

**Diagnóstico de influência  
bayesiano em modelos  
de regressão  
da família  $t$ -assimétrica**

Diego Wesllen da Silva

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

Programa: Estatística

Orientador: Profa. Dra. Márcia D'Elia Branco

São Paulo, fevereiro de 2017

# Diagnóstico de influência bayesiano em modelos de regressão da família $t$ -assimétrica

Esta é a versão original da dissertação/tese elaborada pelo  
candidato Diego Wesllen da Silva, tal como  
submetida à Comissão Julgadora.

# Agradecimentos

À minha amada e querida mãe Maria de Fatima da Silva que não mediu esforços, em toda sua vida, para que eu pudesse chegar onde cheguei, ser o que sou. Jamais conhecerei uma mulher com tamanha fibra e tenacidade. Ela é e sempre será um exemplo para toda minha vida. Cada conquista minha é, acima de tudo, sua conquista também.

Aos meus padrinhos Maria da Graça Richter e Geraldo Botelho Ribeiro por todo apoio e carinho que recebi na minha vida. Sempre provocaram minha curiosidade intelectual e, em uma das inúmeras conversas com eles, lembro-me de ter me conscientizado da beleza da exatidão da matemática.

À minha amada namorada e melhor amiga Karen Yuri Tanaka por todo o sincero e puro companheirismo destes anos juntos. Agradeço imensamente o apoio em minha vida nessa longa trajetória do mestrado. Sua presença motivou-me para enfrentar os momentos difíceis e valorizou cada momento bom.

À minha orientadora Márcia D'Elia Branco pela paciência, orientação e todo conhecimento transferido nestes anos de trabalho. A parceria destes anos reforçou em mim todo apreço que tenho pela estatística. Agradeço também a dedicação em formatar o pensamento científico em mim.

Aos grandes amigos da graduação Caio Jacob Milani, Eric Ossami Endo, Guilherme Feulo, Lucas Baroni Apicella, Marcelo Matheus Gauy, Renato Kechichian e Vinicius Morelli por toda a parceria durante o curso. Agradeço especialmente ao Francisco Felipe Chagas Rangel pela grande amizade nestes anos e ao Alexandre Oliveira que me provocou a cursar a matéria Probabilidade I, responsável despertar meu interesse pela estatística.

Aos amigos com os quais trabalhei Adriana Azevedo Hernandez Perez, Bernardo de Oliveira Guerra Ricca, Filipe Correia Gomes da Silva e Igor Siqueira Cortez que me guiaram nos primeiros usos práticos de estatística que fiz. Agradeço ao Carlos Eduardo Martins Relyas, Gustavo Di Giovanni Bernardo e Marcelo Magalhães Taddeo por serem fontes de inspiração para que eu buscasse este mestrado. Agradeço ao Sérgio Coichev Teixeira por ser um exemplo de que o meu sonho era factível. Agradeço ao Mark Surdutovich pelo apoio no início desta jornada.

Ao Instituto de Matemática e Estatística da Universidade de São Paulo e todos os seus professores, em especial os dos departamentos de matemática e de estatística, por fazerem parte de toda minha formação acadêmica.



# Resumo

SILVA, D. W. **Diagnóstico de influência bayesiano em modelos de regressão da família  $t$ -assimétrica**. 2017. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

O modelo de regressão linear com erros na família de distribuições  $t$ -assimétrica, que contempla as distribuições normal,  $t$ -Student e normal assimétrica como casos particulares, tem sido considerado uma alternativa robusta ao modelo normal. Para concluir qual modelo é, de fato, mais robusto é importante ter um método tanto para identificar uma observação como discrepante quanto aferir a influência que esta observação terá em nossas estimativas. Nos modelos de regressão bayesianos, uma das medidas de identificação de observações discrepantes mais conhecidas é a *conditional predictive ordinate* (CPO). Analisamos a influência destas observações nas estimativas tanto de forma global, isto é, no vetor completo de parâmetros do modelo quanto de forma marginal, apenas nos parâmetros regressores. Consideramos a norma  $L_1$  e a divergência Kullback-Leibler como medidas de influência das observações nas estimativas dos parâmetros. Além disso, encontramos as distribuições condicionais completas de todos os modelos para o uso do algoritmo de Gibbs obtendo, assim, amostras da distribuição *a posteriori* dos parâmetros. Tais amostras são utilizadas no cálculo do CPO e das medidas de divergência estudadas. A principal contribuição deste trabalho é obter as medidas de influência global e marginal calculadas para os modelos  $t$ -Student, normal assimétrico e  $t$ -assimétrico. Na aplicação em dados reais originais e contaminados, observamos que, em geral, o modelo  $t$ -Student é uma alternativa robusta ao modelo normal. Por outro lado, o modelo  $t$ -assimétrico não é, em geral, uma alternativa robusta ao modelo normal. A capacidade de robustificação do modelo  $t$ -assimétrico está diretamente ligada à posição do resíduo do ponto discrepante em relação a distribuição dos resíduos.

**Palavras-chave:** distribuição  $t$ -assimétrica, *conditional predictive ordinate*, CPO, norma  $L_1$ , Kullback-Leibler, medidas de influência.



# Abstract

SILVA, D. W. **Bayesian influence diagnostic in skew- $t$  family linear regression models**. 2017. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

The linear regression model with errors in the skew- $t$  family, which includes the normal, Student- $t$  and skew normal distributions as particular cases, has been considered as a robust alternative to the normal model. To conclude which model is in fact more robust it's important to have a method both to identify an observation as outlier and to assess the influence of this observation in the estimates. In the bayesian regression models, one of the most known measures to identify an outlier is the conditional predictive ordinate (CPO). We analyze the influence of these observations on the estimates both in a global way, that is, in the complete parameter vector of the model and in a marginal way, only in the regressor parameters. We consider the  $L_1$  norm and the Kullback-Leibler divergence as influence measures of the observations on the parameter estimates. Using the bayesian approach, we find the complete conditional distributions of all the models for the usage of the Gibbs sampler thus obtaining samples of the posterior distribution of the parameters. These samples are used in the calculation of the CPO and the studied divergences. The major contribution of this work is to present the global and marginal influence measures calculated for the Student- $t$ , skew normal and skew- $t$  models. In the application on original and contaminated real data, we observed that in general the Student- $t$  model is a robust alternative to the normal model. However, the skew- $t$  model isn't in general a robust alternative to the normal model. The robustification capability of the skew- $t$  model is directly linked to the position of the residual of the outlier in relation to the distribution of the residuals.

**Keywords:** skew- $t$  distribution, conditional predictive ordinate, CPO,  $L_1$  norm, Kullback-Leibler, influence measures.





# Conteúdo

<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivo e Organização do Trabalho . . . . .	2
1.2 Contribuições . . . . .	2
<b>2 Diagnóstico Bayesiano em Modelos de Regressão Simétricos</b>	<b>5</b>
2.1 Medidas de Diagnóstico . . . . .	6
2.1.1 Influência Global . . . . .	7
2.1.2 Influência Marginal . . . . .	10
2.2 Modelo de Regressão Normal . . . . .	11
2.2.1 Estimação dos Parâmetros . . . . .	11
2.2.2 Derivação das Medidas de Diagnóstico . . . . .	12
2.3 Modelo de Regressão t-Student . . . . .	13
2.3.1 Estimação dos Parâmetros . . . . .	13
2.3.2 Derivação das Medidas de Diagnóstico . . . . .	14
2.4 Aplicação . . . . .	16
<b>3 Diagnóstico Bayesiano em Modelos de Regressão Assimétricos</b>	<b>23</b>
3.1 Modelo de Regressão Normal Assimétrico . . . . .	23
3.1.1 Estimação dos Parâmetros . . . . .	23
3.1.2 Derivação das Medidas de Diagnóstico . . . . .	26
3.2 Modelo de Regressão <i>t</i> -Student Assimétrico . . . . .	32
3.2.1 Estimação dos Parâmetros . . . . .	32
3.2.2 Derivação das Medidas de Diagnóstico . . . . .	35
3.3 Aplicação . . . . .	40
3.3.1 Dados originais . . . . .	40
3.3.2 Contaminação da observação 19 . . . . .	46
3.3.3 Contaminação da observação 14 . . . . .	52
3.3.4 Contaminação das observações 14 e 19 . . . . .	58

<b>4</b>	<b>Conclusões</b>	<b>65</b>
<b>A</b>	<b>Códigos em R</b>	<b>67</b>
A.1	Estimação dos parâmetros dos modelos de regressão linear . . . . .	67
A.1.1	Pacotes utilizados . . . . .	67
A.1.2	Modelo Normal . . . . .	67
A.1.3	Modelo $t$ -Student . . . . .	68
A.1.4	Modelo Normal Assimétrico . . . . .	69
A.1.5	Modelo $t$ -assimétrico . . . . .	70
A.2	<i>Conditional Predictive Ordinate</i> e LPML . . . . .	72
A.2.1	Modelo Normal . . . . .	72
A.2.2	Modelo $t$ -Student . . . . .	72
A.2.3	Modelo Normal Assimétrico . . . . .	73
A.2.4	Modelo $t$ -assimétrico . . . . .	73
A.3	Medidas de influência global . . . . .	73
A.3.1	Modelo Normal . . . . .	73
A.3.2	Modelo $t$ -Student . . . . .	74
A.3.3	Modelo Normal Assimétrico . . . . .	74
A.3.4	Modelo $t$ -assimétrico . . . . .	74
A.4	Medidas de influência marginal . . . . .	75
A.4.1	Modelo Normal . . . . .	75
A.4.2	Modelo $t$ -Student . . . . .	75
A.4.3	Modelo Normal Assimétrico . . . . .	76
A.4.4	Modelo $t$ -assimétrico . . . . .	77
	<b>Bibliografia</b>	<b>79</b>

# Lista de Figuras

2.1	Diagrama de dispersão com pontos discrepantes destacados . . . . .	17
2.2	$-\log(CPO)$ para os modelos normal e $t$ -Student . . . . .	18
2.3	Diagrama de dispersão com as retas de regressão estimadas . . . . .	19
2.4	Influência global via norma $L_1$ . . . . .	19
2.5	Influência global via Kullback-Leibler . . . . .	20
2.6	Influência marginal via norma $L_1$ . . . . .	20
2.7	Influência marginal via Kullback-Leibler . . . . .	21
3.1	Diagrama de dispersão dos dados originais . . . . .	41
3.2	Boxplot dos resíduos do modelo de regressão normal nos dados originais . .	41
3.3	$-\log(CPO)$ dos modelos nos dados originais . . . . .	42
3.4	Influência global via norma $L_1$ dos modelos nos dados originais . . . . .	43
3.5	Influência global via divergência de Kullback-Leibler dos modelos nos dados originais . . . . .	44
3.6	Influência marginal via norma $L_1$ dos modelos nos dados originais . . . . .	45
3.7	Influência marginal via divergência de Kullback-Leibler dos modelos nos dados originais . . . . .	46
3.8	Diagrama de dispersão contaminando a observação 19 . . . . .	47
3.9	Boxplot dos resíduos do modelo de regressão normal contaminando a ob- servação 19 . . . . .	47
3.10	$-\log(CPO)$ dos modelos contaminando a observação 19 . . . . .	48
3.11	Influência global via norma $L_1$ dos modelos contaminando a observação 19	49
3.12	Influência global via divergência de Kullback-Leibler dos modelos contami- nando a observação 19 . . . . .	50
3.13	Influência marginal via norma $L_1$ dos modelos contaminando a observação 19 . . . . .	51
3.14	Influência marginal via divergência de Kullback-Leibler dos modelos con- taminando a observação 19 . . . . .	52
3.15	Diagrama de dispersão contaminando a observação 14 . . . . .	53
3.16	Boxplot dos resíduos do modelo de regressão normal contaminando a ob- servação 14 . . . . .	53

3.17 $-\log(CPO)$ dos modelos contaminando a observação 14 . . . . .	54
3.18 Influência global via norma $L_1$ dos modelos contaminando a observação 14 . . . . .	55
3.19 Influência global via divergência de Kullback-Leibler dos modelos contami- nando a observação 14 . . . . .	56
3.20 Influência marginal via norma $L_1$ dos modelos contaminando a observação 14 . . . . .	57
3.21 Influência marginal via divergência de Kullback-Leibler dos modelos con- taminando a observação 14 . . . . .	58
3.22 Diagrama de dispersão contaminando as observações 14 e 19 . . . . .	59
3.23 Boxplot dos resíduos do modelo de regressão normal contaminando as ob- servações 14 e 19 . . . . .	59
3.24 $-\log(CPO)$ dos modelos contaminando as observações 14 e 19 . . . . .	60
3.25 Influência global via norma $L_1$ dos modelos contaminando as observações 14 e 19 . . . . .	61
3.26 Influência global via divergência de Kullback-Leibler dos modelos contami- nando as observações 14 e 19 . . . . .	62
3.27 Influência marginal via norma $L_1$ dos modelos contaminando as observações 14 e 19 . . . . .	63
3.28 Influência marginal via divergência de Kullback-Leibler dos modelos con- taminando as observações 14 e 19 . . . . .	64

# Lista de Tabelas

2.1	Idade em meses em que a criança disse sua primeira palavra e <i>Gesell adaptive score</i> . . . . .	17
2.2	Estimativas dos parâmetros dos modelos . . . . .	18
2.3	Medidas de diagnóstico para as observações 2, 18 e 19 . . . . .	19
3.1	LPML dos modelos nos dados originais . . . . .	40
3.2	LPML dos modelos contaminando a observação 19 . . . . .	47
3.3	LPML dos modelos contaminando a observação 14 . . . . .	53
3.4	LPML dos modelos contaminando as observações 14 e 19 . . . . .	59



# Capítulo 1

## Introdução

O modelo de regressão normal é um dos mais utilizados na análise de variáveis aleatórias contínuas. Porém, a suposição de normalidade para os erros pode nem sempre ser a mais adequada aos dados obtidos. Mantendo a hipótese de simetria para os erros, o modelo de regressão  $t$ -Student tem sido proposto com o objetivo de se obter um modelo mais robusto. Com o modelo  $t$ -Student, podemos obter estimativas dos parâmetros menos influenciadas pela presença de dados muito distantes da média. Para saber se, de fato, nossas estimativas são menos influenciadas por observações discrepantes no modelo  $t$ -Student em relação ao modelo normal é preciso definir algumas medidas de influência. Dessa forma é importante ter um método tanto para identificar uma observação como discrepante quanto aferir a influência que esta terá em nossas estimativas.

Podemos também propor modelos assimétricos. O modelo mais natural para se considerar frente ao modelo normal é o modelo normal assimétrico. A distribuição normal assimétrica foi introduzida por [Azzalini \(1985\)](#) e estudada sob a abordagem bayesiana, por exemplo, por [Bayes \(2005\)](#). Ultimamente o uso da distribuição  $t$ -assimétrica tem sido proposto como uma alternativa robusta ao modelo de regressão normal, ver por exemplo [Azzalini e Genton \(2008\)](#).

A família de distribuições  $t$ -assimétrica é denotada por  $ST(\mu, \sigma^2, \nu, \lambda)$  em que  $\mu$  é um parâmetro de posição,  $\sigma^2$  é um parâmetro de escala,  $\nu$  os graus de liberdade e  $\lambda$  o parâmetro de assimetria ou forma (ver [Godoi \(2007\)](#)). Essa família também contempla, como casos particulares, as distribuições:

- Normal com média  $\mu$  e variância  $\sigma^2$ , denotada por  $N(\mu, \sigma^2)$ . Obtida considerando  $\nu \rightarrow \infty$  e  $\lambda = 0$ ;
- $t$ -Student com posição  $\mu$ , escala  $\sigma^2$  e graus de liberdade  $\nu$ , denotada por  $t(\mu, \sigma^2, \nu)$ . Obtida considerando  $\lambda = 0$ ;
- Normal assimétrica com posição  $\mu$ , escala  $\sigma^2$  e assimetria  $\lambda$ , denotada por  $SN(\mu, \sigma^2, \lambda)$ . Obtida considerando  $\nu \rightarrow \infty$ .

O modelo  $t$ -assimétrico foi introduzido em [Branco e Dey \(2001\)](#) e [Azzalini e Capitanio \(2003\)](#). Sob a abordagem bayesiana foi estudado por [Godoi \(2007\)](#). Também para estes modelos temos o desafio de identificar as observações discrepantes e aferir a influência que estas terão nas estimativas dos parâmetros do modelo.

Nos modelos de regressão bayesianos uma das medidas de identificação de observações discrepantes mais conhecidas é a *conditional predictive ordinate* (CPO). Baseada no

método de validação de modelos *leave-one-out*, esta medida avalia a função densidade de probabilidade em uma observação dadas as outras restantes.

Como medidas de influências das observações nas estimativas dos parâmetros consideraremos a família de divergências introduzidas em Weiss (1996) que possui a norma  $L_1$  e a divergência Kullback-Leibler como casos particulares. Também vale ressaltar que podemos medir a influência nas estimativas tanto de forma global, isto é, no vetor completo de parâmetros a serem estimados; quanto de forma marginal, em parte deste vetor de parâmetros. Ambas medidas serão consideradas neste trabalho.

## 1.1 Objetivo e Organização do Trabalho

O presente trabalho tem como objetivo aplicar os resultados sobre medidas de influência apresentados em Weiss (1996) e estender a aplicação para o modelo de regressão normal feita em Weiss e Cho (1998).

No segundo capítulo, abordamos os modelos de regressão linear baseados nas distribuições simétricas: normal e  $t$ -Student. Inicialmente, definimos as medidas de diagnóstico CPO e divergências global e marginal, aplicáveis a qualquer tipo de distribuição (não necessariamente simétrica). Para ambos modelos apresentamos como obter estimativas dos parâmetros de interesse sob o paradigma bayesiano. Mostramos também como ficam as medidas de diagnóstico para estes modelos exibindo resultados já encontrados em Weiss e Cho (1998) para o modelo normal e estendendo tais resultados para o modelo  $t$ -Student. Ao final do capítulo aplicamos os resultados obtidos em um conjunto de dados que exibe uma aparente normalidade mas que possui observações discrepantes.

No terceiro capítulo, exibimos resultados análogos aos obtidos no capítulo anterior para as distribuições assimétricas: normal assimétrica e  $t$ -assimétrica. A estimação dos parâmetros do modelo normal assimétrico utiliza os resultados obtidos em Bayes (2005). Na estimação dos parâmetros do modelo  $t$ -assimétrico usamos resultados obtidos em Godoi (2007) e, além disso, explicitamos a forma das condicionais completas para implementação do algoritmo de Gibbs. Toda a parte de diagnóstico dos modelos é feita de forma a complementar o trabalho de Weiss e Cho (1998). Ao final desse capítulo, aplicamos os resultados obtidos ao mesmo conjunto de dados estudado no capítulo 2, considerando algumas contaminações para verificar em que condições o modelo  $t$ -assimétrico realmente robustifica o modelo de regressão normal assimétrico.

No quarto capítulo, apresentamos as nossas conclusões e perspectivas para futuros trabalhos.

No apêndice, apresentamos os programas para obter as estimativas e medidas de diagnóstico dos quatro modelos abordados neste trabalho.

## 1.2 Contribuições

A principal contribuição deste trabalho é obter as medidas de diagnóstico global e marginal calculadas para os modelos  $t$ -Student, normal assimétrico e  $t$ -assimétrico. Para tanto, fizemos toda a dedução teórica destas medidas estendendo o trabalho de Weiss e Cho (1998) e também apresentamos como obter estimativas destas medidas por meio da integração de Monte Carlo.

Também fizemos aqui uma adição ao trabalho de Godoi (2007) exibindo as distribuições condicionais completas da distribuição  $t$ -Student possibilitando assim estruturarmos



o algoritmo de Gibbs para obtenção de uma amostra da distribuição *a posteriori*.



## Capítulo 2

# Diagnóstico Bayesiano em Modelos de Regressão Simétricos

O modelo de regressão Normal é um dos modelos mais conhecidos e utilizados, porém sua versatilidade costuma ser superestimada. Uma das fragilidades deste modelo está na hipótese de normalidade dos erros. Uma das características da distribuição Normal, além da sua simetria, é o baixo peso de suas caudas. Assim, sob a suposição de normalidade observações distantes da média possuem baixa probabilidade de ocorrência e, quando estas observações aparecem na amostra, acabam por influenciar muito as estimativas deste modelo.

Uma possível solução para melhorar as estimativas do modelo seria deletar as observações discrepantes (muito distantes da média). Porém, não é óbvio decidir quais observações deveriam ser deletadas já que não existe um único critério para classificar uma observação como discrepante. Uma solução melhor é propor um modelo que seja menos influenciado por observações discrepantes, mas que não despreze completamente estas observações.

O modelo de regressão com erros com distribuição  $t$ -Student tem sido proposto como um modelo mais robusto que o normal, isto é, menos influenciado por observações discrepante, devido ao fato de possuir caudas mais pesadas que a distribuição normal. No entanto, a obtenção das estimativas dos parâmetros do modelo  $t$ -Student apresenta dificuldades que não são encontradas no modelo normal. Sob o paradigma bayesiano, essas estimativas dependem da obtenção da distribuição *a posteriori* e, neste caso, técnicas computacionais são necessárias.

Sob qualquer um destes modelos ainda precisamos conseguir identificar de uma forma precisa quais observações são discrepantes já que uma análise visual pode ser arbitrária. Também precisamos, uma vez identificadas tais observações, saber se estas são de fato influentes nas estimativas dos parâmetros do modelo. Uma medida utilizada para identificação de pontos discrepantes é a *conditional predictive ordinate* (CPO). Neste capítulo mostramos o cálculo desta estatística para os modelos de regressão linear normal e  $t$ -Student. A influência das observações nas estimativas é abordada por meio das medidas de divergência norma  $L_1$  e Kullback-Leibler. Tais medidas de influência são calculadas tanto para os parâmetros dos modelos de uma forma global, isto é, para o conjunto completo de parâmetros do modelo; quanto marginalmente, apenas para os coeficientes regressores.

Ao final do capítulo aplicamos ambos modelos a um conjunto de dados presente em [Weiss e Cho \(1998\)](#) e utilizado em outros trabalhos que tratam de pontos discrepantes. Além de aplicar os modelos, também mostramos como estes são influenciados pelas ob-

servações.

## 2.1 Medidas de Diagnóstico

Uma técnica estatística amplamente utilizada para avaliar a qualidade de um determinado modelo é a validação cruzada. Esta técnica tem como objetivo avaliar a capacidade de previsão de um modelo estatístico a partir de um conjunto de dados fixado. Assim, particiona-se o conjunto de dados em duas partes. A primeira parte, denominada amostra de treinamento, é utilizada para obter as estimativas dos parâmetros do modelo proposto e a segunda, denominada amostra de validação, é utilizada para verificar a qualidade das estimativas deste modelo frente a novos dados.

Considere, inicialmente,  $\mathbf{y} = (y_1, \dots, y_n)$  a amostra observada de um modelo  $Y|\theta$ . Sejam  $I_1, \dots, I_K$  tais que  $\{1, \dots, n\} = \cup_{k=1}^K I_k$  e,  $\forall i \neq j$ ,  $I_i \cap I_j = \emptyset$ , isto é,  $I_1, \dots, I_K$  é uma partição do conjunto de índices  $\{1, \dots, n\}$ . O processo da validação cruzada consiste em  $K$  etapas de estimação de parâmetros e validação destas estimativas. Na  $k$ -ésima etapa separamos nossos dados em dois conjuntos:  $\mathbf{y}_{(I_k)} = (y_j)_{j \notin I_k}$  e  $\mathbf{y}_{I_k} = (y_j)_{j \in I_k}$ . Dessa forma, utilizamos o primeiro conjunto,  $\mathbf{y}_{(I_k)}$ , para obter uma estimativa de  $\theta$  e o segundo conjunto,  $\mathbf{y}_{I_k}$ , para validação desta estimativa. Após as  $K$  etapas, todas as observações foram utilizadas para validar estimativas das quais elas não fizeram parte. Usualmente, os conjuntos  $I_k$  possuem o mesmo tamanho e são escolhidos de forma aleatória. Existem diversos procedimentos para se conduzir tal processo de validação cruzada sendo o método *leave-one-out* um dos mais comuns.

O método *leave-one-out* considera  $K = n$  e assim teremos  $I_i = \{i\}$ , para cada  $i = 1, \dots, n$ . Dessa forma, vamos obter a estimativa de  $\theta$  em  $\mathbf{y}_{(i)} = (y_j)_{j \neq i}$  e validar na observação  $y_i$ . A distribuição preditiva de  $\mathbf{y}_{(i)}$  dada por

$$f(y_i|\mathbf{y}_{(i)}) = \int f(y_i|\theta)f(\theta|\mathbf{y}_{(i)})d\theta \quad (2.1)$$

é a densidade da amostra de validação, neste caso composta por somente  $y_i$ , dado o restante dos dados,  $\mathbf{y}_{(i)}$ . A quantidade  $f(y_i|\mathbf{y}_{(i)})$  é chamada de  $i$ -ésima ordenada preditiva condicional, em inglês *conditional predictive ordinate* ( $CPO_i$ ). Portanto, valores baixos de  $CPO_i$  podem ser associados a valores discrepantes sob o modelo proposto, já que a densidade calculada no ponto  $y_i$  indicará uma região de baixa probabilidade. Uma maneira de comparar a falta de adequação de cada observação em diferentes modelos propostos é por meio da comparação das  $CPO_i$  calculadas sob cada modelo. Perceba que tal comparação é possível pois, como podemos ver na equação (2.1), o  $CPO_i$  é uma medida que depende somente da distribuição preditiva pois todos os parâmetros são integrados.

Note que podemos escrever

$$f(y_i|\mathbf{y}_{(i)}) = \int f(y_i|\mathbf{y}_{(i)}, \theta)f(\theta|\mathbf{y}_{(i)})d\theta = \int f(y_i|\theta)f(\theta|\mathbf{y}_{(i)})d\theta = E_{\theta|\mathbf{y}_{(i)}}[f(y_i|\theta)] \quad (2.2)$$

e, sendo  $\theta^{(1)}, \dots, \theta^{(L)}$  uma amostra de tamanho  $L$  da distribuição *a posteriori* de  $\theta|\mathbf{y}_{(i)}$ , por meio da integração de Monte Carlo, podemos aproximar (2.2) por

$$E_{\theta|\mathbf{y}_{(i)}}[f(y_i|\theta)] \approx \frac{1}{L} \sum_{l=1}^L f(y_i|\theta^{(l)}). \quad (2.3)$$

Porém, a aproximação apresentada em (2.3) necessita a obtenção de uma amostra da distribuição *a posteriori*  $f(\theta|\mathbf{y}_{(i)})$  para cada  $i = 1, \dots, n$  e isto pode ser computacionalmente muito custoso. Podemos utilizar a seguinte caracterização alternativa para o  $CPO_i$ , que simplifica o esforço computacional pois a esperança é calculada com a distribuição *a posteriori* completa de  $\theta|\mathbf{y}$  para todo  $i = 1, \dots, n$

$$\begin{aligned} f(y_i|\mathbf{y}_{(i)}) &= \frac{f(\mathbf{y})}{f(\mathbf{y}_{(i)})} = \frac{\int f(\mathbf{y}|\theta)f(\theta)d\theta}{\int f(\mathbf{y}_{(i)}|\theta)f(\theta)d\theta} = \frac{\int \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})}d\theta}{\int \frac{f(\mathbf{y}_{(i)}|\theta)f(\theta)}{f(\mathbf{y})} \frac{f(\theta|\mathbf{y})}{f(\theta|\mathbf{y}_{(i)})}d\theta} \\ &= \frac{\int f(\theta|\mathbf{y})d\theta}{\int \frac{f(\mathbf{y}_{(i)}|\theta)f(\theta)}{f(\mathbf{y})} \frac{f(\mathbf{y})}{f(\mathbf{y}_{(i)}|f(\theta))} f(\theta|\mathbf{y})d\theta} = \frac{1}{\int \frac{1}{f(\mathbf{y}_{(i)}|\theta)} f(\theta|\mathbf{y})d\theta} \\ &= E_{\theta|\mathbf{y}}[f(y_i|\theta)^{-1}]^{-1} \end{aligned} \quad (2.4)$$

e a equação (2.4) pode ser aproximada por

$$E_{\theta|\mathbf{y}}[f(y_i|\theta)^{-1}]^{-1} \approx \left( \frac{1}{L} \sum_{l=1}^L \frac{1}{f(y_i|\theta^{(l)})} \right)^{-1}, \quad (2.5)$$

onde  $\theta^{(1)}, \dots, \theta^{(L)}$  é uma amostra de tamanho  $L$  de  $\theta|\mathbf{y}$  em que  $\mathbf{y}$  é o vetor com todas as observações. Agora, precisamos de somente uma amostra da distribuição *a posteriori* completa para o cálculo do  $CPO_i$ .

Uma medida resumo natural para os  $CPO_i$  é a *log pseudo marginal likelihood* (LMPL), sugerida em Ibrahim *et al.* (2001), definida por

$$LPML = \sum_{i=1}^n \log(CPO_i). \quad (2.6)$$

Esta estatística pode ser considerada uma medida de ajuste. O modelo que possuir o maior valor de LPML será considerado o mais adequado aos dados.

Utilizaremos o  $CPO_i$  para identificar observações discrepantes. Mais precisamente, utilizaremos a estatística  $-\log(CPO_i)$  pois as observações discrepantes ficam mais evidentes em uma inspeção gráfica.

### 2.1.1 Influência Global

Uma medida de influência de cada observação na estimação de um vetor de parâmetros  $\theta$  é dada pela divergência entre a distribuição *a posteriori*  $f(\theta|\mathbf{y})$  e a distribuição *a posteriori* obtida na ausência da  $i$ -ésima observação  $f(\theta|\mathbf{y}_{(i)})$ . A família de divergências apresentada em Weiss (1996) é dada por

$$D_\theta(g, i) = \int g \left( \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} \right) f(\theta|\mathbf{y})d\theta, \quad (2.7)$$

onde  $g$  é uma função convexa com  $g(1) = 0$ .

Sob a ausência de influência da  $i$ -ésima observação, teríamos que  $f(\theta|\mathbf{y}) = f(\theta|\mathbf{y}_{(i)})$  e assim

$$D_\theta(g, i) = \int g \left( \frac{f(\theta|\mathbf{y})}{f(\theta|\mathbf{y})} \right) f(\theta|\mathbf{y})d\theta = \int g(1)f(\theta|\mathbf{y})d\theta = 0. \quad (2.8)$$

Note também que  $D_\theta(g, i) \geq 0$  pois, sendo  $g$  convexa, pela desigualdade de Jensen,

$$\begin{aligned} D_\theta(g, i) &= \int g\left(\frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})}\right) f(\theta|\mathbf{y}) d\theta = E_{\theta|\mathbf{y}} \left[ g\left(\frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})}\right) \right] \\ &\geq g\left(E_{\theta|\mathbf{y}} \left[ \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} \right]\right) = g\left(\int f(\theta|\mathbf{y}_{(i)}) d\theta\right) = g(1) = 0. \end{aligned} \quad (2.9)$$

Teremos que, se  $g$  for estritamente convexa em 1,  $D_\theta(g, i) = 0$  se, e somente se,  $f(\theta|\mathbf{y}) = f(\theta|\mathbf{y}_{(i)})$  quase certamente. Logo, a divergência  $D_\theta(g, i)$  nos dará uma medida de similaridade entre as distribuições  $f(\theta|\mathbf{y})$  e  $f(\theta|\mathbf{y}_{(i)})$ .

A  $CPO_i$ , definida anteriormente, é relacionada com a divergência da seguinte forma

$$\frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} = \frac{\frac{f(\mathbf{y}_{(i)}|\theta)f(\theta)}{f(\mathbf{y}_{(i)})}}{\frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})}} = \frac{f(\mathbf{y}_{(i)}|\theta)}{f(\mathbf{y}|\theta)} \frac{f(\mathbf{y})}{f(\mathbf{y}_{(i)})} = \frac{1}{f(y_i|\theta)} f(y_i|\mathbf{y}_{(i)}) = \frac{CPO_i}{f(y_i|\theta)} = h_i(\theta) \quad (2.10)$$

e usualmente  $h_i$  é chamada de função de perturbação.

Portanto,

$$D_\theta(g, i) = \int g\left(\frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})}\right) f(\theta|\mathbf{y}) d\theta = E_{\theta|\mathbf{y}} \left[ g\left(\frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})}\right) \right] = E_{\theta|\mathbf{y}} \left[ g\left(\frac{CPO_i}{f(y_i|\theta)}\right) \right]. \quad (2.11)$$

Uma forma aproximada de obter  $D_\theta(g, i)$  é por meio da integração de Monte Carlo. Seja  $\theta^{(1)}, \dots, \theta^{(L)}$ , uma amostra simulada de tamanho  $L$  da distribuição *a posteriori* de  $\theta|\mathbf{y}$ , podemos obter a seguinte aproximação

$$D_\theta(g, i) \approx \frac{1}{L} \sum_{l=1}^n g\left(\frac{CPO_i}{f(y_i|\theta^{(l)})}\right), \quad (2.12)$$

a qual é fácil de ser calculada pois já vimos na equação (2.5) como obter uma estimativa do  $CPO_i$  e  $f(y_i|\theta)$  sempre terá uma forma conhecida. Caso tentássemos trabalhar com a primeira representação de  $D_\theta(g, i)$  apresentada na equação (2.7), precisaríamos obter formas explícitas para  $f(\theta|\mathbf{y})$  e  $f(\theta|\mathbf{y}_{(i)})$  e isto nem sempre é possível de obter.

Neste trabalho vamos nos restringir a duas medidas específicas de divergência: norma  $L_1$  e divergência de Kullback-Leibler.

### Norma $L_1$

A norma  $L_1$  é um caso particular da divergência apresentada na equação (2.7) com  $g(a) = \frac{1}{2}|a - 1|$ . Para que  $g$  seja uma função que define uma divergência precisamos que  $g(1) = 0$  e que  $g$  seja convexa. A primeira condição é satisfeita de forma clara. Para verificar a convexidade, seja  $a_1, a_2 \in \mathbb{R}$  e  $t \in [0, 1]$ , então

$$\begin{aligned} g(ta_1 + (1-t)a_2) &= \frac{1}{2}|ta_1 + (1-t)a_2 - 1| = \frac{1}{2}|t(a_1 - 1) + (1-t)(a_2 - 1)| \\ &\leq t\frac{1}{2}|a_1 - 1| + (1-t)\frac{1}{2}|a_2 - 1| = tg(a_1) + (1-t)g(a_2). \end{aligned} \quad (2.13)$$

A norma  $L_1$  será dada por

$$\begin{aligned} L_{1i,\theta} &= D_\theta(g, i) = \int g \left( \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} \right) f(\theta|\mathbf{y}) d\theta = \int \frac{1}{2} \left| \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} - 1 \right| f(\theta|\mathbf{y}) d\theta \\ &= \frac{1}{2} \int |f(\theta|\mathbf{y}_{(i)}) - f(\theta|\mathbf{y})| d\theta. \end{aligned} \quad (2.14)$$

Uma caracterização alternativa para norma  $L_1$  é

$$L_{1i,\theta} = \sup_B \int_B f(\theta|\mathbf{y}_{(i)}) - f(\theta|\mathbf{y}) d\theta = \sup_B \int_B f(\theta|\mathbf{y}) - f(\theta|\mathbf{y}_{(i)}) d\theta. \quad (2.15)$$

A caracterização apresentada em (2.15) é conhecida como distância variacional total. Podemos interpretá-la como a maior diferença possível entre as funções densidade  $f(\theta|\mathbf{y})$  e  $f(\theta|\mathbf{y}_{(i)})$ . Como Weiss (1996) observa, na equação (2.15) o supremo da primeira igualdade é atingido no conjunto  $B = \{\theta : f(\theta|\mathbf{y}_{(i)}) > f(\theta|\mathbf{y})\}$  enquanto que o supremo da segunda igualdade é atingido no conjunto complementar de  $B$ .

Como, para todo  $x, y \in \mathbb{R}$ , vale que  $|x - y| \leq |x| + |y|$ , teremos que  $L_{1i,\theta} \leq 1$  pois

$$\begin{aligned} L_{1i,\theta} &= \frac{1}{2} \int |f(\theta|\mathbf{y}_{(i)}) - f(\theta|\mathbf{y})| d\theta \leq \frac{1}{2} \int (f(\theta|\mathbf{y}_{(i)}) + f(\theta|\mathbf{y})) d\theta \\ &= \frac{1}{2} \int f(\theta|\mathbf{y}_{(i)}) d\theta + \frac{1}{2} \int f(\theta|\mathbf{y}) d\theta = 1 \end{aligned} \quad (2.16)$$

Logo, podemos ver que na função  $g$  a constante  $\frac{1}{2}$  faz com que o valor máximo de  $L_{1i,\theta}$  seja 1.

### Kullback-Leibler

A divergência de Kullback-Leibler é o caso particular de (2.7) quando  $g(a) = a \log(a)$ . Veja que também claramente  $g(1) = 0$ . Para verificar a convexidade, note que

$$g'(a) = \log(a) + 1 \Rightarrow g''(a) = \frac{1}{a} > 0. \quad (2.17)$$

Logo, sendo a segunda derivada positiva para todo  $a > 0$ ,  $g$  é convexa.

A divergência de Kullback-Leibler fica dada por

$$\begin{aligned} K_{i,\theta} &= D_\theta(g, i) = \int g \left( \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} \right) f(\theta|\mathbf{y}) d\theta \\ &= \int \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} \log \left( \frac{f(\theta|\mathbf{y}_{(i)})}{f(\theta|\mathbf{y})} \right) f(\theta|\mathbf{y}) d\theta \\ &= E_{\theta|\mathbf{y}_{(i)}}[\log(f(\theta|\mathbf{y}_{(i)}))] - E_{\theta|\mathbf{y}_{(i)}}[\log(f(\theta|\mathbf{y}))] \end{aligned} \quad (2.18)$$

Não é possível encontrar de uma forma simples um limite superior para divergência de Kullback-Leibler como fizemos com a norma  $L_1$ . Além disso, ela é de difícil interpretação dificultando assim entender o que o seu valor representa. Há uma grande quantidade de trabalhos focados em dar uma interpretação a esta divergência. Contudo, podemos utilizar outras medidas de divergência que são intrinsecamente mais interpretáveis como, por exemplo, a norma  $L_1$ .

### 2.1.2 Influência Marginal

Nem sempre estamos interessados na influência de uma observação no vetor completo de parâmetros  $\theta$ . Nosso interesse pode estar na influência de uma observação em parte do vetor, ou seja, estamos interessados em uma influência marginal. Denotando por  $\theta = (\theta_1, \theta_2)$  o vetor completo de parâmetros, uma medida para influência da  $i$ -ésima observação nas estimativas de  $\theta_1$  é dada pela divergência

$$D_{\theta_1}(g, i) = \int g \left( \frac{f(\theta_1 | \mathbf{y}_{(i)})}{f(\theta_1 | \mathbf{y})} \right) f(\theta_1 | \mathbf{y}) d\theta_1 \quad (2.19)$$

que possui a definição análoga ao caso global.

Note que

$$\frac{f(\theta_1 | \mathbf{y}_{(i)})}{f(\theta_1 | \mathbf{y})} = \frac{\frac{f(\mathbf{y}_{(i)} | \theta_1) f(\theta_1)}{f(\mathbf{y}_{(i)})}}{\frac{f(\mathbf{y} | \theta_1) f(\theta_1)}{f(\mathbf{y})}} = \frac{f(\mathbf{y})}{f(\mathbf{y}_{(i)})} \frac{1}{\frac{f(\mathbf{y} | \theta_1)}{f(\mathbf{y}_{(i)} | \theta_1)}} = \frac{CPO_i}{f(y_i | \theta_1, \mathbf{y}_{(i)})} = h_{1i}(\theta_1) \quad (2.20)$$

sendo  $h_{1i}$  a função de perturbação neste caso marginal.

A medida de influência no vetor completo de parâmetros será maior do que em um subconjunto do vetor de parâmetros. Supondo  $\theta = (\theta_1, \theta_2)$ , o Teorema 2, p. 742, de Weiss (1996) diz que

$$0 \leq D_{\theta_1}(g, i) \leq D_{\theta}(g, i) \quad (2.21)$$

Logo, a influência marginal será sempre menor que a influência global. Portanto, uma observação pode exercer influência de forma global, mas não de forma marginal.

Para encontrar um cálculo aproximado para  $D_{\theta_1}(g, i)$ , também utilizaremos a integração de Monte Carlo. Veja que precisamos encontrar o valor de  $CPO_i$  e, a princípio, uma forma explícita para  $f(y_i | \theta_1, \mathbf{y}_{(i)})$ . Diferentemente do que ocorre no cálculo da divergência global onde  $f(y_i | \theta)$  está sempre disponível, no cálculo da divergência marginal precisamos da quantidade  $f(y_i | \theta_1, \mathbf{y}_{(i)})$  que nem sempre será fácil encontrar de forma explícita. A relação desenvolvida a seguir facilitará a obtenção de uma estimativa para essa distribuição. Note que,

$$\begin{aligned} f(y_i | \theta_1, \mathbf{y}_{(i)}) &= \int f(y_i | \theta) f(\theta_2 | \theta_1, \mathbf{y}_{(i)}) d\theta_2 = \int f(y_i | \theta) \frac{f(\theta, \mathbf{y}_{(i)})}{f(\theta_1, \mathbf{y}_{(i)})} d\theta_2 \\ &= \int f(y_i | \theta) \frac{f(\mathbf{y}_{(i)} | \theta) f(\theta)}{f(\theta_1, \mathbf{y}_{(i)})} d\theta_2 = \int \frac{f(\mathbf{y} | \theta) f(\theta)}{f(\theta_1, \mathbf{y}_{(i)})} d\theta_2 \\ &\propto \int f(\mathbf{y} | \theta) f(\theta) d\theta_2 \end{aligned} \quad (2.22)$$

Assim, encontramos uma proporcionalidade para  $f(y_i | \theta_1, \mathbf{y}_{(i)})$ , mas precisamos ainda encontrar a constante de padronização integrando o último termo de (2.22) em  $y_i$ .

A seguir apresentamos os modelos de regressão linear normal e  $t$ -Student e derivamos as medidas de diagnóstico aqui apresentadas para cada um dos modelos.



## 2.2 Modelo de Regressão Normal

### 2.2.1 Estimação dos Parâmetros

O modelo de regressão linear com erros de distribuição normal é dado por

$$y_i = x_i^t \boldsymbol{\beta} + \epsilon_i, \quad (2.23)$$

onde, para cada  $i = 1, \dots, n$ ,  $y_i$  é a variável resposta da  $i$ -ésima observação,  $x_i^t = (1, x_{i1}, \dots, x_{ip})$  é o vetor de variáveis explicativas,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^t$  é vetor de parâmetros desconhecidos a serem estimados e  $\epsilon_i \sim N(0, \sigma^2)$  independentes, ou seja, os erros possuem distribuição normal com média 0 variância  $\sigma^2$ .

Para cada  $i$  temos que  $y_i | \boldsymbol{\beta}, \sigma^2 \sim N(x_i^t \boldsymbol{\beta}, \sigma^2)$  e pela independência a verossimilhança deste modelo fica dada por

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2) = \prod_{j=1}^n f(y_j | \boldsymbol{\beta}, \sigma^2) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_j - x_j^t \boldsymbol{\beta})^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta})^2\right) \end{aligned} \quad (2.24)$$

Trataremos este e os outros modelos sob a perspectiva bayesiana de inferência, então é necessário também especificar distribuições *a priori* para  $\boldsymbol{\beta}$  e  $\sigma^2$ . Neste estudo iremos usar distribuições *a priori* não informativas considerando, assim, desconhecimento prévio a respeito dos parâmetros a serem estimados.

A distribuição *a priori* de Jeffreys é uma das mais utilizadas quando se quer trabalhar com distribuições *a priori* não informativas e é definida por

$$f(\theta) \propto \sqrt{\det I(\theta)}, \quad (2.25)$$

onde  $I(\theta)$  é a matriz de informação de Fisher de  $\theta$  (ver [Paulino et al. \(2003\)](#), página 99). Esta distribuição *a priori* possui a propriedade básica de ser invariante sob reparametrizações.

Supondo independência *a priori* entre  $\boldsymbol{\beta}$  e  $\sigma^2$ , e obtendo a distribuição *a priori* de Jeffreys para cada parâmetro, teremos a seguinte distribuição *a priori* conjunta

$$f(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}. \quad (2.26)$$

A distribuição *a posteriori* dos parâmetros do modelo de regressão fica dada por

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) &\propto L(\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) \\ &\propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta})^2\right) \frac{1}{\sigma^2} \\ &\propto \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta})^2\right) \end{aligned} \quad (2.27)$$

e este é o núcleo de uma distribuição Normal-Gama Invertida. Podemos representar tal

distribuição da forma hierárquica a seguir

$$\begin{aligned}\boldsymbol{\beta}|\sigma^2, \mathbf{y} &\sim N_{p+1}(\hat{\boldsymbol{\beta}}, \sigma^2(X^t X)^{-1}) \\ \sigma^2|\mathbf{y} &\sim GI\left(\frac{n - (p + 1)}{2}, \frac{(\mathbf{y} - X\hat{\boldsymbol{\beta}})^t(\mathbf{y} - X\hat{\boldsymbol{\beta}})}{2}\right),\end{aligned}\quad (2.28)$$

onde  $\hat{\boldsymbol{\beta}} = (X^t X)^{-1} X^t \mathbf{y}$ .

Uma forma de obter uma amostra de tamanho  $L$  da distribuição *a posteriori* conjunta de  $(\boldsymbol{\beta}, \sigma^2)|\mathbf{y}$  é utilizar o algoritmo a seguir:

Para cada  $l = 1, \dots, L$

- (a) Simular  $\sigma^{2(l)}$  da distribuição de  $\sigma^2|\mathbf{y}$ .
- (b) Simular  $\boldsymbol{\beta}^{(l)}$  da distribuição de  $\boldsymbol{\beta}|\sigma^{2(l)}, \mathbf{y}$ .

Assim,  $(\boldsymbol{\beta}^{(1)}, \sigma^{2(1)}), \dots, (\boldsymbol{\beta}^{(L)}, \sigma^{2(L)})$  é a amostra desejada

### 2.2.2 Derivação das Medidas de Diagnóstico

Considere  $(\boldsymbol{\beta}^{(1)}, \sigma^{2(1)}), \dots, (\boldsymbol{\beta}^{(L)}, \sigma^{2(L)})$  uma amostra de tamanho  $L$  da distribuição *a posteriori* de  $\boldsymbol{\beta}, \sigma^2|\mathbf{y}$ . Utilizaremos esta amostra em todos os cálculos das medidas de diagnóstico.

Mostramos na equação (2.4) como obter uma estimativa para o  $CPO_i$  em um caso geral. No caso da regressão linear normal teremos

$$\widehat{CPO}_i = \left[ \frac{1}{L} \sum_{l=1}^n \frac{1}{f(y_i|\boldsymbol{\beta}^{(l)}, \sigma^{2(l)})} \right]^{-1} \quad (2.29)$$

onde  $f(y_i|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i^t \boldsymbol{\beta})^2\right)$ , isto é, a função de distribuição normal de média  $x_i^t \boldsymbol{\beta}$  e variância  $\sigma^2$ .

A estimativa para divergência sob o contexto da influência global apresentada em (2.12) ficará dada por

$$D_{\boldsymbol{\beta}, \sigma^2}(\widehat{g}, i) = \frac{1}{L} \sum_{l=1}^n g\left(\frac{\widehat{CPO}_i}{f(y_i|\boldsymbol{\beta}^{(l)}, \sigma^{2(l)})}\right) \quad (2.30)$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (2.29). Para a obtenção da norma  $L_1$  e da divergência Kullback-Leibler, basta substituir em (2.30) a função  $g$  adequada para cada caso.

A influência marginal é medida considerando não todo o vetor de parâmetros  $(\boldsymbol{\beta}, \sigma^2)$  mas parte deste vetor. Neste trabalho, consideramos a influência marginal em  $\boldsymbol{\beta}$ . A obtenção da divergência para o caso da influência marginal possui o desafio de encontrar a distribuição  $f(y_i|\boldsymbol{\beta}, \mathbf{y}_{(i)})$ . No caso normal, teremos

$$\begin{aligned}f(y_i|\boldsymbol{\beta}, \mathbf{y}_{(i)}) &\propto \int f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) d\sigma^2 \propto \\ &\propto \int \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})^t(\mathbf{y} - X\boldsymbol{\beta})\right) d\sigma^2\end{aligned}$$

$$\begin{aligned}
&= \frac{\Gamma(n/2)}{\left[ \frac{(Y-X\boldsymbol{\beta})^t(Y-X\boldsymbol{\beta})}{2} \right]^{n/2}} \\
&\propto [(Y-X\boldsymbol{\beta})^t(Y-X\boldsymbol{\beta})]^{-n/2} \\
&= \left[ \sum_{j \neq i} (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-n/2} \left[ 1 + \frac{(y_i - x_i^t \boldsymbol{\beta})^2}{\sum_{j \neq i} (y_j - x_j^t \boldsymbol{\beta})^2} \right]^{-n/2} \\
&\propto \left[ 1 + \frac{1}{n-1} \left( \frac{y_i - x_i^t \boldsymbol{\beta}}{\sqrt{\frac{1}{n-1} \sum_{j \neq i} (y_j - x_j^t \boldsymbol{\beta})^2}} \right)^2 \right]^{-n/2}
\end{aligned}$$

então  $y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}$  tem distribuição  $t$ -Student com parâmetro de posição  $x_i^t \boldsymbol{\beta}$ , escala  $\frac{1}{n-1} \sum_{j \neq i} (y_j - x_j^t \boldsymbol{\beta})^2$  e graus de liberdade  $n-1$ . Dessa forma, a medida de influência marginal é dada por

$$\widehat{D_{\boldsymbol{\beta}}(g, i)} = \frac{1}{L} \sum_{l=1}^n g \left( \frac{\widehat{CPO}_i}{f(y_i | \boldsymbol{\beta}^{(l)}, \mathbf{y}_{(i)})} \right) \quad (2.31)$$

onde todos os componentes têm forma conhecida. A partir da equação (2.31) conseguimos calcular a norma  $L_1$  e a divergência de Kullback-Leibler para este caso marginal.

## 2.3 Modelo de Regressão t-Student

### 2.3.1 Estimação dos Parâmetros

O modelo de regressão linear com erros de distribuição  $t$ -Student é dado por

$$y_i = x_i^t \boldsymbol{\beta} + \epsilon_i, \quad (2.32)$$

onde, para cada  $i = 1, \dots, n$ ,  $y_i$  é a variável resposta da  $i$ -ésima observação,  $x_i^t = (1, x_{i1}, \dots, x_{ip})$  é o vetor de variáveis explicativas,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^t$  é vetor de parâmetros desconhecidos a serem estimados e  $\epsilon_i \sim t(0, \sigma^2, \nu)$  independentes, ou seja, os erros possuem distribuição  $t$ -Student com posição 0, escala  $\sigma^2$  e graus de liberdade  $\nu$ . Neste trabalho consideramos o parâmetro  $\nu$  fixado.

Para cada  $i$  temos que  $y_i | \boldsymbol{\beta}, \sigma^2 \sim t(x_i^t \boldsymbol{\beta}, \sigma^2, \nu)$  e podemos trabalhar com a seguinte representação hierárquica

$$\begin{aligned}
y_i | \boldsymbol{\beta}, \sigma^2, u_i &\sim N(x_i^t \boldsymbol{\beta}, \sigma^2 / u_i) \\
u_i &\sim \text{Gama}(\nu/2, \nu/2)
\end{aligned} \quad (2.33)$$

Resultando na seguinte função de verossimilhança aumentada

$$\begin{aligned}
L_A(\boldsymbol{\beta}, \sigma^2) &= f(\mathbf{y}, \mathbf{u} | \boldsymbol{\beta}, \sigma^2) = f(\mathbf{y} | \mathbf{u}, \boldsymbol{\beta}, \sigma^2) f(\mathbf{u}) \propto \\
&\propto \frac{1}{\sigma^n} \prod_{i=1}^n \sqrt{u_i} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n u_i (y_i - x_i^t \boldsymbol{\beta})^2 \right) \prod_{i=1}^n u_i^{\frac{\nu}{2}-1} \exp \left( -\frac{\nu}{2} \sum_{i=1}^n u_i \right) \\
&= \frac{1}{\sigma^n} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n u_i (y_i - x_i^t \boldsymbol{\beta})^2 - \frac{\nu}{2} \sum_{i=1}^n u_i \right) \prod_{i=1}^n u_i^{\frac{\nu-1}{2}}.
\end{aligned} \quad (2.34)$$

Vamos considerar a mesma distribuição *a priori* conjunta para  $(\boldsymbol{\beta}, \sigma^2)$ , definida anteriormente para o modelo normal na equação (2.26), ou seja,

$$f(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma^2}. \quad (2.35)$$

Neste caso, a distribuição *a posteriori* aumentada é dada por

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2, \mathbf{u}|\mathbf{y}) &\propto L_A(\boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) \\ &\propto \frac{1}{\sigma^{n+2}} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n u_i (y_i - x_i^t \boldsymbol{\beta})^2 - \frac{\nu}{2} \sum_{i=1}^n u_i \right) \prod_{i=1}^n u_i^{\frac{\nu-1}{2}} \end{aligned} \quad (2.36)$$

que possui as seguintes distribuições condicionais completas

$$\begin{aligned} \boldsymbol{\beta}|\sigma^2, \mathbf{u}, \mathbf{y} &\sim N_{p+1}(\hat{\boldsymbol{\beta}}, \sigma^2(X^t U X)^{-1}) \\ \sigma^2|\boldsymbol{\beta}, \mathbf{u}, \mathbf{y} &\sim GI \left( \frac{n}{2}, \frac{(Y - X\boldsymbol{\beta})^t U (Y - X\boldsymbol{\beta})}{2} \right) \\ u_i|\boldsymbol{\beta}, \sigma^2, \mathbf{y} &\sim Gama \left( \frac{\nu+1}{2}, \frac{y_i - x_i^t \boldsymbol{\beta}}{2\sigma^2} + \frac{\nu}{2} \right) \text{ independentes} \end{aligned} \quad (2.37)$$

com  $\hat{\boldsymbol{\beta}} = (X^t U X)^{-1} X^t U Y$  e  $U = \text{diag}(u_1, \dots, u_n)$ .

Através das condicionais completas, podemos utilizar o amostrador de Gibbs para obter amostras das distribuições *a posteriori* dos parâmetros e, daí, obter a inferência bayesiana para tais parâmetros. No apêndice apresentamos o programa que implementa este amostrador de Gibbs.

### 2.3.2 Derivação das Medidas de Diagnóstico

Considere  $(\boldsymbol{\beta}^{(1)}, \sigma^{2(1)}), \dots, (\boldsymbol{\beta}^{(L)}, \sigma^{2(L)})$  uma amostra de tamanho  $L$  da distribuição *a posteriori* de  $\boldsymbol{\beta}, \sigma^2|\mathbf{y}$ . Utilizaremos esta amostra em todos os cálculos das medidas de diagnóstico.

A estimativa para o  $CPO_i$  no caso da regressão linear *t*-Student é dada por

$$\widehat{CPO}_i = \left[ \frac{1}{L} \sum_{l=1}^L \frac{1}{f(y_i|\boldsymbol{\beta}^{(l)}, \sigma^{2(l)}, \nu)} \right]^{-1} \quad (2.38)$$

onde  $f(y_i|\boldsymbol{\beta}, \sigma^2, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu\sigma}} \left[ 1 + \frac{1}{\nu} \left( \frac{y_i - x_i^t \boldsymbol{\beta}}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}}$ , isto é, a função de distribuição *t*-Student de posição  $x_i^t \boldsymbol{\beta}$ , escala  $\sigma^2$  e grau de liberdade  $\nu$ , que estamos mantendo fixo.

A estimativa da divergência para a influência global é dada por

$$D_{\boldsymbol{\beta}, \sigma^2}(\widehat{g}, i) = \frac{1}{L} \sum_{l=1}^L g \left( \frac{\widehat{CPO}_i}{f(y_i|\boldsymbol{\beta}^{(l)}, \sigma^{2(l)}, \nu)} \right) \quad (2.39)$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (2.38). Para a obtenção da norma  $L_1$  e da divergência Kullback-Leibler basta substituir aqui também a função  $g$  adequada para cada caso.

**Proposição 2.1.** *Uma estimativa de Monte Carlo da divergência para a influência marginal em  $\beta$  no modelo t-Student é dada por*

$$\widehat{D_\beta(g, i)} = \frac{1}{L} \sum_{l=1}^n g \left( \frac{\widehat{CPO}_i}{\hat{f}(y_i | \boldsymbol{\beta}^{(l)}, \mathbf{y}_{(i)})} \right) \quad (2.40)$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (2.38) e

$$\hat{f}(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}) = \frac{\Gamma\left(\frac{n}{2}\right) \frac{1}{L} \sum_{l=1}^L \left[ \sum_{j=1}^n u_j^{(l)} (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}}}{\Gamma\left(\frac{n-1}{2}\right) \sqrt{\pi} \frac{1}{L} \sum_{l=1}^L \frac{1}{\sqrt{u_i^{(l)}}} \left[ \sum_{j \neq i} u_j^{(l)} (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n-1}{2}}} \quad (2.41)$$

onde, para cada  $j = 1, \dots, n$ ,  $u_j^{(1)}, \dots, u_j^{(L)}$  são valores simulados independentes da distribuição  $Gama(\frac{\nu+1}{2}, \frac{\nu}{2})$ .

*Demonstração.* Primeiro iremos obter uma expressão para  $f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)})$ .

Considere  $\mathbf{u} = (u_1, \dots, u_n)$ , onde, para cada  $j = 1, \dots, n$ ,  $u_j \sim Gama(\nu/2, \nu/2)$  independentes com  $f(u_j)$  sendo sua função densidade de probabilidade e assim  $f(\mathbf{u}) = \prod_{j=1}^n f(u_j)$ . Segue de (2.22) que,

$$\begin{aligned} f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}) &\propto \int f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2) f(\boldsymbol{\beta}, \sigma^2) d\sigma^2 \\ &= \int \int f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{u}) f(\mathbf{u} | \boldsymbol{\beta}, \sigma^2) d\mathbf{u} f(\boldsymbol{\beta}, \sigma^2) d\sigma^2 \\ &= \int \int f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{u}) f(\mathbf{u}) d\mathbf{u} f(\boldsymbol{\beta}, \sigma^2) d\sigma^2 \\ &= \int \int f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \mathbf{u}) f(\boldsymbol{\beta}, \sigma^2) d\sigma^2 f(\mathbf{u}) d\mathbf{u} \\ &\propto \int \int \prod_{j=1}^n f(y_j | \boldsymbol{\beta}, \sigma^2, \mathbf{u}) \frac{1}{\sigma^2} d\sigma^2 f(\mathbf{u}) d\mathbf{u} \\ &= \int \int \prod_{j=1}^n \left[ \frac{\sqrt{u_j}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u_j}{2\sigma^2} (y_j - x_j^t \boldsymbol{\beta})^2\right) \right] \frac{1}{\sigma^2} d\sigma^2 f(\mathbf{u}) d\mathbf{u} \\ &\propto \int \int \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta})^2\right) d\sigma^2 \prod_{j=1}^n \sqrt{u_j} f(\mathbf{u}) d\mathbf{u} \\ &= \int \Gamma\left(\frac{n}{2}\right) \left[ \frac{1}{2} \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u} \\ &\propto \int \left[ \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u} \end{aligned}$$

onde  $g(\mathbf{u}) = \prod_{j=1}^n g_j(u_j) = \prod_{j=1}^n \sqrt{u_j} f(u_j)$  e, assim,  $g_j$  é a função densidade de probabilidade de uma distribuição  $Gama(\frac{\nu+1}{2}, \frac{\nu}{2})$ . Vamos obter agora a constante de padronização.

Temos que

$$\begin{aligned}
& \int \left( \int \left[ \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u} \right) dy_i \\
&= \int \int \left[ \sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} \left[ 1 + \frac{u_i (y_i - x_i^t \boldsymbol{\beta})^2}{\sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2} \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u} dy_i \\
&= \int \left( \int \left[ 1 + \frac{u_i (y_i - x_i^t \boldsymbol{\beta})^2}{\sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2} \right]^{-\frac{n}{2}} dy_i \right) \left[ \sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u} \\
&= \int \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \sqrt{\frac{\pi}{u_i}} \sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2 \left[ \sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u} \\
&= \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \sqrt{\pi} \int \frac{1}{\sqrt{u_i}} \left[ \sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n-1}{2}} g(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

Podemos assim concluir que

$$f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}) = \frac{\Gamma(\frac{n}{2}) \int \left[ \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}} g(\mathbf{u}) d\mathbf{u}}{\Gamma(\frac{n-1}{2}) \sqrt{\pi} \int \frac{1}{\sqrt{u_i}} \left[ \sum_{j \neq i} u_j (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n-1}{2}} g(\mathbf{u}) d\mathbf{u}}. \quad (2.42)$$

Uma maneira de calcular (2.42) de forma aproximada é

$$\hat{f}(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}) = \frac{\Gamma(\frac{n}{2}) \frac{1}{L} \sum_{l=1}^L \left[ \sum_{j=1}^n u_j^{(l)} (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n}{2}}}{\Gamma(\frac{n-1}{2}) \sqrt{\pi} \frac{1}{L} \sum_{l=1}^L \frac{1}{\sqrt{u_i^{(l)}}} \left[ \sum_{j \neq i} u_j^{(l)} (y_j - x_j^t \boldsymbol{\beta})^2 \right]^{-\frac{n-1}{2}}} \quad (2.43)$$

onde, para cada  $j = 1, \dots, n$ ,  $u_j^{(1)}, \dots, u_j^{(L)}$  são valores simulados independentes da distribuição  $Gama(\frac{\nu+1}{2}, \frac{\nu}{2})$ .  $\square$

## 2.4 Aplicação

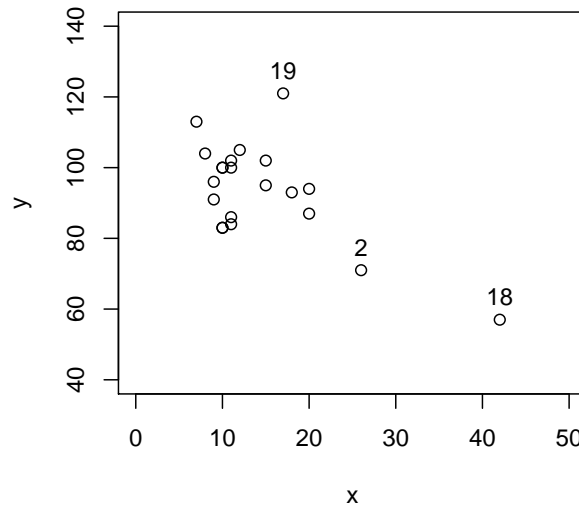
Nesta seção vamos considerar dados de um estudo sobre ocorrência de doença cardíaca cianótica em crianças, conduzido pela *University of California at Los Angeles* (UCLA) e utilizado em Weiss e Cho (1998). Na Tabela 2.1, apresentamos os dados das 21 crianças deste estudo. As seguintes variáveis são consideradas  $X$ : idade em meses em que a criança disse sua primeira palavra e  $Y$ : *Gesell adaptive score*. O *Gesell adaptive score* é uma medida de desenvolvimento cognitivo mensurada em uma idade mais avançada da criança e daí vem o interesse em tentar relacioná-lo com a idade da primeira palavra.

Podemos observar no diagrama de dispersão da Figura 2.1 que existem algumas observações visivelmente discrepantes: 2, 18 e 19. Porém, o diagnóstico visual somente não basta para concluir que estas observações são de fato discrepantes e, mais do que isso, se exercem algum tipo de influência nas estimativas dos parâmetros dos modelos.

**Tabela 2.1:** Idade em meses em que a criança disse sua primeira palavra e Gesell adaptive score

Obs	x	y	Obs	x	y
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	94	20	11	86
			21	10	100

Fonte: Weiss e Cho (1998)

**Figura 2.1:** Diagrama de dispersão com pontos discrepantes destacados

Fonte: Elaborada pelo autor.

Na Tabela 2.2, temos as estimativas dos parâmetros para o modelo normal e do modelo  $t$ -Student com os graus de liberdade fixados em  $\nu = 5$ . As estimativas dos modelos são parecidas, porém a medida  $LPML$  possui um valor maior para o modelo  $t$ -Student, indicando que este modelo ajusta melhor os dados em relação ao normal.

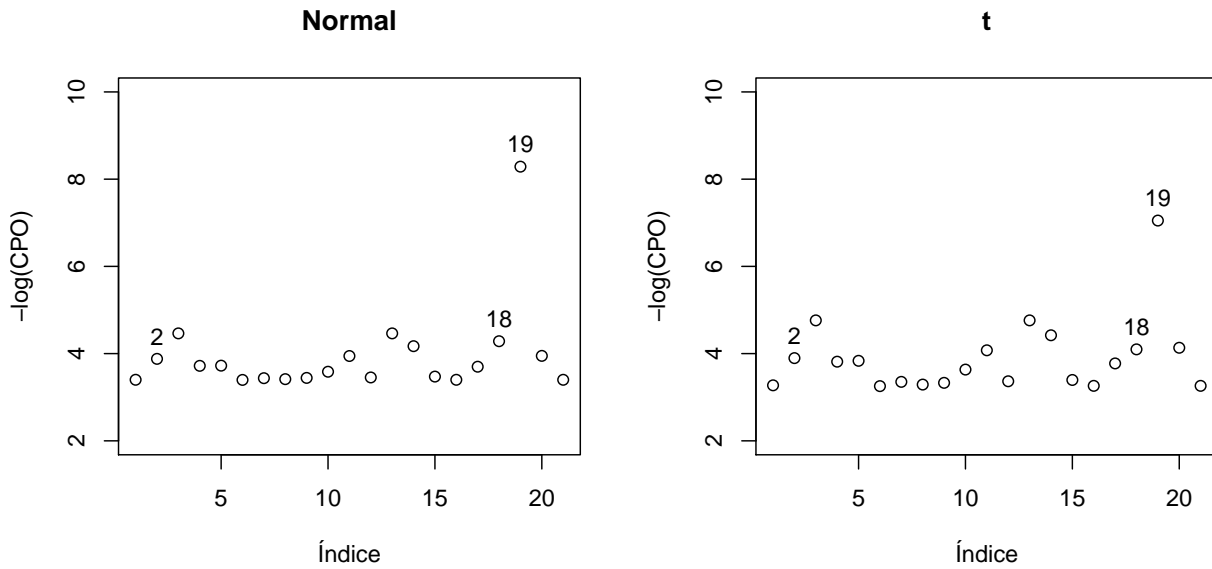
Um alto valor para  $CPO_i$  indica um melhor ajuste do modelo à observação  $y_i$ , enquanto que um baixo valor sugere que  $y_i$  é uma observação discrepante e, possivelmente, influente nas estimativas. Na Figura 2.2, reportamos  $-\log(CPO_i)$  que deixa mais visível as observações discrepantes por terem valores maiores nesta transformação. A observação 19 é a mais discrepante em termos do  $CPO$  para ambos modelos, porém o modelo  $t$ -Student acomoda melhor esta observação. Note também que a observação 18 possui um  $CPO$  não muito elevado não sendo apontada como discrepante segundo esse critério. Note que essa observação se encontra perto da reta de regressão estimada, como se vê na Figura 2.3.

Ao analisarmos a influência nas estimativas de forma global, isto é, no vetor  $(\beta, \sigma^2)$ , verificamos nas Figuras 2.4 e 2.5 que a observação 18 agora também recebe destaque, con-

**Tabela 2.2:** *Estimativas dos parâmetros dos modelos*

Parâmetros	Modelo	
	Normal	$t$ -Student
$\beta_0$	109.82	110.42
$\beta_1$	-1.12	-1.18
$Var(\varepsilon)$	131.79	143.04
$LPML$	-82.99	-82.05

Fonte: Elaborada pelo autor.

**Figura 2.2:**  $-\log(CPO)$  para os modelos normal e  $t$ -Student

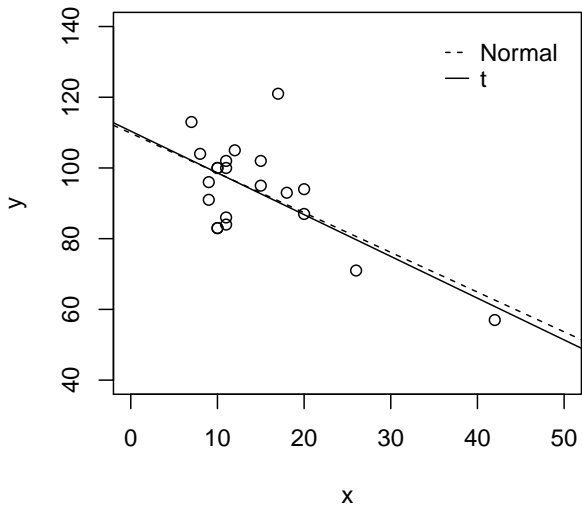
Fonte: Elaborada pelo autor.

siderando ambas medidas de influência global. A observação 18 possui uma forte influência nas estimativas dos modelos enquanto que a observação 19 possui uma influência menor no modelo  $t$ -Student se compararmos com o modelo normal. Dessa forma, pontos como o 19, de resíduo elevado e mais próximo dos demais, mostram como o modelo  $t$ -Student é mais robusto que o normal, isto é, menos afetado por observações discrepantes.

Já na influência nas estimativas de forma marginal (em  $\beta$ ), podemos ver nas figuras 2.6 e 2.7 que a observação 19 sob o modelo normal possui uma menor influência, indicando que esta observação influencia mais as estimativas de  $\sigma^2$ . Sob o modelo  $t$ -Student, as observações possuem influências mais parecidas com o caso global. Entre os modelos, o modelo  $t$ -Student é aquele com as menores medidas de influência para as observações 18 e 19 como podemos verificar na Tabelas 2.3.



Figura 2.3: Diagrama de dispersão com as retas de regressão estimadas



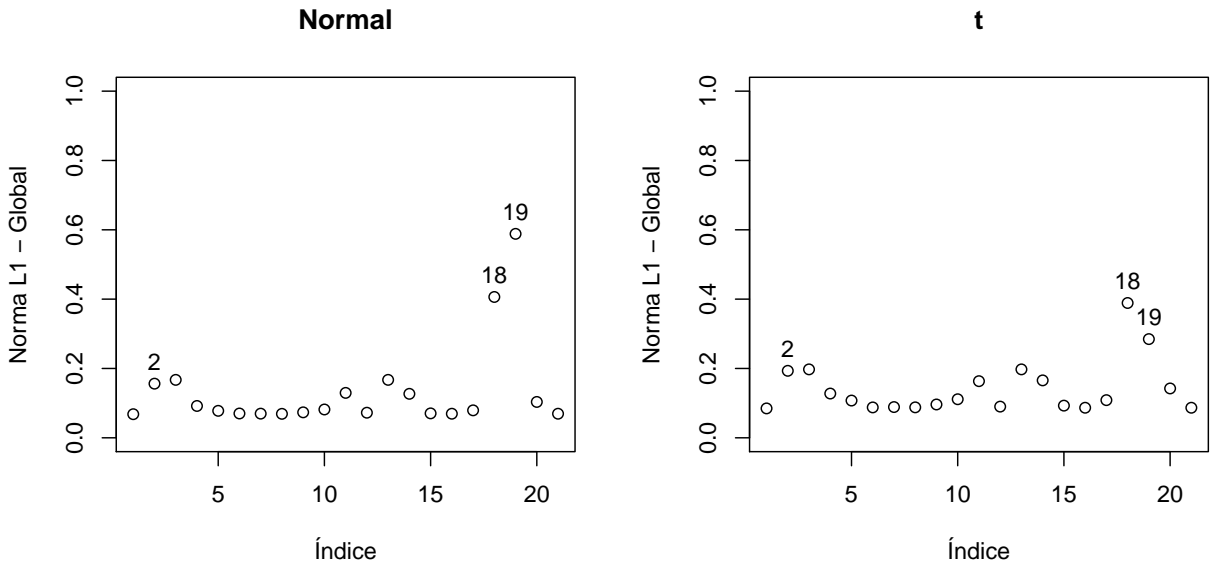
Fonte: Elaborada pelo autor.

Tabela 2.3: Medidas de diagnóstico para as observações 2, 18 e 19

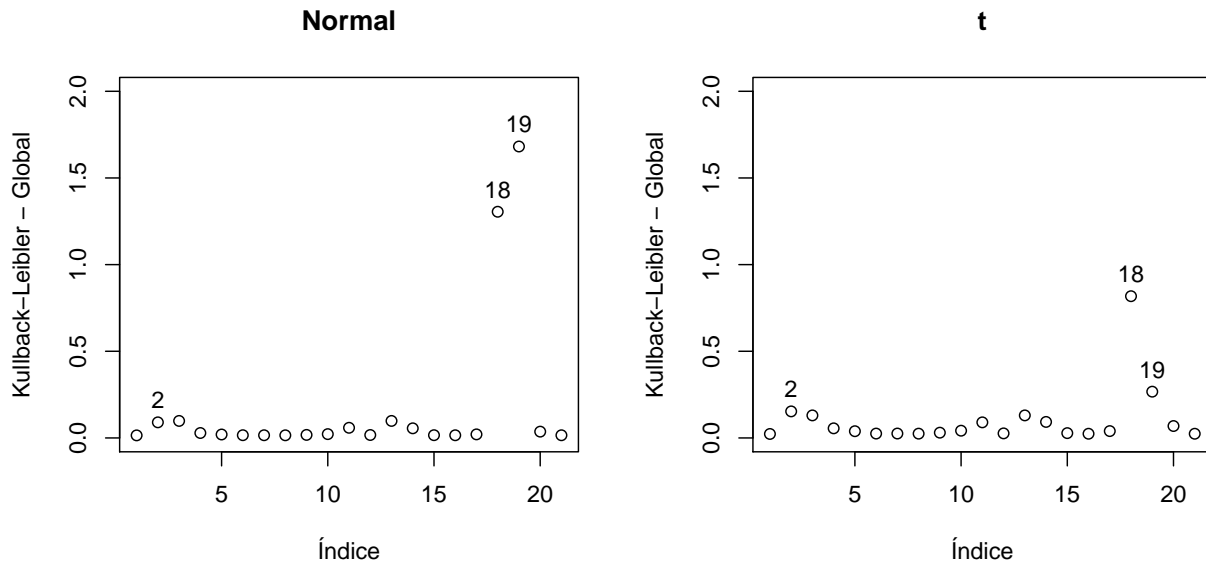
Observação	Modelo	2		18		19	
		Normal	t	Normal	t	Normal	t
Medida	$CPO_i$	0.0207	0.0203	0.0138	0.0167	0.00025	0.00087
	$L_{1i,(\beta,\sigma^2)}$	0.156	0.193	0.406	0.389	0.588	0.285
	$L_{1i,\beta}$	0.155	0.194	0.383	0.384	0.318	0.187
	$K_{1i,(\beta,\sigma^2)}$	0.09	0.153	1.305	0.818	1.682	0.267
	$K_{1i,\beta}$	0.081	0.149	0.724	0.706	0.283	0.189

Fonte: Elaborada pelo autor.

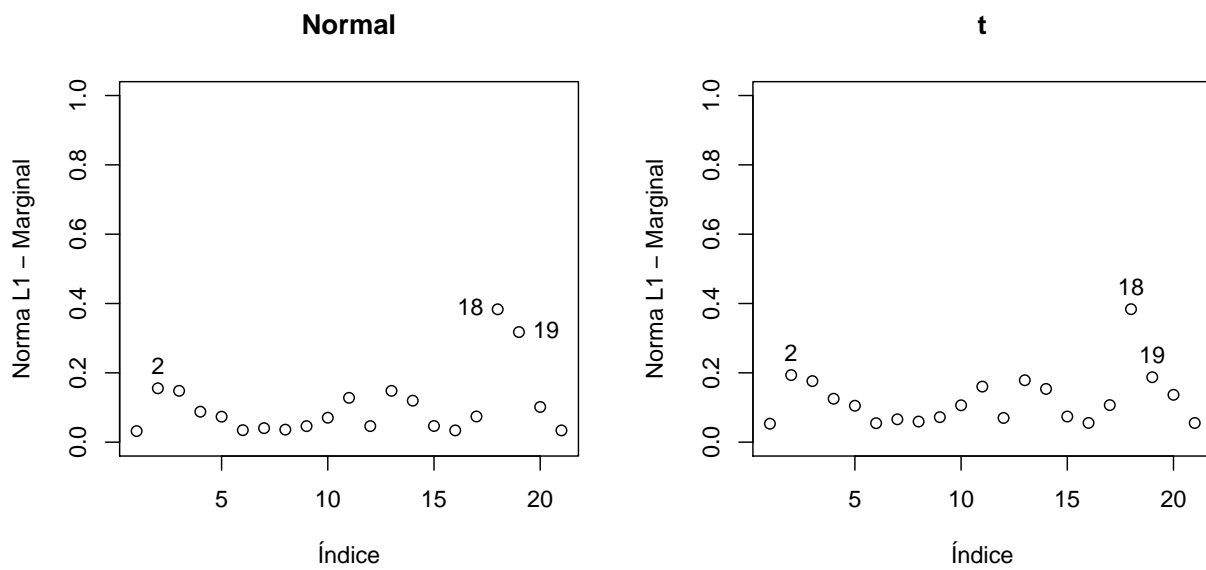
Figura 2.4: Influência global via norma  $L_1$



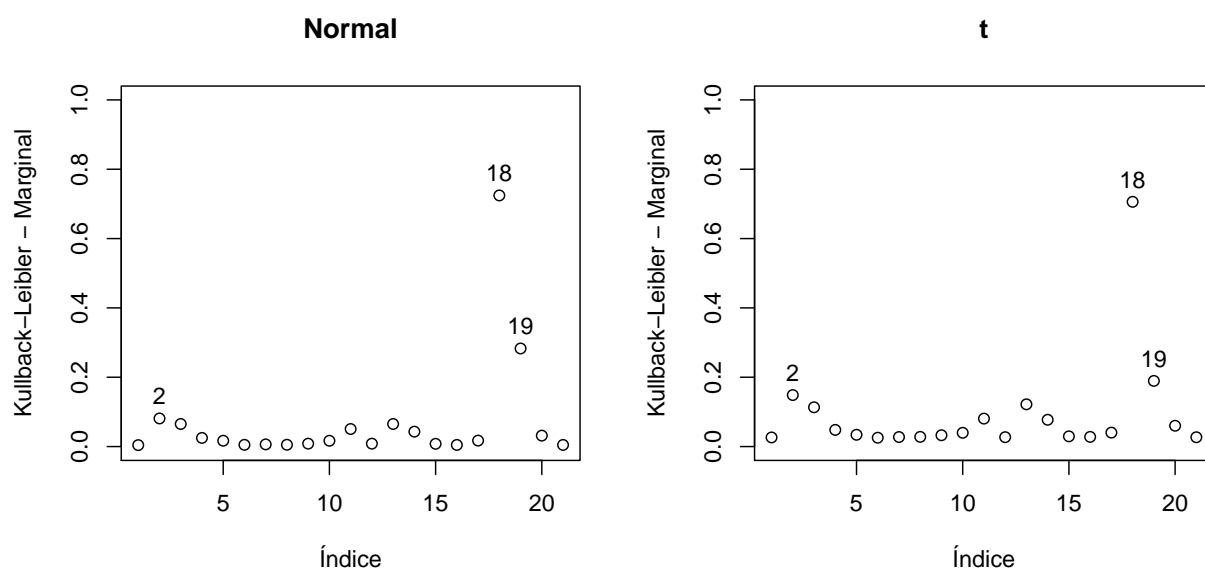
Fonte: Elaborada pelo autor.

**Figura 2.5:** *Influência global via Kullback-Leibler*

Fonte: Elaborada pelo autor.

**Figura 2.6:** *Influência marginal via norma  $L_1$* 

Fonte: Elaborada pelo autor.

**Figura 2.7:** *Influência marginal via Kullback-Leibler*

Fonte: Elaborada pelo autor.



## Capítulo 3

# Diagnóstico Bayesiano em Modelos de Regressão Assimétricos

A hipótese de simetria dos erros presente nos modelos apresentados no capítulo anterior pode não ser sempre adequada. Para analisar dados que não exibam simetria, nos últimos anos novos modelos têm sido propostos como alternativas aos modelos simétricos. O modelo de regressão  $t$ -assimétrico tem sido considerado uma alternativa robusta ao modelo normal (Azzalini e Genton (2008)).

Neste capítulo estudamos dois modelos que possuem erros com distribuição assimétrica: normal e  $t$ -Student. Os modelos normal assimétrico e  $t$ -assimétrico já foram estudados em Bayes (2005) e Godoi (2007), respectivamente, e nestes trabalhos podemos encontrar formas de se obter estimativas dos parâmetros por meio da inferência bayesiana. Aqui complementamos o problema da estimação encontrando as condicionais completas do modelo  $t$ -assimétrico para uso do amostrador de Gibbs.

A identificação das observações discrepantes também será feita por meio do cálculo do CPO para ambos modelos. Neste capítulo mostramos como obter estimativas para tal estatística. A medida de influência global e marginal também será por meio do cálculo das divergências norma  $L_1$  e Kullback-Leibler e mostramos como obter estimativas destas medidas.

### 3.1 Modelo de Regressão Normal Assimétrico

#### 3.1.1 Estimação dos Parâmetros

O modelo de regressão linear com erros de distribuição normal assimétrica é dado por

$$y_i = x_i^t \beta + \epsilon_i, \quad (3.1)$$

onde, para cada  $i = 1, \dots, n$ ,  $y_i$  é a variável resposta da  $i$ -ésima observação,  $x_i^t = (1, x_{i1}, \dots, x_{ip})$  é o vetor de variáveis explicativas,  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^t$  é vetor de parâmetros desconhecidos a serem estimados e  $\epsilon_i \sim SN(0, \sigma^2, \lambda)$  independentes, ou seja, os erros possuem distribuição normal assimétrica com posição 0, escala  $\sigma^2$  e parâmetro de assimetria  $\lambda$ .

Neste caso, das propriedades da distribuição normal assimétrica, segue que  $y_i \sim$

$SN(x_i^t \boldsymbol{\beta}, \sigma^2, \lambda)$  e sua função densidade de probabilidade é

$$f(y_i | \boldsymbol{\beta}, \sigma^2, \lambda) = \frac{2}{\sigma} \phi \left( \frac{y_i - x_i^t \boldsymbol{\beta}}{\sigma} \right) \Phi \left( \lambda \frac{y_i - x_i^t \boldsymbol{\beta}}{\sigma} \right), \quad (3.2)$$

onde  $\phi$  e  $\Phi$  são as funções densidade de probabilidade e distribuição acumulada de uma normal padrão, respectivamente. A esperança desta distribuição é dada por

$$E[y_i] = x_i^t \boldsymbol{\beta} + \sigma \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}}, \quad (3.3)$$

ou seja, a esperança da variável resposta não fica escrita originalmente como uma reta de regressão. Porém, ao considerarmos como intercepto

$$\beta_0^* = \beta_0 + \sigma \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} \quad (3.4)$$

teremos  $E[y_i] = x_i^t \boldsymbol{\beta}^*$ , com  $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1, \dots, \beta_p)$ .

Considere a seguinte representação hierárquica para cada  $y_i$  apresentada em [Bayes \(2005\)](#)

$$\begin{aligned} y_i | v_i &\sim N(x_i^t \boldsymbol{\beta} + \sigma \delta v_i, \sigma^2(1 - \delta^2)) \\ v_i &\sim HN(0, 1) \end{aligned} \quad (3.5)$$

onde  $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$ .

Sob a reparametrização  $\eta = \sigma \delta$  e  $\tau = \sigma \sqrt{1 - \delta^2}$ , reescrevemos o modelo como

$$\begin{aligned} y_i | v_i &\sim N(x_i^t \boldsymbol{\beta} + \eta v_i, \tau^2) \\ v_i &\sim HN(0, 1) \end{aligned} \quad (3.6)$$

com  $i = 1, \dots, n$ .

Sob essa última parametrização a função de verossimilhança aumentada é dada por

$$L_A(\boldsymbol{\beta}, \tau, \eta) = \frac{1}{\tau^n} \exp \left( -\frac{1}{2\tau^2} \sum_{i=1}^n (y_i - x_i^t \boldsymbol{\beta} - \eta v_i)^2 \right) \exp \left( -\frac{1}{2} \sum_{i=1}^n v_i^2 \right) \prod_{i=1}^n \mathbb{I}_{[0, \infty[}(v_i) \quad (3.7)$$

Neste estudo, considerando desconhecimento prévio a respeito dos parâmetros a serem estimados, usamos distribuições *a priori* não informativas. Dessa forma podemos considerar a seguinte distribuição *a priori* de referência apresentada em [Bayes \(2005\)](#) para  $(\boldsymbol{\beta}, \sigma, \lambda)$

$$f(\boldsymbol{\beta}, \sigma, \lambda) \propto \frac{1}{\sigma} f(\lambda) \quad (3.8)$$

com  $\lambda \sim t(0, \sigma_t^2, k)$ . Quando  $k = \frac{1}{2}$  e  $\sigma_t^2 = \frac{\pi^2}{4}$  obtemos a aproximação da distribuição *a priori* de Jeffreys e quando  $k = 2$  e  $\sigma_t^2 = \frac{1}{2}$  obtemos a distribuição *a priori* induzida ao se considerar  $\delta \sim U([-1, 1])$ . Para mais detalhes ver [Bayes \(2005\)](#).

Note que a distribuição  $t$ -Student pode ser representada de forma hierárquica como

$$\begin{aligned}\lambda|\omega &\sim N\left(0, \frac{\sigma_t^2}{\omega}\right) \\ \omega &\sim Gama\left(\frac{k}{2}, \frac{k}{2}\right).\end{aligned}\tag{3.9}$$

Consequentemente,

$$f(\boldsymbol{\beta}, \sigma, \lambda, \omega) \propto \frac{1}{\sigma} \exp\left(-\frac{\lambda^2 \omega}{2\sigma_t^2}\right) \omega^{\frac{k+1}{2}-1} \exp\left(-\frac{k}{2}\omega\right).\tag{3.10}$$

Como estamos trabalhando com a parametrização dada por  $\eta = \sigma\delta$ ,  $\tau = \sigma\sqrt{1-\delta^2}$  e  $\delta = \lambda/\sqrt{1+\lambda^2}$ , precisamos encontrar a distribuição *a priori*  $f(\boldsymbol{\beta}, \tau, \eta, \omega)$ . Primeiramente, notemos que

$$f(\boldsymbol{\beta}, \tau, \eta, \omega) = f(\boldsymbol{\beta})f(\tau, \eta|\omega)f(\omega)\tag{3.11}$$

Logo, podemos considerar somente a seguinte transformação de variáveis aleatórias

$$(\sigma, \eta)|\omega \mapsto (\tau, \eta)|\omega = \left(\sigma \frac{1}{\sqrt{1+\lambda^2}}, \sigma \frac{\lambda}{\sqrt{1+\lambda^2}}\right)|\omega\tag{3.12}$$

e assim teremos

$$f(\tau, \eta|\omega) = f_{(\sigma, \lambda)|\omega}(\sqrt{\eta^2 + \tau^2}, \eta/\tau) \left| \frac{\partial(\sigma, \lambda)}{\partial(\tau, \eta)} \right|\tag{3.13}$$

onde  $f_{(\sigma, \lambda)|\omega}$  é a função densidade de probabilidade conjunta de  $(\sigma, \lambda)|\omega$  e

$$\frac{\partial(\sigma, \lambda)}{\partial(\tau, \eta)} = \begin{vmatrix} \frac{\partial\sigma}{\partial\tau} & \frac{\partial\sigma}{\partial\eta} \\ \frac{\partial\lambda}{\partial\tau} & \frac{\partial\lambda}{\partial\eta} \end{vmatrix}\tag{3.14}$$

Note que

$$\frac{\partial(\sigma, \lambda)}{\partial(\tau, \eta)} = \begin{vmatrix} \frac{\partial\sigma}{\partial\tau} & \frac{\partial\sigma}{\partial\eta} \\ \frac{\partial\lambda}{\partial\tau} & \frac{\partial\lambda}{\partial\eta} \end{vmatrix} = \begin{vmatrix} \frac{\tau}{\sqrt{\eta^2 + \tau^2}} & \frac{\eta}{\sqrt{\eta^2 + \tau^2}} \\ -\frac{\eta}{\tau^2} & \frac{1}{\tau} \end{vmatrix} = \frac{\sqrt{\eta^2 + \tau^2}}{\tau^2}\tag{3.15}$$

com

$$\begin{aligned}f(\sigma) &\propto \frac{1}{\sigma} \\ \lambda|\omega &\sim N\left(0, \frac{\sigma_t^2}{\omega}\right).\end{aligned}\tag{3.16}$$

Logo,

$$\begin{aligned}f(\tau, \eta|\omega) &= f_{(\sigma, \lambda)|\omega}(\sqrt{\eta^2 + \tau^2}, \eta/\tau) \left| \frac{\partial(\sigma, \lambda)}{\partial(\tau, \eta)} \right| = f_{\sigma}(\sqrt{\eta^2 + \tau^2}) f_{\lambda|\omega}(\eta/\tau|\omega) \frac{\sqrt{\eta^2 + \tau^2}}{\tau^2} \\ &\propto \frac{1}{\sqrt{\eta^2 + \tau^2}} \sqrt{\omega} \exp\left(-\frac{\eta^2 \omega}{2\tau^2 \sigma_t^2}\right) \frac{\sqrt{\eta^2 + \tau^2}}{\tau^2} = \frac{\sqrt{\omega}}{\tau^2} \exp\left(-\frac{\eta^2 \omega}{2\tau^2 \sigma_t^2}\right)\end{aligned}\tag{3.17}$$

e, portanto,

$$\begin{aligned} f(\boldsymbol{\beta}, \tau, \eta, \omega) &= f(\boldsymbol{\beta})f(\omega)f(\tau, \eta|\omega) \propto \omega^{\frac{k}{2}-1} \exp\left(-\frac{k}{2}\omega\right) \frac{\omega^{\frac{1}{2}}}{\tau^2} \exp\left(-\frac{\eta^2\omega}{2\tau^2\sigma_t^2}\right) \\ &\propto \frac{1}{\tau^2} \exp\left(-\frac{\eta^2\omega}{2\tau^2\sigma_t^2}\right) \omega^{\frac{k+1}{2}-1} \exp\left(-\frac{k}{2}\omega\right). \end{aligned} \quad (3.18)$$

A distribuição *a posteriori* aumentada é dada por

$$\begin{aligned} f(\boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{v}|\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{v})f(\mathbf{v})f(\boldsymbol{\beta}, \tau, \eta, \omega) = \\ &= L_A(\boldsymbol{\beta}, \tau, \eta)f(\boldsymbol{\beta}, \tau, \eta, \omega). \end{aligned} \quad (3.19)$$

A partir de (3.7), (3.18) e (3.19) conseguimos obtemos as distribuições condicionais completas apresentadas a seguir:

$$\begin{aligned} \boldsymbol{\beta}|\tau, \eta, \omega, \mathbf{v}, \mathbf{y} &\sim N_{p+1}\left((X^t X)^{-1}X^t(Y - \eta V), \tau^2(X^t X)^{-1}\right) \\ \frac{1}{\tau^2}|\boldsymbol{\beta}, \eta, \omega, \mathbf{v}, \mathbf{y} &\sim \text{Gama}\left(\frac{n+1}{2}, \frac{1}{2}\left((Y - \eta V - X\boldsymbol{\beta})^t(Y - \eta V - X\boldsymbol{\beta}) + \frac{\eta^2\omega}{\sigma_t^2}\right)\right) \\ \eta|\boldsymbol{\beta}, \tau, \omega, \mathbf{v}, \mathbf{y} &\sim N\left(\frac{(Y - X\boldsymbol{\beta})^t V}{V^t V + \frac{\omega}{\sigma_t^2}}, \frac{\tau^2}{V^t V + \frac{\omega}{\sigma_t^2}}\right) \\ \omega|\boldsymbol{\beta}, \tau, \eta, \mathbf{v}, \mathbf{y} &\sim \text{Gama}\left(\frac{k+1}{2}, \frac{1}{2}\left(k + \frac{\eta^2}{\tau^2\sigma_t^2}\right)\right) \\ v_i|\boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{v}, \mathbf{y} &\sim N\left(\frac{\eta(y_i - x_i^t \boldsymbol{\beta})}{\eta^2 + \tau^2}, \frac{\tau^2}{\eta^2 + \tau^2}\right) \mathbb{I}_{[0, \infty[}(v_i), \quad i = 1, \dots, n \end{aligned} \quad (3.20)$$

onde

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad (3.21)$$

Por meio das condicionais completas, podemos utilizar o amostrador de Gibbs para obter amostras das distribuições *a posteriori* dos parâmetros e, daí, obter a inferência bayesiana para tais parâmetros. No apêndice apresentamos o programa que implementa este amostrador de Gibbs.

### 3.1.2 Derivação das Medidas de Diagnóstico

Considere  $(\boldsymbol{\beta}^{(1)}, \sigma^{2(1)}, \lambda^{(1)}), \dots, (\boldsymbol{\beta}^{(L)}, \sigma^{2(L)}, \lambda^{(L)})$  uma amostra de tamanho  $L$  simulada da distribuição *a posteriori* conjunta. Esta amostra será utilizada em todos os cálculos das medidas de diagnóstico.

A estimativa para o  $CPO_i$  no caso da regressão linear normal assimétrica será dada por

$$\widehat{CPO}_i = \left[ \frac{1}{L} \sum_{l=1}^n \frac{1}{f(y_i|\boldsymbol{\beta}^{(l)}, \sigma^{2(l)}, \lambda^{(l)})} \right]^{-1} \quad (3.22)$$

onde  $f(y_i|\boldsymbol{\beta}, \sigma^2, \lambda)$  é a função de distribuição normal assimétrica de posição  $x_i^t \boldsymbol{\beta}$ , escala  $\sigma^2$  e assimetria  $\lambda$ .



A estimativa da divergência para a influência global é dada por

$$D_{\widehat{\beta}, \widehat{\sigma^2}, \widehat{\lambda}}(g, i) = \frac{1}{L} \sum_{l=1}^n g \left( \frac{\widehat{CPO}_i}{f(y_i | \beta^{(l)}, \sigma^{2(l)}, \lambda^{(l)})} \right) \quad (3.23)$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (3.22). Para a obtenção da norma  $L_1$  e da divergência Kullback-Leibler basta substituir a função  $g$  adequada para cada caso.

**Proposição 3.1.** *Uma estimativa de Monte Carlo da divergência para a influência marginal no modelo normal assimétrico é dada por*

$$\widehat{D}_{\beta}(g, i) = \frac{1}{L} \sum_{l=1}^n g \left( \frac{\widehat{CPO}_i}{\widehat{f}(y_i | \beta, \mathbf{y}_{(i)})} \right) \quad (3.24)$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (2.38) e

$$\widehat{f}(y_i | \beta, \mathbf{y}_{(i)}) = \frac{\Gamma(\frac{n-1}{2})^{\frac{1}{L}} \sum_{l=1}^L \left[ (\Omega_{(l)} + V_{(l)}) A_{(l)}^{n-1} \right]^{-\frac{1}{2}}}{\Gamma(\frac{n-2}{2}) \sqrt{\pi}^{\frac{1}{L}} \sum_{l=1}^L \left[ \left( \Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2} \right) K_{(l)}^{n-2} \right]^{-\frac{1}{2}}} \quad (3.25)$$

com

$$\begin{aligned} z_i &= y_i - x_i^t \beta \\ A_{(l)} &= \sum_{j=1}^n z_j^2 - \frac{1}{\Omega_{(l)} + V_{(l)}} \left( \sum_{k=1}^n v_k^{(l)} z_k \right)^2 \\ K_{(l)} &= \sum_{j \neq i} z_j^2 - \frac{1}{\Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2}} \left( \sum_{j \neq i} v_j^{(l)} z_j \right)^2 \\ V_{(l)} &= \sum_{j=1}^n v_j^{(l)2} \\ \Omega_{(l)} &= \frac{\omega_{(l)}}{\sigma_t^2} \end{aligned} \quad (3.26)$$

onde, para cada  $j = 1, \dots, n$ ,  $v_j^{(1)}, \dots, v_j^{(L)}$  são valores simulados independentes da distribuição  $HN(0, 1)$  e  $\omega^{(1)}, \dots, \omega^{(L)}$  são valores simulados da distribuição  $\text{Gama}((k+1)/2, k/2)$ .

*Demonstração.* Primeiro, vamos obter uma expressão para  $f(y_i | \beta, \mathbf{y}_{(i)})$ . Segue de (2.22) que:

$$\begin{aligned} f(y_i | \beta, \mathbf{y}_{(i)}) &\propto \int \int \int f(\mathbf{y} | \beta, \tau^2, \eta, \omega) f(\beta, \tau^2, \eta, \omega) d\tau^2 d\eta d\omega \\ &= \int \int \int \int f(\mathbf{y} | \beta, \tau^2, \eta, \omega, \mathbf{v}) f(\mathbf{v} | \beta, \tau^2, \eta, \omega) d\mathbf{v} f(\beta, \tau^2, \eta, \omega) d\tau^2 d\eta d\omega \\ &= \int \int \int \int f(\mathbf{y} | \beta, \tau^2, \eta, \omega, \mathbf{v}) f(\mathbf{v}) f(\beta, \tau^2, \eta, \omega) d\mathbf{v} d\tau^2 d\eta d\omega \end{aligned}$$

$$\begin{aligned}
& \propto \int \int \int \int f(\mathbf{y}|\boldsymbol{\beta}, \tau^2, \eta, \omega, \mathbf{v}) \frac{1}{\tau^2} f(\mathbf{v}) f(\eta|\omega) g(\omega) d\mathbf{v} d\tau^2 d\eta d\omega \\
& \propto \int \int \int \int \frac{1}{\tau^n} \exp \left( \frac{-1}{2\tau^2} \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta} - \eta v_j)^2 \right) \frac{1}{\tau^2} f(\mathbf{v}) \\
& \quad \exp \left( \frac{-\eta^2 \omega}{2\tau^2 \sigma_t^2} \right) g(\omega) d\mathbf{v} d\tau^2 d\eta d\omega \\
& \propto \int \int \int \int \frac{1}{\tau^{n+2}} \exp \left( \frac{-1}{2\tau^2} \left( \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta} - \eta v_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \right) \right) d\tau^2 \\
& \quad f(\mathbf{v}) g(\omega) d\mathbf{v} d\eta d\omega \\
& \propto \int \int \int \left[ \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta} - \eta v_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \right]^{-\frac{n}{2}} d\eta f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega \quad (3.27)
\end{aligned}$$

onde  $f(\mathbf{v})$  é a função densidade de probabilidade de  $\mathbf{v} = (v_1, \dots, v_n)$ , com  $v_i \sim HN(0, 1)$ ,  $i = 1, \dots, n$ , e  $g(\omega)$  é a função densidade de probabilidade de  $\omega \sim Gama(\frac{k+1}{2}, \frac{k}{2})$ .

Fazendo  $\xi = \frac{\sum_{j=1}^n v_j (y_j - x_j^t \boldsymbol{\beta})}{\sum_{j=1}^n v_j^2}$  para isolar  $\eta$  em (3.27) teremos

$$\begin{aligned}
& \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta} - \eta v_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \\
& = \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \sum_{j=1}^n v_j^2 (\xi - \eta)^2 + \frac{\omega \eta^2}{\sigma_t^2} \\
& = \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \\
& \quad + \left( \sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2} \right) \left( \eta - \frac{\sum_{j=1}^n v_j^2 \xi}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} \\
& = \underbrace{\sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}}}_A \\
& \quad + \underbrace{\left( \sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2} \right)}_B \underbrace{\left( \eta - \frac{\sum_{j=1}^n v_j^2 \xi}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} \right)^2}_C \\
& = A + B(\eta - C)^2 \quad (3.28)
\end{aligned}$$

Veja que podemos buscar agora uma forma para o resultado de (3.28) que envolva

uma distribuição de probabilidade conhecida em  $\eta$

$$\begin{aligned}
 [A + B(\eta - C)^2]^{-\frac{n}{2}} &= A^{-\frac{n}{2}} \left[ 1 + \frac{B}{A}(\eta - C)^2 \right]^{-\frac{n}{2}} \\
 &= A^{-\frac{n}{2}} \left[ 1 + \left( \frac{\eta - C}{\sqrt{A}/\sqrt{B}} \right)^2 \right]^{-\frac{n}{2}} \\
 &= A^{-\frac{n}{2}} \left[ 1 + \frac{1}{n-1} \left( \frac{\eta - C}{\sqrt{A}/\sqrt{(n-1)B}} \right)^2 \right]^{-\frac{n}{2}} \\
 &= A^{-\frac{n}{2}} h^*(\eta),
 \end{aligned} \tag{3.29}$$

onde  $h^*$  é o núcleo de uma distribuição  $t(C, \sqrt{A}/\sqrt{(n-1)B}, n-1)$ . Logo, podemos continuar o cálculo de (3.27) com o resultado encontrado em (3.29). Assim

$$\begin{aligned}
 f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}) &\propto \int \int \int \left[ \sum_{j=1}^n (y_j - x_j^t \boldsymbol{\beta} - \eta v_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \right]^{-\frac{n}{2}} d\eta f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega \\
 &= \int \int \int A^{-\frac{n}{2}} h^*(\eta) d\eta f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega \\
 &= \int \int A^{-\frac{n}{2}} \left[ \int h^*(\eta) d\eta \right] f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega \\
 &= \int \int A^{-\frac{n}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right) \sqrt{\pi A}}{\Gamma\left(\frac{n}{2}\right) \sqrt{B}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega \\
 &\propto \int \int A^{-\frac{n-1}{2}} B^{-\frac{1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega.
 \end{aligned} \tag{3.30}$$

Note que na equação (3.30) obtivemos um resultado proporcional para  $f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)})$ , iremos agora calcular a constante de normalização. Dessa forma,

$$\int \int \int A^{-\frac{n-1}{2}} B^{-\frac{1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega dy_i = \int \int B^{-\frac{1}{2}} \left[ \int A^{-\frac{n-1}{2}} dy_i \right] f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega \tag{3.31}$$

$$\text{com } A = \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}}, \quad \xi = \frac{\sum_{j=1}^n v_j (y_j - x_j^t \boldsymbol{\beta})}{\sum_{j=1}^n v_j^2} \text{ e } B = \sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}.$$

Integrando em  $y_i$ , temos

$$\int A^{-\frac{n-1}{2}} dy_i = \int \left[ \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} \right]^{-\frac{n-1}{2}} dy_i \tag{3.32}$$

Podemos trabalhar melhor a fórmula que está sendo integrada deixando o termo  $y_i$

em maior evidência. Assim teremos

$$\begin{aligned} & \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} \\ &= \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \frac{\sum_{k=1}^n v_k (y_k - x_k^t \boldsymbol{\beta})}{\sum_{k=1}^n v_k^2} \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \left( \frac{\sum_{k=1}^n v_k (y_k - x_k^t \boldsymbol{\beta})}{\sum_{k=1}^n v_k^2} \right)^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}}. \end{aligned} \quad (3.33)$$

Para simplificar a notação, considere  $\Omega = \frac{\omega}{\sigma_t^2}$ ,  $V = \sum_{k=1}^n v_k^2$  e  $z_j = y_j - x_j^t \boldsymbol{\beta}$ . Assim

$$\begin{aligned} \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \boldsymbol{\beta}}{v_j} - \xi \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} &= \sum_{j=1}^n \left( z_j - v_j \frac{\sum_{k=1}^n v_k z_k}{V} \right)^2 + \frac{V \Omega \left( \frac{\sum_{k=1}^n v_k z_k}{V} \right)^2}{\Omega + V} \\ &= \underbrace{\sum_{j=1}^n \left( z_j - \frac{v_j}{V} \sum_{k=1}^n v_k z_k \right)^2}_{(*)} + \underbrace{\frac{\Omega/V}{\Omega + V} \left( \sum_{k=1}^n v_k z_k \right)^2}_{(**)} \end{aligned} \quad (3.34)$$

Vamos trabalhar agora com uma das parcelas da soma resultante na equação (3.34), denotadas por (\*) e (\*\*), de forma separada. Para a primeira parcela da soma temos

$$\begin{aligned} (*) &= \sum_{j=1}^n \left( z_j - \frac{v_j}{V} \sum_{k=1}^n v_k z_k \right)^2 = \sum_{j=1}^n \left( z_j^2 + \frac{v_j^2}{V^2} \left( \sum_{k=1}^n v_k z_k \right)^2 - \frac{2v_j z_j}{V} \sum_{k=1}^n v_k z_k \right) \\ &= \sum_{j=1}^n z_j^2 + \frac{(\sum_{k=1}^n v_k z_k)^2}{V^2} \sum_{j=1}^n v_j^2 - \frac{2 \sum_{k=1}^n v_k z_k}{V} \sum_{j=1}^n v_j z_j \\ &= \sum_{j=1}^n z_j^2 + \frac{(\sum_{k=1}^n v_k z_k)^2}{V} - \frac{2 (\sum_{k=1}^n v_k z_k)^2}{V} \\ &= \sum_{j=1}^n z_j^2 - \frac{1}{V} \left( \sum_{k=1}^n v_k z_k \right)^2 \end{aligned} \quad (3.35)$$

Dessa forma, a soma completa será dada por

$$\begin{aligned} (*) + (**) &= \sum_{j=1}^n z_j^2 - \frac{1}{V} \left( \sum_{k=1}^n v_k z_k \right)^2 + \frac{\Omega}{\Omega + V} \frac{1}{V} \left( \sum_{k=1}^n v_k z_k \right)^2 \\ &= \sum_{j=1}^n z_j^2 + \frac{1}{V} \left( \sum_{k=1}^n v_k z_k \right)^2 \left( \frac{\Omega}{\Omega + V} - 1 \right) \\ &= \sum_{j=1}^n z_j^2 - \frac{1}{\Omega + V} \left( \sum_{k=1}^n v_k z_k \right)^2 \\ &= z_i^2 + \sum_{j \neq i} z_j^2 - \frac{1}{\Omega + V} \left( v_i z_i + \sum_{k \neq i} v_k z_k \right)^2 \end{aligned}$$

$$\begin{aligned}
&= z_i^2 + \sum_{j \neq i} z_j^2 - \frac{1}{\Omega + V} \left( v_i^2 z_i^2 + \left( \sum_{k \neq i} v_k z_k \right)^2 + 2v_i z_i \sum_{k \neq i} v_k z_k \right) \\
&= z_i^2 \left[ 1 - \frac{v_i^2}{\Omega + V} \right] + z_i \left[ -\frac{2v_i}{\Omega + V} \sum_{k \neq i} v_k z_k \right] + \sum_{j \neq i} z_j^2 - \frac{1}{\Omega + V} \left( \sum_{k \neq i} v_k z_k \right)^2 \\
&= D(z_i - H)^2 + K
\end{aligned} \tag{3.36}$$

para  $D$ ,  $H$  e  $K$  convenientes, pois no penúltimo passo da equação (3.36) temos um polinômio de segundo grau em  $z_i$ . Logo, retomando o cálculo da equação (3.32),

$$\begin{aligned}
&\int \left[ \sum_{j=1}^n v_j^2 \left( \frac{y_j - x_j^t \beta}{v_j} - \xi \right)^2 + \frac{\sum_{j=1}^n v_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}} \right]^{-\frac{n-1}{2}} dy_i \\
&= \int [D(z_i - H)^2 + K]^{-\frac{n-1}{2}} dy_i \\
&= \int [D(y_i - x_i^t \beta - H)^2 + K]^{-\frac{n-1}{2}} dy_i \\
&= K^{-\frac{n-1}{2}} \int \left[ 1 + \frac{1}{n} \left( \frac{y_i - x_i^t \beta - H}{\sqrt{K}/\sqrt{nD}} \right)^2 \right]^{-\frac{n-1}{2}} dy_i \\
&= \frac{\Gamma\left(\frac{n-2}{2}\right) \sqrt{\pi} K^{-\frac{n-2}{2}}}{\Gamma\left(\frac{n-1}{2}\right) \sqrt{D}}
\end{aligned} \tag{3.37}$$

Com os resultados obtidos nas equações (3.30) e (3.37) conseguimos calcular  $f(y_i | \beta, \mathbf{y}_{(i)})$  por meio da seguinte equação

$$\begin{aligned}
f(y_i | \beta, \mathbf{y}_{(i)}) &= \frac{\Gamma\left(\frac{n-1}{2}\right) \int B^{-\frac{1}{2}} A^{-\frac{n-1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega}{\Gamma\left(\frac{n-2}{2}\right) \sqrt{\pi} \int B^{-\frac{1}{2}} D^{-\frac{1}{2}} K^{-\frac{n-2}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega} \\
&= \frac{\Gamma\left(\frac{n-1}{2}\right) \int [(\Omega + V) A^{n-1}]^{-\frac{1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega}{\Gamma\left(\frac{n-2}{2}\right) \sqrt{\pi} \int \left[ (\Omega + \sum_{k \neq i} v_k^2) K^{n-2} \right]^{-\frac{1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega}
\end{aligned} \tag{3.38}$$

com

$$\begin{aligned}
z_i &= y_i - x_i^t \beta \\
\Omega &= \frac{\omega}{\sigma_t^2} \\
V &= \sum_{k=1}^n v_k^2 \\
D &= 1 - \frac{v_i^2}{\Omega + V} \\
K &= \sum_{j \neq i} z_j^2 - \frac{1}{\Omega + \sum_{k \neq i} v_k^2} \left( \sum_{j \neq i} v_j z_j \right)^2
\end{aligned} \tag{3.39}$$

$$B = \sum_{j=1}^n v_j^2 + \frac{\omega}{\sigma_t^2}$$

$$A = \sum_{j=1}^n z_j^2 - \frac{1}{\Omega + V} \left( \sum_{k=1}^n v_k z_k \right)^2$$

Com o resultado obtido na equação (3.38), via integração de Monte Carlo, podemos calcular de forma aproximada  $f(y_i|\boldsymbol{\beta}, \mathbf{y}_{(i)})$ . Seja  $\mathbf{v}^{(l)} = (v_1^{(l)}, \dots, v_n^{(l)})$ ,  $l = 1, \dots, L$ , tal que  $v_i^{(l)}$  é uma amostra de  $v_i \sim HN(0, 1)$  e  $\omega^{(l)}$ ,  $l = 1, \dots, L$ , uma amostra de  $\omega \sim Gama((k+1)/2, k/2)$ , então

$$\hat{f}(y_i|\boldsymbol{\beta}, \mathbf{y}_{(i)}) = \frac{\Gamma(\frac{n-1}{2})^{\frac{1}{L}} \sum_{l=1}^L \left[ (\Omega_{(l)} + V_{(l)}) A_{(l)}^{n-1} \right]^{-\frac{1}{2}}}{\Gamma(\frac{n-2}{2}) \sqrt{\pi}^{\frac{1}{L}} \sum_{l=1}^L \left[ \left( \Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2} \right) K_{(l)}^{n-2} \right]^{-\frac{1}{2}}} \quad (3.40)$$

com

$$z_i = y_i - x_i^t \boldsymbol{\beta}$$

$$A_{(l)} = \sum_{j=1}^n z_j^2 - \frac{1}{\Omega_{(l)} + V_{(l)}} \left( \sum_{k=1}^n v_k^{(l)} z_k \right)^2$$

$$K_{(l)} = \sum_{j \neq i} z_j^2 - \frac{1}{\Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2}} \left( \sum_{j \neq i} v_j^{(l)} z_j \right)^2 \quad (3.41)$$

$$V_{(l)} = \sum_{j=1}^n v_j^{(l)2}$$

$$\Omega_{(l)} = \frac{\omega^{(l)}}{\sigma_t^2}$$

□

## 3.2 Modelo de Regressão $t$ -Student Assimétrico

### 3.2.1 Estimação dos Parâmetros

O modelo de regressão linear com resposta de distribuição  $t$ -assimétrica é dado por

$$y_i = x_i^t \boldsymbol{\beta} + \epsilon_i, \quad (3.42)$$

onde, para cada  $i = 1, \dots, n$ ,  $y_i$  é a variável resposta da observação  $i$ ,  $x_i^t = (1, x_{i1}, \dots, x_{ip})$  é o vetor de variáveis explicativas,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^t$  é vetor de parâmetros desconhecidos a serem estimados e  $\epsilon_i \sim ST(0, \sigma, \nu, \lambda)$ , ou seja, os erros possuem distribuição Skew  $t$  com parâmetro de posição 0, escala  $\sigma$ , graus de liberdade  $\nu$  (fixado) e assimetria  $\lambda$ .

Note que para cada  $i$ , pelas propriedades da distribuição  $t$ -assimétrica, temos que

$y_i \sim ST(x_i^t \boldsymbol{\beta}, \sigma, \nu, \lambda)$ , cuja função densidade de probabilidade é

$$f(y_i | \boldsymbol{\beta}, \sigma^2, \lambda) = \frac{2}{\sigma} t_\nu \left( \frac{y_i - x_i^t \boldsymbol{\beta}}{\sigma} \right) T_{\nu+1} \left( \lambda \frac{y_i - x_i^t \boldsymbol{\beta}}{\sigma} \sqrt{\frac{\nu+1}{\nu + \sigma^{-2}(y_i - x_i^t \boldsymbol{\beta})^2}} \right) \quad (3.43)$$

com  $t_\gamma$  e  $T_\gamma$  as funções densidade de probabilidade e de distribuição de uma  $t$ -Student padrão com  $\gamma$  graus de liberdade. De forma análoga ao caso normal assimétrico, se quisermos que a esperança de  $y_i$  seja escrita como uma reta de regressão precisamos considerar o seguinte intercepto

$$\beta_0^* = \beta_0 + \sigma \sqrt{\frac{\nu}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}} \frac{\Gamma(\frac{\nu-1}{2})}{(\frac{\nu}{2})} \quad (3.44)$$

e, assim,  $E[y_i] = x_i^t \boldsymbol{\beta}^*$ , com  $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1, \dots, \beta_p)$ .

Podemos utilizar a seguinte representação hierárquica, dada em [Godoi \(2007\)](#), para cada  $y_i$

$$\begin{aligned} y_i | u_i, v_i &\sim N \left( x_i^t \beta + \sigma \delta \frac{v_i}{\sqrt{u_i}}, \frac{\sigma^2(1 - \delta^2)}{u_i} \right) \\ u_i &\sim \text{Gama} \left( \frac{\nu}{2}, \frac{\nu}{2} \right) \\ v_i &\sim HN(0, 1) \end{aligned} \quad (3.45)$$

onde  $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$ .

Considerando a reparametrização  $\eta = \sigma \delta$  e  $\tau = \sigma \sqrt{1 - \delta^2}$  teremos, para cada  $i = 1, \dots, n$ ,

$$\begin{aligned} y_i | u_i, v_i &\sim N \left( x_i^t \beta + \eta \frac{v_i}{\sqrt{u_i}}, \frac{\tau^2}{u_i} \right) \\ u_i &\sim \text{Gama} \left( \frac{\nu}{2}, \frac{\nu}{2} \right) \\ v_i &\sim HN(0, 1). \end{aligned} \quad (3.46)$$

Veja que a variável  $u_i$  está presente tanto no parâmetro de posição quanto de escala da distribuição  $y_i | u_i, v_i$ . Para contornar isto consideraremos mais uma parametrização dada por

$$(u_i, v_i) \mapsto (u_i, t_i) = \left( u_i, \frac{v_i}{\sqrt{u_i}} \right) \quad (3.47)$$

e assim teremos

$$f(u_i, t_i) = f_{u_i, v_i}(u_i, t_i \sqrt{u_i}) \left| \frac{\partial(u_i, v_i)}{\partial(u_i, t_i)} \right| \quad (3.48)$$

onde  $f_{u_i, v_i}$  é a função densidade de probabilidade conjunta de  $(u_i, v_i)$  e

$$\frac{\partial(u_i, v_i)}{\partial(u_i, t_i)} = \begin{vmatrix} \frac{\partial u_i}{\partial u_i} & \frac{\partial u_i}{\partial t_i} \\ \frac{\partial v_i}{\partial u_i} & \frac{\partial v_i}{\partial t_i} \end{vmatrix} \quad (3.49)$$

Note que

$$\frac{\partial(u_i, v_i)}{\partial(u_i, t_i)} = \begin{vmatrix} \frac{\partial u_i}{\partial u_i} & \frac{\partial u_i}{\partial t_i} \\ \frac{\partial v_i}{\partial u_i} & \frac{\partial v_i}{\partial t_i} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \frac{\partial v_i}{\partial u_i} & \sqrt{u_i} \end{vmatrix} = \sqrt{u_i} \quad (3.50)$$

e temos

$$\begin{aligned} u_i &\sim \text{Gama}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \\ v_i &\sim HN(0, 1). \end{aligned} \quad (3.51)$$

Logo,

$$\begin{aligned} f(u_i, t_i) &= f_{u_i, v_i}(u_i, t_i \sqrt{u_i}) \left| \frac{\partial(u_i, v_i)}{\partial(u_i, t_i)} \right| = f_{u_i}(u_i) f_{v_i}(t_i \sqrt{u_i}) \sqrt{u_i} \\ &\propto u_i^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}u_i} \mathbb{I}_{[0, \infty[}(u_i) e^{-\frac{(t_i \sqrt{u_i})^2}{2}} \mathbb{I}_{[0, \infty[}(t_i \sqrt{u_i}) \sqrt{u_i} \\ &= u_i^{\frac{\nu-1}{2}} e^{-\frac{\nu}{2}u_i} e^{-\frac{t_i^2 u_i}{2}} \mathbb{I}_{[0, \infty[}(u_i) \mathbb{I}_{[0, \infty[}(t_i). \end{aligned} \quad (3.52)$$

A função de verossimilhança aumentada será dada por

$$\begin{aligned} L_A(\boldsymbol{\beta}, \tau, \eta) &= \prod_{i=1}^n f(y_i | u_i, t_i) f(u_i, t_i) \\ &\propto \frac{1}{\tau^n} \exp\left(-\frac{1}{2\tau^2} \sum_{i=1}^n u_i (y_i - x_i^t \boldsymbol{\beta} - \eta t_i)^2\right) \prod_{i=1}^n \left[ u_i^{\frac{\nu-1}{2}} e^{-\frac{\nu}{2}u_i} e^{-\frac{u_i t_i^2}{2}} \mathbb{I}_{[0, \infty[}(u_i) \mathbb{I}_{[0, \infty[}(t_i) \right]. \end{aligned} \quad (3.53)$$

Iremos usar distribuições *a priori* vagas neste estudo considerando, assim, desconhecimento prévio a respeito dos parâmetros a serem estimados bem como fizemos no caso do modelo de regressão normal assimétrico. Como  $\nu$  não será objeto de estudo por ser considerado fixo, ele também não terá uma distribuição *a priori*. Dessa forma iremos considerar a seguinte especificação igual a do caso normal assimétrico

$$f(\boldsymbol{\beta}, \tau, \eta, \omega) \propto \frac{1}{\tau^2} \exp\left(-\frac{\eta^2 \omega}{2\tau^2 \sigma_t^2}\right) \omega^{\frac{k+1}{2}-1} \exp\left(-\frac{k}{2}\omega\right) \quad (3.54)$$

A distribuição *a posteriori* aumentada será dada por

$$\begin{aligned} f(\boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{u}, \mathbf{t} | \mathbf{y}) &f(\mathbf{y} | \boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{u}, \mathbf{t}) f(\boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{u}, \mathbf{t}) \\ &= f(\mathbf{y} | \boldsymbol{\beta}, \tau, \eta, \omega, \mathbf{u}, \mathbf{t}) f(\mathbf{u}, \mathbf{t}) f(\boldsymbol{\beta}, \tau, \eta, \omega) \\ &= L_A(\boldsymbol{\beta}, \tau, \eta) f(\boldsymbol{\beta}, \tau, \eta, \omega). \end{aligned} \quad (3.55)$$

A partir de (3.55) obtemos as condicionais completas apresentadas a seguir:

$$\begin{aligned} \boldsymbol{\beta} | \tau, \eta, \omega, \mathbf{u}, \mathbf{t}, \mathbf{y} &\sim N_{p+1} \left( (X^t U X)^{-1} X^t U Z, \tau^2 (X^t U X)^{-1} \right) \\ \frac{1}{\tau^2} | \mu, \eta, \omega, \mathbf{u}, \mathbf{t}, \mathbf{y} &\sim \text{Gama} \left( \frac{n+1}{2}, \frac{1}{2} \left( (Z - X\boldsymbol{\beta})^t U (Z - X\boldsymbol{\beta}) + \frac{\eta^2 \omega}{\sigma_t^2} \right) \right) \\ \eta | \mu, \tau, \omega, \mathbf{u}, \mathbf{t}, \mathbf{y} &\sim N \left( \frac{(Y - X\boldsymbol{\beta})^t U T}{T^t U T + \frac{\omega}{\sigma_t^2}}, \frac{\tau^2}{T^t U T + \frac{\omega}{\sigma_t^2}} \right) \\ \omega | \mu, \tau, \eta, \mathbf{u}, \mathbf{t}, \mathbf{y} &\sim \text{Gama} \left( \frac{k+1}{2}, \frac{1}{2} \left( k + \frac{\lambda^2}{\tau^2 \sigma_t^2} \right) \right) \end{aligned}$$



$$\begin{aligned}
u_i | \mu, \tau, \eta, \omega, \mathbf{t}, \mathbf{y} &\sim \text{Gama} \left( \frac{\nu+1}{2}, \frac{1}{2} \left( t_i^2 + \nu + \frac{1}{\tau^2} (y_i - x_i^t \beta - \eta t_i)^2 \right) \right), \quad i = 1, \dots, n \\
t_i | \mu, \tau, \eta, \omega, \mathbf{u}, \mathbf{y} &\sim N \left( \frac{\eta(y_i - x_i^t \beta)}{\eta^2 + \tau^2}, \frac{\tau^2}{u_i(\eta^2 + \tau^2)} \right) \mathbb{I}_{[0, +\infty[}(t_i), \quad i = 1, \dots, n
\end{aligned} \tag{3.56}$$

onde

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad U = \begin{bmatrix} u_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & u_n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad T = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}, \quad Z = Y - \eta T \tag{3.57}$$

Por meio das condicionais completas, podemos utilizar o amostrador de Gibbs para obter amostras das distribuições *a posteriori* dos parâmetros e, daí, obter a inferência bayesiana para tais parâmetros. No apêndice apresentamos o programa que implementa este amostrador de Gibbs.

### 3.2.2 Derivação das Medidas de Diagnóstico

Considere  $(\beta^{(1)}, \sigma^{2(1)}, \lambda^{(1)}), \dots, (\beta^{(L)}, \sigma^{2(L)}, \lambda^{(L)})$  uma amostra de tamanho  $L$  da distribuição *a posteriori* de  $\beta, \sigma^2, \lambda | \mathbf{y}$  que será utilizada em todos os cálculos das medidas de diagnóstico e consideraremos  $\nu$  fixado.

A estimativa para o  $CPO_i$  neste caso da regressão linear  $t$ -assimétrica será dada por

$$\widehat{CPO}_i = \left[ \frac{1}{L} \sum_{l=1}^n \frac{1}{f(y_i | \beta^{(l)}, \sigma^{2(l)}, \lambda^{(l)})} \right]^{-1} \tag{3.58}$$

onde  $f(y_i | \beta, \sigma^2, \lambda)$  é a função de distribuição  $t$ -assimétrica de posição  $x_i^t \beta$ , escala  $\sigma^2$ , assimetria  $\lambda$  e graus de liberdade  $\nu$ .

A estimativa da divergência para a influência global é dada por

$$D_{\beta, \sigma^2, \lambda}(g, i) = \frac{1}{L} \sum_{l=1}^n g \left( \frac{\widehat{CPO}_i}{f(y_i | \beta^{(l)}, \sigma^{2(l)}, \lambda^{(l)})} \right) \tag{3.59}$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (3.58). Para a obtenção da norma  $L_1$  e da divergência Kullback-Leibler basta substituir a função  $g$  adequada para cada caso.

**Proposição 3.2.** *Uma estimativa de Monte Carlo da divergência para a influência marginal no modelo  $t$ -assimétrico é dada por*

$$\widehat{D}_\beta(g, i) = \frac{1}{L} \sum_{l=1}^n g \left( \frac{\widehat{CPO}_i}{\hat{f}(y_i | \beta, \mathbf{y}_{(i)})} \right) \tag{3.60}$$

onde  $\widehat{CPO}_i$  é a estimativa para o  $CPO_i$  obtida em (2.38) e

$$\hat{f}(y_i | \beta, \mathbf{y}_{(i)}) = \frac{\Gamma(\frac{n-1}{2}) \frac{1}{L} \sum_{l=1}^L \left[ (\Omega_{(l)} + V_{(l)}) A_{(l)}^{n-1} \right]^{-\frac{1}{2}}}{\Gamma(\frac{n-2}{2}) \sqrt{\pi} \frac{1}{L} \sum_{l=1}^L \left[ \left( \Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2} \right) K_{(l)}^{n-2} \right]^{-\frac{1}{2}}} \tag{3.61}$$

com

$$\begin{aligned}
z_j^{(l)} &= \sqrt{u_j^{(l)}}(y_j - x_j^t \boldsymbol{\beta}) \\
v_j^{(l)} &= \sqrt{u_j^{(l)}} t_j^{(l)} \\
A_{(l)} &= \sum_{j=1}^n z_j^{(l)2} - \frac{1}{\Omega_{(l)} + V_{(l)}} \left( \sum_{k=1}^n v_k^{(l)} z_k^{(l)2} \right)^2 \\
K_{(l)} &= \sum_{j \neq i} z_j^{(l)2} - \frac{1}{\Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2}} \left( \sum_{j \neq i} v_j^{(l)} z_j^{(l)2} \right)^2 \\
V_{(l)} &= \sum_{j=1}^n v_j^{(l)2} \\
\Omega_{(l)} &= \frac{\omega_{(l)}}{\sigma_t^2}
\end{aligned} \tag{3.62}$$

sendo  $\mathbf{u}^{(l)}$ ,  $l = 1, \dots, L$ , e, para cada  $i = 1, \dots, n$ ,  $u_i^{(1)}, \dots, u_i^{(L)}$  é uma amostra de tamanho  $L$  de uma distribuição  $\text{Gama}((\nu + 1)/2, \nu/2)$ ,  $\mathbf{v}^{(l)}$ ,  $l = 1, \dots, L$ , e, para cada  $i = 1, \dots, n$ ,  $v_i^{(1)}, \dots, v_i^{(L)}$  é uma amostra de tamanho  $L$  de uma distribuição  $HN(0, 1/u_i)$  e  $\omega^{(1)}, \dots, \omega^{(L)}$  é uma amostra de tamanho  $L$  de uma distribuição  $\text{Gama}((k + 1)/2, k/2)$ .

*Demonstração.* Primeiro, obtenhamos a forma de  $f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)})$  no caso da distribuição  $t$ -assimétrica. Veja que

$$\begin{aligned}
f(y_i | \boldsymbol{\beta}, \mathbf{y}_{(i)}) &= \int f(\mathbf{y} | \boldsymbol{\beta}, \tau^2, \eta, \omega, \mathbf{u}, \mathbf{t}) f(\mathbf{u}, \mathbf{t}) f(\boldsymbol{\beta}, \tau^2, \eta, \omega) d\mathbf{u} d\mathbf{t} d\tau^2 d\eta d\omega \\
&\propto \int f(\mathbf{y} | \boldsymbol{\beta}, \tau^2, \eta, \omega, \mathbf{u}, \mathbf{t}) \frac{1}{\tau^2} f(\mathbf{u}, \mathbf{t}) f(\eta | \omega) g(\omega) d\mathbf{u} d\mathbf{t} d\tau^2 d\eta d\omega \\
&\propto \int \frac{1}{\tau^n} \exp \left( \frac{-1}{2\tau^2} \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta} - \eta t_j)^2 \right) \frac{1}{\tau^2} \\
&\quad f(\mathbf{u}, \mathbf{t}) \exp \left( \frac{-\eta^2 \omega}{2\tau^2 \sigma_t^2} \right) g(\omega) d\mathbf{u} d\mathbf{t} d\tau^2 d\eta d\omega \\
&\propto \int \int \frac{1}{\tau^{n+2}} \exp \left( \frac{-1}{2\tau^2} \left( \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta} - \eta t_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \right) \right) d\tau^2 \\
&\quad f(\mathbf{u}, \mathbf{t}) g(\omega) d\mathbf{u} d\mathbf{t} d\eta d\omega \\
&\propto \int \left[ \sum_{j=1}^n u_j (y_j - x_j^t \boldsymbol{\beta} - \eta t_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \right]^{-\frac{n}{2}} d\eta f(\mathbf{u}, \mathbf{t}) f(\omega) d\mathbf{u} d\mathbf{t} d\omega
\end{aligned} \tag{3.63}$$

onde  $f(\mathbf{u}, \mathbf{t}) = \prod_{i=1}^n f(u_i, t_i)$  e  $f(u_i, t_i)$  é a função densidade de probabilidade deduzida na equação 3.52 e  $g(\omega)$  é a função densidade de probabilidade de  $\omega \sim \text{Gama}(\frac{k+1}{2}, \frac{k}{2})$ . Podemos escrever de forma hierárquica a distribuição  $(\mathbf{u}, \mathbf{t})$

$$\begin{aligned}
t_i | u_i &\sim HN(0, 1/u_i) \\
u_i &\sim \text{Gamma} \left( \frac{\nu + 1}{2}, \frac{\nu}{2} \right).
\end{aligned} \tag{3.64}$$

Fazendo  $\xi = \frac{\sum_{j=1}^n u_j t_j (y_j - x_j^t \beta)}{\sum_{j=1}^n u_j t_j^2}$  para isolar  $\eta$  em (3.63) teremos

$$\begin{aligned}
& \sum_{j=1}^n u_j (y_j - x_j^t \beta - \eta t_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \\
&= \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2 + \sum_{j=1}^n u_j t_j^2 (\xi - \eta)^2 + \frac{\omega \eta^2}{\sigma_t^2} \\
&= \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2 + \\
&\quad + \left( \sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2} \right) \left( \eta - \frac{\sum_{j=1}^n u_j t_j^2 \xi}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}} \right)^2 + \frac{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}} \\
&= \underbrace{\sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2}_{A} + \underbrace{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \xi^2}_{C} + \\
&\quad + \left( \underbrace{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}}_B \right) \left( \eta - \underbrace{\frac{\sum_{j=1}^n u_j t_j^2 \xi}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}}}_C \right)^2 \\
&= A + B(\eta - C)^2
\end{aligned} \tag{3.65}$$

De forma análoga ao caso normal assimétrico, visto na equação (3.29),

$$[A + B(\eta - C)^2]^{-\frac{n}{2}} = A^{-\frac{n}{2}} h^*(\eta) \tag{3.66}$$

onde  $h^*$  é o núcleo de uma distribuição  $t(C, \sqrt{A}/\sqrt{(n-1)B}, n-1)$ . Logo, analogamente ao resultado da equação (3.30),

$$\begin{aligned}
f(y_i | \beta, \mathbf{y}_{(i)}) &= \int \left[ \sum_{j=1}^n u_j (y_j - x_j^t \beta - \eta t_j)^2 + \frac{\omega \eta^2}{\sigma_t^2} \right]^{-\frac{n}{2}} d\eta f(\mathbf{u}, \mathbf{t}) g(\omega) d\mathbf{u} d\mathbf{t} d\omega \\
&\propto \int A^{-\frac{n-1}{2}} B^{-\frac{1}{2}} f(\mathbf{u}, \mathbf{t}) g(\omega) d\mathbf{u} d\mathbf{t} d\omega
\end{aligned} \tag{3.67}$$

A constante de normalização para proporcionalidade encontrada na equação (3.67) será calculada por meio da equação a seguir

$$\int \int A^{-\frac{n-1}{2}} B^{-\frac{1}{2}} f(\mathbf{u}, \mathbf{t}) f(\omega) d\mathbf{u} d\mathbf{t} d\omega dy_i = \int B^{-\frac{1}{2}} \left[ \int A^{-\frac{n-1}{2}} dy_i \right] f(\mathbf{u}, \mathbf{t}) f(\omega) d\mathbf{u} d\mathbf{t} d\omega \tag{3.68}$$

com  $A = \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2 + \frac{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}}$ ,  $\xi = \frac{\sum_{j=1}^n u_j t_j (y_j - x_j^t \beta)}{\sum_{j=1}^n u_j t_j^2}$  e  $B = \sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}$ .

Iremos integrar analiticamente (3.68) somente em  $y_i$  pois, nas outras variáveis,  $\mathbf{u}$ ,  $\mathbf{t}$  e

$\omega$ , iremos utilizar a integração de Monte Carlo e assim teremos

$$\int A^{-\frac{n+1}{2}} dy_i = \int \left[ \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2 + \frac{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}} \right]^{-\frac{n+1}{2}} dy_i \quad (3.69)$$

Podemos trabalhar melhor a fórmula que está sendo integrada deixando o termo  $y_i$  em maior evidência. Assim teremos

$$\begin{aligned} & \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2 + \frac{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}} \\ &= \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \frac{\sum_{k=1}^n u_k t_k (y_k - x_k^t \beta)}{\sum_{k=1}^n u_k t_k^2} \right)^2 + \frac{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \left( \frac{\sum_{k=1}^n u_k t_k (y_k - x_k^t \beta)}{\sum_{k=1}^n u_k t_k^2} \right)^2}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}} \end{aligned} \quad (3.70)$$

Para simplificar a notação, façamos  $z_j = \sqrt{u_j}(y_j - x_j^t \beta)$ ,  $v_j = \sqrt{u_j} t_j$ ,  $\Omega = \frac{\omega}{\sigma_t^2}$  e  $V = \sum_{k=1}^n u_k t_k^2$ ,

$$\begin{aligned} & \sum_{j=1}^n u_j t_j^2 \left( \frac{y_j - x_j^t \beta}{t_j} - \xi \right)^2 + \frac{\sum_{j=1}^n u_j t_j^2 \frac{\omega}{\sigma_t^2} \xi^2}{\sum_{j=1}^n u_j t_j^2 + \frac{\omega}{\sigma_t^2}} \\ &= \sum_{j=1}^n \left( \sqrt{u_j}(y_j - x_j^t \beta) - v_j \frac{\sum_{k=1}^n v_k \sqrt{u_k}(y_k - x_k^t \beta)}{V} \right)^2 + \frac{V \Omega \left( \frac{\sum_{k=1}^n v_k \sqrt{u_k}(y_k - x_k^t \beta)}{V} \right)^2}{\Omega + V} \\ &= \sum_{j=1}^n \left( z_j - \frac{v_j}{V} \sum_{k=1}^n v_k z_k \right)^2 + \frac{\Omega/V}{\Omega + V} \left( \sum_{k=1}^n v_k z_k \right)^2 \end{aligned} \quad (3.71)$$

Note que obtemos uma forma análoga a equação (3.34) no caso da distribuição normal assimétrica, porém, aqui,  $v_j = \sqrt{u_j} t_j$  e  $z_j = \sqrt{u_j}(y_j - x_j^t \beta)$ . Podemos assim concluir que

$$f(y_i | \beta, \mathbf{y}_{(i)}) = \frac{\Gamma\left(\frac{n-1}{2}\right) \int [(\Omega + V) A^{n-1}]^{-\frac{1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega}{\Gamma\left(\frac{n-2}{2}\right) \sqrt{\pi} \int \left[ (\Omega + \sum_{k \neq i} v_k^2) K^{n-2} \right]^{-\frac{1}{2}} f(\mathbf{v}) g(\omega) d\mathbf{v} d\omega} \quad (3.72)$$

com

$$\begin{aligned} z_j &= \sqrt{u_j}(y_j - x_j^t \beta) \\ v_j &= \sqrt{u_j} t_j \\ K &= \sum_{j \neq i} z_j^2 - \frac{1}{\Omega + \sum_{k \neq i} v_k^2} \left( \sum_{j \neq i} v_j z_j \right)^2 \\ A &= \sum_{j=1}^n z_j^2 - \frac{1}{\Omega + V} \left( \sum_{k=1}^n v_k z_k \right)^2 \\ \Omega &= \frac{\omega}{\sigma_t^2} \end{aligned} \quad (3.73)$$

$$V = \sum_{k=1}^n u_k t_k^2$$

Com o resultado obtido na equação (3.72), via integração de Monte Carlos, podemos calcular de forma aproximada  $f(y_i|\boldsymbol{\beta}, \mathbf{y}_{(i)})$ , sendo  $\mathbf{u}^{(l)} = (u_1^{(l)}, \dots, u_n^{(l)})$ ,  $l = 1, \dots, L$ , tal que  $u_i^{(l)}$  é uma amostra de  $Gama((\nu + 1)/2, \nu/2)$ ,  $\mathbf{t}^{(l)} = (t_1^{(l)}, \dots, t_n^{(l)})$ ,  $l = 1, \dots, L$ , tal que  $t_i^{(l)}$  é uma amostra de  $HN(0, 1/u_i)$  e  $\omega^{(l)}$ ,  $l = 1, \dots, L$ , uma amostra de  $\omega \sim Gama((k + 1)/2, k/2)$ ,

$$\hat{f}(y_i|\boldsymbol{\beta}, \mathbf{y}_{(i)}) = \frac{\Gamma(\frac{n-1}{2})^{\frac{1}{L}} \sum_{l=1}^L \left[ (\Omega_{(l)} + V_{(l)}) A_{(l)}^{n-1} \right]^{-\frac{1}{2}}}{\Gamma(\frac{n-2}{2}) \sqrt{\pi}^{\frac{1}{L}} \sum_{l=1}^L \left[ \left( \Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2} \right) K_{(l)}^{n-2} \right]^{-\frac{1}{2}}} \quad (3.74)$$

com

$$\begin{aligned} z_j^{(l)} &= \sqrt{u_j^{(l)}} (y_j - x_j^t \boldsymbol{\beta}) \\ v_j^{(l)} &= \sqrt{u_j^{(l)}} t_j^{(l)} \\ A_{(l)} &= \sum_{j=1}^n z_j^{(l)2} - \frac{1}{\Omega_{(l)} + V_{(l)}} \left( \sum_{k=1}^n v_k^{(l)} z_k^{(l)2} \right)^2 \\ K_{(l)} &= \sum_{j \neq i} z_j^{(l)2} - \frac{1}{\Omega_{(l)} + \sum_{k \neq i} v_k^{(l)2}} \left( \sum_{j \neq i} v_j^{(l)} z_j^{(l)2} \right)^2 \\ V_{(l)} &= \sum_{j=1}^n v_j^{(l)2} \\ \Omega_{(l)} &= \frac{\omega_{(l)}}{\sigma_t^2} \end{aligned} \quad (3.75)$$

□

### 3.3 Aplicação

Nesta seção iremos considerar os dados apresentados na seção 2.4 tanto em sua forma original quanto contaminando algumas observações. As contaminações propostas têm como objetivo explorar os possíveis casos de robustez do modelo  $t$ -assimétrico.

#### 3.3.1 Dados originais

Primeiro vamos avaliar o modelo que melhor ajusta os dados, considerando agora os modelos assimétricos. Na tabela 3.1, podemos observar que o modelo que melhor ajusta ao dados é o  $t$ -Student. Embora os resíduos sob o modelo normal apresentados na figura 3.2 exibam assimetria a esquerda, a observação 19 (discrepante) está na direção oposta desta assimetria. Neste caso, um modelo simétrico de caudas pesadas ajustou melhor os dados do que os modelos assimétricos propostos.

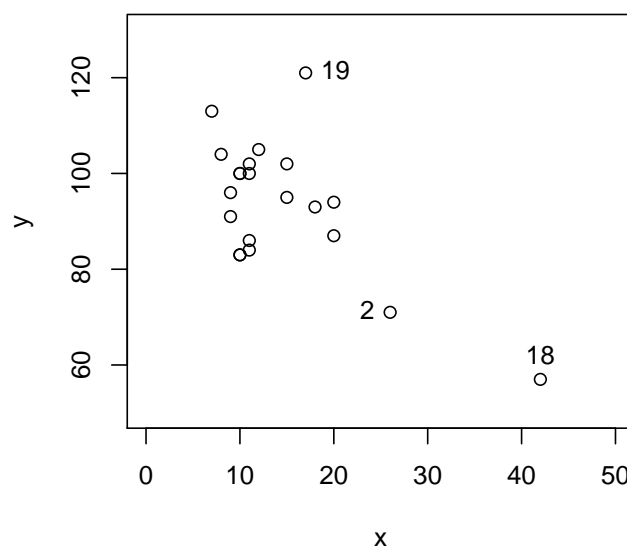
Em todos os modelos estudados a observação 19 é considerada discrepante por possuir a estatística  $-\log(CPO)$  muito alta como podemos ver na figura 3.3. Contudo, no modelo  $t$ -Student, a observação 19 não possui uma influência global de destaque como nos demais modelos, considerando tanto a norma  $L_1$  quanto a divergência de Kullback-Leibler (figuras 3.4 e 3.5, respectivamente).

Analisando agora a influência marginal, isto é, na distribuição *a posteriori*  $\beta|\mathbf{y}$ , observamos nas figuras 3.6 e 3.7 que nos modelos simétricos a observação 19 não possui uma influência de tanto destaque. Já nos modelos assimétricos, a observação 19 ainda se mostra influente quando consideramos a norma  $L_1$ . Considerando a divergência Kullback-Leibler, principalmente para o modelo  $t$ -assimétrico, temos gráficos pouco conclusivos. Como as divergências são calculadas por meio de diversas estimativas de Monte Carlo, podemos encontrar alguma instabilidade nas estimativas obtidas. Porém, de forma geral, a observação 19 possui maior influência marginal nos modelos assimétricos do que nos modelos simétricos.

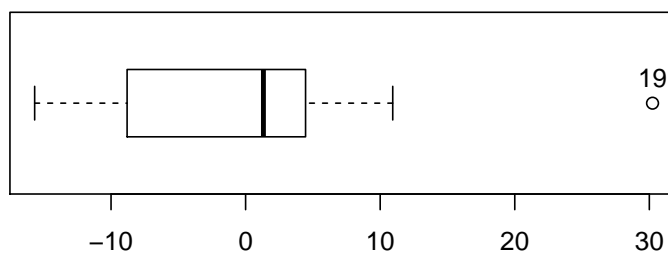
**Tabela 3.1:** *LPML dos modelos nos dados originais*

Normal	$t$ -Student	Normal assimétrica	$t$ -assimétrica
-82.75	-81.75	-82.5	-84.82

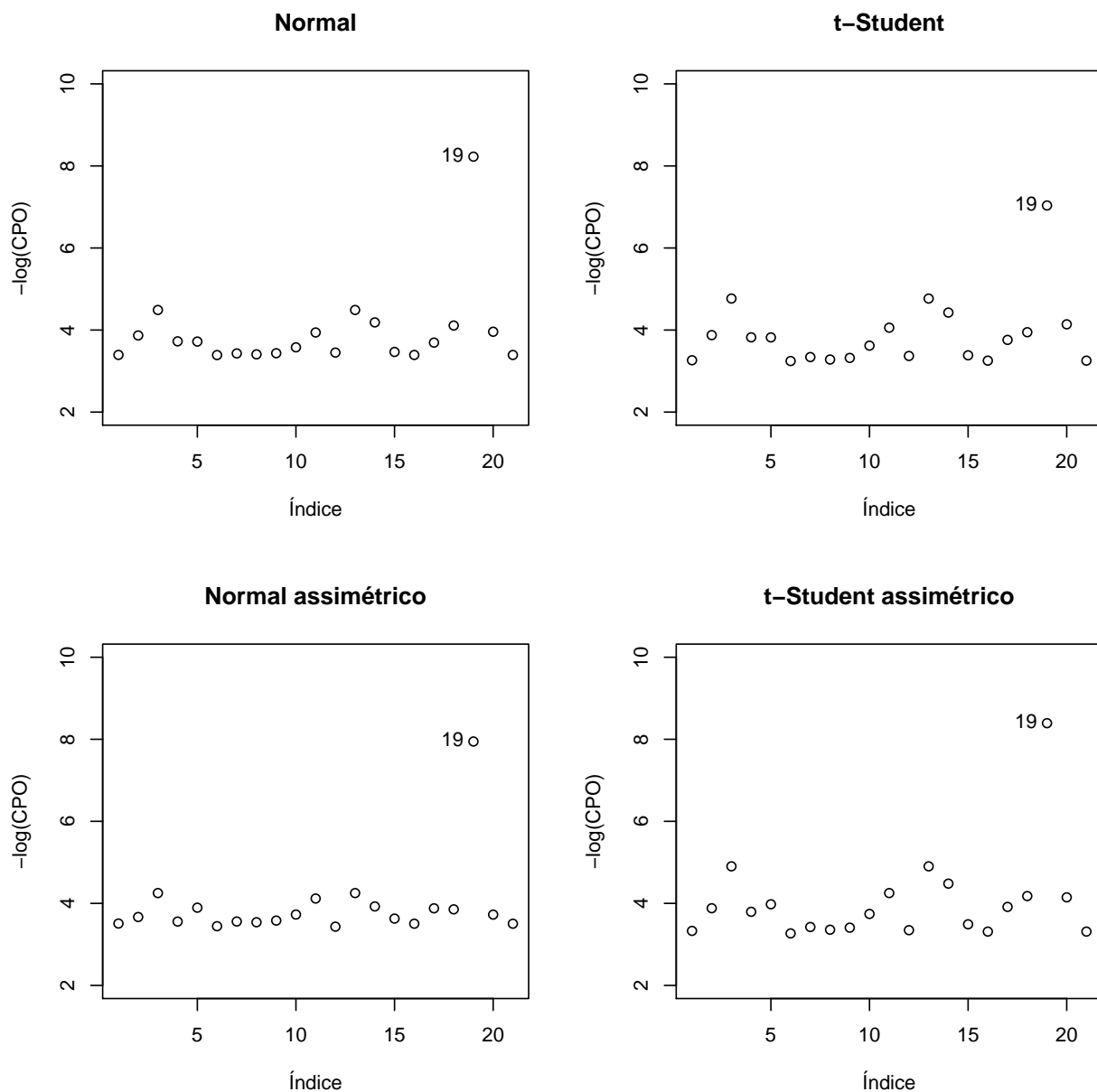
Fonte: Elaborada pelo autor.

**Figura 3.1:** *Diagrama de dispersão dos dados originais*

Fonte: Elaborada pelo autor.

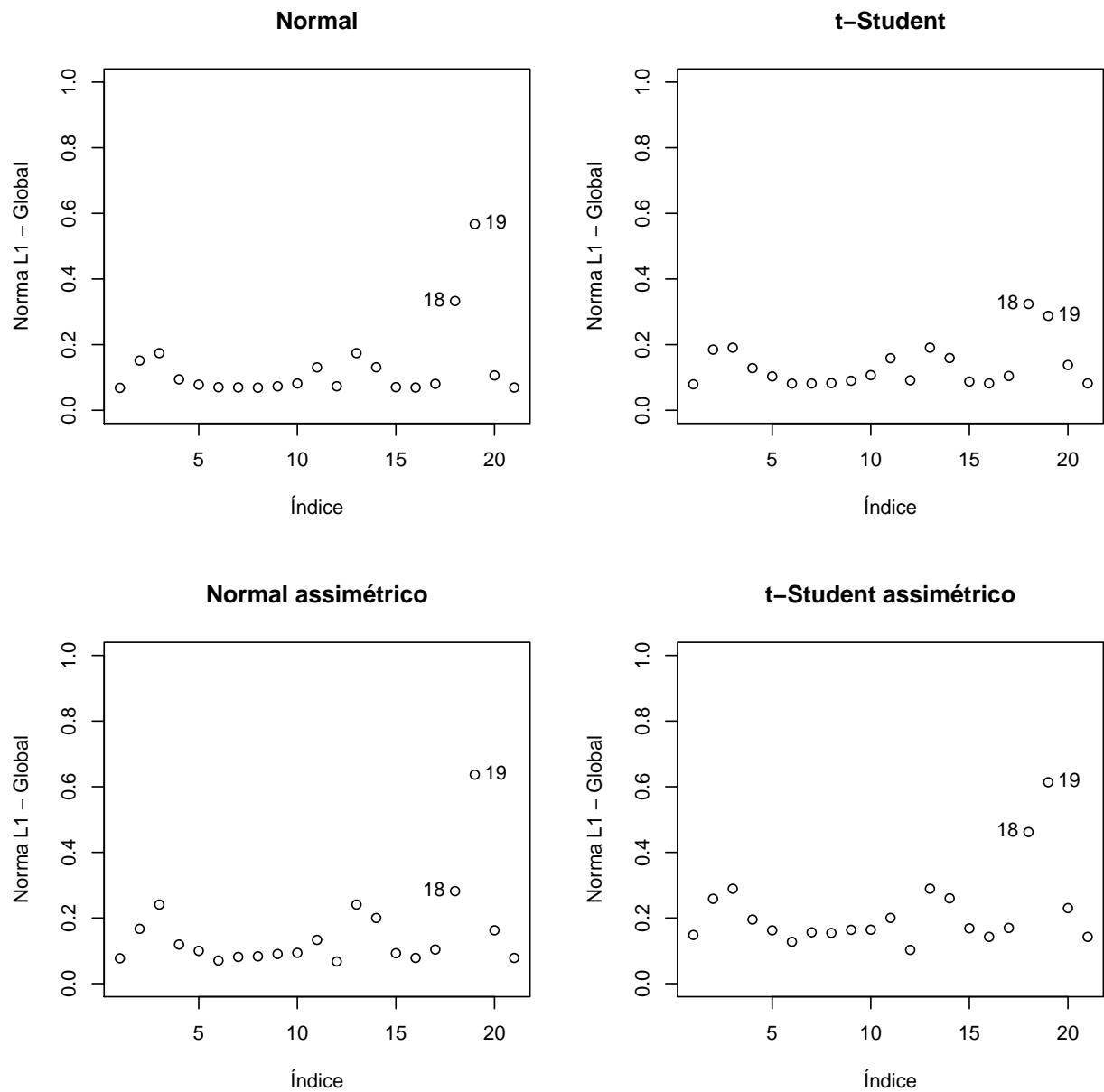
**Figura 3.2:** *Boxplot dos resíduos do modelo de regressão normal nos dados originais*

Fonte: Elaborada pelo autor.

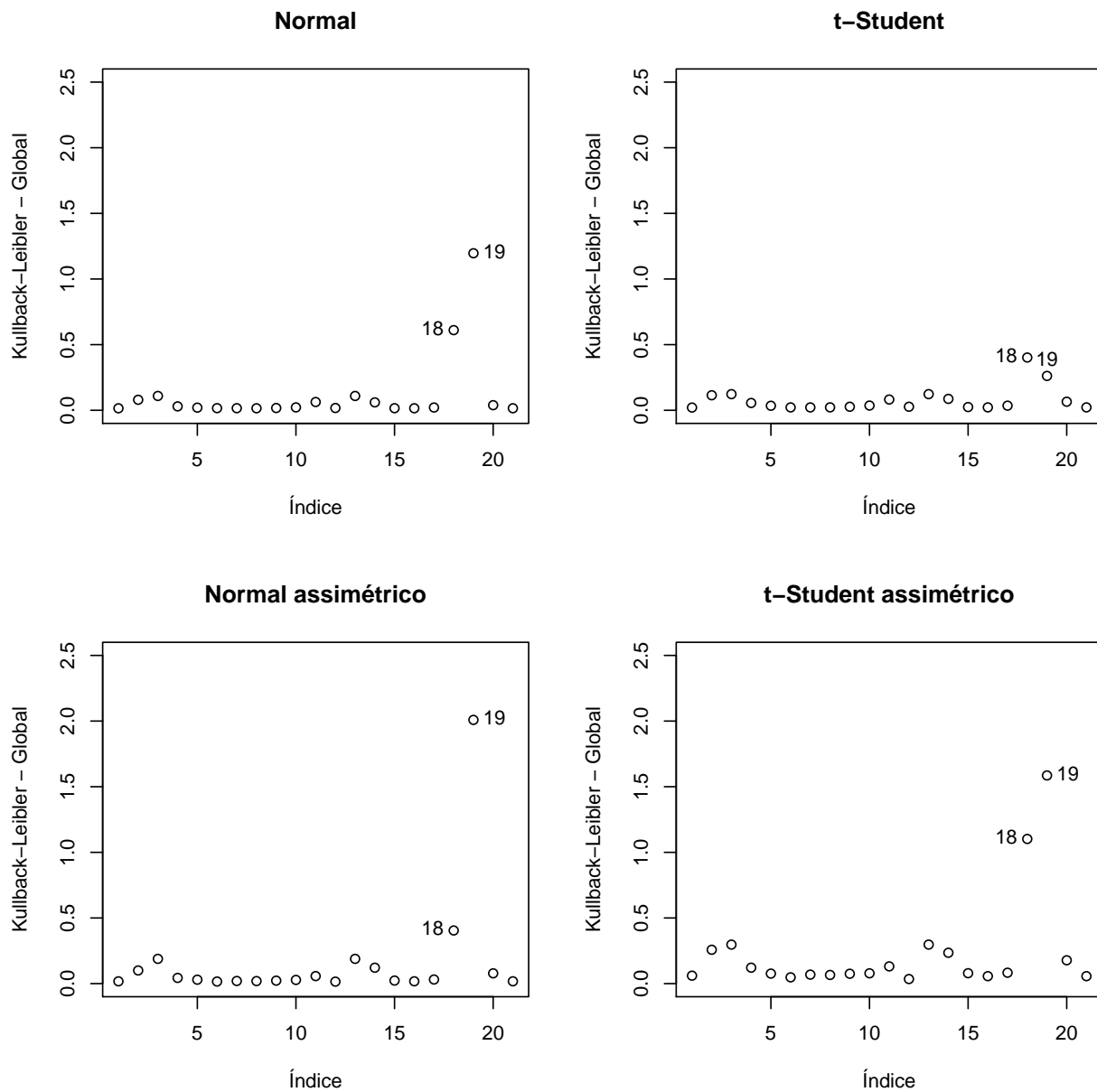
**Figura 3.3:**  $-\log(CPO)$  dos modelos nos dados originais

Fonte: Elaborada pelo autor.



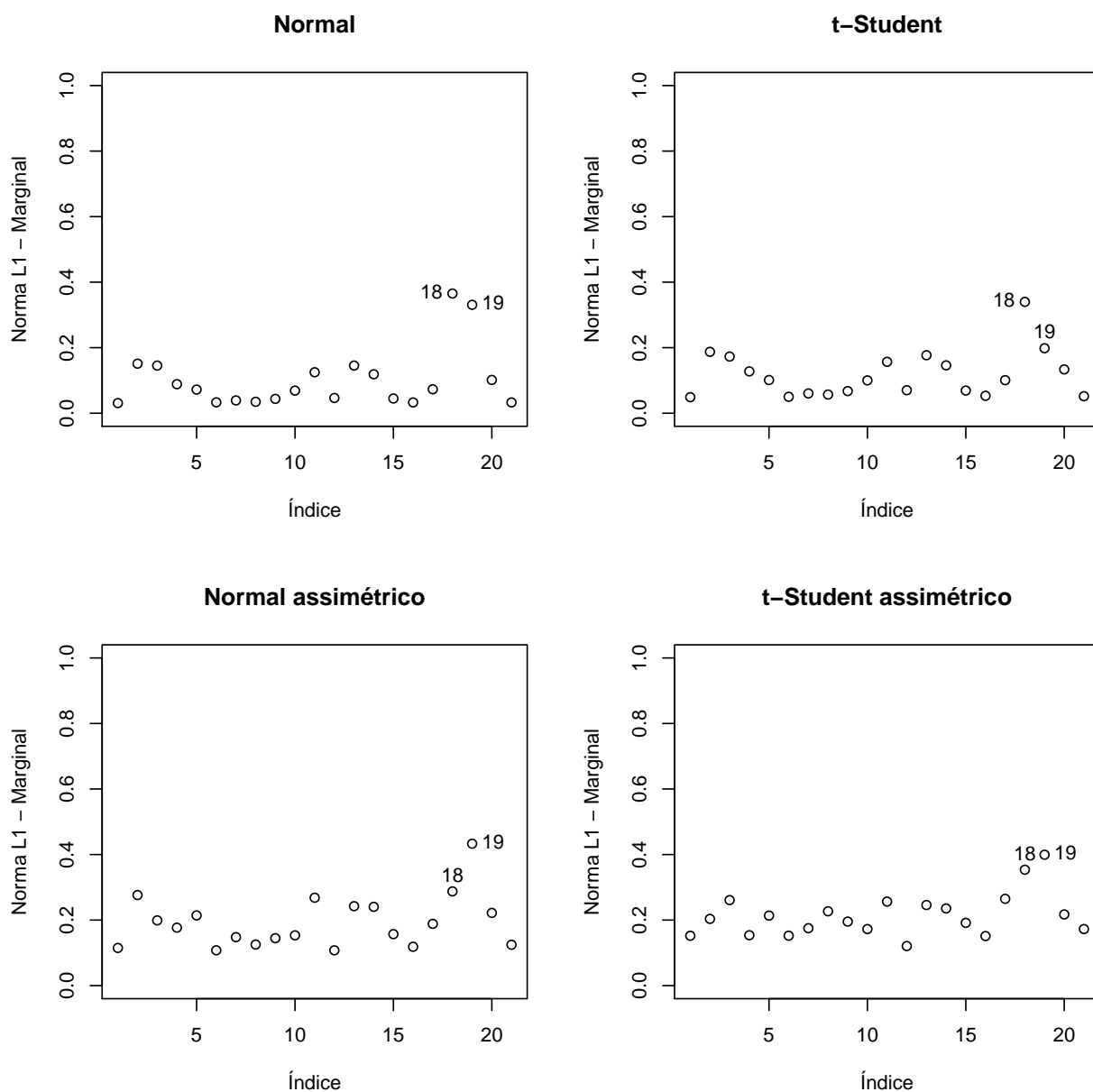
**Figura 3.4:** *Influência global via norma  $L_1$  dos modelos nos dados originais*

Fonte: Elaborada pelo autor.

**Figura 3.5:** *Influência global via divergência de Kullback-Leibler dos modelos nos dados originais*

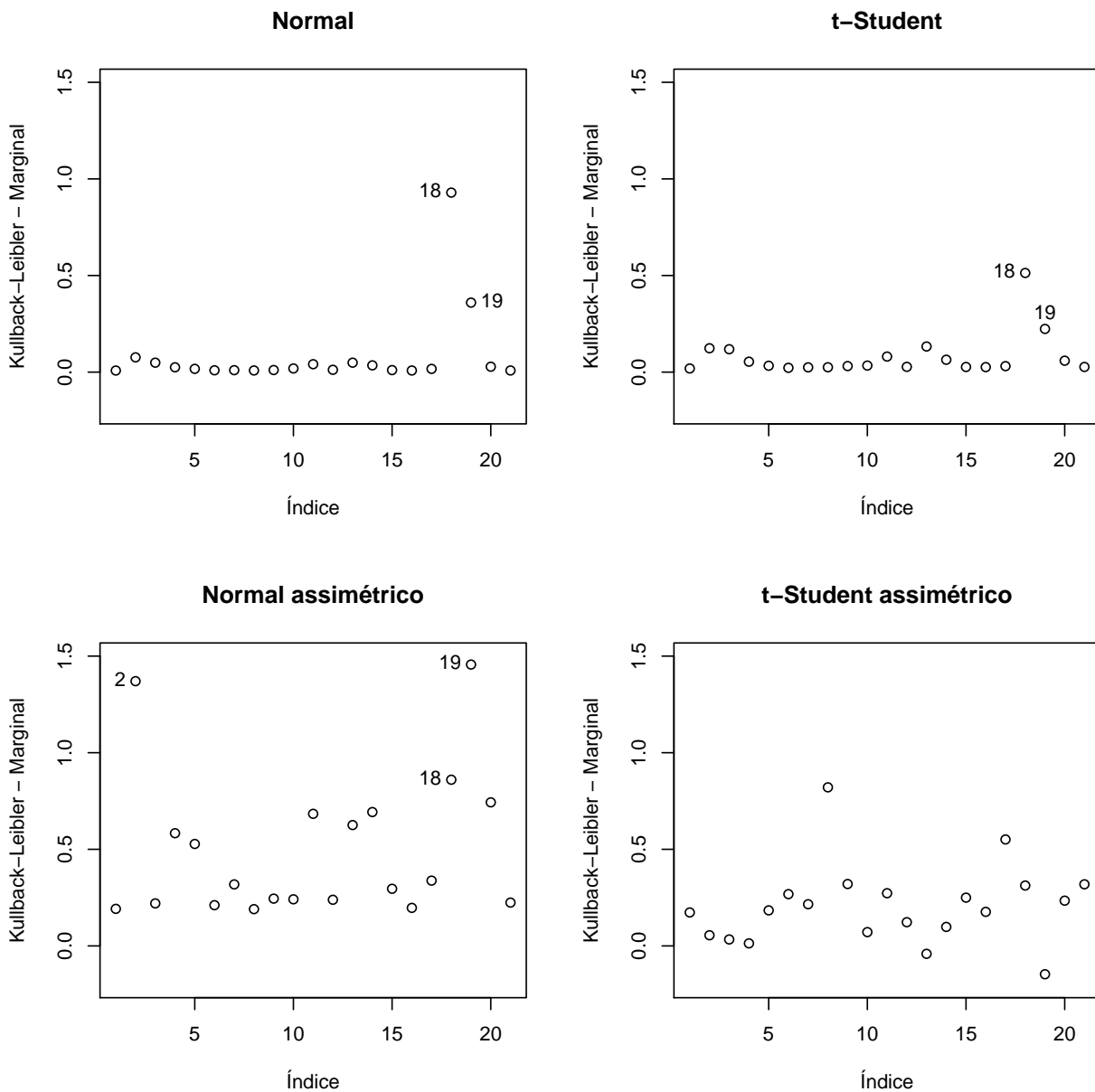
Fonte: Elaborada pelo autor.

**Figura 3.6:** *Influência marginal via norma  $L_1$  dos modelos nos dados originais*



Fonte: Elaborada pelo autor.

**Figura 3.7:** *Influência marginal via divergência de Kullback-Leibler dos modelos nos dados originais*



Fonte: Elaborada pelo autor.

### 3.3.2 Contaminação da observação 19

Nesta aplicação, promovemos a contaminação da observação 19 definindo  $y_{19} = 60$ . Agora podemos observar na figura que o resíduo da distribuição 19 está de acordo com a assimetria negativa da distribuição dos resíduos. Logo, podemos ver na tabela 3.2 que, de fato, os modelos assimétricos possuem um ajuste melhor que os modelos simétricos.

A observação 19 é a mais discrepante em termos do  $-\log(CPO)$  em todos os modelos (figura 3.10). Nas figuras 3.11 e 3.12 podemos ver que a observação 19 possui uma menor influência global quando comparamos o modelo normal com o modelo  $t$ -Student e o modelo normal assimétrico com o modelo  $t$ -assimétrico. Logo, a suposição de uma cauda mais pesada robustifica os modelos de cauda mais leve. Há um destaque adicional para a

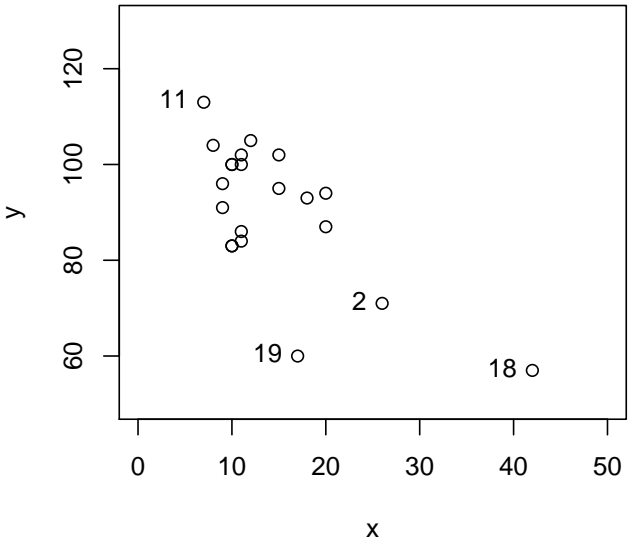
observação 11 que, apesar de não ser uma observação discrepante, exerce uma influência nos modelos assimétricos por estar localizado na cauda de menor peso destas distribuições. Pelas figuras 3.13 e 3.14 notamos que a influência marginal da observação 19 nos modelos simétricos é bem reduzida. Já nos modelos assimétricos, vemos que a influência da observação 19 é maior no modelo normal assimétrico do que no modelo *t*-assimétrico. Logo, há um indício que o modelo *t*-assimétrico é menos influenciado marginalmente.

Tabela 3.2: LPML dos modelos contaminando a observação 19

Normal	<i>t</i> -Student	Normal assimétrica	<i>t</i> -assimétrica
-81.6	-81.38	-78.73	-79.7

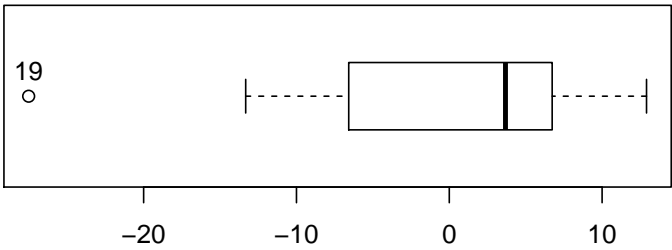
Fonte: Elaborada pelo autor.

Figura 3.8: Diagrama de dispersão contaminando a observação 19

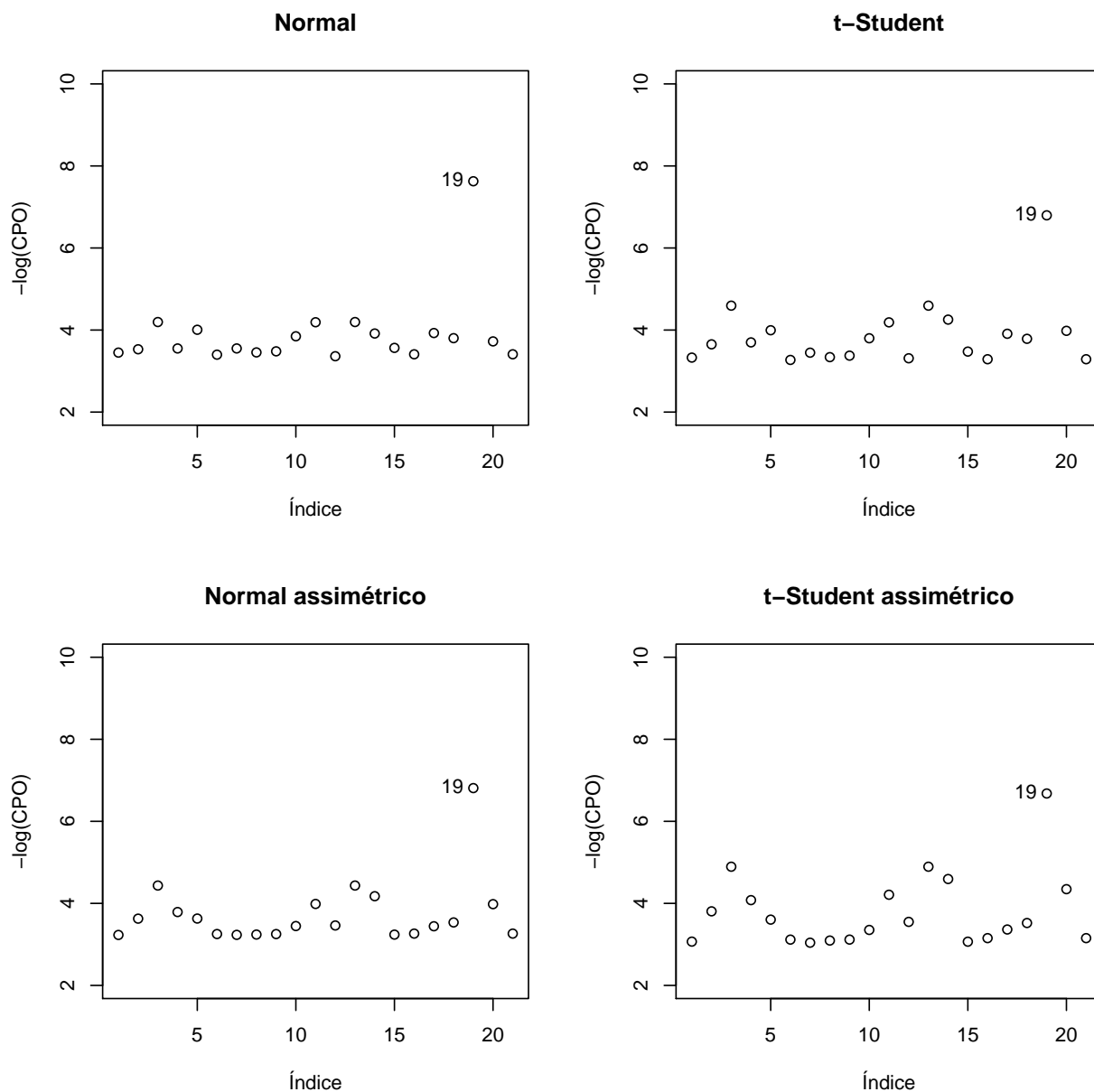


Fonte: Elaborada pelo autor.

Figura 3.9: Boxplot dos resíduos do modelo de regressão normal contaminando a observação 19

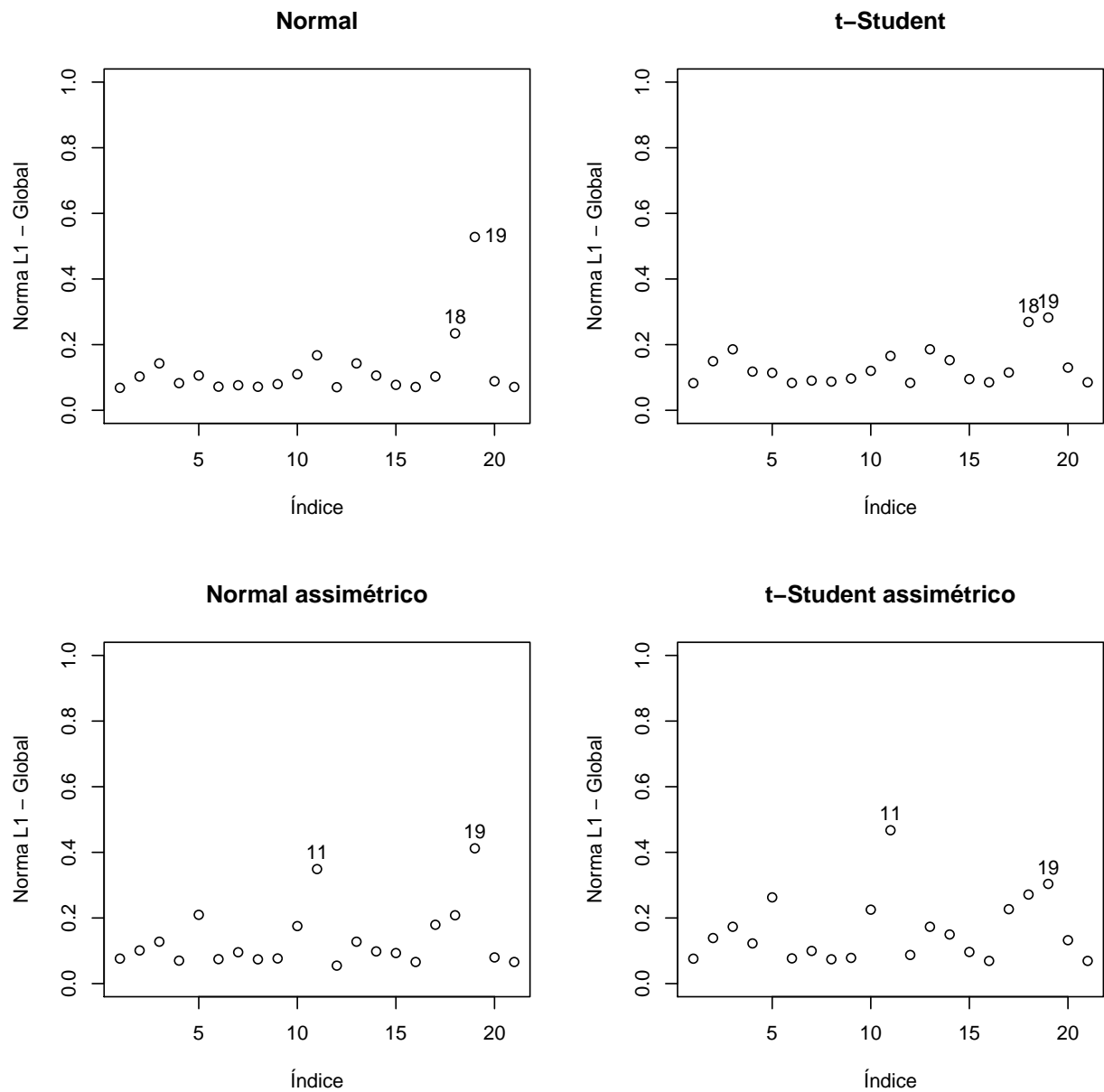


Fonte: Elaborada pelo autor.

**Figura 3.10:**  $-\log(CPO)$  dos modelos contaminando a observação 19

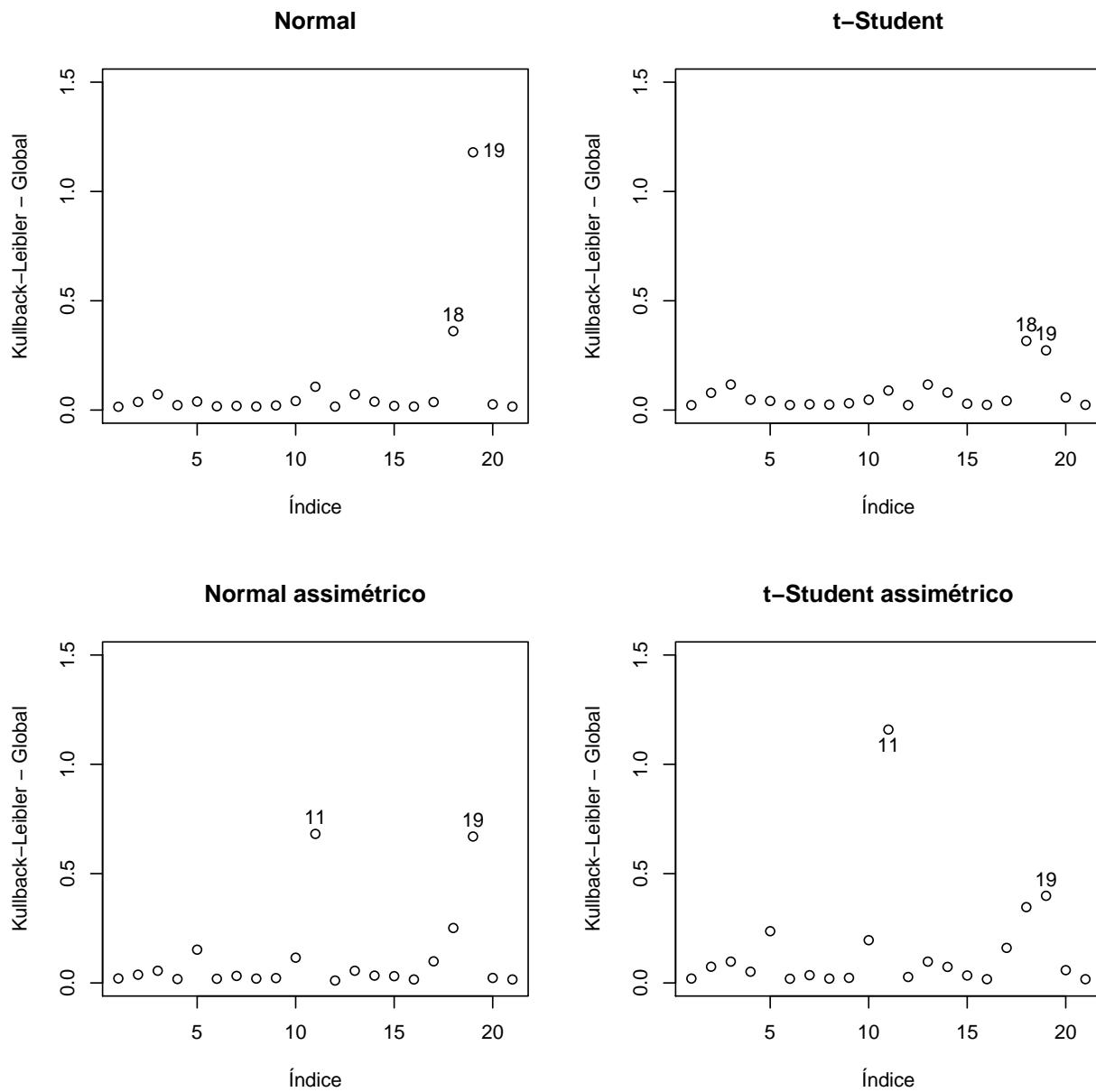
Fonte: Elaborada pelo autor.

**Figura 3.11:** *Influência global via norma  $L_1$  dos modelos contaminando a observação 19*



Fonte: Elaborada pelo autor.

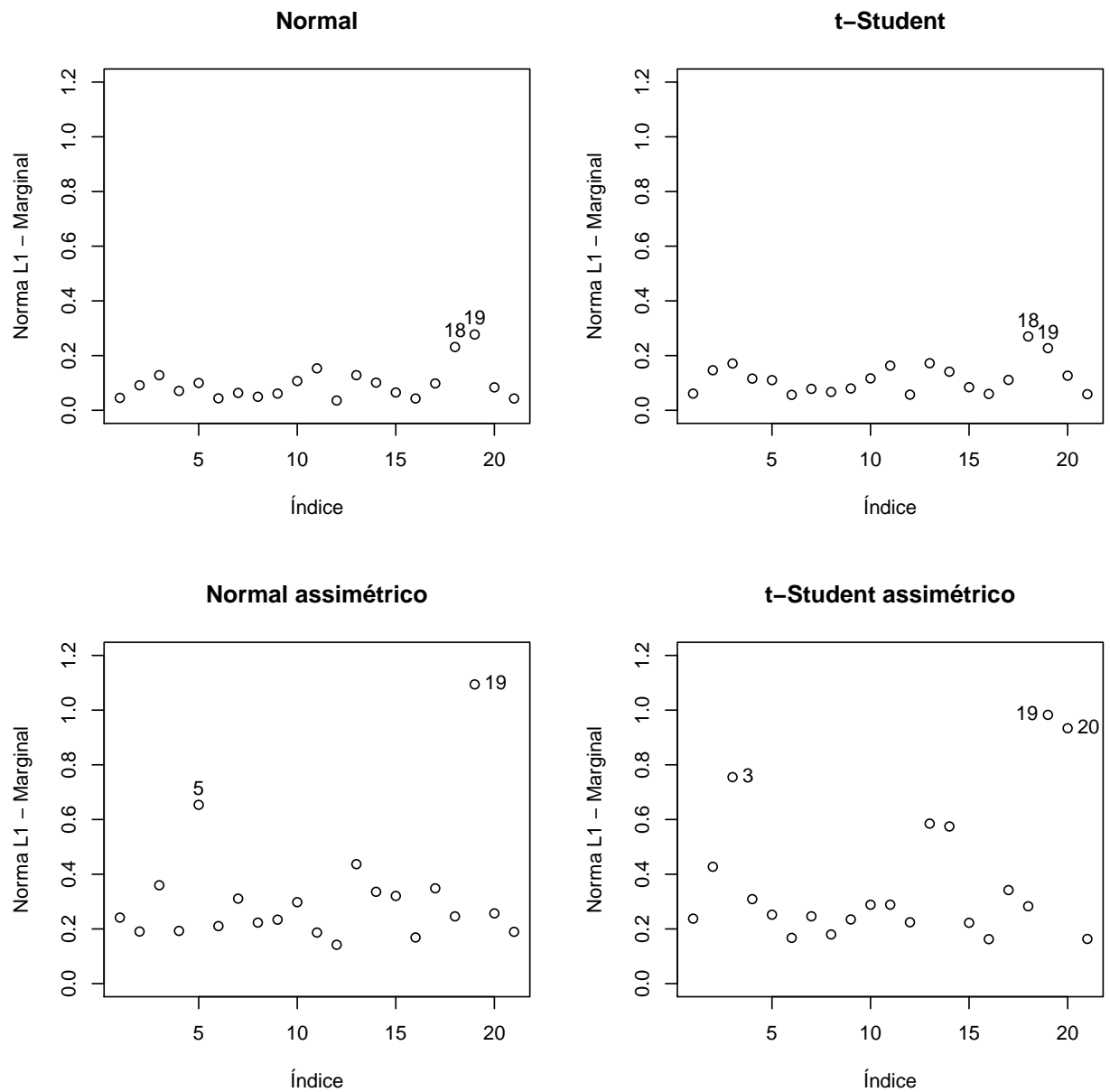
**Figura 3.12:** *Influência global via divergência de Kullback-Leibler dos modelos contaminando a observação 19*



Fonte: Elaborada pelo autor.

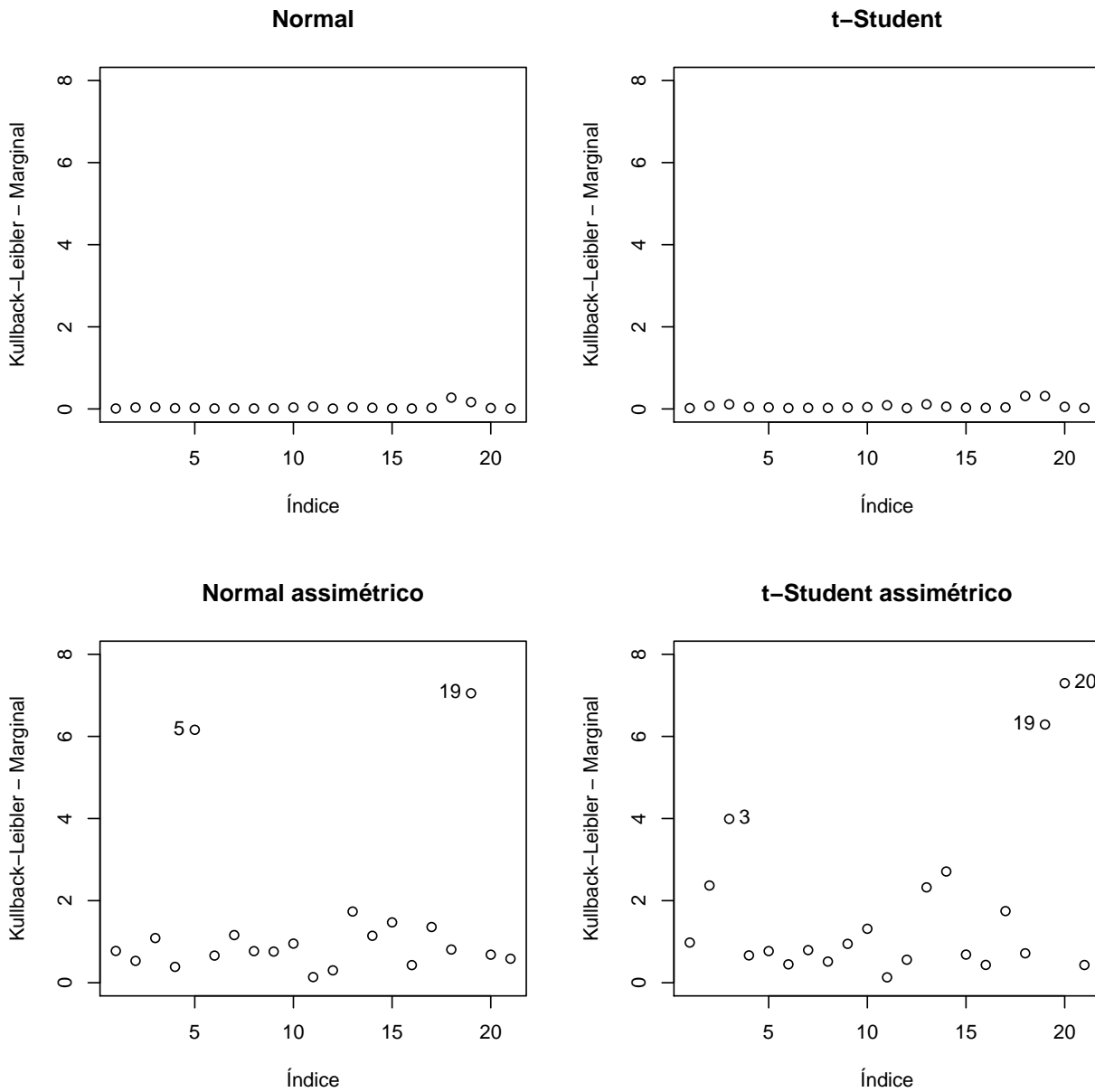


**Figura 3.13:** *Influência marginal via norma  $L_1$  dos modelos contaminando a observação 19*



Fonte: Elaborada pelo autor.

**Figura 3.14:** *Influência marginal via divergência de Kullback-Leibler dos modelos contaminando a observação 19*



Fonte: Elaborada pelo autor.

### 3.3.3 Contaminação da observação 14

Nesta aplicação contaminamos a observação 14 fazendo  $y_{14} = 60$ . Com esta contaminação, podemos observar as observações 14 e 19 com resíduo alto (figura 3.16). A existência de resíduos altos em ambos lados da distribuição justifica o melhor ajustes dos modelos simétricos frente aos modelos assimétricos (tabela 3.3).

Vemos na figura 3.17 que em quase todos os modelos a observação 14 é um pouco mais discrepante que a 19. Porém, no modelo  $t$ -assimétrico, a observação 19 é mais discrepante que a 14 por estar localizada na cauda mais leve desta distribuição. Podemos ver nas figuras 3.18 e 3.19 que o modelo  $t$ -Student é aquele em que as observações 14, 18 e 19 são menos influentes de forma global. Por outro lado, a observação 19 exerce uma grande

influência no modelo  $t$ -assimétrico.

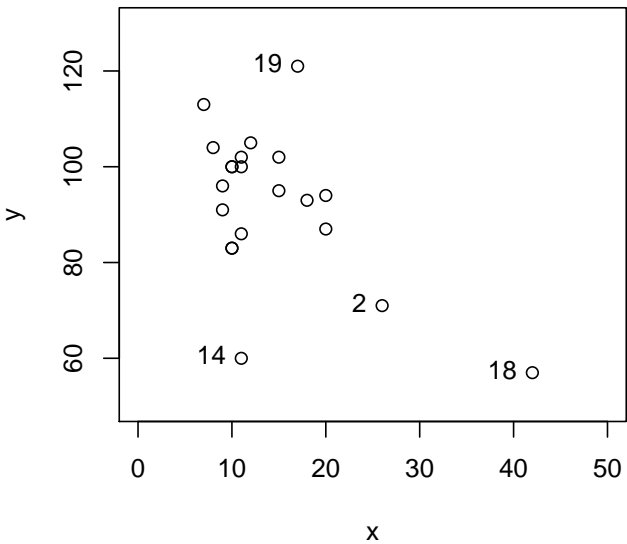
Analisando a influência marginal vemos nas figuras 3.20 e 3.21 que a observação 19 exerce uma influência menor nos modelos simétricos do que nos modelos assimétricos. Isto pode acontecer por ela estar localizada na cauda de menor peso destas distribuições assimétricas.

**Tabela 3.3:** LPML dos modelos contaminando a observação 14

Normal	$t$ -Student	Normal assimétrica	$t$ -assimétrica
-87.49	-85.31	-87.8	-90.68

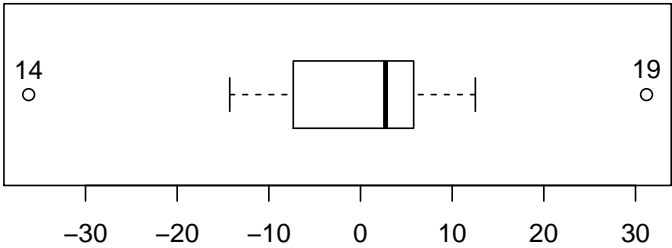
Fonte: Elaborada pelo autor.

**Figura 3.15:** Diagrama de dispersão contaminando a observação 14

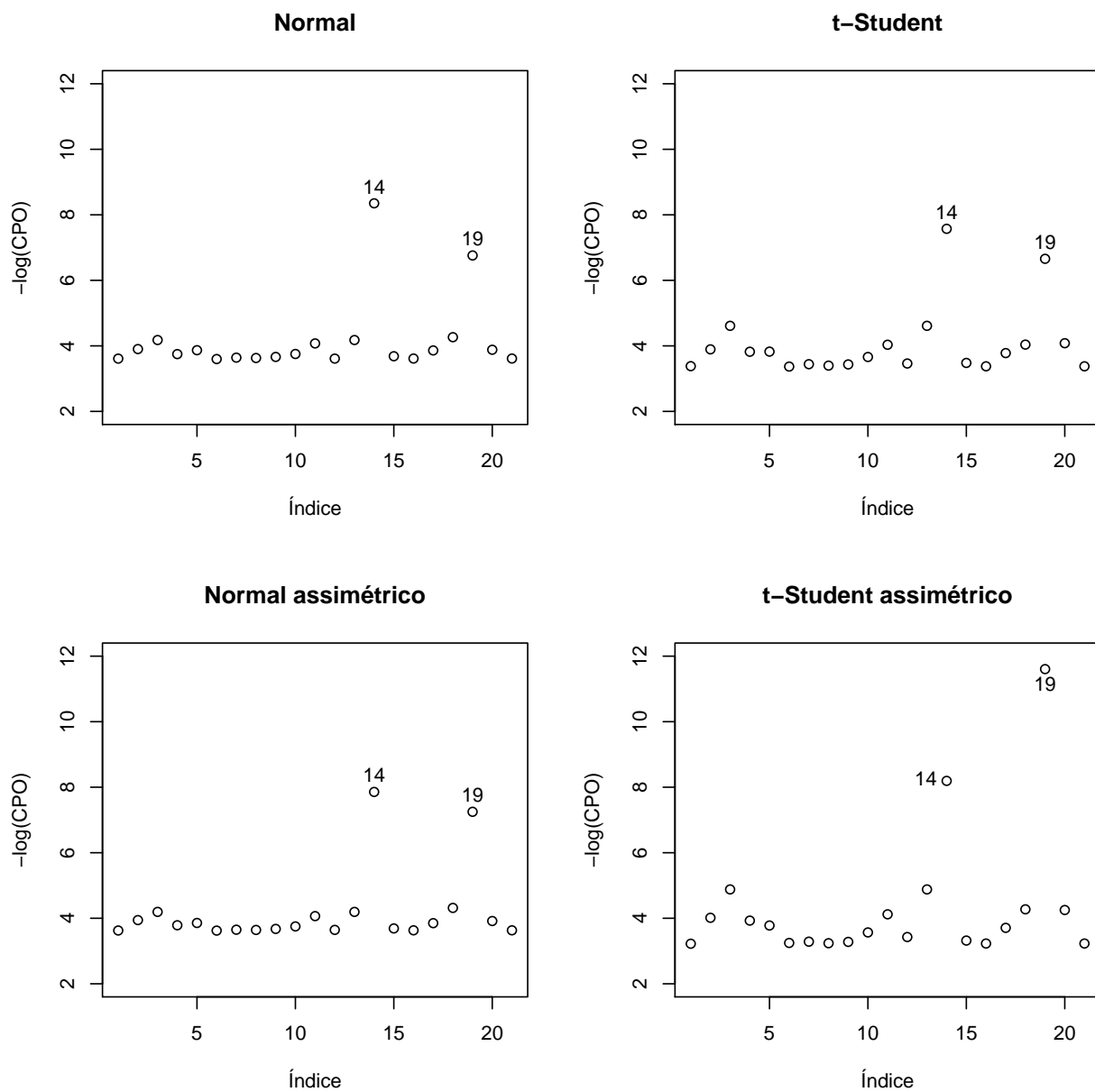


Fonte: Elaborada pelo autor.

**Figura 3.16:** Boxplot dos resíduos do modelo de regressão normal contaminando a observação 14

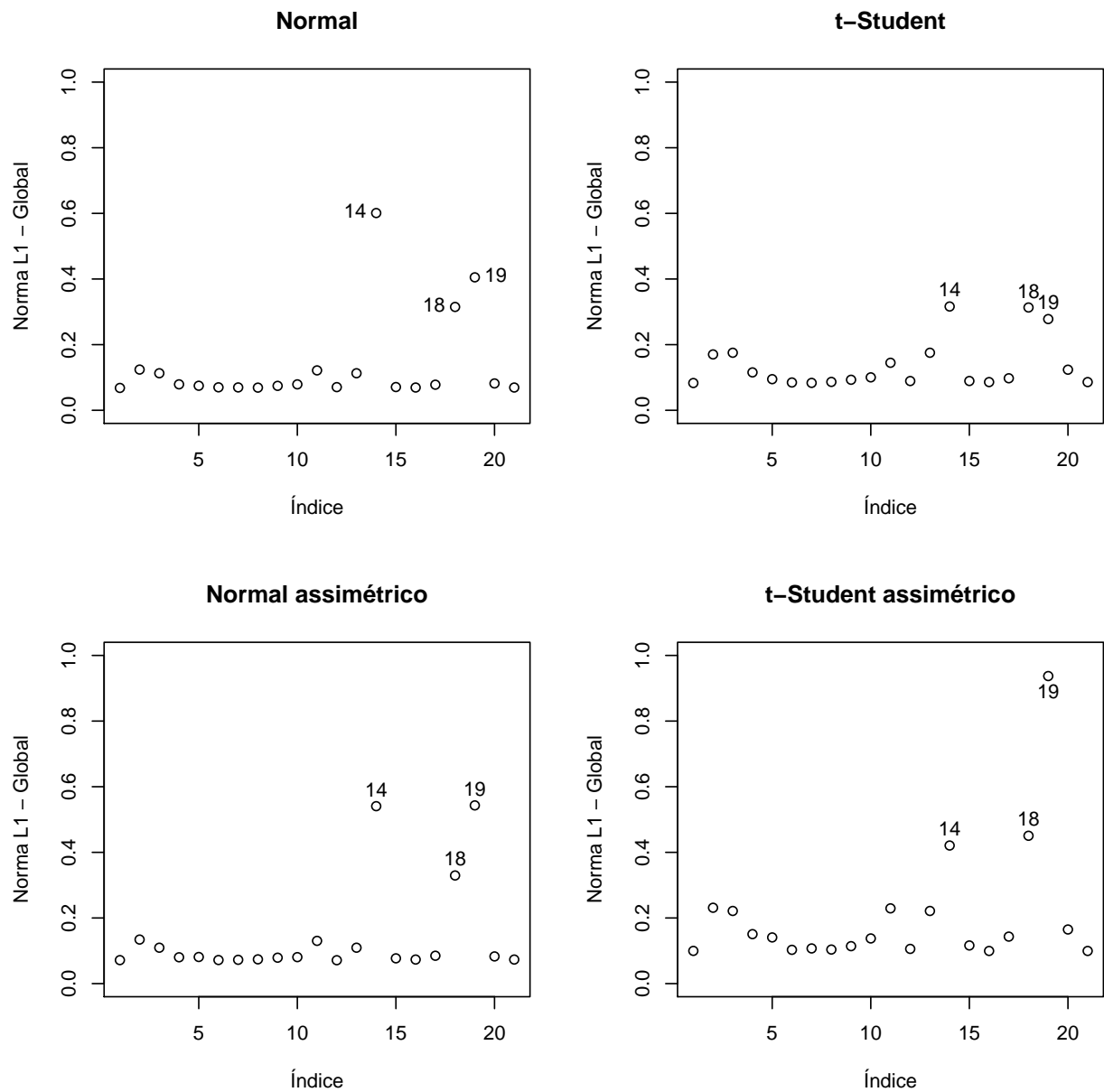


Fonte: Elaborada pelo autor.

**Figura 3.17:**  $-\log(CPO)$  dos modelos contaminando a observação 14

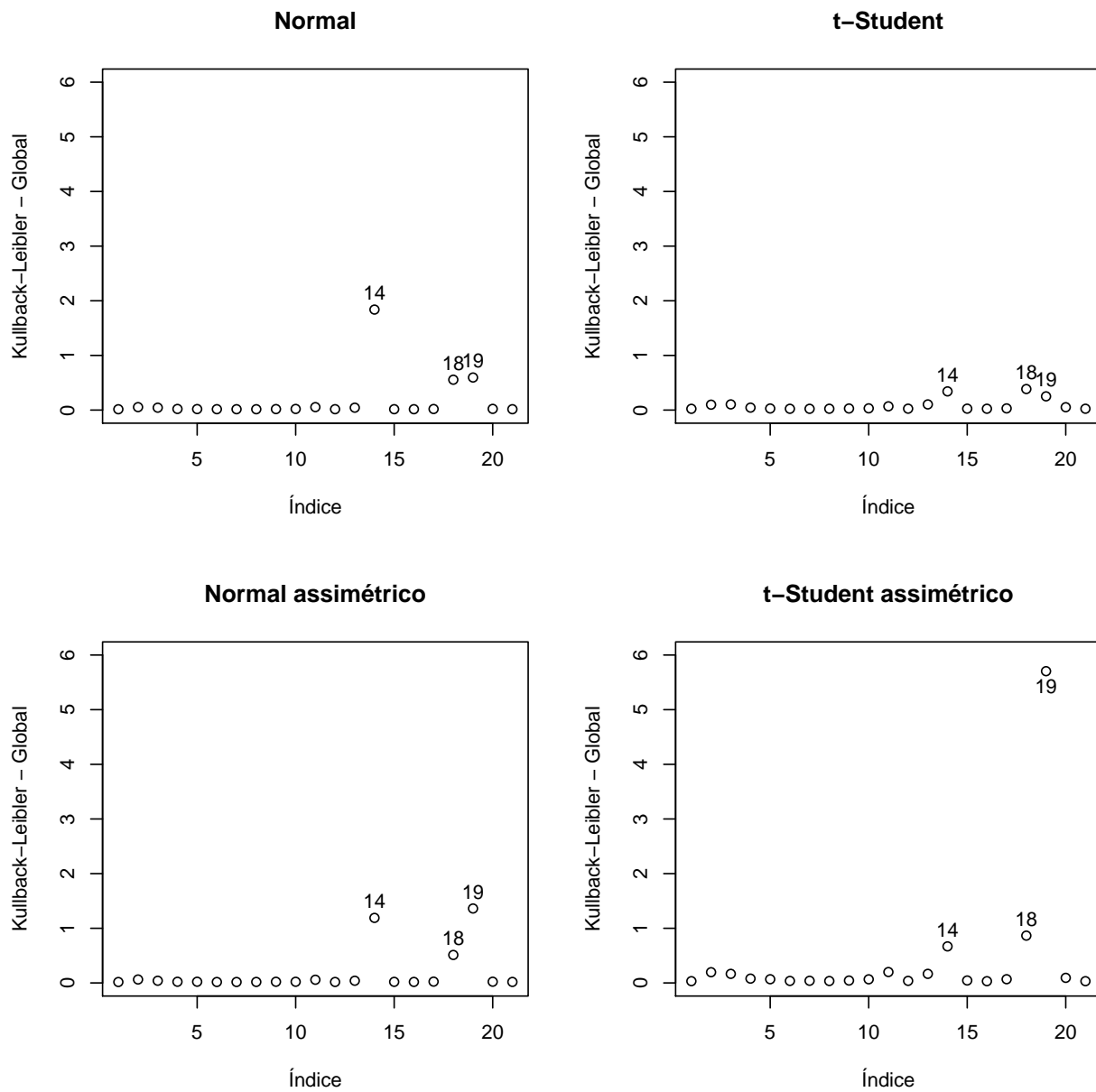
Fonte: Elaborada pelo autor.

**Figura 3.18:** *Influência global via norma  $L_1$  dos modelos contaminando a observação 14*



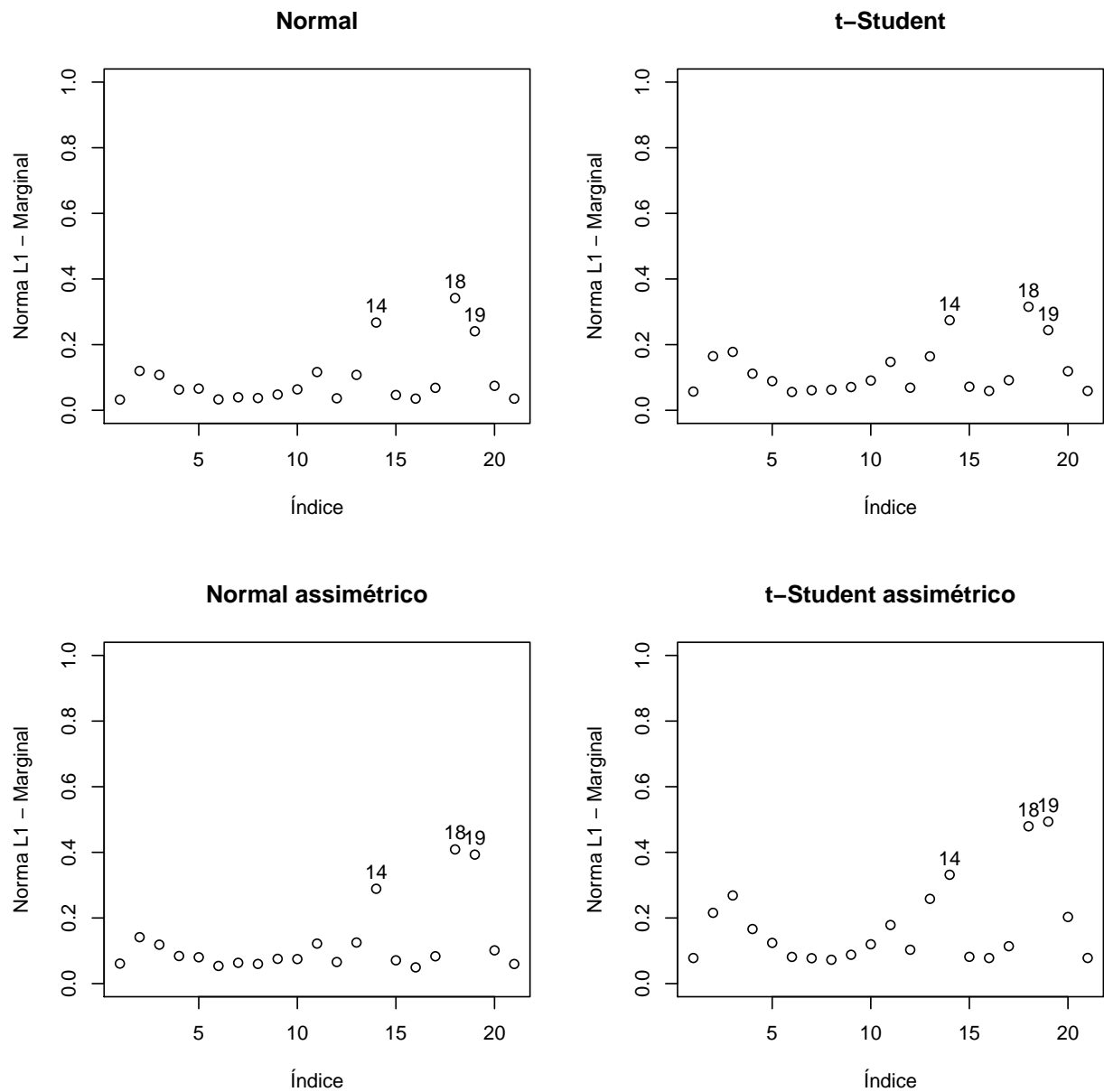
Fonte: Elaborada pelo autor.

**Figura 3.19:** *Influência global via divergência de Kullback-Leibler dos modelos contaminando a observação 14*



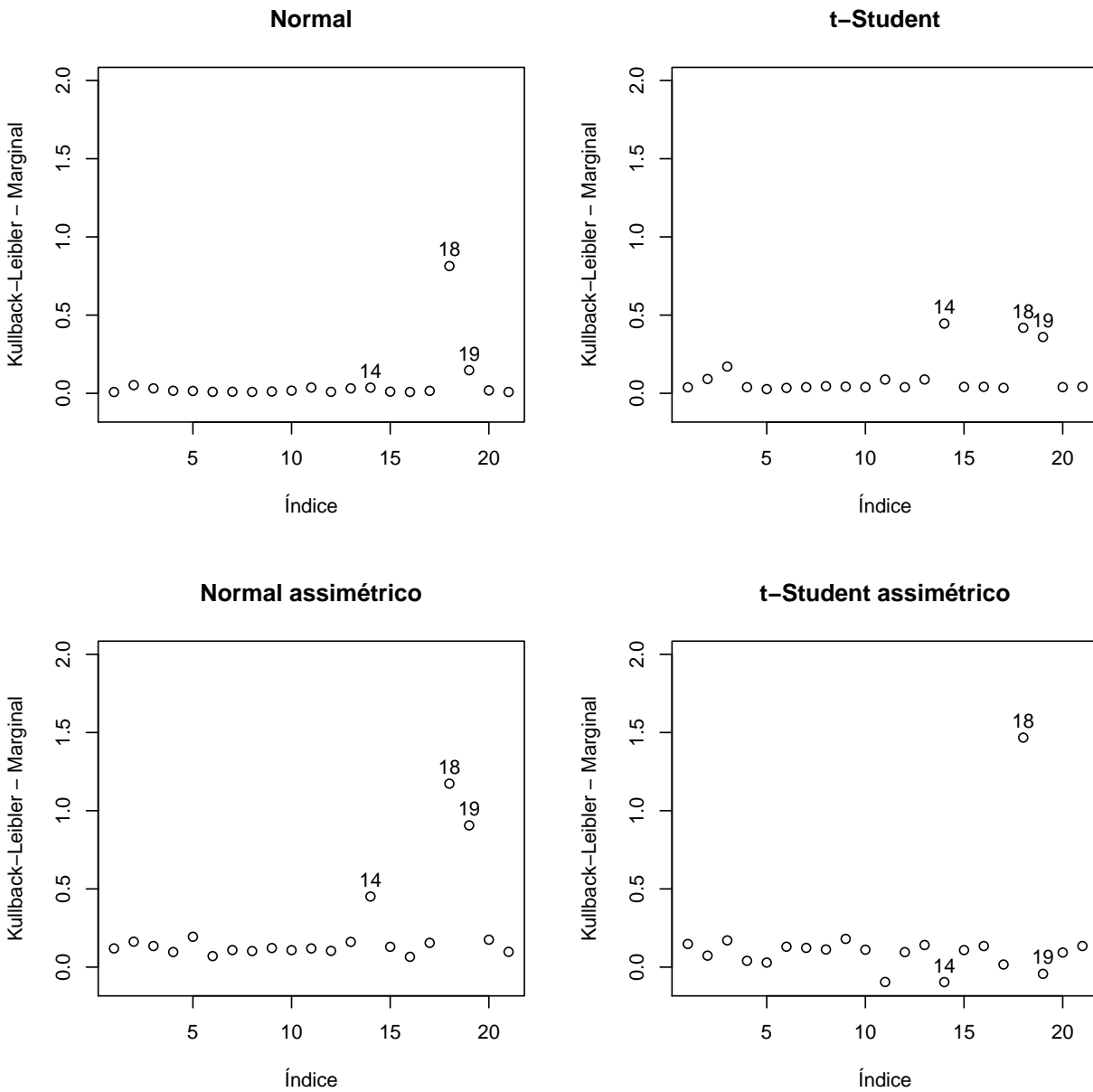
Fonte: Elaborada pelo autor.

**Figura 3.20:** *Influência marginal via norma  $L_1$  dos modelos contaminando a observação 14*



Fonte: Elaborada pelo autor.

**Figura 3.21:** *Influência marginal via divergência de Kullback-Leibler dos modelos contaminando a observação 14*



Fonte: Elaborada pelo autor.

### 3.3.4 Contaminação das observações 14 e 19

Nesta última aplicação promovemos a contaminação das observações 14 e 19 simultaneamente, fazendo  $y_{14} = y_{19} = 60$ . Assim, agora, as duas observações de maior resíduo estão na direção da assimetria da distribuição dos resíduos (figura 3.23). Neste caso, os modelos assimétricos mostram um ajuste muito superior aos simétricos (tabela 3.4).

As observações 14 e 19 são discrepantes em todos os modelos propostos (figura 3.24) sendo a observação 14, de maior resíduo, a mais discrepante. Vemos nas figuras 3.25 e 3.26 que os modelos  $t$ -Student e  $t$ -assimétrico são aqueles menos influenciados de forma global pela observação 14.

Apesar de apresentarem influência global, as observações 14 e 19 possuem uma baixa



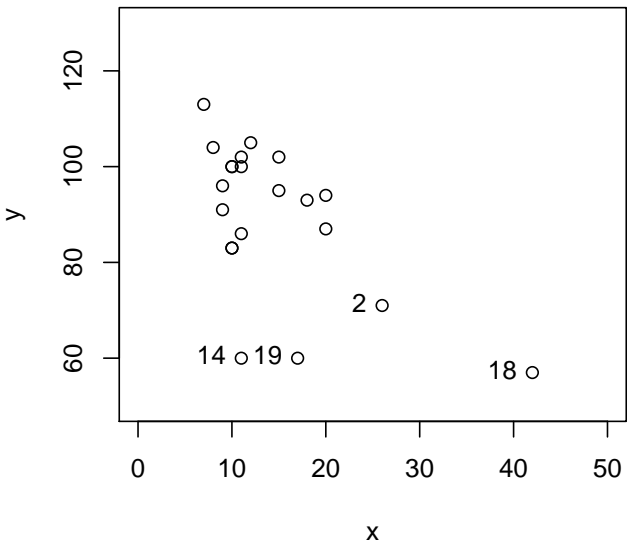
influência marginal nos modelos simétricos. Já sob os modelos assimétricos, a observação 14 aparenta exercer uma maior influência sob o modelo  $t$ -assimétrico do que sob o modelo normal assimétrico. Contudo, como os valores obtidos tanto para norma  $L_1$  quanto para a divergência Kullback-Leibler aparentam uma maior instabilidade nos modelos assimétricos do que nos modelos simétricos (a norma  $L_1$  excede o domínio  $[0, 1]$ , por exemplo), devemos ter cuidado ao considerar estes resultados.

**Tabela 3.4:** LPML dos modelos contaminando as observações 14 e 19

Normal	$t$ -Student	Normal assimétrica	$t$ -assimétrica
-86.15	-84.95	-80.92	-81.58

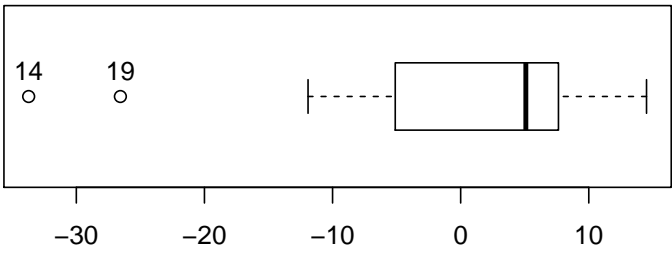
Fonte: Elaborada pelo autor.

**Figura 3.22:** Diagrama de dispersão contaminando as observações 14 e 19

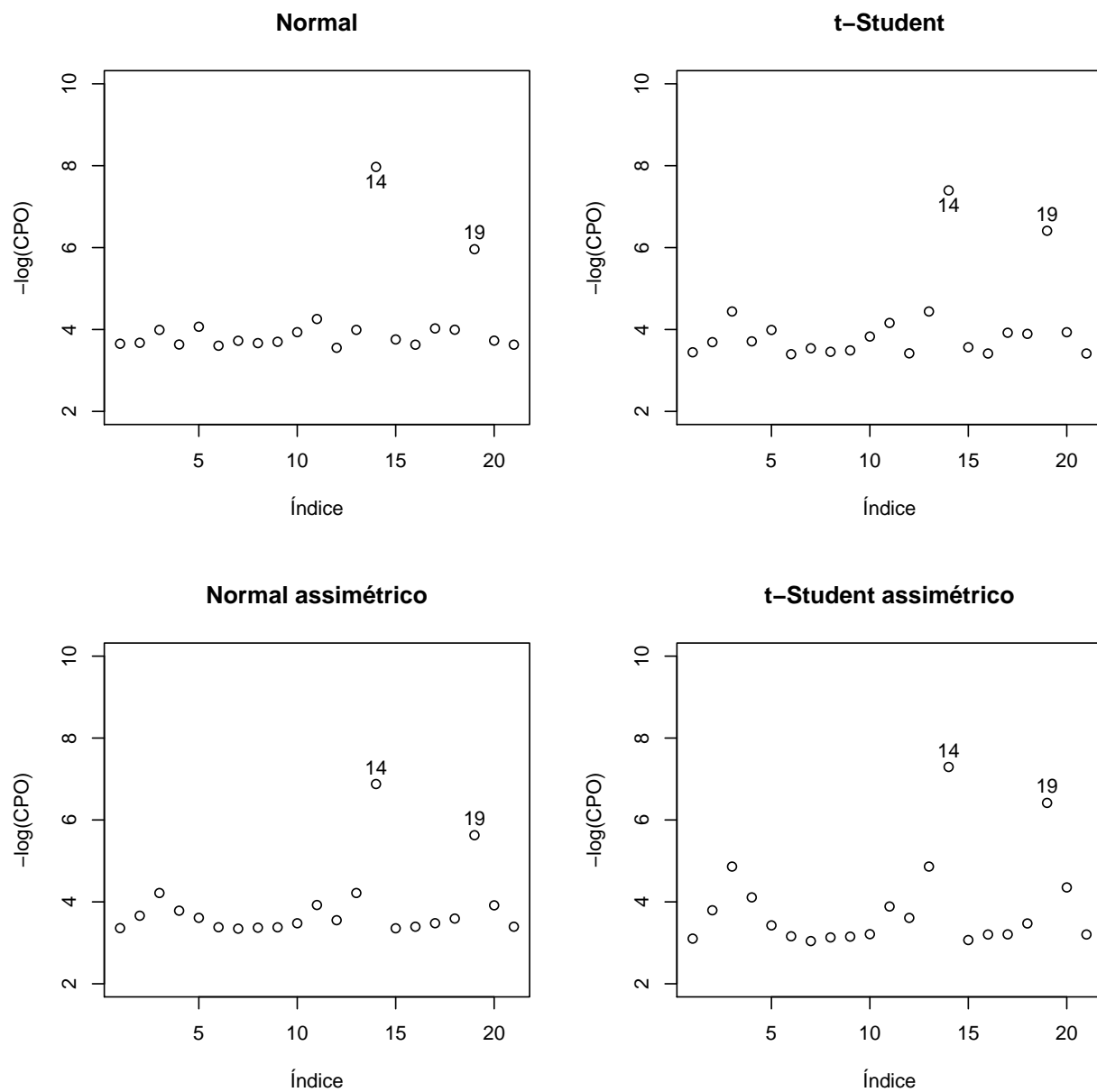


Fonte: Elaborada pelo autor.

**Figura 3.23:** Boxplot dos resíduos do modelo de regressão normal contaminando as observações 14 e 19

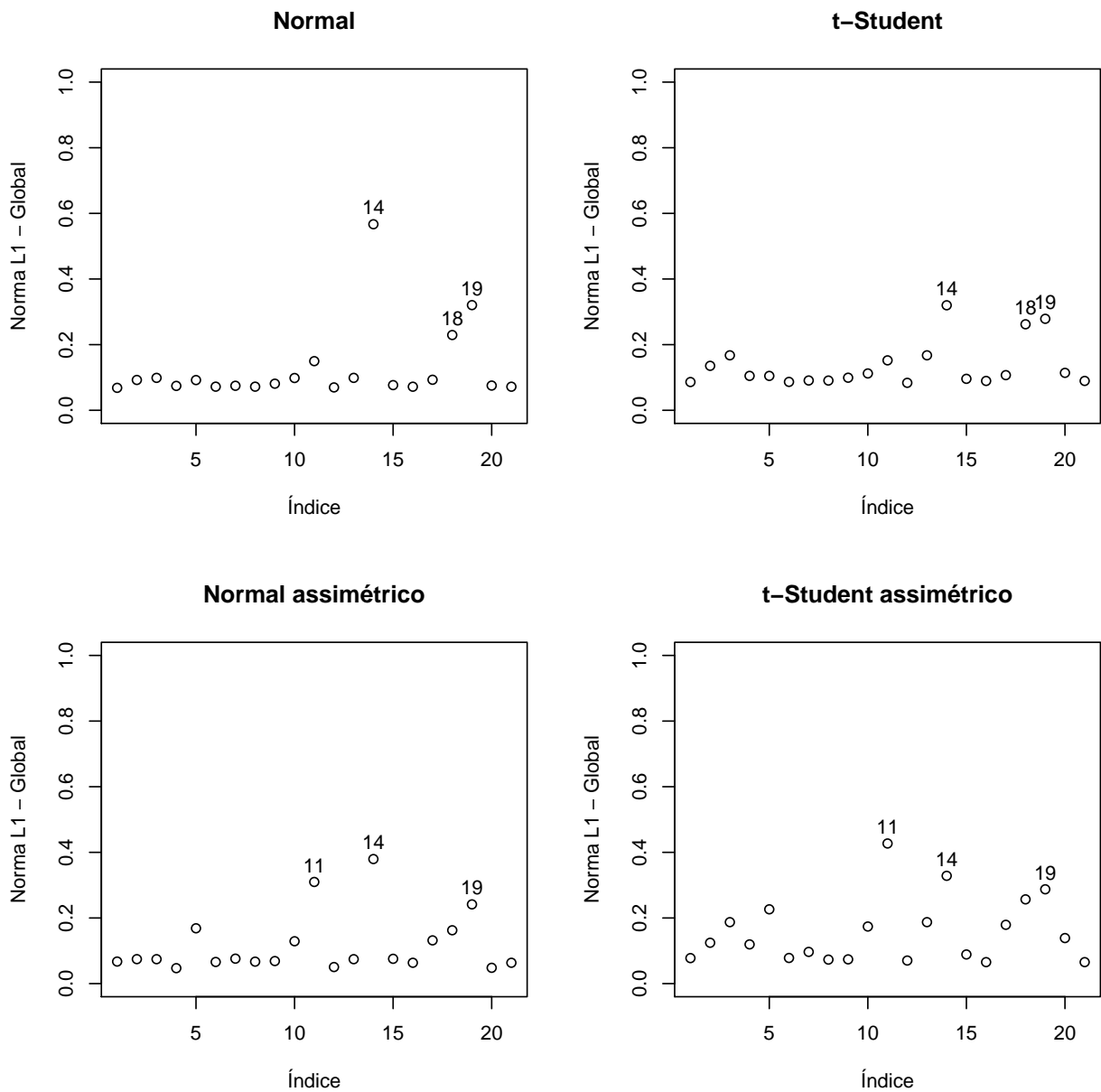


Fonte: Elaborada pelo autor.

**Figura 3.24:**  $-\log(CPO)$  dos modelos contaminando as observações 14 e 19

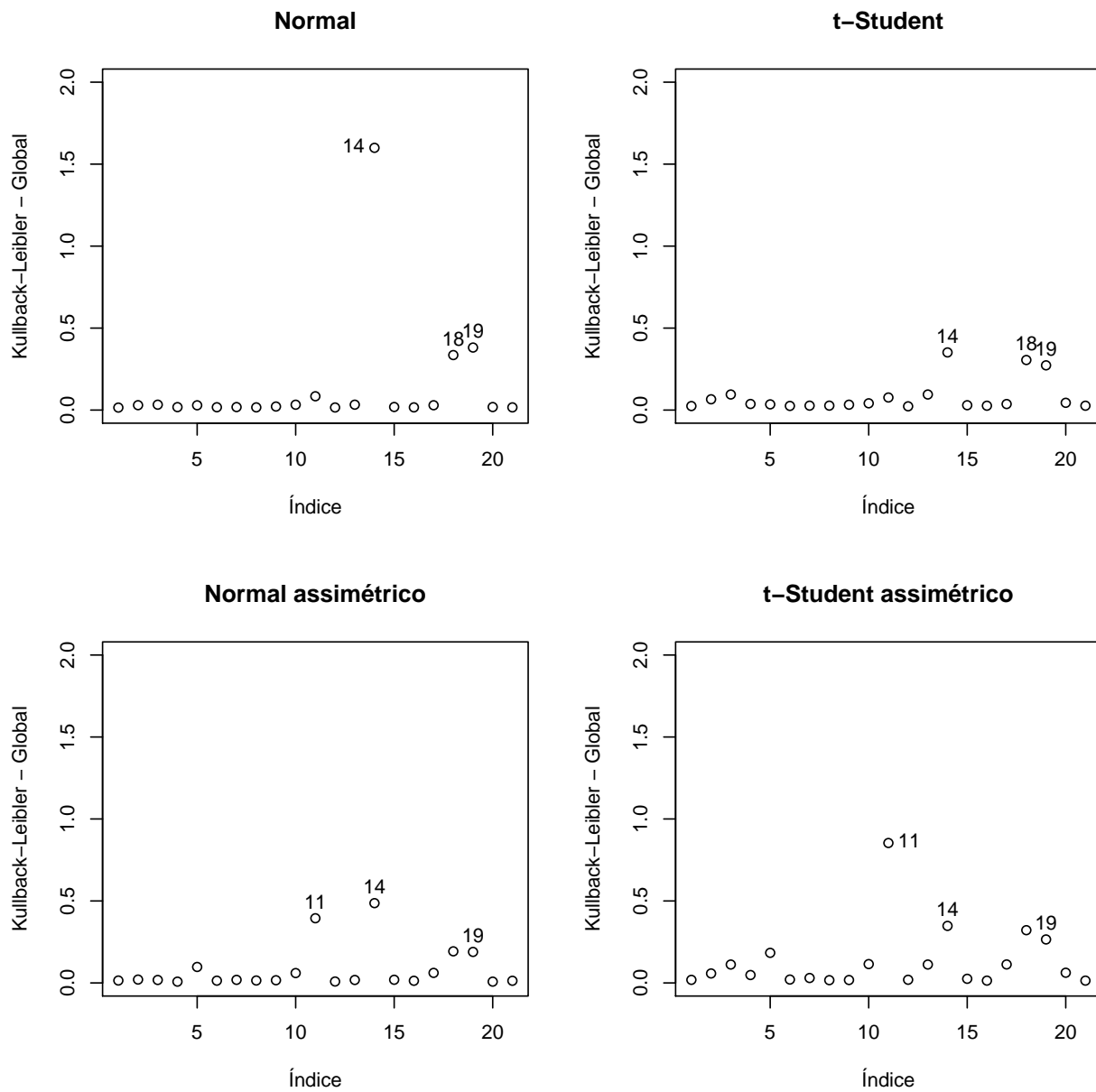
Fonte: Elaborada pelo autor.

**Figura 3.25:** *Influência global via norma  $L_1$  dos modelos contaminando as observações 14 e 19*



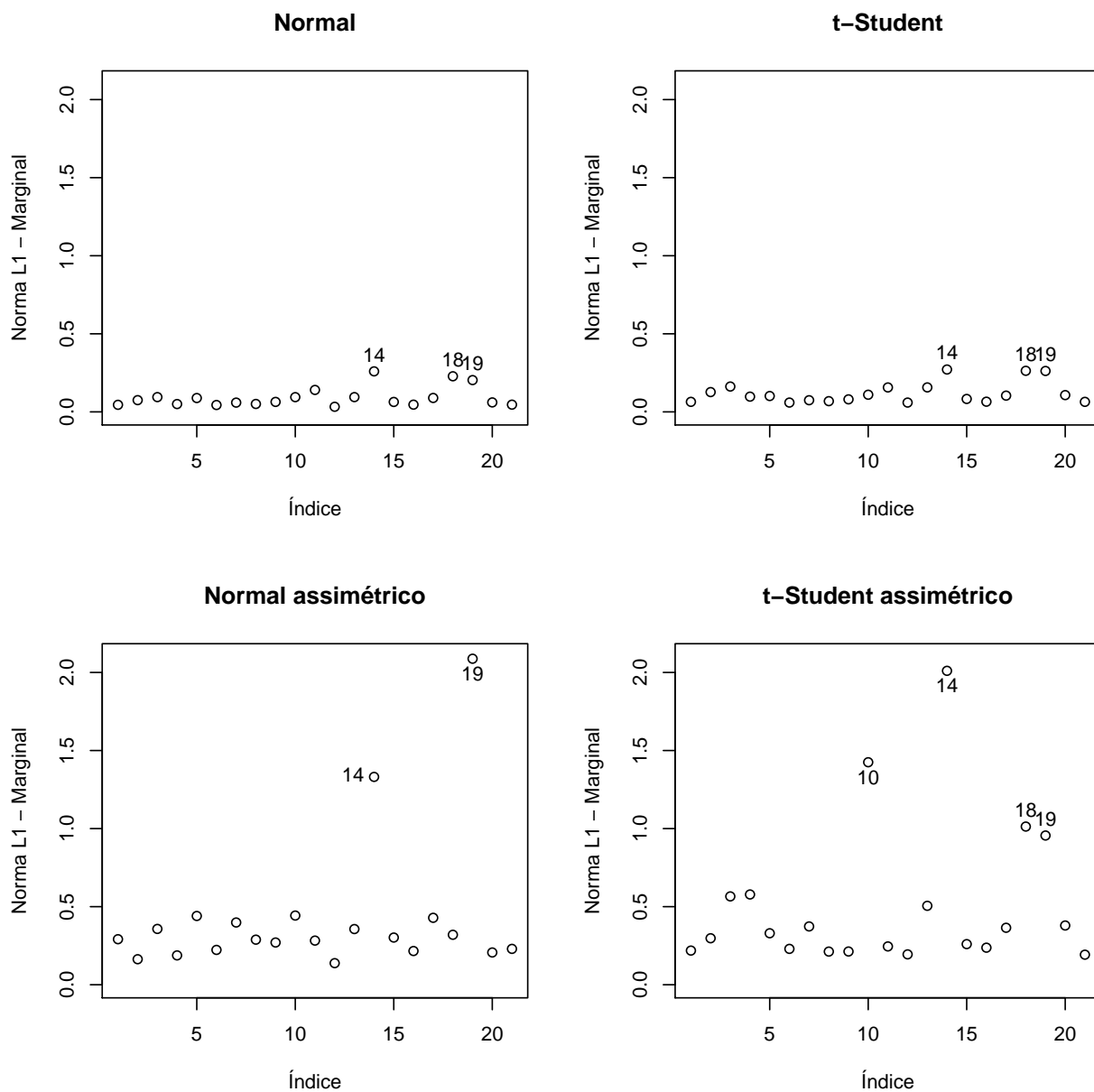
Fonte: Elaborada pelo autor.

**Figura 3.26:** *Influência global via divergência de Kullback-Leibler dos modelos contaminando as observações 14 e 19*



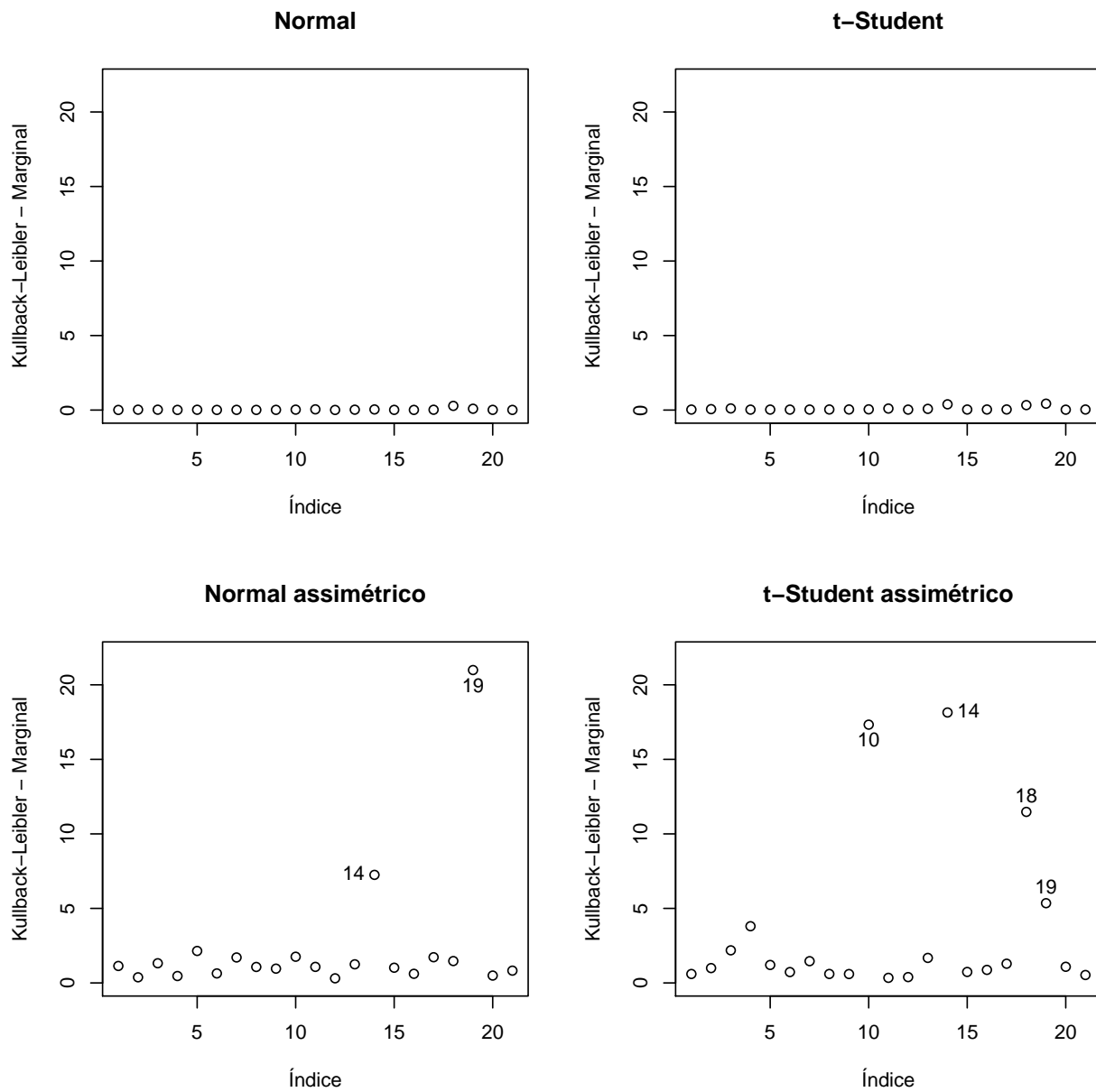
Fonte: Elaborada pelo autor.

**Figura 3.27:** *Influência marginal via norma  $L_1$  dos modelos contaminando as observações 14 e 19*



Fonte: Elaborada pelo autor.

**Figura 3.28:** *Influência marginal via divergência de Kullback-Leibler dos modelos contaminando as observações 14 e 19*



Fonte: Elaborada pelo autor.

# Capítulo 4

## Conclusões

Neste trabalho estudamos modelos de regressão linear com erros na família de distribuição  $t$ -assimétrica, apresentando como obter as estimativas dos parâmetros e analisar a influência de observações discrepantes nestas estimativas sob a abordagem bayesiana.

Apresentamos os algoritmos e os programas para obtenção das estimativas dos parâmetros destes modelos. A obtenção da amostra das distribuições *a posteriori* dos parâmetros dos modelos normal e  $t$ -Student já são bem conhecidos e a obtenção para o caso normal assimétrico foi apresentada em [Bayes \(2005\)](#). Para o modelo  $t$ -assimétrico, complementamos o trabalho de [Godoi \(2007\)](#) explicitando as condicionais completas para aplicação do algoritmo de Gibbs.

Introduzimos o *conditional predictive ordinate* (CPO) como uma estatística para identificação de observações discrepantes sob o modelo proposto. Tal estatística é de fácil obtenção uma vez que se tenha uma amostra da distribuição *a posteriori* dos parâmetros do modelo proposto e que se conheça a função densidade de probabilidade deste modelo.

Para aferir a influência das observações nas estimativas definimos inicialmente a família de divergências apresentada por [Weiss \(1996\)](#) considerando os casos particulares norma  $L_1$  e divergência de Kullback-Leibler como objetos deste estudo. Abordando tanto a influência global, quanto a marginal, generalizamos os resultados obtidos para a distribuição normal de [Weiss e Cho \(1998\)](#) para as distribuições  $t$ -Student, normal assimétrica e  $t$ -assimétrica.

Observamos que, em geral, o modelo  $t$ -Student é uma alternativa robusta ao modelo normal. A melhor aplicação do modelo  $t$ -Student é quando os resíduos possuem uma certa simetria ou que pontos discrepantes ocorram em ambas caudas da distribuição. Contudo, nas aplicações estudadas, as estimativas dos coeficientes regressores foram pouco influenciadas em ambos modelos.

O modelo  $t$ -assimétrico não é, em geral, uma alternativa robusta ao modelo normal. Vimos que a capacidade de robustificação do modelo  $t$ -assimétrico está diretamente ligada à posição do resíduo do ponto discrepante em relação a distribuição dos resíduos. Quando comparamos as versões assimétricas dos modelos normal e  $t$ -Student, supondo que ambos possuem assimetria positiva (sem perda de generalidade) a cauda direita da  $t$ -Student será mais pesada que a da normal, enquanto que a cauda esquerda da normal será mais pesada que a da  $t$ -Student. Posto isto, o modelo  $t$ -assimétrico será mais robusto que o normal assimétrico caso os pontos discrepantes estejam na cauda de maior peso.

As principais conclusões deste trabalho baseiam-se na análise de influência global pois o cálculo da influência marginal mostrou-se instável. Esta instabilidade pode estar associada as sucessivas aproximações de Monte Carlo que compõe o cálculo das estimativas.

A seguir apresentamos algumas possibilidades de pesquisas futuras relacionadas ao

diagnóstico bayesiano em modelos de regressão linear da família  $t$ -Student.

- Considerar os graus de liberdade,  $\nu$ , da distribuição desconhecidos e obter as condicionais completas explicitamente para este caso.
- Estudar a sensibilidade do cálculo das estimativas de influência de acordo com o tamanho da amostra de Monte Carlo considerada.
- Estudar outras divergências que não norma  $L_1$  e divergência de Kullback-Leibler.
- Para os modelos normal e  $t$ -Student, obter estimativas de influência marginal para as componentes do vetor  $\beta$  e para  $\sigma^2$ .
- Para os modelos normal assimétrico e  $t$ -assimétrico, obter estimativas de influência marginal para as componentes do vetor  $\beta$ , para  $\sigma^2$  e para  $\lambda$ .



# Apêndice A

## Códigos em R

### A.1 Estimação dos parâmetros dos modelos de regressão linear

#### A.1.1 Pacotes utilizados

```
library(sn)
library(truncnorm)
library(mvtnorm)
library(Rcpp)
library(MCMCpack)
library(TeachingDemos)
library(microbenchmark)
library(foreach)
library(ggplot2)
library(snow)
library(parallel)
library(LaplacesDemon)
```

#### A.1.2 Modelo Normal

```
#Estimador
modelo.Normal = function(y,x,n.sim,seed){
  tx=Sys.time()
  n.obs=length(y)
  set.seed(seed)
  X=cbind(rep(1,n.obs),x)
  p=dim(X)[2]
  b.hat=solve(t(X)%*%X)%*%t(X)%*%y
  k=n.obs-p
  s.2=as.numeric(t(y-X%*%b.hat)%*%(y-X%*%b.hat)/k)
  sigma.2.sim=rinvgamma(n.sim,k/2,k*s.2/2)
  beta.sim <-sapply(sigma.2.sim,function(x) mvrnorm(1,b.hat,
    x*solve(t(X)%*%X)))
  b0 = beta.sim[1,]
```

```

b1= beta.sim[2,]
print(Sys.time()-tx)
output = list("beta" = beta.sim, "sigma" = sqrt(sigma.2.
  sim))
return(output)
invisible()
gc()
}

# Teste
set.seed(345)
n.teste = 100000
x=rnorm(n.teste)
y = 2 + 3*x + rnorm(n.teste,0,1.25)
fit.N = modelo.Normal(y,x,1000,123)

mean(fit.N$beta[1,])
mean(fit.N$beta[2,])
mean(fit.N$sigma)

```

### A.1.3 Modelo *t*-Student

```

#Estimador
gibbs.T = function(y,x,n.sim,v,seed){
  tx=Sys.time()
  set.seed(seed)
  b.ini = c(0,0)
  sig.2.ini = 1
  n.obs=length(x)
  X=cbind(rep(1,n.obs),x)
  B = rbind(rep(b.ini[1],n.sim),rep(b.ini[2],n.sim))
  s=rep(sig.2.ini,n.sim)
  L = matrix(rep(1,n.sim*n.obs),nrow=n.obs)

  for(k in 1:(n.sim-1)){

    L.aux=function(x,y){rgamma(1, shape = x, rate = y)}
    L[,k+1]=mapply(L.aux,v/2 + 1/2,(y-B[1,k]-B[2,k]*X[,2])
      ^2/(2*s[k]) + v/2)

    Sig = diag(L[,k+1])
    B.hat= solve(t(X)%*%Sig%*%X)%*%t(X)%*%Sig%*%y
    V.b = s[k]*solve(t(X)%*%Sig%*%X)
    B[,k+1] = rmvnorm(1,B.hat,V.b)
    s[k+1] = rinvgamma(1, n.obs/2, 1/2 * t(y-X%*%B[,k+1])%*%
      Sig%*%(y-X%*%B[,k+1]))
  }
}

```

```

output = list("beta" = B, "sigma" = sqrt(s), "lambda"=L)
print(Sys.time()-tx)
return(output)
invisible()
gc()
}

# Teste
set.seed(345)
n.teste = 500
x=rnorm(n.teste)
y = 2 + 3*x + 1.25*rt(n.teste,5)
fit.T = gibbs.T(y,x,1000,5,123)

mean(fit.T$beta[1,-c(1:50)])
mean(fit.T$beta[2,-c(1:50)])
mean(fit.T$sigma[-c(1:50)])

```

#### A.1.4 Modelo Normal Assimétrico

```

#Estimador
gibbs.SN = function(y,x,n.sim,sig.t.2,k,seed){
  tx=Sys.time()
  b.ini = c(0,0)
  sig.2.ini = 1
  set.seed(seed)
  n.obs=length(y)
  X=cbind(rep(1,n.obs),x)
  B = rbind(rep(b.ini[1],n.sim),rep(b.ini[2],n.sim))
  eta=rep(1,n.sim)
  tau=rep(1,n.sim)
  w=rep(1,n.sim)
  u=matrix(rep(1,n.sim*n.obs),ncol=n.sim)

  for(i in 2:n.sim){

    inv.Xt.X = chol2inv(chol(t(X)%*%X))
    B.hat= inv.Xt.X %*% t(X)%*%(y-eta[i-1]*u[,i-1])
    V.b = inv.Xt.X*tau[i-1]^2
    B[,i] = rmvnorm(1,B.hat,V.b)

    aux.1=(t(u[,i-1])%*%u[,i-1]+w[i-1]/sig.t.2)
    eta[i]=rnorm(1,mean=t(u[,i-1])%*%(y-X%*%B[,i])/aux.1,sd=
      tau[i-1]/sqrt(aux.1))

    tau[i]=1/sqrt(rgamma(1,shape=(n.obs+1)/2,rate=1/2*(t(y-X
      %*%B[,i]-eta[i]*u[,i-1])%*%(y-X%*%B[,i]-eta[i]*u[,i-1])

```

```

-1]))+eta[i]^2*w[i-1]/sig.t.2)))

w[i]=rgamma(1,shape=(k+1)/2,rate=1/2*(k+eta[i]^2/(tau[i]^2*sig.t.2)))

u.aux=function(x,y){rtruncnorm(1, a=0, b=Inf, mean = x,
  sd = y)}
u[,i]=mapply(u.aux,eta[i]*(y-X%%B[,i])/(eta[i]^2+tau[i]^2),tau[i]/sqrt(eta[i]^2+tau[i]^2))
}

sigma=sqrt(eta^2+tau^2)
lambda=eta/tau
B.correct = B
B.correct[1,]=B[1,]+sqrt(2/pi)*sigma*lambda/sqrt(1+lambda^2)
output = list(
  "beta" = B,
  "beta.correct" = B.correct,
  "sigma" = sigma,
  "tau" = tau,
  "lambda"=lambda,
  "eta"=eta,
  "U"=u
)
print(Sys.time()-tx)
return(output)
invisible()
gc()
}

# Teste
set.seed(123)
n.teste = 1000
x=rnorm(n.teste)
y = 2 + 3*x + 1.25*rsn(n.teste, xi=0, omega=0.5, alpha=5)
fit.SN = gibbs.SN(y,x,300,pi^2/4,1/2,123)
mean(fit.SN$beta[1,-c(1:50)])
mean(fit.SN$beta[2,-c(1:50)])
mean(fit.SN$sigma[-c(1:50)])
mean(fit.SN$lambda[-c(1:50)])

```

### A.1.5 Modelo *t*-assimétrico

```

#Estimador
gibbs.ST = function(y,x,n.sim,v,sig.t.2,k,seed){
  tx=Sys.time()

```

```

set.seed(seed)
n.obs=length(y)
b.ini = c(0,0)
sig.2.ini = 1
X=cbind(rep(1,n.obs),x)
B = rbind(rep(b.ini[1],n.sim),rep(b.ini[2],n.sim))
eta=rep(1,n.sim)
tau=rep(1,n.sim)
w=rep(1,n.sim)
u=matrix(rep(1,n.sim*n.obs),ncol=n.sim)
t=matrix(rep(1,n.sim*n.obs),ncol=n.sim)

for(i in 2:n.sim){

  inv.X.u.X = chol2inv(chol(t(X*u[,i-1])%*%X))

  B.hat= inv.X.u.X %*% t(X*u[,i-1])%*%(y-eta[i-1]*t[,i-1])
  V.b = inv.X.u.X*tau[i-1]^2

  B[,i] = rmvnorm(1,B.hat,V.b)

  aux.1=(t(t[,i-1])%*%diag(u[,i-1])%*%t[,i-1]+w[i-1]/sig.t
    .2)
  eta[i]=rnorm(1,mean=t(t[,i-1])%*%diag(u[,i-1])%*%(y-X%*%
    B[,i])/aux.1,sd=tau[i-1]/sqrt(aux.1))

  tau[i]=1/sqrt(rgamma(1,shape=n.obs/2+1/2,rate=1/2*(t(y-X
    %*%B[,i]-eta[i]*t[,i-1])%*%diag(u[,i-1])%*%(y-X%*%B[,
    i]-eta[i]*t[,i-1))+eta[i]^2*w[i-1]/sig.t.2)))

  w[i]=rgamma(1,shape=(k+1)/2,rate=1/2*(k+eta[i]^2/(tau[i]
    ^2*sig.t.2)))

  u.aux=function(x,y){rgamma(1, shape = x, rate = y)}
  u[,i]=mapply(u.aux,(v+1)/2,1/2*(t[,i-1]^2+v+(1/tau[i]^2)
    *(y-X%*%B[,i]-eta[i]*t[,i-1])^2))

  t.aux=function(x,y){rtruncnorm(1, a=0, b=Inf, mean = x,
    sd = y)}
  t[,i]=mapply(t.aux,eta[i]*(y-X%*%B[,i])/(eta[i]^2+tau[i]
    ^2),tau[i]/sqrt(u[,i]*(eta[i]^2+tau[i]^2)))
}

sigma=sqrt(eta^2+tau^2)
lambda=eta/tau
B.correct = B
B.correct[1,]=B[1,]+sqrt(v/pi)*sigma*lambda/sqrt(1+lambda
  ^2)*gamma((v-1)/2)/gamma(v/2)

```

```

output = list(
  "beta" = B,
  "beta.correct" = B.correct,
  "sigma" = sigma,
  "tau" = tau,
  "lambda"=lambda,
  "eta"=eta,
  "T"=t,
  "U"=u
)
print(Sys.time()-tx)
return(output)
invisible()
gc()
}

# Teste
set.seed(123)
n.teste = 1000
x=rnorm(n.teste)
y = 2 + 3*x + 1.25*sn::rst(n.teste, xi=0, omega=0.5, alpha
  =5, nu=5)
fit.ST = gibbs.ST(y,x, 300, 5, pi^2/4, 1/2, 123)
mean(fit.ST$beta[1,-c(1:50)])
mean(fit.ST$beta[2,-c(1:50)])
mean(fit.ST$sigma[-c(1:50)])

```

## A.2 *Conditional Predictive Ordinate* e LPML

### A.2.1 Modelo Normal

```

n.obs=length(y)
CPO.N = c()
for(i in 1:n.obs){
  CPO.N[i]=1/mean(1/sn::dst(y[i],xi=b0.N+b1.N*x[i],omega=sig
    .N,alpha=0,nu=Inf))
}

LPML.N = sum(log(CPO.N))

```

### A.2.2 Modelo *t*-Student

```

CPO.t = c()
for(i in 1:n.obs){
  CPO.t[i]=1/mean(1/sn::dst(y[i],xi=b0.t+b1.t*x[i],omega=sig
    .t,alpha=0,nu=5))
}

```

```
}

LPML.t = sum(log(CPO.t))
```

### A.2.3 Modelo Normal Assimétrico

```
CPO.SN = c()
for(i in 1:n.obs){
  CPO.SN[i]=1/mean(mapply(function(x1,x2,x3,x4,x5) 1/sn::dst
    (x1,xi=x2,omega=x3,alpha=x4,nu=x5),y[i],b0.SN+b1.SN*x[i]
    ),sig.SN,lambda.SN,Inf))
}

LPML.SN = sum(log(CPO.SN))
```

### A.2.4 Modelo *t*-assimétrico

```
CPO.St = c()
for(i in 1:n.obs){
  CPO.St[i]=1/mean(mapply(function(x1,x2,x3,x4,x5) 1/sn::dst
    (x1,xi=x2,omega=x3,alpha=x4,nu=x5),y[i],b0.St+b1.St*x[i]
    ),sig.St,lambda.St,5))
}

LPML.St = sum(log(CPO.St))
```

## A.3 Medidas de influência global

### A.3.1 Modelo Normal

```
h.N.global = matrix(rep(-1,length(y)*length(sig.N)),ncol=
  length(sig.N))
for(i in 1:length(y)){
  for(l in 1:length(sig.N)){
    h.N.global[i,l] = CPO.N[i]/sn::dst(y[i],xi=b0.N[l]+b1.N[
      l]*x[i],omega=sig.N[l],alpha=0,nu=Inf)
  }
}

#Norma L1
L1.N.global = rowMeans(abs(h.N.global-1)/2)
#Kullback-Leibler
K1.N.global = rowMeans(h.N.global*log(h.N.global))
```

### A.3.2 Modelo *t*-Student

```

L.sim = length(b0.t)

h.t.global = matrix(rep(-1,n.obs*L.sim),ncol=L.sim)
for(i in 1:n.obs){
  for(l in 1:L.sim){
    h.t.global[i,l] = CPO.t[i]/sn::dst(y[i],xi=b0.t[l]+b1.t[
      l]*x[i],omega=sig.t[l],alpha=0,nu=5)
  }
}

#Norma L1
L1.t.global = rowMeans(abs(h.t.global -1)/2))
#Kullback-Leibler
K1.t.global = rowMeans(h.t.global *log(h.t.global ))

```

### A.3.3 Modelo Normal Assimétrico

```

L.sim = length(b0.SN)
h.SN.global = matrix(rep(-1,n.obs*L.sim),ncol=L.sim)
for(i in 1:n.obs){
  for(l in 1:L.sim){
    h.SN.global[i,l] = CPO.SN[i]/sn::dst(y[i],xi=b0.SN[l]+b1
      .SN[l]*x[i],omega=sig.SN[l],alpha=lambda.SN[l],nu=Inf
    )
  }
}

#Norma L1
L1.SN.global = rowMeans(abs(h.SN.global-1)/2))
#Kullback-Leibler
K1.SN.global=rowMeans(h.SN.global*log(h.SN.global))

```

### A.3.4 Modelo *t*-assimétrico

```

h.St.global = matrix(rep(-1,n.obs*L.sim),ncol=L.sim)
for(i in 1:n.obs){
  for(l in 1:L.sim){
    h.St.global[i,l] = CPO.St[i]/sn::dst(y[i],xi=b0.St[l]+b1
      .St[l]*x[i],omega=sig.St[l],alpha=lambda.St[l],nu=5)
  }
}

#Norma L1
L1.St.global = rowMeans(abs(h.St.global-1)/2))

```



```
#Kullback-Leibler
K1.St.global=rowMeans(h.St.global*log(h.St.global))
```

## A.4 Medidas de influência marginal

### A.4.1 Modelo Normal

```
h.N.marg = matrix(rep(-1,n.obs*length(b0.N)),ncol=length(b0.N))
for(i in 1:n.obs){
  for(l in 1:length(b0.N)){
    h.N.marg[i,l] = CPO.N[i]/sn::dst(y[i],xi=b0.N[l]+b1.N[l]*x[i],omega=sqrt((t(y-b0.N[l]-b1.N[l]*x)%*(y-b0.N[l]-b1.N[l]*x)-(y[i]-b0.N[l]-b1.N[l]*x[i])^2)/(n.obs-1)),alpha=0,nu=n.obs-1)
  }
}

#Norma L1
L1.N.marg = rowMeans(abs(h.N.marg-1)/2)
#Kullback-Leibler
K1.N.marg = rowMeans(h.N.marg*log(h.N.marg))
```

### A.4.2 Modelo *t*-Student

```
f.t.inf = function(i,y,x,beta,n.sim,nu){
  n.obs=length(y)
  u=matrix(rgamma(n.sim*n.obs,shape=(nu+1)/2,rate=nu/2),ncol=n.sim)
  X=cbind(rep(1,length(x)),x)
  num=gamma(n.obs/2)*sum((apply(u,2,function(x) t(y-X%*beta)%*diag(x)%*(y-X%*beta))^-0.5)^(n.obs))
  den=gamma((n.obs-1)/2)*sqrt(pi)*sum(1/sqrt(u[i,])*(apply(u[-i,],2,function(x) t(y[-i]-X[-i,]%*beta)%*diag(x)%*(y[-i]-X[-i,]%*beta))^-0.5)^(n.obs-1))
  return(num/den)
}

beta.sim.t = cbind(b0.t,b1.t)
nu=5
n.sim=1000
n.obs=length(y)
cl=makeCluster(8)
clusterSetRNGStream(cl, iseed=12345)
clusterExport(cl,c("beta.sim.t","y","f.t.inf","x","n.sim","nu"))
```

```

tx=Sys.time()
par.f.t=sapply(1:n.obs, function(arg2) parApply(cl,beta.sim.
  t,1,function(arg1) f.t.inf(arg2,y,x,arg1,n.sim,nu)))
Sys.time()-tx
stopCluster(cl)
L.sim = amostra

h.t.marg = matrix(rep(-1,n.obs*L.sim),ncol=L.sim)
for(i in 1:n.obs){
  for(l in 1:L.sim){
    h.t.marg[i,l] = CPO.t[i]/par.f.t[l,i]
  }
}

#Norma L1
L1.t.marg = rowMeans(abs(h.t.marg-1)/2)
#Kullback-Leibler
K1.t.marg = rowMeans(h.t.marg*log(h.t.marg))

```

### A.4.3 Modelo Normal Assimétrico

```

fi.marg.SN = function(i,y,x,beta,n.sim,sig.t.2,k){
  n.obs=length(y)
  X = cbind(rep(1,size=length(x)),x)
  w = rgamma(n.sim , shape = (k+1)/2 , rate = k/2)
  v = matrix(rtruncnorm(n.sim*n.obs, a=0, b=Inf, mean = 0,
    sd = 1),ncol = n.sim)
  Omega = w/sig.t.2
  V = colSums(v^2)
  B = Omega + V
  z = y - X%%beta
  A = sum(z^2) - colSums(apply(v,2, function(x) x*z))^2/(
    Omega+V)
  D = 1 - v[i,]^2/(Omega+V)
  K = sum(z[-i]^2)-1/(Omega+colSums(v[-i,]^2))*colSums(apply
    (v[-i,],2, function(x) x*z[-i]))^2
  num = gamma((n.obs-1)/2)*sum(B^(-0.5)*(A^(-0.5)))^(n.obs-1)
  )
  den = gamma((n.obs-2)/2)*sqrt(pi)*sum(B^(-0.5)*D^(-0.5)*(K
    ^(-0.5)))^(n.obs-2))
  return(num/den)
}

n.sim = 1000
sig.t.2 = pi^2/4
k = 1/2

```

```

beta.sim.SN = cbind(b0.SN,b1.SN)

cl=makeCluster(8)
clusterSetRNGStream(cl, iseed=12345)
clusterExport(cl,c("beta.sim.SN","y","fi.marg.SN","x","n.sim",
  "sig.t.2","k","rtruncnorm","n.obs"))
tx=Sys.time()
par.f.SN=sapply(1:n.obs, function(arg2) parApply(cl,beta.sim
  .SN,1,function(arg1) fi.marg.SN(arg2,y,x,arg1,n.sim,sig.t
  .2,k)))
Sys.time()-tx
stopCluster(cl)

h.SN.marg = matrix(rep(-1,n.obs*L.sim),ncol=L.sim)
for(i in 1:n.obs){
  for(l in 1:L.sim){
    h.SN.marg[i,l] = CPO.SN[i]/par.f.SN[l,i]
  }
}

#Norma L1
L1.SN.marg = rowMeans(abs(h.SN.marg-1)/2)
#Kullback-Leibler
K1.SN.marg = rowMeans(h.SN.marg*log(h.SN.marg))

```

#### A.4.4 Modelo *t*-assimétrico

```

fi.marg.St = function(i,y,x,beta,v,n.sim,sig.t.2,k){
  n.obs=length(y)
  X = cbind(rep(1,size=length(x)),x)
  w = rgamma(n.sim , shape = (k+1)/2 , rate = k/2)
  u.aux = rgamma(n.sim*n.obs, shape = v/2 , rate = v/2)
  u = matrix(u.aux,ncol = n.sim)
  t = matrix(sapply(u.aux,function(x) rtruncnorm(1, a=0, b=
    Inf, mean = 0, sd = 1/sqrt(x))),ncol = n.sim)
  v=sqrt(u)*t
  Omega = w/sig.t.2
  V = colSums(v^2)
  B = Omega + V
  z = apply(sqrt(u),2,function(x) x*(y - X%%beta))
  A = colSums(z^2)- colSums(v*z)^2/(Omega+V)
  D = 1 - v[i,]^2/(Omega+V)
  K = colSums(z[-i,]^2)-1/(Omega+colSums(v[-i,]^2))*colSums(
    v[-i,]*z[-i,])^2
  num = gamma((n.obs-1)/2)*sum(B^(-0.5)*(A^(-0.5))^(n.obs-1)
    )
}

```

```

    den = gamma((n.obs-2)/2)*sqrt(pi)*sum(B^(-0.5)*D^(-0.5)*(K
      ^(-0.5))^(n.obs-2))
    return(num/den)
  }

n.sim = 1000
sig.t.2 = pi^2/4
k = 1/2
v = 5

beta.sim.St = cbind(b0.St,b1.St)

cl=makeCluster(8)
clusterSetRNGStream(cl, iseed=12345)
clusterExport(cl,c("beta.sim.St","y","fi.marg.St","x","n.sim
  ","v","sig.t.2","k","rtruncnorm","n.obs"))
tx=Sys.time()
par.f.St=sapply(1:n.obs, function(arg2) parApply(cl,beta.sim
  .St,1,function(arg1) fi.marg.St(arg2,y,x,arg1,v,n.sim,sig
  .t.2,k)))
Sys.time()-tx
stopCluster(cl)

h.St.marg = matrix(rep(-1,n.obs*L.sim),ncol=L.sim)
for(i in 1:n.obs){
  for(l in 1:L.sim){
    h.St.marg[i,l] = CPO.St[i]/par.f.St[l,i]
  }
}

#Norma L1
L1.St.marg = rowMeans(abs(h.St.marg-1)/2)
#Kullback-Leibler
K1.St.marg=rowMeans(h.St.marg*log(h.St.marg))

```

# Bibliografia

- Azzalini(1985)** A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178. Citado na pág. [1](#)
- Azzalini e Capitanio(2003)** A. Azzalini e A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc., B*, 65:367–389. Citado na pág. [1](#)
- Azzalini e Genton(2008)** A. Azzalini e M. G. Genton. Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76:106–129. Citado na pág. [1](#), [23](#)
- Bayes(2005)** C. L. Bayes. Inferência bayesiana no modelo normal assimétrico. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil. Citado na pág. [1](#), [2](#), [23](#), [24](#), [65](#)
- Branco e Dey(2001)** M. D. Branco e D. K. Dey. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79:99–113. Citado na pág. [1](#)
- Godoi(2007)** L. G. Godoi. A distribuição t-assimétrica univariada: propriedades e inferência. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil. Citado na pág. [1](#), [2](#), [23](#), [33](#), [65](#)
- Ibrahim et al.(2001)** J. G. Ibrahim, M.-H. Chen e D. Sinha. *Bayesian Survival Analysis*. Springer. Citado na pág. [7](#)
- Paulino et al.(2003)** C. D. Paulino, M. A. A. Turkman e B. Murteira. *Estatística Bayesiana*. Fundação Calouste Gulbenkian. Citado na pág. [11](#)
- Weiss(1996)** R. E. Weiss. An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:739–250. Citado na pág. [2](#), [7](#), [9](#), [10](#), [65](#)
- Weiss e Cho(1998)** R. E. Weiss e M. Cho. Bayesian marginal influence assessment. *Journal of Statistical Planning and Inference*, 71:163–177. Citado na pág. [2](#), [5](#), [16](#), [17](#), [65](#)