0.0144	0.0584	0.9340	0.0038	0.0359	0.9440
		$\hat{\beta}_1$	0.0036	0.2717	0.9520	-0.0052	0.1233	0.9432	0.0029	0.0584	0.9504
		$\hat{\beta}_2$	-0.0023	0.2090	0.9570	0.0013	0.1237	0.9508	0.0016	0.0577	0.9488
	0.5	$\hat{\alpha}$	-0.0350	0.0139	0.9526	-0.0167	0.0069	0.9518	-0.0084	0.0035	0.9526
		$\hat{\beta}_0$	0.0015	0.1765	0.9190	0.0032	0.0559	0.9340	-0.0018	0.0355	0.9440
		$\hat{\beta}_1$	0.0036	0.2717	0.9540	-0.0052	0.1233	0.9488	0.0029	0.0584	0.9502
		$\hat{\beta}_2$	-0.0023	0.2090	0.9570	0.0013	0.1237	0.9508	0.0016	0.0577	0.9488
	0.75	$\hat{\alpha}$	-0.0350	0.0139	0.9526	-0.0167	0.0069	0.9518	-0.0084	0.0035	0.9526
		$\hat{\beta}_0$	-0.0220	0.1824	0.9190	-0.0081	0.0593	0.9340	-0.0074	0.0381	0.9440
		$\hat{\beta}_1$	0.0036	0.2717	0.9520	-0.0052	0.1233	0.9474	0.0029	0.0584	0.9468
		$\hat{\beta}_2$	-0.0023	0.2090	0.9570	0.0013	0.1237	0.9508	0.0016	0.0577	0.9488

**Tabela 3.4:** Sumário de estatísticas para os resíduos GCS considerando o modelo de regressão quantílica SBS.

$\alpha$	$q$		$n = 50$			$n = 100$			$n = 200$		
			CN-BS	SL-BS	$t_\nu$ -BS	CN-BS	SL-BS	$t_\nu$ -BS	CN-BS	SL-BS	$t_\nu$ -BS
0.2	0.25	MN	1.0007	0.2249	1.0002	1.0003	3.4373	1.0000	1.0002	2.7172	1.0001
		MD	0.6980	0.0000	0.6964	0.6956	0.7111	0.6952	0.6941	0.2370	0.6940
		SD	1.0008	0.7546	1.0007	0.9994	4.4814	0.9993	1.0000	3.8889	0.9998
		SK	1.6378	4.0803	1.6428	1.7848	1.0108	1.7883	1.8767	1.2936	1.8774
		CK	3.1251	17.7753	3.1549	4.1353	-0.4425	4.1577	4.8594	0.4671	4.8577
	0.5	MN	1.0004	6.9641	0.9999	1.0003	1.1698	1.0000	1.0002	2.8319	1.0001
		MD	0.6974	4.5335	0.6968	0.6956	0.0004	0.6952	0.6941	0.5567	0.6940
		SD	0.9997	6.6111	0.9995	0.9994	2.5135	0.9993	1.0000	3.8723	0.9999
		SK	1.6322	0.5019	1.6373	1.7848	2.3886	1.7883	1.8767	1.2494	1.8774
		CK	3.0959	-1.2344	3.1217	4.1354	4.9818	4.1578	4.8594	0.3308	4.8576
	0.75	MN	1.0004	3.4180	0.9999	1.0003	5.4234	1.0000	1.0002	13.7272	1.0001
		MD	0.6974	1.1057	0.6968	0.6956	3.1935	0.6952	0.6941	14.9827	0.6940
		SD	0.9997	4.3039	0.9995	0.9994	5.6753	0.9993	1.0000	6.0437	0.9999
		SK	1.6322	0.9856	1.6373	1.7849	0.5817	1.7883	1.8768	-0.5336	1.8774
		CK	3.0961	-0.4622	3.1218	4.1356	-1.1559	4.1579	4.8596	-0.6473	4.8578
0.5	0.25	MN	1.0002	6.2318	0.9996	1.0001	5.0268	0.9998	1.0001	1.0149	1.0000
		MD	0.6976	5.4538	0.6970	0.6957	3.8996	0.6953	0.6941	0.2530	0.6939
		SD	0.9978	4.8210	0.9976	0.9984	4.3076	0.9983	0.9995	1.6239	0.9993
		SK	1.6229	0.4127	1.6271	1.7793	0.6738	1.7821	1.8734	2.3450	1.8738
		CK	3.0526	-1.0874	3.0718	4.1048	-0.6312	4.1224	4.8365	6.2482	4.8344
	0.5	MN	1.0001	3.4666	0.9996	1.0001	3.4666	0.9998	1.0001	3.2460	1.0000
		MD	0.6976	2.4009	0.6970	0.6957	2.4009	0.6953	0.6941	2.0450	0.6939
		SD	0.9978	3.3556	0.9976	0.9984	3.3556	0.9983	0.9995	3.3414	0.9993
		SK	1.6229	0.9513	1.6271	1.7793	0.9513	1.7821	1.8734	1.1899	1.8738
		CK	3.0526	0.0021	3.0717	4.1049	0.0021	4.1224	4.8366	0.6885	4.8344
	0.75	MN	1.0001	1.3331	0.9996	1.0001	0.7800	0.9998	1.0001	2.4466	1.0000
		MD	0.6976	0.4540	0.6970	0.6957	0.2002	0.6953	0.6941	1.0708	0.6939
		SD	0.9978	1.9673	0.9976	0.9984	1.3305	0.9983	0.9995	3.0264	0.9993
		SK	1.6229	1.9690	1.6271	1.7794	2.6832	1.7822	1.8734	1.4214	1.8738
		CK	3.0527	3.8322	3.0718	4.1049	8.5732	4.1224	4.8366	1.2562	4.8345
1	0.25	MN	0.9993	1.7884	0.9989	0.9997	0.8007	0.9994	0.9997	2.6806	0.9997
		MD	0.6986	1.0537	0.6979	0.6961	0.3817	0.6957	0.6942	1.7059	0.6941
		SD	0.9921	2.0649	0.9918	0.9956	1.1435	0.9953	0.9978	2.7691	0.9976
		SK	1.5961	1.7839	1.5976	1.7637	2.6123	1.7642	1.8639	1.6326	1.8633
		CK	2.9246	3.3013	2.9240	4.0165	8.8216	4.0183	4.7733	2.5941	4.7663
	0.5	MN	0.9993	1.3375	0.9989	0.9997	1.4969	0.9994	0.9997	0.6229	0.9997
		MD	0.6986	0.7459	0.6979	0.6961	0.8467	0.6957	0.6942	0.2700	0.6941
		SD	0.9921	1.6592	0.9918	0.9956	1.8269	0.9953	0.9978	0.9235	0.9976
		SK	1.5961	1.9805	1.5976	1.7637	2.0812	1.7642	1.8639	2.7777	1.8633
		CK	2.9246	4.4033	2.9239	4.0165	5.0290	4.0183	4.7733	10.5620	4.7663
	0.75	MN	0.9993	1.9122	0.9989	0.9997	3.4851	0.9994	0.9997	0.9887	0.9997
		MD	0.6986	1.0797	0.6979	0.6961	2.3979	0.6957	0.6942	0.4844	0.6941
		SD	0.9921	2.2636	0.9918	0.9956	3.2461	0.9953	0.9978	1.3787	0.9976
		SK	1.5961	1.7622	1.5976	1.7637	1.2772	1.7642	1.8639	2.6184	1.8633
		CK	2.9247	3.0441	2.9240	4.0165	1.1415	4.0184	4.7733	8.8776	4.7663

**Tabela 3.5:** Sumário de estatísticas para os resíduos RQ considerando o modelo de regressão quantílica SBS.

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$		
		CN-BS	SL-BS	$t_\nu$ -BS	CN-BS	SL-BS	$t_\nu$ -BS	CN-BS	SL-BS	$t_\nu$ -BS
0.25	MN	-0.0007	3.6621	-0.0003	-0.0004	-0.0222	-0.0001	-0.0002	0.5812	-0.0002
	MD	-0.0018	4.4962	-0.0002	-0.0010	0.0091	-0.0007	-0.0002	0.8356	-0.0001
	SD	1.0093	2.1911	1.0093	1.0045	3.1032	1.0044	1.0022	3.1333	1.0021
	SK	-0.0054	-0.9419	-0.0057	-0.0023	-0.0018	-0.0028	-0.0035	-0.1484	-0.0033
	CK	-0.2145	-0.1194	-0.2044	-0.1149	-1.4337	-0.1129	-0.0611	-1.4539	-0.0606
0.2	MN	-0.0006	-2.1173	-0.0001	-0.0004	2.2921	-0.0001	-0.0002	0.2326	-0.0002
	MD	-0.0012	-2.2758	-0.0008	-0.0010	3.4442	-0.0007	-0.0002	0.2018	-0.0001
	SD	1.0092	2.6909	1.0092	1.0045	2.8829	1.0044	1.0022	2.9603	1.0021
	SK	-0.0027	0.3635	-0.0030	-0.0023	-0.7669	-0.0028	-0.0035	-0.0082	-0.0033
	CK	-0.2150	-1.0596	-0.2074	-0.1149	-0.7860	-0.1129	-0.0611	-1.3435	-0.0606
0.75	MN	-0.0006	-0.2725	-0.0001	-0.0004	-1.1287	-0.0001	-0.0002	-4.5000	-0.0002
	MD	-0.0012	-0.4018	-0.0008	-0.0010	-1.7211	-0.0007	-0.0002	-4.9832	-0.0001
	SD	1.0092	2.8834	1.0092	1.0045	3.1633	1.0044	1.0022	1.5547	1.0021
	SK	-0.0027	0.0591	-0.0030	-0.0023	0.3594	-0.0029	-0.0035	1.2891	-0.0033
	CK	-0.2149	-1.3973	-0.2074	-0.1149	-1.2754	-0.1129	-0.0611	1.2414	-0.0606
0.25	MN	-0.0006	-2.3923	-0.0001	-0.0004	-1.9300	-0.0001	-0.0001	0.7844	-0.0001
	MD	-0.0012	-2.5996	-0.0007	-0.0010	-2.0355	-0.0006	-0.0001	0.7682	0.0000
	SD	1.0088	1.8164	1.0088	1.0042	1.8304	1.0042	1.0020	1.8929	1.0020
	SK	-0.0021	0.3022	-0.0023	-0.0023	0.2409	-0.0027	-0.0035	-0.0576	-0.0033
	CK	-0.2258	-1.0039	-0.2194	-0.1212	-0.8613	-0.1196	-0.0649	-0.7816	-0.0645
0.5	MN	-0.0006	-0.2474	-0.0001	-0.0004	-1.2189	-0.0001	-0.0001	-1.1277	-0.0000
	MD	-0.0012	-0.3091	-0.0007	-0.0010	-1.3267	-0.0006	-0.0001	-1.1230	0.0000
	SD	1.0088	1.9014	1.0088	1.0042	1.8326	1.0042	1.0020	1.7712	1.0020
	SK	-0.0021	0.1187	-0.0023	-0.0023	0.2699	-0.0027	-0.0035	0.0708	-0.0033
	CK	-0.2258	-0.7672	-0.2194	-0.1212	-0.6692	-0.1196	-0.0649	-0.6857	-0.0645
0.75	MN	-0.0006	0.3435	-0.0001	-0.0004	0.9719	-0.0001	-0.0001	-0.4675	-0.0000
	MD	-0.0012	0.3910	-0.0007	-0.0010	0.9259	-0.0006	-0.0001	-0.3988	0.0000
	SD	1.0088	1.8131	1.0088	1.0042	1.7567	1.0042	1.0020	1.9726	1.0020
	SK	-0.0021	-0.1176	-0.0023	-0.0023	-0.0706	-0.0027	-0.0035	-0.0207	-0.0033
	CK	-0.2258	-0.7264	-0.2194	-0.1212	-0.6657	-0.1196	-0.0649	-0.8315	-0.0645
0.25	MN	-0.0007	-0.4018	-0.0002	-0.0004	0.5441	-0.0000	0.0000	-1.0096	0.0001
	MD	-0.0012	-0.3610	-0.0008	-0.0009	0.4899	-0.0007	0.0001	-0.9013	0.0002
	SD	1.0071	1.4377	1.0074	1.0034	1.3864	1.0035	1.0016	1.4481	1.0016
	SK	-0.0004	-0.0746	-0.0004	-0.0020	0.1054	-0.0022	-0.0034	-0.2055	-0.0032
	CK	-0.2586	-0.2157	-0.2562	-0.1399	-0.0067	-0.1402	-0.0758	-0.2410	-0.0763
1	MN	-0.0007	-0.0396	-0.0002	-0.0004	-0.1923	-0.0000	0.0000	0.8391	0.0001
	MD	-0.0012	-0.0447	-0.0008	-0.0009	-0.1726	-0.0007	0.0001	0.7316	0.0002
	SD	1.0071	1.4219	1.0074	1.0034	1.4142	1.0035	1.0016	1.4253	1.0016
	SK	-0.0004	0.0318	-0.0004	-0.0020	-0.0342	-0.0022	-0.0034	0.2282	-0.0032
	CK	-0.2587	-0.0933	-0.2562	-0.1399	-0.0014	-0.1402	-0.0758	-0.1461	-0.0763
0.75	MN	-0.0007	-0.4664	-0.0002	-0.0004	-1.4258	-0.0000	0.0000	0.3369	0.0001
	MD	-0.0012	-0.4011	-0.0008	-0.0009	-1.3223	-0.0007	0.0001	0.2991	0.0002
	SD	1.0071	1.4966	1.0074	1.0034	1.4880	1.0035	1.0016	1.4284	1.0016
	SK	-0.0005	-0.1262	-0.0004	-0.0020	-0.1390	-0.0022	-0.0034	0.0640	-0.0032
	CK	-0.2586	-0.2693	-0.2562	-0.1399	-0.5011	-0.1402	-0.0758	0.0339	-0.0763

**Tabela 3.6:** Taxa de acerto considerando os critérios de informação AIC,BIC, AICc, HIC e o REQM para o modelo de regressão quantílica.

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		CN-BS	SL-BS	$t_\nu$ -BS	CN-BS	SL-BS	$t_\nu$ -BS	CN-BS	SL-BS	$t_\nu$ -BS	
0.2	0.25	AIC	0.5224	0.0000	0.4678	0.4124	0.0000	0.5620	0.3164	0.0000	0.6264
		BIC	0.5224	0.0000	0.4678	0.4124	0.0000	0.5620	0.3164	0.0000	0.6264
		AICc	0.5224	0.0000	0.4678	0.4124	0.0000	0.5620	0.3164	0.0000	0.6264
		HIC	0.5224	0.0000	0.4678	0.4124	0.0000	0.5620	0.3164	0.0000	0.6264
		REQM	0.1884	0.1786	0.6360	0.2026	0.1062	0.7298	0.1634	0.0424	0.8248
	0.5	AIC	0.5232	0.0000	0.4532	0.4126	0.0000	0.5688	0.3166	0.0000	0.6302
		BIC	0.5232	0.0000	0.4532	0.4126	0.0000	0.5688	0.3166	0.0000	0.6302
		AICc	0.5232	0.0000	0.4532	0.4126	0.0000	0.5688	0.3166	0.0000	0.6302
		HIC	0.5232	0.0000	0.4532	0.4126	0.0000	0.5688	0.3166	0.0000	0.6302
		REQM	0.2300	0.3948	0.3802	0.2618	0.4170	0.3170	0.2634	0.4146	0.3138
	0.75	AIC	0.5244	0.0000	0.4656	0.4126	0.0000	0.5536	0.3168	0.0000	0.6266
		BIC	0.5244	0.0000	0.4656	0.4126	0.0000	0.5536	0.3168	0.0000	0.6266
		AICc	0.5244	0.0000	0.4656	0.4126	0.0000	0.5536	0.3168	0.0000	0.6266
		HIC	0.5244	0.0000	0.4656	0.4126	0.0000	0.5536	0.3168	0.0000	0.6266
		REQM	0.1914	0.2014	0.6084	0.2026	0.1066	0.7284	0.1520	0.0444	0.8374
0.5	0.25	AIC	0.5312	0.0000	0.4526	0.4160	0.0000	0.5598	0.3170	0.0000	0.6270
		BIC	0.5312	0.0000	0.4526	0.4160	0.0000	0.5598	0.3170	0.0000	0.6270
		AICc	0.5312	0.0000	0.4526	0.4160	0.0000	0.5598	0.3170	0.0000	0.6270
		HIC	0.5312	0.0000	0.4526	0.4160	0.0000	0.5598	0.3170	0.0000	0.6270
		REQM	0.2050	0.1924	0.6356	0.2088	0.1030	0.7080	0.1768	0.0518	0.8156
	0.5	AIC	0.5314	0.0000	0.4458	0.4160	0.0000	0.5512	0.3170	0.0000	0.6280
		BIC	0.5314	0.0000	0.4458	0.4160	0.0000	0.5512	0.3170	0.0000	0.6280
		AICc	0.5314	0.0000	0.4458	0.4160	0.0000	0.5512	0.3170	0.0000	0.6280
		HIC	0.5314	0.0000	0.4458	0.4160	0.0000	0.5512	0.3170	0.0000	0.6280
		REQM	0.2436	0.4040	0.3486	0.2720	0.3968	0.3184	0.2936	0.3994	0.3174
	0.75	AIC	0.5314	0.0000	0.4458	0.4160	0.0000	0.5534	0.3170	0.0000	0.6272
		BIC	0.5314	0.0000	0.4458	0.4160	0.0000	0.5534	0.3170	0.0000	0.6272
		AICc	0.5314	0.0000	0.4458	0.4160	0.0000	0.5534	0.3170	0.0000	0.6272
		HIC	0.5314	0.0000	0.4458	0.4160	0.0000	0.5534	0.3170	0.0000	0.6272
		REQM	0.1832	0.2318	0.6450	0.2026	0.1410	0.6966	0.1618	0.0642	0.8188
1	0.25	AIC	0.5510	0.0000	0.4314	0.4252	0.0000	0.5494	0.3204	0.0000	0.6166
		BIC	0.5510	0.0000	0.4314	0.4252	0.0000	0.5494	0.3204	0.0000	0.6166
		AICc	0.5510	0.0000	0.4314	0.4252	0.0000	0.5494	0.3204	0.0000	0.6166
		HIC	0.5510	0.0000	0.4314	0.4252	0.0000	0.5494	0.3204	0.0000	0.6166
		REQM	0.1936	0.2056	0.6140	0.2208	0.1454	0.6748	0.1960	0.0692	0.7626
	0.5	AIC	0.5510	0.0000	0.4372	0.4252	0.0000	0.5460	0.3204	0.0000	0.6234
		BIC	0.5510	0.0000	0.4372	0.4252	0.0000	0.5460	0.3204	0.0000	0.6234
		AICc	0.5510	0.0000	0.4372	0.4252	0.0000	0.5460	0.3204	0.0000	0.6234
		HIC	0.5510	0.0000	0.4372	0.4252	0.0000	0.5460	0.3204	0.0000	0.6234
		REQM	0.2552	0.3924	0.3788	0.2962	0.3912	0.3208	0.3202	0.3982	0.2942
	0.75	AIC	0.5510	0.0000	0.4330	0.4252	0.0000	0.5368	0.3204	0.0000	0.6216
		BIC	0.5510	0.0000	0.4330	0.4252	0.0000	0.5368	0.3204	0.0000	0.6216
		AICc	0.5510	0.0000	0.4330	0.4252	0.0000	0.5368	0.3204	0.0000	0.6216
		HIC	0.5510	0.0000	0.4330	0.4252	0.0000	0.5368	0.3204	0.0000	0.6216
		REQM	0.1988	0.2670	0.5378	0.2092	0.2022	0.6022	0.1754	0.0784	0.7606

### 3.4.2 Estudo de simulação 2

Considere o modelo de regressão quantílica com o seguinte componente sistemático:

$$\log(Q_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}, \quad i = 1, 2, \dots, n,$$

onde os valores dos coeficientes  $\beta_j$ ,  $j = 0, 1, \dots, 3$  não fixados na hipótese nula são iguais 1, enquanto que as covariáveis do modelo são simuladas de acordo com uma distribuição Uniforme(0,1). Neste estudo foram consideradas as estatísticas  $S_W, S_{LR}, S_R, S_T$  na análise do tamanho e do poder do teste para testar  $H_0 : \beta_{3-\kappa+1} = \dots = \beta_3 = 0$ , com  $\kappa = 1, 3$ . A ideia aqui é verificar se a performance dos testes utilizados variam de acordo com o número de parâmetros fixados na hipótese nula. Para o estudo de simulação do tamanho do teste, foram considerados os níveis de significância de 1%, 5% e 10%. O procedimento do estudo de Monte Carlo para obtenção do tamanho e do poder dos testes seguem de acordo com o Algoritmo 4.

---

**Algoritmo 4** Etapas para obtenção do tamanho e poder do teste.

---

- 1: Fixar o número de parâmetros  $\kappa$  sob  $H_0$  e o nível de significância nominal.
  - 2: Fixar os parâmetros e gerar 5000 amostras de tamanho  $n$  baseadas em uma variável latente  $U$  com densidade  $h_U(u)$  das Equações (3.10), (3.11) ou (3.13), de acordo com o modelo postulado em  $H_0$  para o tamanho do teste, ou de acordo com o modelo postulado em  $H_1$  para o poder do teste.
  - 3: Estimar os parâmetros do modelo pelo método da máxima verossimilhança para cada amostra.
  - 4: Calcular as estatísticas de teste  $S_W, S_{LR}, S_R$  e  $S_T$  para cada amostra.
  - 5: Calcular o valor crítico  $\chi_{\kappa, 1-\alpha}^2$  com base no nível de significância  $\alpha$  e no número de parâmetros  $\kappa$  fixados sob  $H_0$ .
  - 6: Obter o tamanho/poder do teste calculando a proporção de réplicas da estatística de teste ( $S_W, S_{LR}, S_R$  e  $S_T$ ) maior que o valor crítico  $\chi_{\kappa, 1-\alpha}^2$ .
- 

Nas Tabelas 3.7-3.9 temos as taxas de rejeição sob a hipótese nula  $H_0 : \beta_3 = 0$  para os modelos CN-BS, SL-BS e  $t_\nu$ -BS, respectivamente. De forma geral, a taxa de rejeição das estatísticas de teste não mudam de acordo com o valor do parâmetro de forma  $\alpha$  fixado para os três modelos de regressão avaliados. Também é possível notar que a medida que aumentamos o tamanho da amostra o tamanho do teste para todas as estatísticas de teste analisados tendem a se aproximar dos seus respectivos valores nominais, com destaque para as estatísticas Wald e gradiente, apresentando taxas de rejeição mais próximas dos níveis de significância nominal fixados a priori mesmo com tamanhos de amostra menores. Por outro lado, a estatística Score,

além de ser a estatística de teste que apresenta a maior taxa de rejeição dentre todas as estatísticas de teste analisadas, a taxa de rejeição para esta estatística, diferentemente das outras analisadas, aumenta quando o quantil é fixado em  $q = 0.25, 0.75$ . Nas Tabelas 3.10-3.12 temos as taxas de rejeição sob a hipótese nula  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$  para os modelos CN-BS, SL-BS e  $t_\nu$ -BS, respectivamente. De forma análoga ao caso em que  $\kappa = 1$ , a taxa de rejeição das estatísticas de teste não mudam de acordo com o valor do parâmetro de forma  $\alpha$  fixado e as estatísticas Wald e gradiente apresentaram maior controle sobre o erro do tipo I, se comparado com as demais estatísticas analisadas, por outro lado, a estatística Score é a que apresenta a maior taxa de rejeição dentre todas analisadas.

Na Figura 3.2 temos as curvas do poder do teste para os modelos CN-BS, SL-BS e  $t_\nu$ -BS considerando as hipóteses  $H_0 : \beta_{3-\kappa+1} = \dots = \beta_3 = 0$  contra  $H_1 : \beta_{3-\kappa+1} = \dots = \beta_3 = \delta$ , com  $|\delta| = 0, 1, 2, 3, 4$  e  $\kappa = 1, 3$ , onde  $\pi(\delta)$  denota a função poder. O nível de significância considerado foi de 1%, com tamanho amostral e quantil  $n = 100$  e  $q = 0.5$ , respectivamente. Vale ressaltar que os resultados observados sobre outras configurações são similares. De um modo geral, podemos observar que não há mudanças no desempenho das estatísticas em relação ao poder, bem como para quaisquer dos modelos CN-BS, SL-BS e  $t_\nu$ -BS, de tal forma que temos  $\pi(\delta) = 1$  para  $|\delta| \geq 1$ .

**Tabela 3.7:** Taxa de rejeição sob a hipótese nula para testar  $H_0 : \beta_3 = 0$ , para o modelo de regressão CN-BS, com  $\nu = 0.1$ ,  $\delta = 0.3$ .

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
0.2	0.25	$S_W$	0.0126	0.0516	0.1026	0.0150	0.0566	0.1024	0.0078	0.0488	0.1040
		$S_{LR}$	0.0214	0.0768	0.1324	0.0196	0.0728	0.1260	0.0120	0.0668	0.1262
		$S_R$	0.0386	0.1080	0.1764	0.0204	0.0782	0.1356	0.0142	0.0630	0.1232
		$S_T$	0.0120	0.0618	0.1156	0.0114	0.0502	0.1038	0.0100	0.0534	0.1026
	0.5	$S_W$	0.0126	0.0514	0.1026	0.0150	0.0566	0.1022	0.0078	0.0488	0.1038
		$S_{LR}$	0.0186	0.0786	0.1386	0.0196	0.073	0.1260	0.0120	0.0668	0.1262
		$S_R$	0.0258	0.0894	0.1486	0.0198	0.0740	0.1354	0.0116	0.0600	0.1102
		$S_T$	0.0086	0.0486	0.1012	0.0082	0.0504	0.1020	0.0126	0.0544	0.1040
	0.75	$S_W$	0.0126	0.0514	0.1024	0.0148	0.0566	0.1020	0.0078	0.0490	0.1042
		$S_{LR}$	0.0186	0.0786	0.1386	0.0196	0.0728	0.1260	0.0120	0.0668	0.1262
		$S_R$	0.0380	0.1068	0.1824	0.0198	0.0758	0.1370	0.0164	0.0736	0.1320
		$S_T$	0.0124	0.0544	0.1112	0.0088	0.0508	0.1012	0.0108	0.0506	0.1038
0.5	0.25	$S_W$	0.0130	0.0482	0.0964	0.0142	0.0554	0.1024	0.0074	0.0488	0.1014
		$S_{LR}$	0.0178	0.0780	0.1374	0.0200	0.0716	0.1252	0.0114	0.0660	0.1236
		$S_R$	0.0378	0.1104	0.1834	0.0236	0.0760	0.1420	0.0120	0.0602	0.1112
		$S_T$	0.0128	0.0592	0.1110	0.0108	0.0480	0.1018	0.0086	0.0526	0.0940
	0.5	$S_W$	0.0130	0.0482	0.0964	0.0142	0.0554	0.1024	0.0074	0.0488	0.1012
		$S_{LR}$	0.0178	0.0780	0.1374	0.0200	0.0716	0.1252	0.0114	0.0660	0.1236
		$S_R$	0.0216	0.0834	0.1486	0.0202	0.0704	0.1236	0.0122	0.0560	0.1056
		$S_T$	0.0062	0.0468	0.0982	0.0122	0.0518	0.1076	0.0092	0.0524	0.1044
	0.75	$S_W$	0.0130	0.0482	0.0964	0.0142	0.0556	0.1024	0.0074	0.0488	0.1012
		$S_{LR}$	0.0178	0.0780	0.1374	0.0200	0.0716	0.1252	0.0114	0.0660	0.1236
		$S_R$	0.0442	0.1186	0.1932	0.0200	0.0716	0.1342	0.0150	0.0668	0.1180
		$S_T$	0.0144	0.0604	0.1122	0.0108	0.0556	0.1102	0.0134	0.0588	0.1158
1	0.25	$S_W$	0.0108	0.0418	0.0830	0.0122	0.0500	0.0932	0.0062	0.0460	0.0980
		$S_{LR}$	0.0180	0.0728	0.1322	0.0176	0.0632	0.1160	0.0110	0.0634	0.1168
		$S_R$	0.0424	0.1252	0.1890	0.0256	0.0834	0.1422	0.0158	0.0624	0.1162
		$S_T$	0.0092	0.0544	0.1114	0.0084	0.0556	0.1104	0.0108	0.0494	0.1066
	0.5	$S_W$	0.0108	0.0418	0.0830	0.0122	0.0500	0.0932	0.0062	0.0460	0.0980
		$S_{LR}$	0.0180	0.0728	0.1322	0.0178	0.0632	0.1160	0.0110	0.0634	0.1168
		$S_R$	0.0238	0.0898	0.1594	0.0182	0.0724	0.1286	0.0136	0.0582	0.1096
		$S_T$	0.0086	0.0484	0.1018	0.0070	0.0444	0.0972	0.0096	0.0458	0.0928
	0.75	$S_W$	0.0108	0.0418	0.0830	0.0122	0.0500	0.0932	0.0062	0.0460	0.0980
		$S_{LR}$	0.0180	0.0728	0.1322	0.0176	0.0632	0.1160	0.0110	0.0634	0.1168
		$S_R$	0.0384	0.1166	0.1834	0.0210	0.0774	0.1342	0.0120	0.0600	0.1154
		$S_T$	0.0130	0.0538	0.1070	0.0112	0.0516	0.1052	0.0106	0.0520	0.1066

**Tabela 3.8:** Taxa de rejeição sob a hipótese nula para testar  $H_0 : \beta_3 = 0$ , para o modelo de regressão SL-BS, com  $\nu = 4$ .

$\alpha$	$q$		$n = 50$			$n = 100$			$n = 200$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
0.2	0.25	$S_W$	0.0116	0.0478	0.0898	0.0112	0.0504	0.0996	0.0104	0.0490	0.0980
		$S_{LR}$	0.0146	0.0640	0.1220	0.0114	0.0560	0.1100	0.0120	0.0534	0.1118
		$S_R$	0.0366	0.1116	0.1876	0.0234	0.0774	0.1334	0.0126	0.0650	0.1178
		$S_T$	0.0078	0.0522	0.1034	0.0122	0.0554	0.1060	0.0098	0.0600	0.1130
	0.5	$S_W$	0.0106	0.0490	0.0908	0.0112	0.0502	0.0996	0.0104	0.0490	0.0980
		$S_{LR}$	0.0170	0.0634	0.1170	0.0124	0.0608	0.1112	0.0108	0.0484	0.1006
		$S_R$	0.0284	0.0926	0.1524	0.0160	0.0672	0.1192	0.0106	0.0510	0.1022
		$S_T$	0.0088	0.0452	0.0998	0.0102	0.0500	0.1016	0.0102	0.0516	0.0966
	0.75	$S_W$	0.0106	0.0490	0.0908	0.0112	0.0502	0.0996	0.0104	0.0490	0.0978
		$S_{LR}$	0.0148	0.0672	0.1272	0.0116	0.0602	0.1114	0.0132	0.0534	0.1004
		$S_R$	0.0310	0.0980	0.1684	0.0220	0.0766	0.1306	0.0148	0.0604	0.1220
		$S_T$	0.0082	0.0506	0.1080	0.0086	0.0482	0.1014	0.0114	0.0520	0.0996
0.5	0.25	$S_W$	0.0100	0.0452	0.0852	0.0106	0.0486	0.0980	0.0102	0.0482	0.0970
		$S_{LR}$	0.0158	0.0668	0.1232	0.0122	0.0566	0.1082	0.0100	0.0530	0.1064
		$S_R$	0.0386	0.1120	0.1766	0.0288	0.0892	0.1500	0.0152	0.0608	0.1144
		$S_T$	0.0078	0.0486	0.0984	0.0092	0.0514	0.0990	0.0086	0.0470	0.1004
	0.5	$S_W$	0.0100	0.0452	0.0852	0.0106	0.0486	0.0980	0.0120	0.0498	0.0918
		$S_{LR}$	0.0122	0.0570	0.1136	0.0136	0.0580	0.1162	0.0098	0.0502	0.0998
		$S_R$	0.0258	0.0884	0.1528	0.0168	0.0674	0.1250	0.0130	0.0630	0.1110
		$S_T$	0.0080	0.0466	0.1030	0.0076	0.0492	0.1016	0.0090	0.0532	0.1022
	0.75	$S_W$	0.0100	0.0452	0.0852	0.0106	0.0486	0.0980	0.0080	0.0446	0.0926
		$S_{LR}$	0.0142	0.0590	0.1112	0.0126	0.0644	0.1264	0.0108	0.0562	0.1096
		$S_R$	0.0348	0.1076	0.1820	0.0248	0.0836	0.1450	0.0162	0.0606	0.1132
		$S_T$	0.0086	0.0522	0.1080	0.0104	0.0492	0.1022	0.0102	0.0450	0.0938
1	0.25	$S_W$	0.0072	0.0310	0.0696	0.0088	0.0422	0.0852	0.0076	0.0434	0.0926
		$S_{LR}$	0.0140	0.0624	0.1206	0.0108	0.0546	0.1038	0.0088	0.0486	0.095
		$S_R$	0.0402	0.1184	0.1794	0.0196	0.0730	0.1378	0.0178	0.0616	0.1146
		$S_T$	0.0084	0.0510	0.1066	0.0124	0.0606	0.1114	0.0094	0.0528	0.1042
	0.5	$S_W$	0.0072	0.0310	0.0696	0.0088	0.0420	0.0852	0.0076	0.0434	0.0926
		$S_{LR}$	0.0108	0.0530	0.1098	0.0144	0.0574	0.1122	0.0110	0.0442	0.0944
		$S_R$	0.0246	0.0874	0.1500	0.0186	0.0684	0.1252	0.0130	0.0642	0.1162
		$S_T$	0.0080	0.0474	0.0986	0.0076	0.0476	0.0998	0.0080	0.0470	0.0994
	0.75	$S_W$	0.0072	0.0310	0.0696	0.0088	0.0422	0.0852	0.0076	0.0434	0.0926
		$S_{LR}$	0.0124	0.0568	0.1102	0.0130	0.0550	0.1050	0.0100	0.0454	0.0894
		$S_R$	0.0344	0.1048	0.1744	0.0230	0.0816	0.1424	0.0144	0.0596	0.1218
		$S_T$	0.0104	0.0638	0.1130	0.0110	0.0524	0.1036	0.0130	0.0520	0.1066

**Tabela 3.9:** Taxa de rejeição sob a hipótese nula para testar  $H_0 : \beta_3 = 0$ , para o modelo de regressão  $t_\nu$ -BS, com  $\nu = 11$ .

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
0.2	0.25	$S_W$	0.0126	0.0534	0.1036	0.0156	0.0546	0.1062	0.0070	0.0502	0.1052
		$S_{LR}$	0.0162	0.0696	0.1296	0.0182	0.0664	0.1230	0.0092	0.0598	0.1164
		$S_R$	0.0332	0.0966	0.1634	0.0204	0.0786	0.1428	0.0140	0.0666	0.1222
		$S_T$	0.0096	0.0534	0.1100	0.0086	0.0544	0.1032	0.0104	0.0490	0.1054
	0.5	$S_W$	0.0126	0.0536	0.1038	0.0156	0.0546	0.1062	0.0070	0.0502	0.1052
		$S_{LR}$	0.0162	0.0696	0.1296	0.0182	0.0662	0.1230	0.0092	0.0598	0.1162
		$S_R$	0.0210	0.0812	0.1446	0.0150	0.0646	0.1208	0.0126	0.0568	0.1070
		$S_T$	0.0086	0.0480	0.1010	0.0116	0.0484	0.1036	0.0122	0.0494	0.1054
	0.75	$S_W$	0.0126	0.0536	0.1038	0.0156	0.0546	0.1064	0.0070	0.0502	0.1052
		$S_{LR}$	0.0162	0.0694	0.1296	0.0184	0.0662	0.1232	0.0092	0.0598	0.1160
		$S_R$	0.0424	0.1118	0.1776	0.0202	0.0742	0.1394	0.0184	0.0676	0.1274
		$S_T$	0.0106	0.0548	0.1082	0.0102	0.0536	0.1028	0.0098	0.0520	0.1052
0.5	0.25	$S_W$	0.0124	0.0514	0.0980	0.0138	0.0560	0.1024	0.0066	0.0482	0.1046
		$S_{LR}$	0.0160	0.0666	0.1274	0.0178	0.0630	0.1216	0.0088	0.0576	0.1126
		$S_R$	0.0320	0.1006	0.1736	0.0212	0.0748	0.1400	0.0146	0.0578	0.1162
		$S_T$	0.0120	0.0608	0.1174	0.0112	0.0562	0.1036	0.0104	0.0528	0.0974
	0.5	$S_W$	0.0124	0.0514	0.0980	0.0138	0.0560	0.1024	0.0066	0.0482	0.1046
		$S_{LR}$	0.0160	0.0666	0.1274	0.0178	0.0630	0.1216	0.0088	0.0576	0.1126
		$S_R$	0.0218	0.084	0.1466	0.0144	0.0630	0.1178	0.0100	0.0576	0.1072
		$S_T$	0.0092	0.0530	0.1026	0.0114	0.0504	0.1034	0.0098	0.0510	0.0980
	0.75	$S_W$	0.0124	0.0514	0.0980	0.0138	0.0560	0.1024	0.0066	0.0482	0.1046
		$S_{LR}$	0.0160	0.0666	0.1274	0.0178	0.0630	0.1216	0.0088	0.0576	0.1126
		$S_R$	0.0326	0.1030	0.1688	0.0180	0.0690	0.1326	0.0126	0.0648	0.1234
		$S_T$	0.0106	0.0576	0.1136	0.0146	0.0586	0.1050	0.0104	0.0528	0.1044
1	0.25	$S_W$	0.0104	0.0426	0.0846	0.0120	0.0510	0.0952	0.0064	0.0438	0.0992
		$S_{LR}$	0.0136	0.0630	0.1230	0.0154	0.0584	0.1144	0.0078	0.0508	0.1072
		$S_R$	0.0338	0.1046	0.1722	0.0196	0.0744	0.1364	0.0124	0.0592	0.1208
		$S_T$	0.0112	0.0586	0.1092	0.0120	0.0526	0.1014	0.0112	0.0548	0.1054
	0.5	$S_W$	0.0104	0.0426	0.0846	0.0120	0.0510	0.0952	0.0064	0.0438	0.0992
		$S_{LR}$	0.0136	0.0630	0.1230	0.0154	0.0584	0.1144	0.0078	0.0508	0.1072
		$S_R$	0.0246	0.0888	0.1424	0.0188	0.0654	0.1188	0.0122	0.0566	0.1104
		$S_T$	0.0070	0.0498	0.1016	0.0094	0.0492	0.1016	0.0108	0.0502	0.1028
	0.75	$S_W$	0.0104	0.0426	0.0846	0.0120	0.0510	0.0952	0.0064	0.0438	0.0992
		$S_{LR}$	0.0136	0.0630	0.1230	0.0154	0.0584	0.1144	0.0078	0.0508	0.1072
		$S_R$	0.0360	0.1148	0.1822	0.0250	0.0818	0.1356	0.0140	0.0632	0.1212
		$S_T$	0.0124	0.0582	0.1082	0.0120	0.0568	0.1076	0.0096	0.0470	0.1002

**Tabela 3.10:** Taxa de rejeição sob a hipótese nula para testar  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , para o modelo de regressão CN-BS, com  $\nu = 0.1$ ,  $\delta = 0.3$ .

$\alpha$	$q$		$n = 50$			$n = 100$			$n = 200$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
0.2	0.25	$S_W$	0.0156	0.0570	0.1082	0.0124	0.0540	0.1046	0.0108	0.0512	0.1028
		$S_{LR}$	0.0210	0.0792	0.1402	0.0202	0.0778	0.1344	0.0188	0.0742	0.1362
		$S_R$	0.0388	0.1204	0.1972	0.0232	0.0836	0.1488	0.0150	0.0662	0.1214
		$S_T$	0.0066	0.0476	0.1024	0.0102	0.0472	0.0978	0.0090	0.0468	0.1012
	0.5	$S_W$	0.0190	0.0608	0.1078	0.0158	0.0558	0.1034	0.0110	0.0478	0.0956
		$S_{LR}$	0.0200	0.0850	0.1402	0.0146	0.0726	0.1290	0.0142	0.0666	0.1220
		$S_R$	0.0294	0.1032	0.1704	0.0188	0.0706	0.1266	0.0140	0.0646	0.1284
		$S_T$	0.0066	0.0446	0.0944	0.0094	0.0458	0.0958	0.0084	0.0460	0.1000
	0.75	$S_W$	0.0184	0.0614	0.1066	0.0128	0.0536	0.1046	0.0142	0.0504	0.0998
		$S_{LR}$	0.0212	0.0764	0.1332	0.0166	0.0700	0.1320	0.0172	0.0728	0.1390
		$S_R$	0.0496	0.1384	0.2166	0.0284	0.1004	0.1636	0.0190	0.0610	0.1178
		$S_T$	0.0074	0.0436	0.1002	0.0106	0.0450	0.1010	0.0092	0.0470	0.1008
0.5	0.25	$S_W$	0.0170	0.0544	0.0990	0.0124	0.0510	0.0994	0.0094	0.0492	0.1022
		$S_{LR}$	0.0192	0.0798	0.1410	0.0166	0.0662	0.1226	0.0142	0.0710	0.1230
		$S_R$	0.0664	0.1646	0.2450	0.0228	0.0814	0.1430	0.0134	0.0596	0.1160
		$S_T$	0.0074	0.0464	0.0980	0.0102	0.0474	0.0978	0.0086	0.0470	0.1002
	0.5	$S_W$	0.0128	0.0568	0.1044	0.0114	0.0536	0.1014	0.0104	0.0514	0.1002
		$S_{LR}$	0.0250	0.0832	0.1364	0.0166	0.0656	0.1252	0.0144	0.0638	0.1182
		$S_R$	0.0266	0.0994	0.1678	0.0176	0.0668	0.1266	0.0138	0.0638	0.1202
		$S_T$	0.0068	0.0446	0.0944	0.0094	0.0458	0.0956	0.0084	0.0458	0.0996
	0.75	$S_W$	0.0154	0.0636	0.1052	0.0136	0.0532	0.1004	0.0118	0.0530	0.0996
		$S_{LR}$	0.0166	0.0746	0.1338	0.0154	0.0782	0.1350	0.0166	0.0728	0.1300
		$S_R$	0.0406	0.1244	0.1922	0.0258	0.0834	0.1410	0.0130	0.0594	0.1104
		$S_T$	0.0076	0.0430	0.1008	0.0108	0.0444	0.0992	0.0088	0.0468	0.1012
1	0.25	$S_W$	0.0128	0.0452	0.0866	0.0108	0.0460	0.0862	0.0084	0.0448	0.0954
		$S_{LR}$	0.0170	0.0710	0.1252	0.0150	0.0642	0.1160	0.0146	0.0632	0.1136
		$S_R$	0.0700	0.1586	0.2410	0.0224	0.0798	0.1482	0.0142	0.0668	0.1256
		$S_T$	0.0074	0.0460	0.0988	0.0104	0.0472	0.0980	0.0088	0.0462	0.1016
	0.5	$S_W$	0.0122	0.0478	0.0908	0.0130	0.0470	0.0938	0.0104	0.0488	0.0944
		$S_{LR}$	0.0206	0.0772	0.1318	0.0138	0.0622	0.1220	0.0144	0.0578	0.1072
		$S_R$	0.0292	0.1054	0.1740	0.0186	0.0718	0.1276	0.0140	0.0644	0.1208
		$S_T$	0.0068	0.0444	0.0948	0.0094	0.0462	0.0962	0.0084	0.0454	0.1004
	0.75	$S_W$	0.0194	0.0534	0.0922	0.0106	0.0486	0.0888	0.0092	0.0514	0.1010
		$S_{LR}$	0.0208	0.0726	0.1276	0.0138	0.0622	0.1150	0.0158	0.0642	0.1130
		$S_R$	0.0454	0.1348	0.2070	0.0208	0.0866	0.1506	0.0126	0.0592	0.1128
		$S_T$	0.0078	0.0430	0.1012	0.0110	0.0462	0.1022	0.0086	0.0464	0.1014

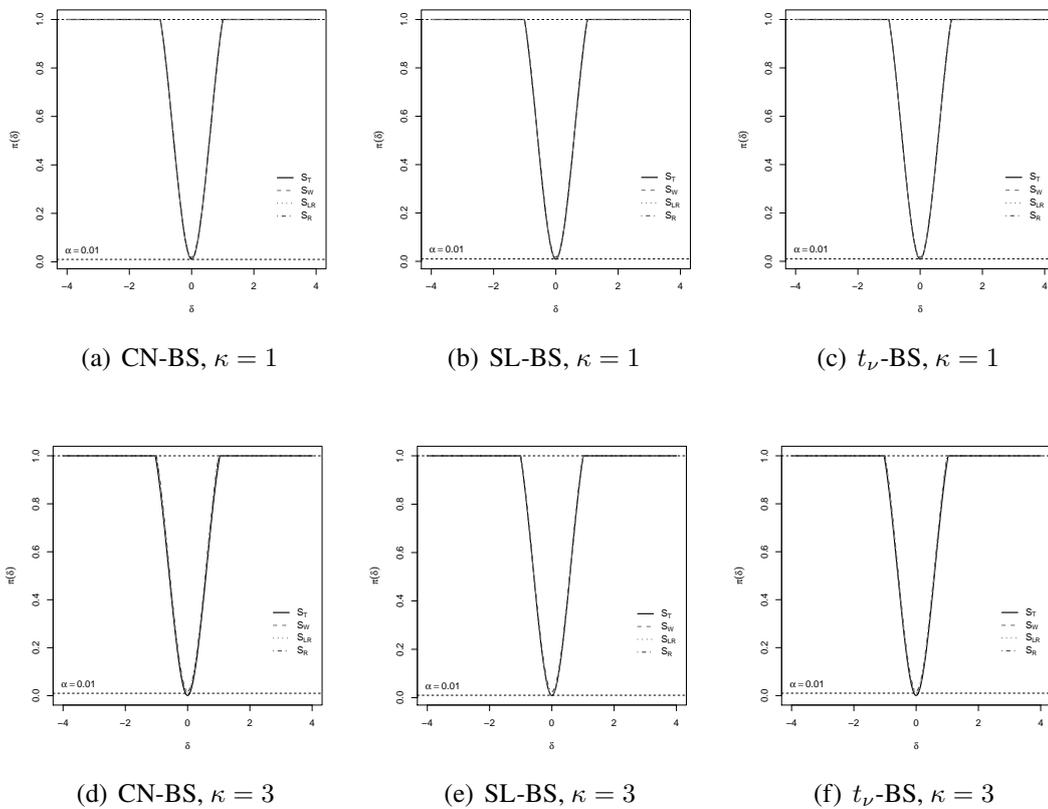
**Tabela 3.11:** Taxa de rejeição sob a hipótese nula para testar  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , para o modelo de regressão SL-BS, com  $\nu = 4$ .

$\alpha$	$q$		$n = 50$			$n = 100$			$n = 200$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
0.2	0.25	$S_W$	0.0146	0.0584	0.1000	0.0118	0.0506	0.0974	0.0096	0.0468	0.0956
		$S_{LR}$	0.0160	0.0658	0.1250	0.0130	0.0616	0.1170	0.0126	0.0542	0.1038
		$S_R$	0.0446	0.1258	0.2072	0.0222	0.0892	0.1520	0.0170	0.0702	0.1286
		$S_T$	0.0072	0.0436	0.0916	0.0078	0.0436	0.0960	0.0090	0.0524	0.1010
	0.5	$S_W$	0.0148	0.0496	0.0902	0.0118	0.0506	0.0974	0.0096	0.0468	0.0954
		$S_{LR}$	0.0154	0.0666	0.1284	0.0164	0.0632	0.1190	0.0118	0.0548	0.1092
		$S_R$	0.0664	0.1626	0.2434	0.0188	0.0736	0.1406	0.0148	0.0664	0.1254
		$S_T$	0.0066	0.0426	0.0914	0.0082	0.0494	0.1004	0.0108	0.0450	0.0948
	0.75	$S_W$	0.0150	0.0496	0.0902	0.0118	0.0508	0.0974	0.0096	0.0468	0.0952
		$S_{LR}$	0.0150	0.0676	0.1236	0.0118	0.0574	0.1106	0.0104	0.0572	0.1054
		$S_R$	0.0422	0.1240	0.1978	0.0260	0.0882	0.1554	0.0176	0.0640	0.1146
		$S_T$	0.0070	0.0474	0.1028	0.0062	0.0458	0.0926	0.0086	0.0482	0.0960
0.5	0.25	$S_W$	0.0136	0.0446	0.0826	0.0114	0.0470	0.0916	0.0096	0.0454	0.0904
		$S_{LR}$	0.0160	0.0708	0.1290	0.0152	0.0544	0.1066	0.0102	0.0500	0.1042
		$S_R$	0.0418	0.1262	0.2064	0.0270	0.0890	0.1538	0.0158	0.0684	0.1182
		$S_T$	0.0058	0.0424	0.095	0.0078	0.0488	0.0974	0.0094	0.0486	0.1000
	0.5	$S_W$	0.0136	0.0446	0.0826	0.0114	0.0470	0.0916	0.0096	0.0454	0.0902
		$S_{LR}$	0.0144	0.0666	0.1210	0.0134	0.0584	0.1090	0.0130	0.0566	0.1048
		$S_R$	0.0436	0.1268	0.2104	0.0332	0.1036	0.1690	0.0130	0.0602	0.1174
		$S_T$	0.0042	0.0438	0.0994	0.0104	0.0456	0.0928	0.0112	0.0476	0.1008
	0.75	$S_W$	0.0136	0.0446	0.0824	0.0114	0.0470	0.0916	0.0096	0.0454	0.0902
		$S_{LR}$	0.0156	0.0608	0.1194	0.0124	0.0574	0.1072	0.0120	0.0566	0.1064
		$S_R$	0.0366	0.1098	0.1852	0.0260	0.0896	0.1560	0.0178	0.0694	0.1288
		$S_T$	0.0062	0.0460	0.0984	0.0070	0.0408	0.0944	0.0094	0.0562	0.1030
1	0.25	$S_W$	0.0088	0.0310	0.0608	0.0102	0.0366	0.0738	0.0088	0.0394	0.0834
		$S_{LR}$	0.0130	0.0618	0.1130	0.0090	0.0504	0.0992	0.0100	0.0450	0.0946
		$S_R$	0.0450	0.1288	0.2100	0.0270	0.0942	0.1568	0.0122	0.0638	0.1268
		$S_T$	0.0080	0.0478	0.0928	0.0080	0.0508	0.1056	0.0112	0.0486	0.0966
	0.5	$S_W$	0.0088	0.0308	0.0608	0.0102	0.0366	0.0738	0.0088	0.0394	0.0834
		$S_{LR}$	0.0130	0.0578	0.1120	0.0114	0.0500	0.0992	0.0084	0.0428	0.0934
		$S_R$	0.0338	0.1080	0.1830	0.0206	0.0762	0.1396	0.0114	0.0602	0.1142
		$S_T$	0.0064	0.0460	0.0974	0.0076	0.0470	0.0974	0.0088	0.0502	0.1040
	0.75	$S_W$	0.0088	0.0308	0.0608	0.0102	0.0366	0.0738	0.0088	0.0394	0.0834
		$S_{LR}$	0.0144	0.0594	0.1148	0.0084	0.0468	0.0894	0.0094	0.0528	0.0978
		$S_R$	0.0586	0.1494	0.2310	0.0216	0.0826	0.1496	0.0128	0.0644	0.1218
		$S_T$	0.0054	0.0436	0.0916	0.0088	0.0468	0.0934	0.0086	0.0498	0.0994

**Tabela 3.12:** Taxa de rejeição sob a hipótese nula para testar  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ , para o modelo de regressão  $t_\nu$ -BS, com  $\nu = 11$ .

$\alpha$	$q$	$n = 50$			$n = 100$			$n = 200$			
		1%	5%	10%	1%	5%	10%	1%	5%	10%	
0.2	0.25	$S_W$	0.0160	0.0630	0.1128	0.0112	0.0522	0.0982	0.0104	0.0494	0.0984
		$S_{LR}$	0.0178	0.0712	0.1304	0.0156	0.0668	0.1214	0.0138	0.0604	0.1164
		$S_R$	0.0452	0.1308	0.2092	0.0154	0.0704	0.1290	0.0150	0.0692	0.1254
		$S_T$	0.0086	0.0498	0.1016	0.0068	0.0430	0.0918	0.0082	0.0478	0.0962
	0.5	$S_W$	0.0156	0.0662	0.1232	0.0160	0.0566	0.1074	0.0120	0.0538	0.1060
		$S_{LR}$	0.0192	0.0790	0.1418	0.0146	0.0660	0.1240	0.0130	0.0580	0.1102
		$S_R$	0.0422	0.1188	0.1908	0.0150	0.0662	0.1230	0.0122	0.0568	0.1126
		$S_T$	0.0058	0.0394	0.0886	0.0094	0.0454	0.0934	0.0096	0.0506	0.1020
	0.75	$S_W$	0.0214	0.0674	0.1206	0.0138	0.0572	0.1084	0.0116	0.0580	0.1032
		$S_{LR}$	0.0180	0.0676	0.1364	0.0146	0.0594	0.1156	0.0138	0.0630	0.1176
		$S_R$	0.0604	0.1486	0.2198	0.0168	0.0668	0.1270	0.0156	0.0618	0.1204
		$S_T$	0.0092	0.0522	0.1064	0.0086	0.0486	0.0982	0.0092	0.0452	0.0950
0.5	0.25	$S_W$	0.0144	0.0612	0.1110	0.0112	0.0492	0.0958	0.0110	0.0554	0.1026
		$S_{LR}$	0.0150	0.0678	0.1310	0.0164	0.0674	0.1220	0.0112	0.0576	0.1062
		$S_R$	0.0432	0.1368	0.2080	0.0194	0.0806	0.1418	0.0164	0.0668	0.1248
		$S_T$	0.0086	0.0502	0.1046	0.0068	0.0438	0.0916	0.0080	0.0454	0.0922
	0.5	$S_W$	0.0188	0.0614	0.1116	0.0128	0.0544	0.0970	0.0106	0.0518	0.0996
		$S_{LR}$	0.0156	0.0694	0.1284	0.0140	0.0586	0.1118	0.0108	0.0626	0.1174
		$S_R$	0.0410	0.1124	0.1846	0.0210	0.0758	0.1430	0.0124	0.0608	0.1186
		$S_T$	0.0060	0.0456	0.0988	0.0094	0.0500	0.1024	0.0066	0.0452	0.0948
	0.75	$S_W$	0.0172	0.0548	0.1032	0.0100	0.0566	0.1038	0.0108	0.0494	0.0972
		$S_{LR}$	0.0144	0.0720	0.1354	0.0142	0.0590	0.1112	0.0126	0.0624	0.1218
		$S_R$	0.0474	0.1302	0.2032	0.0180	0.0736	0.1328	0.0138	0.0624	0.1186
		$S_T$	0.0084	0.0496	0.1016	0.0086	0.0436	0.0912	0.0084	0.0490	0.0946
1	0.25	$S_W$	0.0118	0.0478	0.0898	0.0118	0.0420	0.0864	0.0106	0.0446	0.0874
		$S_{LR}$	0.0138	0.0604	0.1128	0.0134	0.0580	0.1058	0.0138	0.0550	0.1016
		$S_R$	0.0544	0.1374	0.2124	0.0156	0.0708	0.1356	0.0144	0.0638	0.1252
		$S_T$	0.0086	0.0514	0.1050	0.0068	0.0440	0.0916	0.0098	0.0544	0.1016
	0.5	$S_W$	0.0102	0.0430	0.0848	0.0102	0.0480	0.0940	0.0100	0.0452	0.0936
		$S_{LR}$	0.0160	0.0664	0.1232	0.0136	0.0552	0.1074	0.0116	0.0498	0.0970
		$S_R$	0.0196	0.0844	0.1572	0.0122	0.0698	0.1370	0.0134	0.0592	0.1218
		$S_T$	0.0080	0.0452	0.0998	0.0086	0.0508	0.0974	0.0090	0.0472	0.0960
	0.75	$S_W$	0.0144	0.0484	0.0882	0.0094	0.0464	0.0940	0.0110	0.0482	0.0974
		$S_{LR}$	0.0178	0.0696	0.1236	0.0096	0.0526	0.1000	0.0122	0.0528	0.1074
		$S_R$	0.0416	0.1196	0.1964	0.0224	0.0820	0.1416	0.0114	0.0592	0.1132
		$S_T$	0.0072	0.0490	0.0982	0.0082	0.0518	0.1028	0.0110	0.0510	0.0990

**Figura 3.2:** Poder do teste das estatísticas  $S_W$ ,  $S_{LR}$ ,  $S_R$  e  $S_T$  para os modelos CN-BS, SL-BS e  $t_\nu$ -BS considerando o nível de significância de 1%,  $n = 100$  e  $q = 0.5$ .



### 3.5 Exemplo ilustrativo: Seguro de acidentes pessoais

O conjunto de dados encontrado em De Jong, Heller et al. (2008) contém informações a respeito de 767 reivindicações de seguros de danos pessoais ocorridos entre Janeiro de 1998 e Janeiro de 1999. O conjunto de dados em questão conta com as variáveis: ‘montante’ (valor pago por uma apólice em dólares australianos), ‘legrep’ (com representação legal ou não), ‘mês’ (mês de ocorrência do acidente) e ‘optime’ (tempo operacional em % com amplitude de [0,100]), onde o montante é a variável resposta e as demais variáveis são as covariáveis. Devido ao fato da falta de correlação entre a covariável mês e a variável resposta, essa variável foi removida da análise.

Inicialmente foi feita uma análise descritiva da variável resposta, a Tabela 3.13 apresenta um sumário de estatísticas descritivas para a variável montante, em dólares australianos. O sumário de estatísticas em questão conta com o desvio padrão (SD), coeficiente de variação (CV), coeficiente de assimetria (CS), o coeficiente de curtose (CK) e a amplitude (Range), ainda

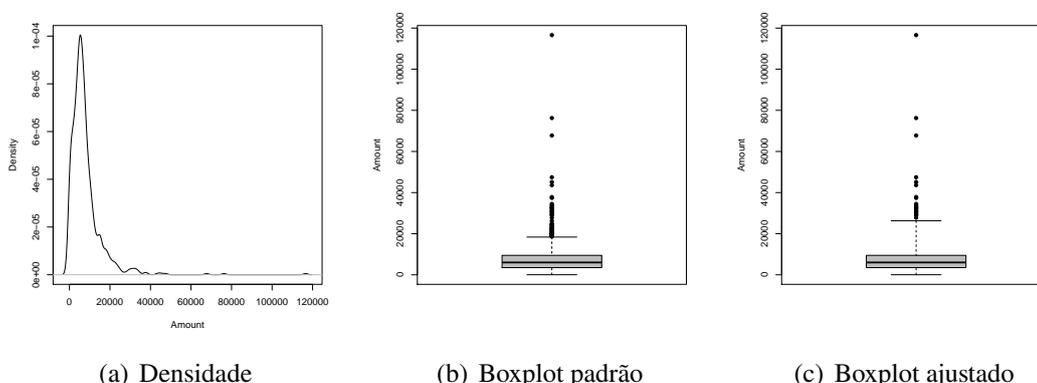
pela Tabela 3.13, temos que o valor pago, em média para as apólices é de 7820.59 em dólares australianos, entretanto com desvio padrão de 8339.26 e coeficiente de variação de 106.63% fica evidente o elevado grau de dispersão desses dados, além disso, pelo coeficiente de assimetria, os dados do montante também apresentam elevada assimetria à direita (5.09) e curtose (47.77), por conta disso uma medida de tendência central mais adequada é a mediana, a qual é dada por 7820.59.

Na Figura 3.3 estão dispostos o gráfico da estimativa da densidade, boxplot padrão e o boxplot ajustado do montante pago em apólices de seguros, pela Figura 3.3 podemos observar claramente como os dados se dispõem de forma assimétrica à direita pelo gráfico da estimativa da densidade e dos dois tipos de boxplots da variável montante.

**Tabela 3.13:** Sumário de estatísticas descritivas para a variável montante.

Mediana	Média	SD	CV	CS	CK	Range	Min	Max	<i>n</i>
6000	7820.59	8339.26	106.63%	5.09	47.77	116556.7	30	116586.7	767

**Figura 3.3:** Gráfico de frequência por estimativa de kernel (a), boxplot padrão (b) e boxplot ajustado (c) para a variável montante, em dólares australianos.



Supondo que a variável resposta, o montante pago, segue uma distribuição  $T_i \sim \text{QSBS}(\alpha, Q_i; H)$ , então, considere o modelo de regressão quantílica a ser ajustado com função de ligação log, de tal forma que o componente sistemático é dado por:

$$\log(Q_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \quad i = 1, 2, \dots, 767, \tag{3.22}$$

onde  $\beta_0, \beta_1$  e  $\beta_2$  são o intercepto e os coeficientes referentes as covariáveis *optime* e *legrep*, respectivamente.

A estimação do vetor de parâmetros  $\nu$  indexados aos modelos é feita de forma análoga a Paula et al. (2012). Primeiro, selecionamos um grid de valores para  $\nu$ , com  $\nu \in [1, 20]$  para os modelos SL-BS e  $t_\nu$ -BS e  $\nu \in [0.01, 0.99]$ ,  $\delta \in [0.01, 0.99]$  para o modelo CN-BS. Após a escolha do grid de valores para os parâmetros, tomamos as estimativas iniciais dos parâmetros  $\alpha, \beta_0, \beta_1, \beta_2$  utilizando o modelo log-BS; ver (Rieck e Nedelman, 1991), então calculamos a função de log-verossimilhança para cada valor do grid e escolhemos  $\nu$  que maximiza a função de log-verossimilhança. Uma vez selecionados os valores de  $\nu$ , os parâmetros do modelo são estimados utilizando o algoritmo EM. Na Tabela 3.14 temos a média dos critérios de informação e o REQM para os modelos CN-BS, SL-BS e  $t_\nu$ -BS para os quantis  $q = 0.01, 0.02, \dots, 0.99$ , como podemos observar pela Tabela 3.14, o modelo CN-BS apresenta as menores médias dos dos critérios de informação e do REQM, portanto, segundo esses critérios, o modelo ajustado é o modelo CN-BS. Na Figura 3.4 podemos observar as variações nas estimativas dos parâmetros do modelo CN-BS para os quantis  $q = 0.01, 0.02, \dots, 0.99$ , como podemos notar, as estimativas  $\hat{\beta}_1$  e  $\hat{\beta}_2$  diminuem a partir do quantil  $q = 0.8$  aproximadamente, indicando que para os maiores valores pagos, o impacto do tempo operacional e do fato da representação ser diminui. Na Tabela 3.15 temos as estimativas de máxima verossimilhança dos parâmetros e seus respectivos erro padrão do modelo de regressão quantílica considerando os quantis  $q = 0.25, 0.5, 0.75$ . Para os quantis analisados, podemos observar que o impacto do tempo operacional e da representação legal é maior para o quantil  $q = 0.75$ , o mesmo acontece com as estimativas  $\hat{\alpha}$  e  $\hat{\beta}_0$ .

**Tabela 3.14:** Média dos critérios AIC, BIC, AICc, HIC e REQM dos modelos de regressão quantílica SBS considerando os quantis  $q = 0.01, 0.02, \dots, 0.98, 0.99$ .

Modelo	AIC	BIC	AICc	HIC	REQM
CN-BS	8204.634	8223.204	8204.686	8211.781	10612.825
SL-BS	8220.061	8238.631	8220.114	8227.209	11059.936
$t_\nu$ -BS	8213.111	8231.681	8213.163	8220.259	11085.714

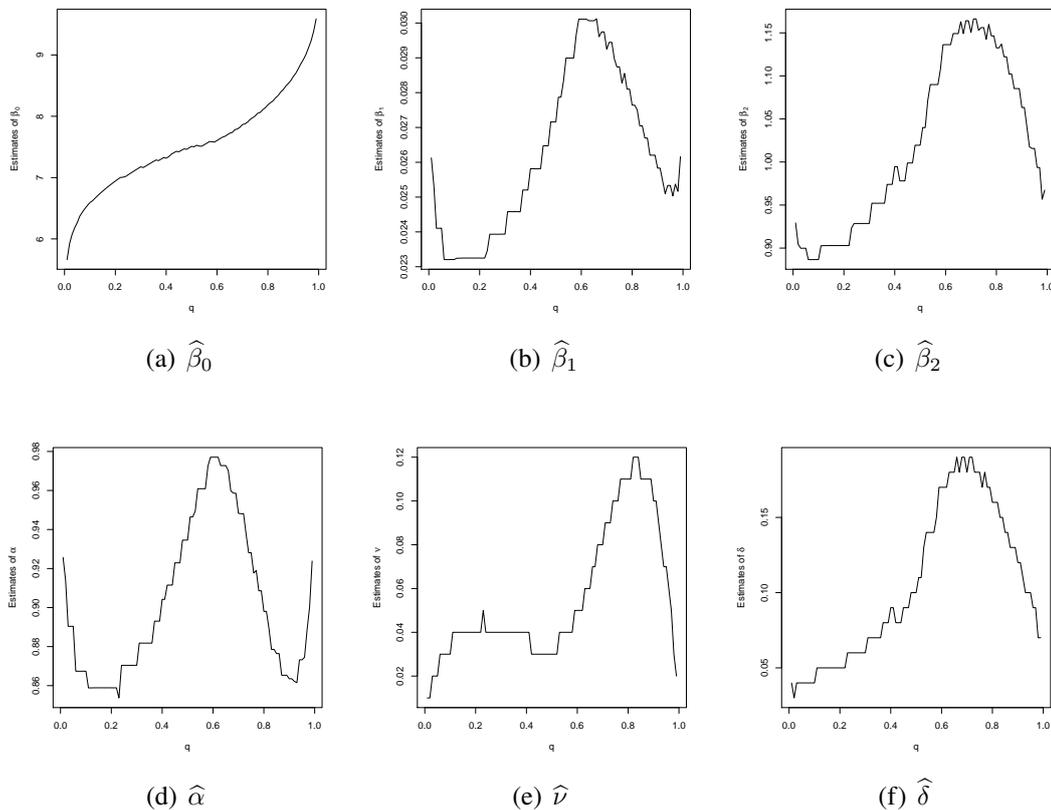
**Tabela 3.15:** Estimativas e erro padrão (em parênteses) dos parâmetros do modelo de regressão quantílica CN-BS considerando os quantis  $q = 0.25, 0.5, 0.75$ .

$q$	Estimativa(Erro-padrão)					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\nu}$	$\hat{\delta}$
0.25	7.0447(0.0543)	0.0239(0.0033)	0.9286(0.0638)	0.8704(0.0217)	0.04	0.06
0.5	7.5092(0.0543)	0.0272(0.0033)	1.0195(0.0671)	0.9356(0.0220)	0.03	0.1
0.75	8.0053(0.0614)	0.0287(0.0030)	1.1561(0.0694)	0.9281(0.0217)	0.1	0.18

Na Tabela 3.16 estão dispostos as estatísticas observadas e seu respectivo  $p$ -valor para testar  $H_0 : \beta_1 = 0$  e  $H_0 : \beta_2 = 0$ , onde foram consideradas as estatísticas de teste  $S_W, S_{LR}, S_R$  e  $S_T$ . Considerando o nível de significância de 5%, rejeitamos a hipótese nula em todos os casos testados na Tabela 3.16, indicando que as covariáveis tempo operacional e a representação legal são significantes para o modelo de regressão quantílica. Além disso, é possível observar que, dentre os quantis  $q = 0.25, 0.5, 0.75$ , as estatísticas observadas aumentam de acordo com o quantil fixado, reforçando mais ainda a ideia de que o impacto dessas covariáveis aumentam a medida que  $q$  se aproxima de 0.75.

Na Figura 3.5 estão dispostos os gráficos (QQ) dos resíduos GCS e RQ para o modelo CN-BS considerando  $q = 0.25, 0.5, 0.75$ . Pelos gráficos (QQ) para os resíduos GCS podemos observar que os pontos ficam dispostos dentro da região delimitada pelas bandas de confiança, indicando que o modelo CN-BS apresenta um bom ajuste aos dados, por outro lado, os gráficos (QQ) para os resíduos RQ apresenta alguns pontos mais extremos fora da região delimitada pelas bandas de confiança.

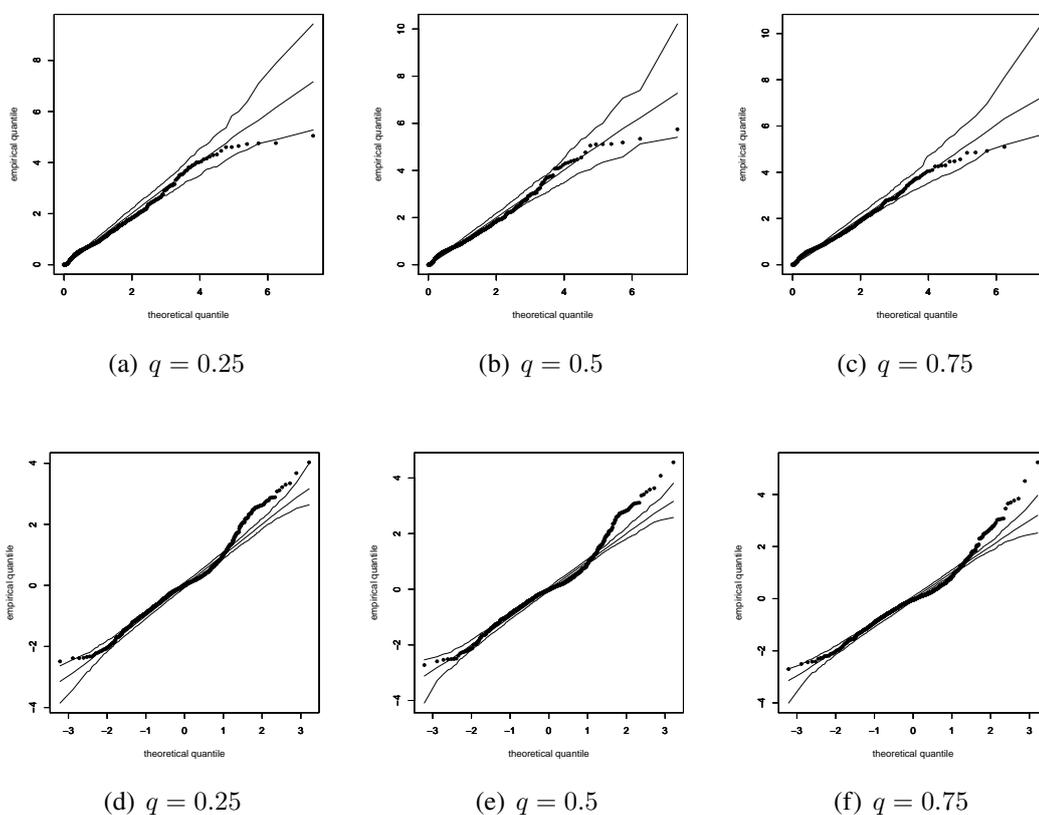
**Figura 3.4:** Estimativas de máxima verossimilhança dos parâmetros do modelo de regressão quantílica CN-BS para os quantis  $q = 0.01, 0.02, \dots, 0.99$ .



**Tabela 3.16:** Estatísticas observadas e  $p$ -valor (em parênteses) para testar  $H_0$  considerando os parâmetros do modelo de regressão quantílica CN-BS para os quantis  $q = 0.25, 0.5, 0.75$ .

$q$	Hipótese	Estatística de teste			
		$S_W$	$S_{LR}$	$S_R$	$S_T$
0.25	$H_0 : \beta_1 = 0$	53.471(< 0.0001)	61.971(< 0.0001)	149.8(< 0.0001)	144.23(< 0.0001)
	$H_0 : \beta_2 = 0$	212.12(< 0.0001)	13.438(0.0002)	36.392(< 0.0001)	40.174(< 0.0001)
0.5	$H_0 : \beta_1 = 0$	68.507(< 0.0001)	159.35(< 0.0001)	202.04(< 0.0001)	165.38(< 0.0001)
	$H_0 : \beta_2 = 0$	230.97(< 0.0001)	17.197(< 0.0001)	42.467(< 0.0001)	47.923(< 0.0001)
0.75	$H_0 : \beta_1 = 0$	90.682(< 0.0001)	231.28(< 0.0001)	262.37(< 0.0001)	190.9(< 0.0001)
	$H_0 : \beta_2 = 0$	275.56(< 0.0001)	36.009(< 0.0001)	40.564(< 0.0001)	51.648(< 0.0001)

**Figura 3.5:** Gráficos QQ dos resíduos GCS ((a), (b) e (c)) e RQ ((d), (e) e (f)) do modelo de regressão quantílica CN-BS para os quantis  $q = 0.25, 0.5, \dots, 0.75$ .



### 3.6 Conclusões

Neste estudo foi proposto uma nova classe de modelos de regressão quantílica baseado na família de distribuição mistura-escala Birnbaum-Saunders reparametrizadas pelo quantil. A grande relevância do trabalho está na proposta de um novo modelo de regressão quantílica que apresenta uma nova alternativa capaz de prover um ajuste bastante flexível para dados estritamente positivos e assimétricos. Os resultados obtidos nos estudos de simulação de Monte

Carlo mostram que (a) as estimativas de máxima verossimilhança, em termos do viés, EQM e CP, têm bom desempenho em amostras de tamanho  $n = 50$ ; (b) a taxa de acerto dos critérios de informação varia de modelo para modelo, mostrando melhores resultados para o modelo  $t_\nu$ -BS; (c) os resíduos GCS e RQ têm boa adequação com as respectivas distribuições de referência, isto é,  $\text{Exp}(1)$  e  $N(0,1)$ , com exceção do modelo SL-BS; e (d) os testes Wald, razão de verossimilhança, Score e Gradiente apresentam desempenhos similares, em termos do poder do teste, entretanto, os testes Wald e Gradiente apresentam taxas de rejeição sob a hipótese nula mais próxima dos níveis de significância nominal, indicando assim, que estes testes apresentam maior controle sobre o erro do tipo I. Desse modo, os testes Wald e Gradiente são preferíveis em relação aos demais testes analisados. Por fim, fizemos uma aplicação utilizando um conjunto de dados reais relacionados a seguros de acidentes pessoais. A aplicação mostrou a flexibilidade do modelo proposto apresentando o efeito das covariáveis na variável resposta para diferentes quantis.

# Capítulo 4

## Considerações Finais

Neste trabalho, dois modelos de regressão quantílica para dados positivos assimétricos à direita foram propostos. O primeiro é baseado nas distribuições log-simétricas e o segundo nas distribuições mistura-escala Birnbaum-Saunders, em ambos os casos uma reparametrização pelo quantil foi introduzida. A proposta desses modelos de regressão quantílica é de grande relevância não só na estatística mas também em outras áreas do conhecimento, visto que a análise de modelos de regressão é muito utilizada em diversas áreas. Devido ao fato das distribuições base desses modelos (log-simétricas e mistura-escala Birnbaum-Saunders) contarem com aplicações em diversas áreas, os dois modelos de regressão quantílica podem vir a ser muito úteis em diferentes contextos práticos, em que a modelagem de dados estritamente positivos é necessária. Como parte de pesquisas futuras, os seguintes aspectos podem ser explorados: (i) propor modelos multivariados; (ii) fazer um estudo de diagnóstico de influência local e global; e (iii) propor modelos para séries temporais. Trabalhos considerando esses problemas estão em andamento e espera-se publicação dos respectivos resultados em artigos futuros.

# Bibliografia

- Akaike, Hirotugu (1974). “A new look at the statistical model identification”. *IEEE transactions on automatic control* 19.6, pp. 716–723.
- Andrews, David F e Mallows, Colin L (1974). “Scale mixtures of normal distributions”. *Journal of the Royal Statistical Society: Series B (Methodological)* 36.1, pp. 99–102.
- Baker, Stuart G (1992). “A simple method for computing the observed information matrix when using the EM algorithm with categorical data”. *Journal of Computational and Graphical Statistics* 1.1, pp. 63–76.
- Balakrishnan, Narayanaswamy et al. (2009). “Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm”. *SORT-Statistics and Operations Research Transactions*.
- Birnbaum, Zygmunt W e Saunders, Sam C (1969a). “A new family of life distributions”. *Journal of applied probability*, pp. 319–327.
- (1969b). “Estimation for a family of life distributions with applications to fatigue”. *Journal of Applied probability*, pp. 328–347.
- Bozdogan, Hamparsum (1987). “Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions”. *Psychometrika* 52.3, pp. 345–370.
- Cordeiro, Gauss M e Lemonte, Artur J (2011). “The  $\beta$ -Birnbaum–Saunders distribution: an improved distribution for fatigue life modeling”. *Computational Statistics & Data Analysis* 55.3, pp. 1445–1461.
- Cox, D. R. e Snell, E. (1968). “A general definition of residuals”. *Journal of the Royal Statistical Society B* 2, pp. 248–275.

- Dasilva, Alan et al. (2020). “[Invited tutorial] Birnbaum–Saunders regression models: a comparative evaluation of three approaches”. *Journal of Statistical Computation and Simulation*, pp. 1–19.
- De Jong, Piet, Heller, Gillian Z et al. (2008). “Generalized linear models for insurance data”. *Cambridge Books*.
- Dempster, Arthur P, Laird, Nan M e Rubin, Donald B (1977). “Maximum likelihood from incomplete data via the EM algorithm”. *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1, pp. 1–22.
- Dunn, P. e Smyth, G. (1996). “Randomized quantile residuals”. *Journal of Computational and Graphical Statistics* 5, pp. 236–244.
- Efron, B. e Hinkley, D. V. (1978). “Assessing the accuracy of the maximum likelihood estimator: Observed vs. expected Fisher information”. *Biometrika* 65, pp. 457–487.
- Efron, Bradley e Olshen, Richard A (1978). “How broad is the class of normal scale mixtures?” *The Annals of Statistics*, pp. 1159–1164.
- Fang, K. T. e Anderson, T. W. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*. New York, US: Allerton Press.
- Fernández, Carmen e Steel, Mark FJ (1999). “Multivariate Student-t regression models: Pitfalls and inference”. *Biometrika* 86.1, pp. 153–167.
- Ferreira, M., Gomes, Maria e Leiva, Víctor (jun. de 2012). “On an extreme value version of the Birnbaum-Saunders distribution”. *Revstat - Statistical Journal* 10, pp. 181–210.
- Gneiting, Tilmann (1997). “Normal scale mixtures and dual probability densities”. *Journal of Statistical Computation and Simulation* 59.4, pp. 375–384.
- Hannan, Edward J e Quinn, Barry G (1979). “The determination of the order of an autoregression”. *Journal of the Royal Statistical Society: Series B (Methodological)* 41.2, pp. 190–195.
- Hao, L. e Naiman, D.Q. (2007). *Quantile Regression*. California , US: Sage Publications.
- Jamshidian, Mortaza e Jennrich, Robert I. (1997). “Acceleration of the EM Algorithm by Using Quasi-Newton Methods”. *Journal of the Royal Statistical Society. Series B (Methodological)* 59.3, pp. 569–587. ISSN: 00359246. URL: <http://www.jstor.org/stable/2346010>.

- Johnson, N. L., Kotz, S. e Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Vol. 1. New York, US: Wiley.
- (1995). *Continuous Univariate Distributions*. Vol. 2. New York, US: Wiley.
- Kano, Yutaka, Berkane, Maia e Bentler, Peter M. (1993). “Statistical Inference Based on Pseudo-Maximum Likelihood Estimators in Elliptical Populations”. *Journal of the American Statistical Association* 88.421, pp. 135–143.
- Koenker, Roger (2004). “Quantile regression for longitudinal data”. *Journal of Multivariate Analysis* 91.1, pp. 74–89.
- Koenker, Roger e Bassett Jr, Gilbert (1978). “Regression quantiles”. *Econometrica: journal of the Econometric Society*, pp. 33–50.
- Koenker, Roger e Hallock, Kevin F (2001). “Quantile regression”. *Journal of economic perspectives* 15.4, pp. 143–156.
- Lachos, Victor H e Labra, FV (2007). *Skew-normal/independent distributions, with applications*. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Física.
- Lachos, Victor H et al. (2017). “Scale mixtures log-Birnbaum–Saunders regression models with censored data: a Bayesian approach”. *Journal of Statistical Computation and Simulation* 87.10, pp. 2002–2022.
- Lange, Kenneth (2010). *Numerical analysis for statisticians*. Springer Science & Business Media.
- Lange, Kenneth e Sinsheimer, Janet S (1993). “Normal/independent distributions and their applications in robust regression”. *Journal of Computational and Graphical Statistics* 2.2, pp. 175–198.
- Lange, Kenneth L, Little, Roderick JA e Taylor, Jeremy MG (1989). “Robust statistical modeling using the t distribution”. *Journal of the American Statistical Association* 84.408, pp. 881–896.
- Lawless, Jerald F (2011). *Statistical models and methods for lifetime data*. Vol. 362. John Wiley & Sons.
- Lee, Elisa T e Wang, John (2003). *Statistical methods for survival data analysis*. Vol. 476. John Wiley & Sons.

- Leiva, V. et al. (2020). “Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regression”. *Stochastic Environmental Research and Risk Assessment* 34, pp. 1457–1471.
- Leiva, Víctor, Sanhueza, Antonio e Angulo, José M (2009). “A length-biased version of the Birnbaum-Saunders distribution with application in water quality”. *Stochastic Environmental Research and Risk Assessment* 23.3, pp. 299–307.
- Leiva, Víctor et al. (2008a). “Generalized Birnbaum-Saunders distributions applied to air pollutant concentration”. *Environmetrics: The official journal of the International Environmetrics Society* 19.3, pp. 235–249.
- Leiva, Víctor et al. (2008b). “Lifetime analysis based on the generalized Birnbaum–Saunders distribution”. *Computational Statistics & Data Analysis* 52.4, pp. 2079–2097.
- Leiva, Victor et al. (2010). “A skewed sinh-normal distribution and its properties and application to air pollution”. *Communications in Statistics—Theory and Methods* 39.3, pp. 426–443.
- Louis, Thomas A (1982). “Finding the observed information matrix when using the EM algorithm”. *Journal of the Royal Statistical Society: Series B (Methodological)* 44.2, pp. 226–233.
- Lucas, A. (1997a). “Robustness of the student  $t$  based M-estimator”. *Communications in Statistics: Theory and Methods* 41, pp. 1165–1182.
- Lucas, André (1997b). “Robustness of the student  $t$  based M-estimator”. *Communications in Statistics-Theory and Methods* 26.5, pp. 1165–1182.
- Marchant, Carolina et al. (2013). “Air contaminant statistical distributions with application to PM10 in Santiago, Chile”. Em: *Reviews of Environmental Contamination and Toxicology Volume 223*. Springer, pp. 1–31.
- McCullagh, P. e Nelder, J. A. (1983). *Generalized Linear Models*. London, UK: Chapman e Hall.
- McLachlan, Geoffrey J e Krishnan, Thriyambakam (2007). *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons.
- Meilijson, Isaac (1989). “A fast improvement to the EM algorithm on its own terms”. *Journal of the Royal Statistical Society: Series B (Methodological)* 51.1, pp. 127–138.
- Nocedal, J. e Wright, S.J. (2006). *Numerical Optimization*. New York, US: Springer-Verlag.

- Oakes, David (1999). “Direct calculation of the information matrix via the EM”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.2, pp. 479–482.
- Paula, Gilberto A et al. (2012). “Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance”. *Applied Stochastic Models in Business and Industry* 28.1, pp. 16–34.
- Rao, C Radhakrishna (1948). “Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation”. *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 44. 1. Cambridge University Press, pp. 50–57.
- Rao, CR (2005). “Score test: historical review and recent developments”. Em: *Advances in ranking and selection, multiple comparisons, and reliability*. Springer, pp. 3–20.
- Rieck, James R e Nedelman, Jerry R (1991). “A log-linear model for the Birnbaum-Saunders distribution”. *Technometrics* 33.1, pp. 51–60.
- Sakamoto, Yosiyuki, Ishiguro, Makio e Kitagawa, Genshiro (1986). “Akaike information criterion statistics”. *Dordrecht, The Netherlands: D. Reidel* 81.
- Sánchez, L. et al. (2020a). “Birnbaum-Saunders quantile regression and its diagnostics with application to economic data”. *Applied Stochastic Models in Business and Industry*, pages in press available at <http://doi.org/10.1002/asmb.2556>.
- (2020b). “Birnbaum-Saunders quantile regression models with application to spatial data”. *Mathematics* 8, p. 1000.
- Saulo, H. e Leão, J. (2017). “On log-symmetric duration models applied to high frequency financial data”. *Economics Bulletin* 37, pp. 1089–1097.
- Schwarz, Gideon (1978). “Estimating the dimension of a model”. *The annals of statistics* 6.2, pp. 461–464.
- Spiegelhalter, David J et al. (2002). “Bayesian measures of model complexity and fit”. *Journal of the royal statistical society: Series b (statistical methodology)* 64.4, pp. 583–639.
- Sugiura, Nariaki (1978). “Further analysts of the data by akaike’s information criterion and the finite corrections: Further analysts of the data by akaike’s”. *Communications in Statistics-Theory and Methods* 7.1, pp. 13–26.
- Taylor, Julian e Verbyla, Arūnas (2004). “Joint modelling of location and scale parameters of the t distribution”. *Statistical Modelling* 4.2, pp. 91–112.

- Terrell, George R (2002). “The gradient statistic”. *Computing Science and Statistics* 34.34, pp. 206–215.
- Vanegas, L. e Paula, G. A. (2017). “Log-symmetric regression models under the presence of non-informative left-or right-censored observations”. *TEST* 26, pp. 405–428.
- Vanegas, L. H. e Paula, G. A. (2016a). “An extension of log-symmetric regression models: R codes and applications”. *Journal of Statistical Simulation and Computation* 86, pp. 1709–1735.
- (2016b). “Log-symmetric distributions: statistical properties and parameter estimation”. *Brazilian Journal of Probability and Statistics* 30, pp. 196–220.
- Vanegas, Luis Hernando e Paula, Gilberto A (2015). “A semiparametric approach for joint modeling of median and skewness”. *Test* 24.1, pp. 110–135.
- Ventura, M. et al. (2019). “Log-symmetric regression models: information criteria and application to movie business and industry data”. *Applied Stochastic Models in Business and Industry* DOI: 10.1002/asmb.2433.
- Wald, A. (1947). *Sequential Analysis*. New York, US: Wiley.
- Walker, Stephen G e Gutiérrez-Peña, Eduardo (2007). “Bayesian parametric inference in a nonparametric framework”. *Test* 16.1, pp. 188–197.
- Weisberg, Sanford (2014). *Applied Linear Regression*. Fourth Edition. Hoboken, New Jersey: John Wiley & Sons.
- West, Mike (1987). “On scale mixtures of normal distributions”. *Biometrika* 74.3, pp. 646–648.
- Wilks, Samuel S (1938). “The large-sample distribution of the likelihood ratio for testing composite hypotheses”. *The annals of mathematical statistics* 9.1, pp. 60–62.
- Yu, Keming, Lu, Zudi e Stander, Julian (2003). “Quantile regression: applications and current research areas”. *Journal of the Royal Statistical Society: Series D (The Statistician)* 52.3, pp. 331–350.
- Yu, Keming e Moyeed, Rana A (2001). “Bayesian quantile regression”. *Statistics & Probability Letters* 54.4, pp. 437–447.