
Modelos de regressão linear heteroscedásticos
com erros t-Student: uma abordagem bayesiana
objetiva.

Aline Campos Reis de Souza

Aline Campos Reis de Souza

**Modelos de regressão linear heteroscedásticos com erros
t-Student: uma abordagem bayesiana objetiva.**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestra em Estatística – Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Vicente Garibay Cancho

UFSCar – São Carlos
Abril de 2016

Aline Campos Reis de Souza

Heteroscedastics linear regression models with Student-t errors: an objective bayesian analysis.

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação - ICMC-USP and to the Departamento de Estatística – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Joint Graduate Program in Statistics DEs-UFSCar/ICMC-USP. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Vicente Garibay Cancho

UFSCar – São Carlos
April 2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

S729m Souza, Aline Campos Reis de
Modelos de regressão linear heteroscedásticos com
erros t-Student : uma abordagem bayesiana objetiva /
Aline Campos Reis de Souza. -- São Carlos : UFSCar,
2016.
80 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2016.

1. Distribuição a priori de Jeffreys. 2.
Inferência robusta. 3. Erros t-Student. 4. Modelos
de regressão linear heteroscedásticos. I. Título.

ALINE CAMPOS REIS DE SOUZA

**MODELOS DE REGRESSÃO LINEAR HETEROSCEDÁSTICOS COM ERROS T-STUDENT: UMA
ABORDAGEM BAYESIANA OBJETIVA**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Aprovada em 18 de fevereiro de 2016.

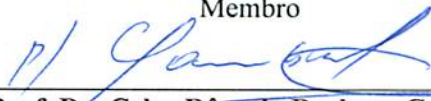
COMISSÃO JULGADORA:



Prof. Dr. Vicente Garibay Cancho
ICMC-USP
Presidente



Prof. Dr. Victor Hugo Lachos Davila
UNICAMP
Membro



Prof. Dr. Celso Rômulo Barbosa Cabral
UFAM - por videoconferência
Membro

Agradecimentos

Agradeço a Deus pela minha vida, saúde e por todas as oportunidades que me fizeram chegar até aqui.

À minha mãe Regina, por servir de exemplo e inspiração e ao meu pai Edson por apoiar, incentivar e investir em mim desde sempre.

Aos meus irmãos Thais e Matheus e aos familiares Milton, Silvania e Yoshida, pessoas pelas quais tenho muito carinho e afeto.

Ao professor Vicente Garibay Cancho, pela orientação deste trabalho e por todo conhecimento compartilhado.

À Willian Luis de Oliveira, pelo apoio e paciência.

Aos professores Reiko Aoki e Celso Rômulo pelas sugestões dadas.

Ao ICMC-USP e a todos os funcionários que fizeram parte da minha formação.

E finalmente, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro concedido para realização deste trabalho.

Resumo

Neste trabalho, apresentamos uma extensão da análise bayesiana objetiva feita em Fonseca *et al.* (2008), baseada nas distribuições a priori de Jeffreys para o modelo de regressão linear com erros t-Student, para os quais consideramos a suposição de heteoscedasticidade. Mostramos que a distribuição a posteriori dos parâmetros do modelo regressão gerada pela distribuição a priori é própria. Através de um estudo de simulação, avaliamos as propriedades frequentistas dos estimadores bayesianos e comparamos os resultados com outras distribuições a priori encontradas na literatura. Além disso, uma análise de diagnóstico baseada na medida de divergência Kullback-Leiber é desenvolvida com a finalidade de estudar a robustez das estimativas na presença de observações atípicas. Finalmente, um conjunto de dados reais é utilizado para o ajuste do modelo proposto.

Palavras chave: distribuição *a priori* de Jeffreys, inferência robusta, erros t-Student, modelos de regressão linear heteroscedásticos.

Abstract

In this work , we present an extension of the objective bayesian analysis made in Fonseca *et al.* (2008), based on Jeffreys priors for linear regression models with Student t errors, for which we consider the heteroscedasticity assumption. We show that the posterior distribution generated by the proposed Jeffreys prior, is proper. Through simulation study , we analyzed the frequentist properties of the bayesian estimators obtained. Then we tested the robustness of the model through disturbances in the response variable by comparing its performance with those obtained under another prior distributions proposed in the literature. Finally, a real data set is used to analyze the performance of the proposed model . We detected possible influential points through the Kullback -Leibler divergence measure, and used the selection model criterias EAIC, EBIC, DIC and LPML in order to compare the models.

Keywords: Jeffreys prior, robust inference, Student t errors, heteroscedastic linear regression models.

Lista de Figuras

1.1	Comportamento das funções densidade e distribuição acumulada da t-Student padrão, considerando diferentes graus de liberdade, e das funções densidade e distribuição acumulada da distribuição normal padrão.	4
1.2	Variação dos quantis 2.5% e 97.5% da distribuição t-Student com relação a variação do parâmetro grau de liberdade.	4
2.1	Gráfico dos resíduos bayesianos padronizados contra os valores ajustados, considerando o modelo linear normal homocedástico.	19
2.2	Gráfico dos resíduos bayesianos padronizados contra as variáveis explicativas pontuação no trabalho e anos de serviço na empresa, considerando o modelo linear normal homocedástico.	19
2.3	Gráfico da medida de divergência de K-L (a) e do logaritmo da ordenada preditiva condicional (b).	20
2.4	Distribuições <i>a priori</i> de Jeffreys para o parâmetro grau de liberdade do modelo linear heteroscedástico t-Student, considerando (a) $k=2$ e (b) $k=3$	32

- 3.1 \sqrt{EQM} relativo simulado com amostras de tamanho $n = 30$, obtido utilizando as distribuições *a priori* de Jeffreys(linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul). 45
- 3.2 \sqrt{EQM} relativo simulado com amostras de tamanho $n = 100$, obtido utilizando as distribuições *a priori* de Jeffreys(linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul). 46
- 3.3 Probabilidade de cobertura com amostras de tamanho $n = 30$, obtida utilizando as distribuições *a priori* de Jeffreys(linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul). 52
- 3.4 Probabilidade de cobertura com amostras de tamanho $n = 100$, obtida utilizando as distribuições *a priori* de Jeffreys(linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul). 52
- 3.5 Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Jeffreys, considerando o conjunto de dados na gerado Seção 3.2, sem perturbações nas observações. 54

3.6	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Geweke com $\lambda = 1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.	54
3.7	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Geweke com $\lambda = 0.1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.	55
3.8	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Geweke com $\lambda = 0.005$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.	55
3.9	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Jeffreys, considerando o conjunto de dados na gerado Seção 3.2.	58
3.10	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Geweke com $\lambda = 1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.	58
3.11	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Geweke com $\lambda = 0.1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.	59

3.12	Histogramas, gráficos das cadeias geradas e de autocorrelação para o modelo linear heteroscedástico t-Student com distribuição <i>a priori</i> exponencial com $\lambda = 0.005$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.	59
3.13	Estimativa Monte Carlo da medida de divergência de K-L, considerando as distribuições <i>a priori</i> de Jeffreys (a) e exponenciais com parâmetros $\lambda = 1$ (b), $\lambda = 0.1$ (c) e $\lambda = 0.005$ (d).	61
3.14	Logaritmo das ordenadas preditivas condicionais para o conjunto de dados simulado, considerando as distribuições <i>a priori</i> de Jeffreys (a) e exponenciais com parâmetros $\lambda = 1$ (b), $\lambda = 0.1$ (c) e $\lambda = 0.005$ (d). . .	62
3.15	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição <i>a priori</i> de Jeffreys, considerando o conjunto de dados da Seção 2.3.	64
3.16	Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico com erros normal e distribuição <i>a priori</i> de Jeffreys, considerando o conjunto de dados da Seção 2.3.	64
3.17	Estimativa Monte Carlo da medida de divergência de K-L e logaritmo da ordenada preditiva condicional, considerando os modelos heteroscedásticos com erros de distribuição t-Student (a) e (c) e de distribuição normal (b) e (d).	67

Lista de Tabelas

3.1	$\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição <i>a priori</i> de Jeffreys para os parâmetros do modelo linear heteroscedástico t-Student.	41
3.2	$\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição <i>a priori</i> exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 1$	42
3.3	$\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição <i>a priori</i> exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.1$	43
3.4	$\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição <i>a priori</i> exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.005$	44
3.5	Viés relativo e probabilidade de cobertura, considerando distribuição <i>a priori</i> de Jeffreys para os parâmetros do modelo linear heteroscedástico t-Student.	47
3.6	Viés relativo e probabilidade de cobertura, considerando a distribuição <i>a priori</i> exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 1$	48

3.7	Viés relativo e probabilidade de cobertura, considerando a distribuição <i>a priori</i> exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.1$	49
3.8	a distribuição <i>a priori</i> exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.05$	50
3.9	Estimativas pontuais, desvios padrão <i>a posteriori</i> e intervalos de credibilidade 95% para os parâmetros dos modelo linear heteroscedástico, considerando diferentes distribuições <i>a priori</i>	56
3.10	Estimativas dos critérios de seleção bayesianos para os modelos ajustados.	57
3.11	Estimativas pontuais, desvios padrão <i>a posteriori</i> e intervalos de credibilidade 95% para os parâmetros dos modelo linear heteroscedástico, considerando diferentes distribuições <i>a priori</i>	60
3.12	Estimativas dos critérios de seleção bayesianos para os modelos ajustados.	61
3.13	Estimativas pontuais, desvios padrão <i>a posteriori</i> e intervalos de credibilidade 95% para os parâmetros dos modelo linear heteroscedástico, considerando erros com distribuição t-Student e distribuição normal. . .	65
3.14	Estimativas dos critérios de seleção bayesianos para os modelos ajustados	65
3.15	Porcentagem das observações contidas em seus respectivos intervalos preditivos.	66
3.16	Mudança percentual nos valores das estimativas.	68

Sumário

1	Introdução	1
1.1	Distribuição t-Student: definição e propriedades	3
1.1.1	Momentos	5
1.2	Inferência Bayesiana Objetiva	7
1.2.1	Distribuição <i>a priori</i> de Jeffreys	8
1.2.2	Distribuição <i>a posteriori</i>	10
1.2.3	Algoritmo Metropolis-Hastings	11
1.3	Critérios de seleção de modelos e diagnóstico de influência bayesianos .	11
1.3.1	Critérios de seleção de modelos	11
1.3.2	Diagnóstico de influência bayesiano	14
1.4	Organização dos capítulos	16
2	Modelo de regressão linear heteroscedástico com erros t-Student	17
2.1	Motivação: dados de pesquisa salarial	18
2.2	Formulação do Modelo	21
2.3	Distribuições <i>a priori</i> e <i>a posteriori</i>	25

2.3.1	Geração de amostras MCMC da distribuição <i>a posteriori</i>	35
3	Estudo de simulação e aplicação	37
3.1	Propriedades Frequentistas	40
3.2	Robustez das estimativas em relação à perturbações na variável resposta	53
3.3	Aplicação dados de pesquisa salarial	63
4	Considerações finais e propostas futuras	69
4.1	Propostas futuras	70
A	Derivadas parciais de primeira e segunda ordem da Seção 2.2	71
A.1	Derivadas parciais de primeira ordem	71
A.2	Derivadas parciais de segunda ordem	73
	Referências Bibliográficas	75

Capítulo 1

Introdução

Em diversas áreas da ciência como medicina, bioquímica, agricultura, entre outras, e em aplicações de caráter industrial ou financeiro, a investigação da relação entre variáveis pode ser de grande interesse. A análise de regressão estuda tais relações e, por este motivo, vem se apresentando ao longo dos anos como uma área de pesquisa de grande relevância na Estatística.

Nos modelos de regressão, é comum assumir que os erros possuem distribuição normal. No entanto, para alguns conjuntos de dados, esta suposição pode não ser a mais adequada, de forma que o modelo obtido pode ser sensível a presença de outliers.

As distribuições definidas como mistura de escala normal, introduzida por Andrews & Mallows (1974), aparecem como boa alternativa na direção de uma modelagem mais robusta. A esta classe, pertencem distribuições como a t-Student, exponencial potência, slash, normal contaminada, entre outras.

Tratando-se de modelos lineares com erros t-Student, uma ampla teoria encontra-se desenvolvida na literatura. Sob o enfoque clássico, Zellner (1976) investiga as consequências de se assumir erros t-Student, ao invés da suposição usual de normalidade. No entanto, não considera a estimação dos graus de liberdade, devido as dificuldades

em se estimar este parâmetro. Geweke (1993) apresenta uma abordagem bayesiana, reportando sobre a dificuldade de uma análise não informativa. Branco *et al.* (2000), por sua vez, aconselha o uso de distribuições *a priori* próprias para a estimação dos graus de liberdade, a fim de evitar que a distribuição *a posteriori* gerada seja imprópria. Fernandez & Steel (1999) apresentam condições sobre a distribuição *a priori* em modelos de regressão linear, com erros pertencentes a classe de mistura de escala da normal, de forma que a distribuição *a posteriori* obtida seja própria.

Em situações nas quais o conhecimento prévio a respeito dos parâmetros é vago, ou até mesmo inexistente, é aconselhável o uso de distribuições *a priori* não informativas. Paulino *et al.* (2003) menciona que estes tipos de distribuições possibilitam a dedução de crenças *a posteriori* a partir de um conhecimento escasso e podem, além disso, desempenhar um papel de referência, permitindo a comparação com resultados da inferência clássica e a verificação da influência de distribuições *a priori* subjetivas quando a informação prévia existe. Fonseca *et al.* (2008), utiliza uma abordagem bayesiana objetiva na inferência sobre os modelos lineares, fazendo uso de distribuições *a priori* de Jeffreys.

Outra suposição usualmente atribuída aos modelos de regressão é a de homocedasticidade dos erros, sob a qual tais componentes variam de maneira constante, para todas as observações. Quando esta hipótese não é satisfeita, um modelo heteroscedástico deve ser proposto, com parâmetros de dispersão que se adequem a natureza dos dados.

Cysneiros *et al.* (2007) trata, através de metodologias clássicas, de modelos heteroscedásticos com erros pertencentes a classe de distribuição simétrica obtendo, desta forma, uma modelagem mais robusta.

Neste trabalho estendemos a análise bayesiana objetiva feita em Fonseca *et al.* (2008) para os modelos lineares heteroscedásticos com erros t-Student, em que o grau de liberdade é desconhecido. Para isto, é proposto o uso das distribuições *a priori* de

Jeffreys, tendo como objetivos principais: a análise da distribuição *a posteriori* gerada, por meio de resultados que garantam que esta é própria, e o estudo dos estimadores bayesianos obtidos, verificando as suas propriedades frequentistas, o seu desempenho comparado ao uso de outras distribuições *a priori* propostas na literatura, a robustez das estimativas obtidas mediante a presença de outliers, e seu ajuste a um conjunto de dados reais.

1.1 Distribuição t-Student: definição e propriedades

Definição 1.1.1. *Uma variável aleatória X é dita ter distribuição t-Student univariada com parâmetro de locação $\mu \in \mathbb{R}$, parâmetro de escala $\sigma^2 > 0$ e $\nu > 0$ graus de liberdade, denotada por $X \sim t-St_\nu(\mu, \sigma^2)$, se sua função densidade de probabilidade é dada por*

$$f(x|\mu, \sigma^2) = \frac{\Gamma(\frac{\nu+1}{2})\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})(\pi\sigma)^{\frac{1}{2}}} \left[\nu + \frac{(x - \mu)^2}{\sigma^2} \right]^{-\frac{\nu+1}{2}},$$

em que Γ denota a função gama, definida por $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$.

Quando $\mu = 0$ e $\sigma = 1$ dizemos que X tem distribuição t-Student padrão.

O grau de liberdade, também referido como parâmetro de forma, oferece à distribuição t-Student uma flexibilidade bastante atrativa em diversas aplicações, em particular na análise de regressão. Na medida em que $\nu \rightarrow \infty$, a função densidade definida em (1.1) se aproxima da função densidade da normal, o que pode ser observado na Figura 1.1.

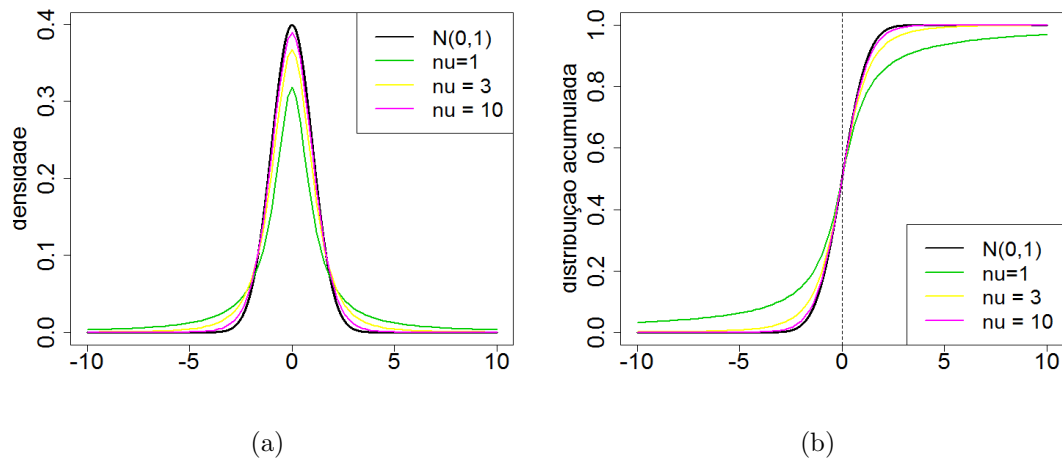


Figura 1.1: Comportamento das funções densidade e distribuição acumulada da t-Student padrão, considerando diferentes graus de liberdade, e das funções densidade e distribuição acumulada da distribuição normal padrão.

Por outro lado, quando $\nu \rightarrow 0$, a função densidade (1.1) tem caudas pesadas comparada à distribuição normal. Na Figura 1.2 ilustramos este fato, analisando o comportamento dos quantis 2.5% e 97.5% da distribuição t-Student com relação a variação do parâmetro grau de liberdade, funções que denotaremos por $q_{2.5\%}(\nu)$ e $q_{97.5\%}(\nu)$ respectivamente.

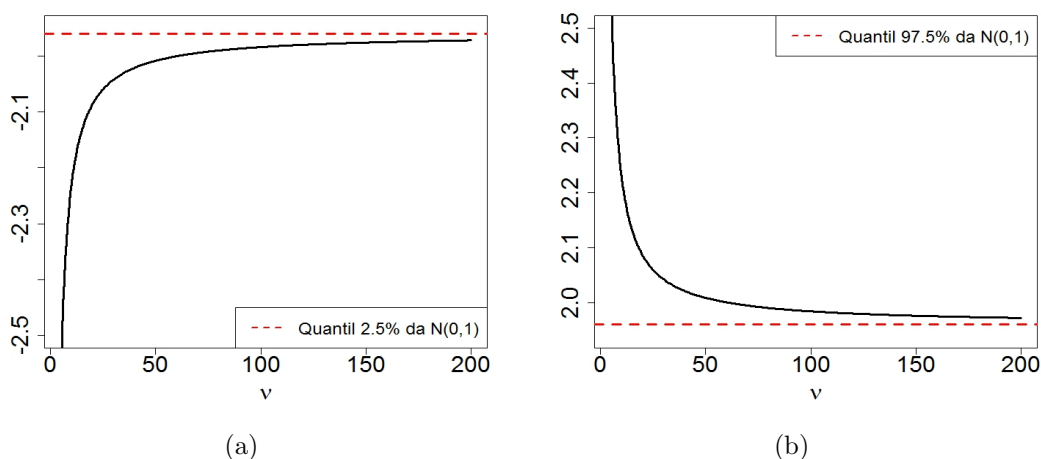


Figura 1.2: Variação dos quantis 2.5% e 97.5% da distribuição t-Student com relação a variação do parâmetro grau de liberdade.

Observe que, quando $\nu \rightarrow \infty$, $q_{2.5\%}(\nu)$ e $q_{97.5\%}(\nu)$ tendem para a aproximação numérica dos quantis 2.5% e 97.5% da distribuição normal padrão, ou seja, convergem para os valores -1.96 e 1.96, respectivamente. Por outro lado, quando $\nu \rightarrow 0$, nota-se um aumento em módulo de ambos os quantis, indicando que, para valores pequenos do grau de liberdade, há uma maior concentração de probabilidade nas caudas da distribuição.

1.1.1 Momentos

Proposição 1.1.1. *Seja X uma variável aleatória com distribuição $t-St_\nu(0, 1)$, então*

(i) $E[X^\alpha] = 0$, se $0 \leq \alpha \leq \nu$ e α for ímpar

(ii) $E[X^\alpha] = \frac{\nu^\alpha \Gamma(1/2 + \alpha) \Gamma(\nu/2 - \alpha)}{\Gamma(1/2) \Gamma(\nu/2)}$, se $0 \leq \alpha \leq \nu$ e α for par

Demonstração. A demonstração pode ser encontrada em Fonseca *et al.* (2008). □

O próximo lema será de grande utilidade no capítulo subsequente.

Lema 1.1.2. *Seja X variável aleatória tal que $X \sim t-St_\nu(0, 1)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ o vetor aleatório cujos elementos se definem como $Y_i = \sqrt{\phi_i}X + \mu_i$, com $\phi_i > 0$ e $\mu_i \in \mathbb{R}$.*

Denotemos por a_s o seguinte valor esperado

$$a_s = \mathbb{E} \left\{ \sum_{i=1}^n \left[\nu + \frac{(Y_i - \mu_i)^2}{\phi_i} \right]^{-s} \right\}.$$

Nestas condições, temos que

$$(i) \quad a_s = n \frac{\Gamma(\frac{\nu+1}{2}) \Gamma(\frac{\nu+2s}{2})}{\Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu+2s+1}{2})} \nu^{-s},$$

$$(ii) \quad \mathbb{E} \left\{ \sum_{i=1}^n (Y_i - \mu_i)^q \left[\nu + \frac{(Y_i - \mu_i)^2}{\phi_i} \right]^{-s} \right\} = a_s \sum_{i=1}^n \left\{ \left(\phi_i \frac{\nu}{\nu + 2s} \right)^{\frac{q}{2}} \mathbb{E}[W_i^q] \right\},$$

em que $W_i \sim t-St_{\nu+2s}(0, 1)$.

1.1. DISTRIBUIÇÃO T-STUDENT: DEFINIÇÃO E PROPRIEDADES

Demonstração. Como em Fonseca *et al.* (2008), denotemos por $c_i(\nu, \phi_i) = \frac{\Gamma(\frac{\nu+1}{2})\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})\sqrt{\pi}\phi_i}$. Para o item (i) obtemos,

$$\begin{aligned} \mathbb{E} \left\{ \sum_{i=1}^n \left[\nu + \frac{(Y_i - \mu)^2}{\phi_i} \right]^{-s} \right\} &= \sum_{i=1}^n \left\{ \int_{-\infty}^{\infty} c_i(\nu, \phi_i) \left[\nu + \frac{(y_i - \mu_i)^2}{\phi_i} \right]^{-\frac{\nu+1+2s}{2}} dy_i \right\} \\ &= \sum_{i=1}^n \left\{ \int_{-\infty}^{\infty} c_i(\nu, \phi_i) \left(\frac{\nu+2s}{\nu} \right)^{\frac{\nu+1+2s}{2}} \left[(\nu+2s) + \frac{(y_i - \mu_i)^2}{\phi_i \frac{\nu}{\nu+2s}} \right]^{-\frac{\nu+1+2s}{2}} dy_i \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\Gamma(\frac{\nu+1}{2}) \Gamma(\frac{\nu+2s}{2}) \nu^{-s}}{\Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu+2s+1}{2})} \int_{-\infty}^{\infty} f_{Y_i} \left(y_i | \nu+2s, \mu_i, \phi_i \frac{\nu}{\nu+2s} \right) dy_i \right\}, \end{aligned}$$

em $f_{Y_i} \left(y_i | \nu+2s, \mu_i, \phi_i \frac{\nu}{\nu+2s} \right) = c_i \left(\nu+2s, \phi_i \frac{\nu}{\nu+2s} \right) \left[(\nu+2s) + \frac{(y_i - \mu_i)^2}{\phi_i \frac{\nu}{\nu+2s}} \right]^{-\frac{\nu+1+2s}{2}}$, é a densidade da distribuição t-Student com parâmetro de locação μ_i , parâmetro de escala $\phi_i \frac{\nu}{\nu+2s}$ e $\nu+2s$ graus de liberdade. Assim,

$$\mathbb{E} \left\{ \sum_{i=1}^n \left[\nu + \frac{(Y_i - \mu)^2}{\phi_i} \right]^{-s} \right\} = n \frac{\Gamma(\frac{\nu+1}{2}) \Gamma(\frac{\nu+2s}{2})}{\Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu+2s+1}{2})} \nu^{-s}.$$

Considerando agora o item (ii), temos que

$$\begin{aligned} \mathbb{E} \left\{ \sum_{i=1}^n (Y_i - \mu_i)^q \left[\nu + \frac{(Y_i - \mu)^2}{\phi_i} \right]^{-s} \right\} &= \sum_{i=1}^n \left\{ a_s \int_{-\infty}^{\infty} (Y_i - \mu_i)^q f_{Y_i} \left(y_i | \nu+2s, \mu_i, \phi_i \frac{\nu}{\nu+2s} \right) dy_i \right\}, \\ &= a_s \mathbb{E} \sum_{i=1}^n \{(Y_i - \mu_i)^q\}, \\ &= a_s \mathbb{E} \sum_{i=1}^n \left\{ \left(\phi \frac{\nu}{\nu+2s} \right)^{\frac{q}{2}} W_i^q \right\}, \end{aligned}$$

com $W_i = \frac{Y_i - \mu_i}{\phi \frac{\nu}{\nu+2s}}$ e $W_i \sim \text{t-St}_{\nu+2s}(0, 1)$. □

Em particular, para $s = 1$ temos que $a_1 = \frac{n}{\nu+1}$ e para $s = 2$, $a_2 = \frac{n(\nu+2)}{\nu(\nu+1)(\nu+3)}$.

1.2 Inferência Bayesiana Objetiva

A metodologia bayesiana tem sido amplamente utilizada na teoria e em práticas estatísticas ao longo dos últimos anos. Sua implementação exige, além da função de verossimilhança, a especificação da informação *a priori* a respeito dos parâmetros, a qual se pretende incorporar às análises. A quantificação do conhecimento prévio é feita através de uma distribuição *a priori*, que por sua vez pode ser obtida por meio de esquemas e métodos de elicitacão (veja por exemplo, Paulino *et al.*, 2003).

Quando a informação disponível é escassa, situação caracterizada como estado de “ignorância *a priori*”, a distribuição *a priori* deve ser derivada a partir de métodos bayesianos objetivos, isto é, através de formulação matemática baseada no modelo e nos dados obtidos (Bernardo, 2009).

A elicitacão de distribuições *a priori* minimamente informativas, referidas como distribuições não informativas, pode ser uma tarefa bastante complexa, uma vez que a falta de informação é um conceito passível de múltiplas interpretações. No entanto, na tentativa de formalizar matematicamente a idéia de “conhecimento vago”, alguns métodos para obtenção de distribuições *a priori* não informativas foram propostos na literatura.

O mais intuitivo deles é o método de Bayes-Laplace, que assume equiprobabilidade dos parâmetros, mas que em espaços paramétricos ilimitados viola o axioma de probabilidade total unitária, gerando distribuições *a priori* denominadas distribuições impróprias. Além disso, se mostra incoerente sob transformações bijetoras do parâmetro: se $\pi(\theta)$ é a distribuição *a priori* para θ definida por

$$\pi(\theta) = k, \quad \forall \theta \in \Theta,$$

em que k é uma constante e Θ o espaço paramétrico, então pelo Teorema da Transformação, a distribuição *a priori* para uma transformação bijetora $\phi \equiv \phi(\theta)$ é dada por

$$\pi(\phi) = \pi(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|,$$

perdendo sua uniformidade em Θ e conseqüentemente, a idéia inicial de representação da falta de informação.

O método proposto por Jeffreys (Jeffreys, 1998) é bastante utilizado na obtenção de distribuições *a priori* não informativas, e como veremos, possui a vantagem de ser invariante sob transformações 1-1.

1.2.1 Distribuição *a priori* de Jeffreys

Seja X variável aleatória com função densidade dada por $f(X|\theta)$, em que $\theta \in \Theta \subset \mathbb{R}$. A informação de Fisher para o parâmetro θ , é definida por

$$\mathcal{I}(\theta) = E_{X|\theta} \left[\left(\frac{\partial \ln f(X|\theta)}{\partial \theta} \right)^2 \right]. \quad (1.1)$$

Esta quantidade mensura a informação que a variável aleatória X carrega a respeito do parâmetro θ , o qual indexa a família de distribuições que a descreve. Sob certas condições de regularidade, é possível mostrar que a informação de Fisher pode ser calculada de maneira equivalente (veja por exemplo, Lehmann, 1998) através da seguinte expressão

$$\mathcal{I}(\theta) = -E_{X|\theta} \left[\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2} \right]. \quad (1.2)$$

A classe de distribuições *a priori* de Jeffreys é obtida baseada nesta medida de informação, e é definida como segue.

Definição 1.2.1 (Distribuição *a priori* de Jeffreys uniparamétrica). *A distribuição a*

priori de Jeffreys para $\theta \in \mathbb{R}$ é dada por $\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)}$.

Considerando agora o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \Theta^m \subset \mathbb{R}^m$, temos que a matriz de informação de Fisher é calculada como

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}) & \mathcal{I}_{12}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{1m}(\boldsymbol{\theta}) \\ \mathcal{I}_{21}(\boldsymbol{\theta}) & \mathcal{I}_{22}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{2m}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{m1}(\boldsymbol{\theta}) & \mathcal{I}_{m2}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{mm}(\boldsymbol{\theta}) \end{pmatrix}, \quad (1.3)$$

com $\mathcal{I}_{ij}(\boldsymbol{\theta}) = -E_{X|\theta} \left[\frac{\partial^2 \ln f(X|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]$, para $i, j = 1, \dots, m$. Neste caso, a distribuição *a priori* de Jeffreys é definida da seguinte maneira.

Definição 1.2.2 (Distribuição *a priori* de Jeffreys multiparamétrica). *A distribuição a priori de Jeffreys para $\boldsymbol{\theta} \in \Theta^m \subset \mathbb{R}^m$ é dada por $\pi(\boldsymbol{\theta}) \propto \sqrt{\det \mathcal{I}(\boldsymbol{\theta})}$.*

Por fim, como mencionado anteriormente, uma propriedade muito importante desta classe de distribuições é sua invariância sob transformações 1-1. De fato, consideremos o caso uniparamétrico em que $\phi \equiv \phi(\theta)$ é uma transformação bijetora de θ . Desta

forma,

$$\begin{aligned}
 \pi(\phi) &\propto \pi(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \\
 &= \sqrt{\mathcal{I}(\theta(\phi)) \left(\frac{d\theta}{d\phi} \right)^2} \\
 &= \sqrt{E \left[\left(\frac{d \ln f(X|\theta(\phi))}{d\theta} \right)^2 \right] \left(\frac{d\theta}{d\phi} \right)^2} \\
 &= \sqrt{E \left[\left(\frac{d \ln f(X|\theta(\phi))}{d\theta} \frac{d\theta}{d\phi} \right)^2 \right]} \\
 &= \sqrt{E \left[\left(\frac{d \ln f(X|\phi)}{d\phi} \right)^2 \right]} \\
 &= \sqrt{I(\phi)}.
 \end{aligned}$$

1.2.2 Distribuição *a posteriori*

Seja X variável aleatória com distribuição amostral $f(x|\theta)$. Vimos que na abordagem bayesiana o conhecimento prévio a respeito da quantidade θ é incorporado à análise por meio da distribuição *a priori* $\pi(\theta)$. Sua atualização, dada a observação $X = x$, é feita através do Teorema de Bayes. Assim, a distribuição *a posteriori* de $\theta|x$ é obtida por

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)}.$$

O denominador $f(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$ não depende de θ e funciona como constante normalizadora da distribuição *a posteriori*. Quando esta integral não pode ser calculada analiticamente, aproximações numéricas ou métodos de simulação devem ser considerados para obtenção de $\pi(\theta|x)$.

Os métodos Monte Carlo via cadeias de Markov (MCMC) são bastante eficientes na resolução de problemas complexos, em particular na inferência bayesiana, fornecendo

a distribuição *a posteriori* de $\theta|x$ através de simulação estocástica. Entre eles, mencionamos o algoritmo Metropolis-Hastings, descrito a seguir, que consiste na construção de uma cadeia de Markov cuja distribuição de equilíbrio é $\pi(\theta|x)$.

1.2.3 Algoritmo Metropolis-Hastings

No contexto bayesiano, dado o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$, a distribuição *a posteriori* $\pi(\boldsymbol{\theta}|x)$ é simulada através dos seguintes passos.

- 1 Inicialize o contador $i = 0$, e forneça um ponto inicial $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_m^{(0)})$;
- 2 Gere um candidato a transição da cadeia $\boldsymbol{\theta}^{(prop)}$ de uma distribuição proposta $q(\cdot|\boldsymbol{\theta})$;
- 3 Compute a probabilidade de aceitação $\alpha = \min\left(1, \frac{\pi(\boldsymbol{\theta}^{(prop)}|x)q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}^{(prop)})}{\pi(\boldsymbol{\theta}^{(i)}|x)q(\boldsymbol{\theta}^{(prop)}|\boldsymbol{\theta}^{(i)})}\right)$;
- 4 Gere u da distribuição $U(0, 1)$ e atualize $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(prop)}$ se $u < \alpha$ e rejeite caso contrário, fazendo $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)}$;
- 5 Repita os passos 2, 3 e 4 até que o processo atinja a distribuição de equilíbrio.

1.3 Critérios de seleção de modelos e diagnóstico de influência bayesianos

1.3.1 Critérios de seleção de modelos

Uma questão a ser considerada em qualquer análise estatística é a avaliação e escolha, dentre um conjunto de modelos, daquele que melhor represente a situação em

1.3. CRITÉRIOS DE SELEÇÃO DE MODELOS E DIAGNÓSTICO DE INFLUÊNCIA BAYESIANOS

estudo. Do ponto de vista bayesiano, trabalhos como os de Gelfand (1996) e Carlin & Louis (1997), estudam a adequabilidade e seleção de modelos usando as distribuições preditivas *a posteriori*.

Destá forma, sejam $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_m)$ duas amostras independentes dado $\boldsymbol{\theta}$, observadas de uma variável aleatória X , com função densidade de probabilidade $f(x|\boldsymbol{\theta})$. Suponha que \mathbf{x} seja utilizada na estimação dos parâmetros da distribuição *a posteriori* $\pi(\boldsymbol{\theta}|\mathbf{x})$ e \mathbf{y} seja utilizada como uma amostra de validação, ou seja, aplicada para verificar a acurácia das predições do modelo ajustado. A distribuição preditiva de \mathbf{y} é dada por

$$p(\mathbf{y}|\mathbf{x}) = \int f(y|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.$$

Obtêm-se assim uma medida de adequabilidade para o modelo estudado, na qual podemos avaliar se as predições dadas por $p(\mathbf{y}|\mathbf{x})$ estão de acordo com a amostra \mathbf{y} : através dos valores médios e das variâncias desta distribuição, por exemplo, podemos definir o resíduo bayesiano padronizado, cuja expressão é dada por,

$$r_i = \frac{y_i - E(Y_i|\mathbf{x})}{\sqrt{\text{Var}(Y_i|\mathbf{x})}}, \quad i = 1, \dots, m,$$

e verificar se existem possíveis afastamentos das suposições iniciais.

Alternativamente, na impossibilidade de se obter duas amostras independentes da variável X , ou de se poder particioná-la em duas, uma reamostragem pode ser feita para a obtenção de uma distribuição preditiva, da seguinte maneira. Sejam \mathcal{D} e $\mathcal{D}_{(-i)}$ os dados completos e com a i -ésima observação excluída, respectivamente. A distribuição preditiva condicional para a i -ésima observação, denotada CPO_i , é dada por

$$CPO_i \doteq p(x_i|\mathcal{D}_{(-i)}) = \int f(x_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{D}_{(-i)})d\boldsymbol{\theta} = \left\{ \int \frac{\pi(\boldsymbol{\theta}|\mathcal{D})}{f(x_i|\boldsymbol{\theta})}d\boldsymbol{\theta} \right\}^{-1}.$$

1.3. CRITÉRIOS DE SELEÇÃO DE MODELOS E DIAGNÓSTICO DE INFLUÊNCIA BAYESIANOS

Esta quantidade, também denominada ordenada preditiva condicional, é bastante útil para validação de modelos. Maiores valores de CPO_i indicam melhores ajustes.

Para modelos mais complexos, nos quais não existe uma forma fechada para CPO_i , a estimativa Monte Carlo para esta medida, dada uma amostra $(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(Q)})$ da distribuição *a posteriori* (Chen *et al.*, 2012), é dada por

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i | \boldsymbol{\theta}^{(q)})} \right\}^{-1}.$$

Como em Cancho *et al.* (2010) utilizamos a estatística *log pseudo marginal likelihood* (LPML),

$$LPML = \sum_{i=1}^n \widehat{CPO}_i.$$

Na seleção dos modelos, maiores valores de LPML indicam o melhores ajustes.

Utilizaremos também como critérios de seleção de modelos o *Deviance Information Criterion* (DIC), o *Expected Akaike Information Criterion* (EAIC) e o *Expected Bayesian Information Criterion* (EBIC).

Desta forma, denotemos por $d(\boldsymbol{\theta}) = -2 \log(f(\mathcal{D} | \boldsymbol{\theta}))$ a função desvio, cuja média *a posteriori* pode ser aproximada por $\bar{D} = \sum_{q=1}^Q d(\boldsymbol{\theta}^{(q)})/Q$. E por p_d o número de parâmetros efetivos, definido por $p_d = \bar{D} - \hat{D}$, em que $\hat{D} = d(\bar{\boldsymbol{\theta}})$. A estimativa MCMC para o critério DIC é calculada como

$$\widehat{DIC} = \bar{D} + p_d = \hat{D} + 2p_d,$$

e quanto menor este valor, melhor o ajuste do modelo aos dados.

Analogamente, os critérios de seleção EAIC e EBIC também podem ser aproximados por meio de

$$\widehat{EAIC} = \bar{d} + 2\#(\boldsymbol{\theta}),$$

e

$$\widehat{EBIC} = \bar{d} + \#(\boldsymbol{\theta}) \log(n),$$

em que $\#(\boldsymbol{\theta})$ é o número de parâmetros do modelo. Menores valores destes critérios indicam melhor ajuste do modelo aos dados.

Outra maneira de selecionar e verificar a adequabilidade de modelos é através da construção de intervalos preditivos (Gelfand, 1996). Para cada $i = 1, \dots, n$, simula-se uma amostra $(x_{i1}^*, \dots, x_{iq}^*)$, das distribuições preditivas condicionais correspondentes. Estas distribuições podem ser obtidas através das amostras MCMC da distribuição *a posteriori* gerando-se amostras Monte Carlo, $(\boldsymbol{\theta}_1^{**}, \dots, \boldsymbol{\theta}_q^{**})$ de $\pi(\boldsymbol{\theta}|\mathcal{D}_{(-i)})$ e, para cada $\boldsymbol{\theta}_j^{**}$, obtendo-se uma amostra de $f(x_i|\boldsymbol{\theta}_j^{**})$.

A partir da amostra das distribuições preditivas condicionais, um intervalo preditivo de probabilidade $(1 - \alpha)$ é construído. Neste trabalho consideramos intervalos com caudas equiprováveis. Se o modelo analisado for adequado, é esperado que $(1 - \alpha) \times 100\%$ dos valores observados caiam dentro de seus respectivos intervalos preditivos.

1.3.2 Diagnóstico de influência bayesiano

Em muitas situações, os conjuntos de dados analisados contêm observações que diferem substancialmente das demais. Tais observações discrepantes podem causar grandes distorções no ajuste do modelo, o que pode levar a resultados equivocados. Nestes casos é importante se investigar qual a influência desses tipos de pontos nas estimativas obtidas. Peng & Dey (1995) apresentam duas distintas abordagens bayesianas para detectar observações influentes no ajuste de modelos de regressão, uma é baseada na distribuição *a posteriori* e a segunda baseada na distribuição preditiva. Quatro medidas específicas são propostas, entre elas, destacamos a divergência Kullback-Leibler (K-L).

Análise de influência caso a caso

Uma maneira comum de avaliar a influência de uma observação no ajuste de um modelo é por meio da deleção de casos Cook (1986).

Suponha que $K(P, P_{(-i)})$ denota a divergência K-L entre P e $P_{(-i)}$, em que P denota a distribuição a *posteriori* de θ para os dados completos e $P_{(-i)}$ é a distribuição a *posteriori* de θ sem o i -ésimo caso. Especificamente,

$$K(P, P_{(-i)}) = \int \pi(\theta|\mathcal{D}) \log \left\{ \frac{\pi(\theta|\mathcal{D})}{\pi(\theta|\mathcal{D}_{(-i)})} \right\} d\theta, \quad (1.4)$$

$K(P, P_{(-i)})$ mede o efeito de omitir o i -ésimo caso dos dados completos na distribuição a *posteriori* de θ . Note que $K(P, P_{(-i)}) \neq K(P_{(-i)}, P)$ em geral. Após alguma álgebra, pode-se demonstrar uma expressão simplificada para $K(P, P_{(-i)})$ dada por,

$$K(P, P_{(-i)}) = \log E_{\theta} \left[\frac{L(\theta|\mathcal{D}_{(-i)})}{L(\theta|\mathcal{D})} \middle| \mathcal{D} \right] + E_{\theta} \left[\log \left\{ \frac{L(\theta|\mathcal{D})}{L(\theta|\mathcal{D}_{(-i)})} \right\} \middle| \mathcal{D} \right], \quad (1.5)$$

em que, $E_{\theta}[\cdot|\mathcal{D}]$ representa a média a *posteriori* de θ , $L(\theta|\mathcal{D}) = \prod_{k=1}^n f(t_k|\theta)$ a função verossimilhança para os dados completos e $L(\theta|\mathcal{D}_{(-i)}) = \prod_{k=1, k \neq i}^n f(t_k|\theta)$ a função verossimilhança sem a i -ésima observação. A equação (1.5) pode ser reescrita como

$$\begin{aligned} K(P, P_{(-i)}) &= \log E_{\theta}[\{f(y_i|\theta)\}^{-1}|\mathcal{D}] + E_{\theta}[\log\{f(y_i|\theta)\}|\mathcal{D}], \\ &= -\log(CPO_i) + E_{\theta}[\log\{f(y_i|\theta)\}|\mathcal{D}]. \end{aligned} \quad (1.6)$$

Da Equação (1.6) uma estimativa de Monte Carlo para $K(P, P_{(-i)})$ considerando uma amostra de tamanho Q da distribuição a *posteriori* de $p(\theta|\mathcal{D})$ é dada por

$$K(\widehat{P}, \widehat{P}_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q} \sum_{q=1}^Q \log[f(y_i|\theta^{(q)})], \quad (1.7)$$

sendo $(\widehat{CPO}_i) = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\theta^{(q)})} \right\}^{-1}$.

Uma vez calculada a medida de divergência, é necessário se estabelecer um ponto de corte, acima do qual classificaremos uma observação influente no ajuste do modelo. Neste trabalho, como em Yiqi (2012) e em Garay (2014), utilizaremos a proposta de Peng & Dey (1995), e consideraremos sob a medida de divergência de Kullback-Leibler, pontos influentes quando $K(P, P_{(-i)}) > 0.14$.

1.4 Organização dos capítulos

No Capítulo 2, desenvolvemos uma análise bayesiana objetiva para o modelo de regressão linear heteroscedástico, com erros t-Student. Derivamos as expressões para o cálculo da distribuição *a priori* de Jeffreys e mostramos que a distribuição *a posteriori* gerada é própria.

No Capítulo 3, apresentamos, através de um estudo de simulação, as propriedades frequentistas do estimador bayesiano obtido e seu desempenho na presença de outliers. Apresentamos ainda, uma aplicação com um conjunto de dados reais analisado em Cysneiros *et al.* (2007), na qual detectamos os possíveis pontos influentes através da medida de divergência de Kullback-Leibler e utilizamos os critérios de seleção EAIC, EBIC, DIC e LPML para comparação de modelos.

Finalmente, no Capítulo 4 encontram-se as considerações e tópicos para sequência do trabalho, com propostas de continuidade, são listadas.

Capítulo 2

Modelo de regressão linear heteroscedástico com erros t-Student

Em geral os modelos de regressão linear partem das suposições de normalidade e homocedasticidade dos erros, isto é, consideram que estas componentes são normalmente distribuídas, com variância constante ao longo de todas as observações.

No entanto, em conjuntos de dados em que este padrão de variabilidade não é verificado, e mediante a presença de outliers, o uso dessas hipóteses pode ser inapropriado, comprometendo a qualidade dos ajustes. Neste caso, uma alternativa para se obter uma modelagem mais robusta é a utilização de distribuições com caudas mais pesadas do que as normal, como por exemplo, a distribuição t-Student.

Esta escolha, porém, deve ser feita com cautela, principalmente quando se deseja estimar o parâmetro grau de liberdade. Geweke (1993) e Branco *et al.* (2000) relatam sobre esta dificuldade nas abordagens bayesianas, tendo em vista que distribuições *a priori* impróprias para o grau de liberdade gerarão, necessariamente, distribuições *a*

posteriori impróprias. Em contrapartida, o uso de distribuições *a priori* próprias para o mesmo não fornece garantias teóricas de que a distribuição *a posteriori* também possua tal propriedade.

Neste sentido, o trabalho de Fernandez & Steel (1999) mostra sua importância ao apresentar uma classe de distribuições *a priori* para modelos de regressão linear, com erros pertencentes a classe de mistura de escala da normal, sob a qual as distribuições *a posteriori* geradas são próprias.

Fonseca *et al.* (2008), utiliza uma abordagem bayesiana objetiva na inferência sobre os modelos lineares com erros t-Student, fazendo o uso de distribuições *a priori* de Jeffreys. Apresentamos neste capítulo uma extensão deste trabalho, na qual consideramos heteroscedasticidade dos dados. A estrutura de variância foi motivada pelo conjunto de dados reais descrito a seguir, previamente analisado em Cysneiros *et al.* (2007) sob uma perspectiva clássica, considerando erros com distribuições simétricas.

2.1 Motivação: dados de pesquisa salarial

O conjunto de dados apresentado em Chatterjee & Hadi (2009) é referente a um estudo que relaciona o salário mensal de 31 funcionários de uma empresa, de acordo com as variáveis explicativas pontuação no trabalho (x_1), sexo (x_2), anos de serviço na empresa (x_3) e performance (x_4), considerando nesta última as escalas 1 = insatisfatório a 5 = ótimo.

O ajuste do modelo linear com erros normais a estes dados, feito em Cysneiros *et al.* (2007), mostra que apenas as variáveis pontuação no trabalho e anos de serviço na empresa são significativas.

Ajustamos o modelo a este conjunto de dados, utilizando uma abordagem bayesiana com distribuição *a priori* de Jeffreys. O gráfico dos resíduos bayesianos padroni-

2.1. MOTIVAÇÃO: DADOS DE PESQUISA SALARIAL

zados contra os valores ajustados, ilustrado na Figura 2.1, mostra um decrescimento da variância dos resíduos a medida que os valores ajustados aumentam, o que é um indicativo de heteroscedasticidade.

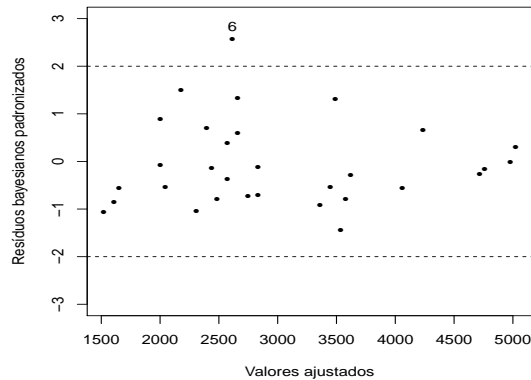


Figura 2.1: Gráfico dos resíduos bayesianos padronizados contra os valores ajustados, considerando o modelo linear normal homocedástico.

Além disso, a análise do gráfico dos resíduos contra as variáveis explicativas, apresentado na Figura 2.2, sugere que a variância dos erros dependa da variável pontuação no trabalho.

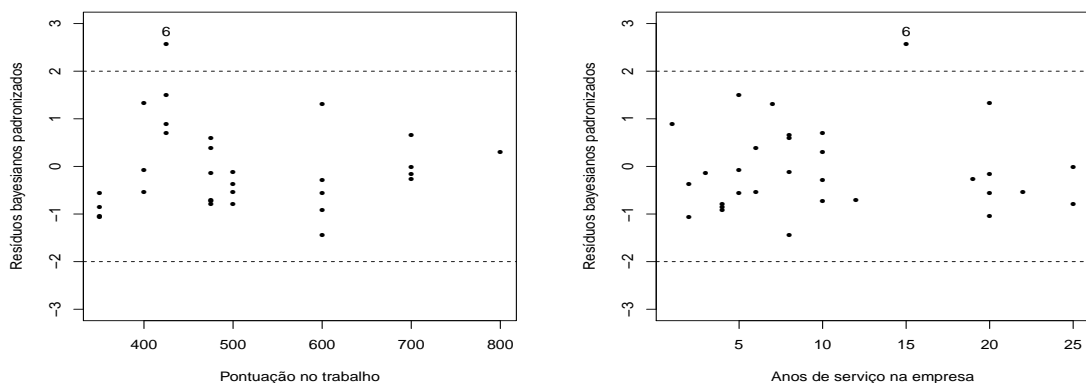


Figura 2.2: Gráfico dos resíduos bayesianos padronizados contra as variáveis explicativas pontuação no trabalho e anos de serviço na empresa, considerando o modelo linear normal homocedástico.

2.1. MOTIVAÇÃO: DADOS DE PESQUISA SALARIAL

Na tentativa de explicar a estrutura heteroscedástica detectada no gráfico dos resíduos, Cysneiros *et al.* (2007) sugere a seguinte parametrização para as variâncias dos erros

$$\sigma_i^2 = \exp\{\gamma_0 + (x_{1i} - \bar{x}_1)\gamma_1\}, \quad i = 1, \dots, 31. \quad (2.1)$$

Utilizando a distribuição *a priori* de Jeffreys, ajustamos novamente o modelo linear aos dados apresentados, considerando heteroscedasticidade dos erros e variâncias parametrizadas como em (2.1). Mais uma vez, o modelo se mostrou sensível a presença da observação 6.

Na Figura 2.3(a) verificamos que o valor da medida K-L para a observação 6 ultrapassa o ponto de corte, indicando que esta pode ser influente. Já na Figura 2.3(b), é possível ver que o logaritmo da ordenada preditiva condicional é relativamente menor nesta observação, em comparação com as restantes.

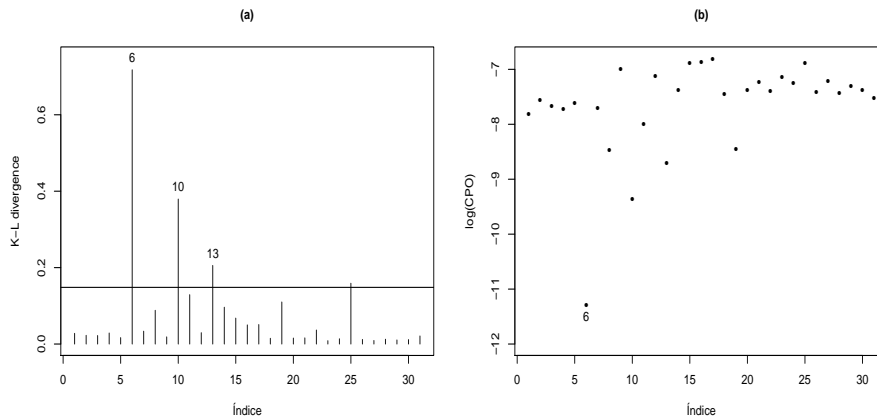


Figura 2.3: Gráfico da medida de divergência de K-L (a) e do logaritmo da ordenada preditiva condicional (b).

Para contornar este problema, distribuições mais robustas para os erros devem ser propostas, a fim de acomodar pontos atípicos sem sofrer alterações nos ajustes. Cysneiros *et al.* (2007) propõe o uso de distribuições simétricas para estas componentes, classe a qual pertence a distribuição t-Student.

Motivados por estes resultados, utilizamos o conjunto de dados em questão para o ajuste do modelo linear heteroscedástico com erros t-Student, cujos parâmetros de dispersão são dados por (2.1) e o grau de liberdade é desconhecido.

2.2 Formulação do Modelo

Seja $\mathbf{y} = (y_1, \dots, y_n)^\top$ o vetor n -dimensional de observações, obtido através do seguinte modelo de regressão linear

$$\mathbf{y} = \mathbf{X}^\top \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.2)$$

em que $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ é a matriz das variáveis explicativas de dimensão $p \times n$ e posto completo, cuja i -ésima coluna é definida por $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip-1})^\top$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^\top$ é o vetor p -dimensional dos coeficientes de regressão e $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ é o vetor n -dimensional dos erros aleatórios, cujos elementos ϵ_i 's são independentes e seguem distribuição t-Student com parâmetro de locação zero, $\nu > 0$ graus de liberdade e parâmetro de escala ϕ_i , para $i = 1, \dots, n$.

Para a especificação das variâncias dos erros heteroscedásticos, assumimos que cada ϕ_i é parametrizado como $\phi_i = \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})$, em que $\mathbf{z}_i = (1, z_{i1}, \dots, z_{ik-1})^\top$ é o vetor k -dimensional de variáveis explicativas e $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_{k-1})^\top$ é o vetor k -dimensional de parâmetros.

Nos referiremos à Equação (2.2), sob as hipóteses consideradas acima, como modelo linear heteroscedástico t-Student.

2.2. FORMULAÇÃO DO MODELO

Denotemos por $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \nu)^\top$ o vetor de parâmetros do modelo linear heteroscedástico t-Student. Dadas as observações (y_1, \dots, y_n) e sob a suposição de que os erros aleatórios são independentes, temos que a função log-verossimilhança de $\boldsymbol{\theta}$ pode ser escrita como

$$l(\boldsymbol{\theta}) = n \left[\log \Gamma \left(\frac{\nu + 1}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) + \frac{\nu}{2} \log \nu - \frac{1}{2} \log \pi \right] - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^\top \boldsymbol{\gamma} - \frac{\nu + 1}{2} \sum_{i=1}^n \log \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]. \quad (2.3)$$

As derivadas parciais de primeira e segunda ordem da log-verossimilhança de $\boldsymbol{\theta}$ com respeito aos parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ e ν , são elementares para a realização de alguns procedimentos inferenciais clássicos e bayesianos. Estas quantidades fornecem, respectivamente, as componentes da função escore

$$U(\boldsymbol{\theta}) = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}, \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}, \frac{\partial l(\boldsymbol{\theta})}{\partial \nu} \right)^\top, \quad (2.4)$$

e da matriz Hessiana de $l(\boldsymbol{\theta})$,

$$H(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^\top} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \nu} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \nu} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \nu \partial \boldsymbol{\beta}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \nu \partial \boldsymbol{\gamma}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \nu \partial \nu} \end{pmatrix}, \quad (2.5)$$

e suas expressões, expostas em forma matricial, podem ser verificadas com maiores detalhes no Apêndice A.

Utilizando (2.5) obtemos a matriz de informação de Fisher de $\boldsymbol{\theta}$ associada ao modelo linear heteroscedástico t-Student, calculando o valor esperado $E[-H(\boldsymbol{\theta})]$ como segue. Denotemos por $H_{i,j}$ e $\mathcal{I}_{i,j}$ os elementos (i, j) das matrizes $H(\boldsymbol{\theta})$ e $\mathcal{I}(\boldsymbol{\theta})$, respectivamente, e por W_i a variável aleatória dada por $W_i = (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) / \phi_i$. Baseados no Lema 1.1.2

2.2. FORMULAÇÃO DO MODELO

e nas derivadas parciais obtidas no Apêndice A, temos que, da equação (A.4)

$$\begin{aligned}
\mathcal{I}_{1,1} &= - \mathbb{E}[H_{1,1}] \\
&= - 2a_2(\nu+1) \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \mathbb{E}\{W_i^2\} + a_1(\nu+1) \mathbb{E}\left\{ \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right\} \\
&= - 2 \frac{(\nu+1)(\nu+2)}{\nu(\nu+1)(\nu+3)} \frac{\nu+4}{\nu+2} \frac{\nu}{(\nu+4)} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} + \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \\
&= \frac{\nu+1}{\nu+3} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})}, \tag{2.6}
\end{aligned}$$

em que $\mathcal{I}_{1,1}$ é uma matriz quadrada de dimensão p . Da equação (A.5), obtemos

$$\begin{aligned}
\mathcal{I}_{1,2} &= - \mathbb{E}[H_{1,2}] \\
&= - a_2(\nu+1) \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i \mathbf{z}_i^\top}{\exp(\frac{1}{2} \mathbf{z}_i^\top \boldsymbol{\gamma})} \mathbb{E}\{W_i^3\} \right\} + a_1(\nu+1) \sum_{i=1}^n \{ \mathbf{x}_i \mathbf{z}_i^\top \mathbb{E}\{W_i\} \} \\
&= \mathbf{0}_{p \times k} \tag{2.7}
\end{aligned}$$

em que $\mathcal{I}_{1,2}$ é uma matriz de zeros, de dimensão $p \times k$. Da equação (A.6), obtemos

$$\begin{aligned}
\mathcal{I}_{1,3} &= - \mathbb{E}[H_{1,3}] \\
&= - a_2(\nu+1) \sum_{i=1}^n \{ \mathbf{x}_i \mathbb{E}\{W_i\} \} + a_1 \sum_{i=1}^n \{ \mathbf{x}_i \mathbb{E}\{W_i\} \} \\
&= \mathbf{0}_{p \times 1}, \tag{2.8}
\end{aligned}$$

em que $\mathcal{I}_{1,3}$ é uma matriz de zeros, de dimensão $p \times 1$. Da equação (A.8), obtemos

$$\begin{aligned}
\mathcal{I}_{2,2} &= - \mathbb{E}[H_{2,2}] \\
&= - a_2 \frac{\nu+1}{2} \sum_{i=1}^n \{ \mathbf{z}_i \mathbf{z}_i^\top \mathbb{E}\{W_i^4\} \} + a_1 \frac{\nu+1}{2} \sum_{i=1}^n \{ \mathbf{z}_i \mathbf{z}_i^\top \mathbb{E}\{W_i^2\} \} \\
&= - \frac{\nu+2}{\nu(\nu+1)(\nu+3)} \frac{\nu+1}{2} \frac{\nu^2}{(\nu+4)^2} \frac{3(\nu+4)^2}{\nu(\nu+2)} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top \\
&\quad + \frac{1}{\nu+1} \frac{\nu+1}{2} \frac{\nu}{\nu+2} \frac{\nu+2}{\nu} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top \\
&= \frac{1}{2} \frac{\nu}{\nu+3} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top, \tag{2.9}
\end{aligned}$$

2.2. FORMULAÇÃO DO MODELO

em que $\mathcal{I}_{2,2}$ é uma matriz de dimensão $k \times k$. Da equação (A.10), obtemos

$$\begin{aligned}
\mathcal{I}_{2,3} &= - \mathbb{E}[H_{2,3}] \\
&= - a_1 \frac{1}{2} \sum_{i=1}^n \{z_i \mathbb{E}\{W_i^2\}\} + a_2 \frac{(\nu+1)}{2} \sum_{i=1}^n \{z_i \mathbb{E}\{W_i^2\}\} \\
&= - \frac{1}{2} \frac{1}{\nu+1} \frac{\nu+2}{\nu} \frac{\nu}{\nu+2} \sum_{i=1}^n z_i \\
&\quad + \frac{\nu+2}{\nu(\nu+1)(\nu+3)} \frac{\nu+1}{2} \frac{\nu+4}{\nu+2} \frac{\nu}{\nu+4} \sum_{i=1}^n z_i \\
&= - \frac{1}{(\nu+1)(\nu+3)} \sum_{i=1}^n z_i, \tag{2.10}
\end{aligned}$$

em que $\mathcal{I}_{2,3}$ é uma matriz de dimensão $k \times 1$. E, finalmente, da equação (A.12) obtemos

$$\begin{aligned}
\mathcal{I}_{3,3} &= - \mathbb{E}[H_{3,3}] \\
&= - \frac{n}{4} \left\{ \psi' \left(\frac{\nu+1}{2} \right) - \psi' \left(\frac{\nu}{2} \right) + \frac{2}{\nu} \right\} + \sum_{i=1}^n a_1 - \frac{(\nu+1)}{2} \sum_{i=1}^n a_2 \\
&= \frac{n}{4} \left\{ \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) \right\} - \frac{n}{\nu} + \frac{n}{(\nu+1)} - \frac{(\nu+1)}{2} \frac{n(\nu+2)}{\nu(\nu+1)(\nu+3)} \\
&= \frac{n}{4} \left\{ \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)} \right\} \tag{2.11}
\end{aligned}$$

em que $\mathcal{I}_{3,3}$ é um escalar. Os elementos restantes são obtidos considerando a simetria da matriz Hessiana $H(\boldsymbol{\theta})$.

Desta forma, a matriz de informação de Fisher $\mathcal{I}(\boldsymbol{\theta})$ é dada por

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\nu+1}{\nu+3} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} & \mathbf{0}_{p \times k} & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{k \times p} & \frac{1}{2} \frac{\nu}{(\nu+3)} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top & - \frac{1}{(\nu+1)(\nu+3)} \sum_{i=1}^n \mathbf{z}_i \\ \mathbf{0}_{1 \times p} & - \frac{1}{(\nu+1)(\nu+3)} \sum_{i=1}^n \mathbf{z}_i^\top & \frac{n}{4} \left\{ \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)} \right\} \end{pmatrix}, \tag{2.12}$$

em que $\psi'(z) = \frac{d^2}{dz^2} \ln \Gamma(z)$ denota a função trigama.

A especificação dos parâmetros de escala ϕ_i associados aos erros aleatórios é um fator importante na modelagem dos dados e, como vemos, define de maneira direta as expressões de quantidades como a informação de Fisher.

Um caso particular da parametrização utilizada pode ser obtido fazendo $\mathbf{z}_i = 1 \in \mathbb{R}$.

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

Como consequência os erros teriam variância constante dada por $\text{Var}(\epsilon_i) = \frac{\nu}{\nu-2}\phi$, com $\phi = \exp(\gamma_0)$, caracterizando homocedasticidade destas componentes.

A matriz de informação de Fisher do vetor $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \phi, \nu)^\top$, nestas condições, pode ser facilmente obtida via regra da cadeia e é dada por

$$\mathcal{I}(\boldsymbol{\xi}) = \begin{pmatrix} \frac{\nu+1}{\nu+3} \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\phi} & \mathbf{0}_{p \times 1} & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{1 \times p} & \frac{n}{2\phi^2} \frac{\nu}{(\nu+3)} & -\frac{n}{\phi(\nu+1)(\nu+3)} \\ \mathbf{0}_{1 \times p} & -\frac{n}{\phi(\nu+1)(\nu+3)} & \frac{n}{4} \left\{ \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)} \right\} \end{pmatrix}, \quad (2.13)$$

O resultado apresentado é o mesmo exposto em Valejos & Steel (2013) e similar¹ ao obtido em Fonseca *et al.* (2008), ambos relacionados ao modelo linear homocedástico com erros t-Student. Como veremos, este fato é importante para a coerência das distribuições *a priori* de Jeffreys obtidas a seguir, pois no caso particular mencionado não gerará distribuições *a priori* conflitantes com as encontradas na literatura.

2.3 Distribuições *a priori* e *a posteriori*

O uso da distribuição t-Student na modelagem dos erros dos modelos de regressão tem se popularizado como alternativa para comportar dados com observações aberrantes. No entanto, a escolha de uma distribuição *a priori* para o parâmetro grau de liberdade pode ser bastante desafiadora.

Em Fonseca *et al.* (2008) o modelo linear t-Student homocedástico é considerado e duas distribuições *a priori* objetivas são propostas para $\boldsymbol{\xi}^*$: a primeira é baseada na regra de Jeffreys e a segunda é uma de suas variantes, a distribuição *a priori* de Jeffreys independente, na qual aplicamos a regra de Jeffreys para o parâmetro de interesse, assumindo que os restantes são constantes conhecidas. Considerando a parametrização

¹Em Valejos & Steel (2013) calcula-se a informação de Fisher associada ao modelo linear t-Student homocedástico, considerando-se o vetor de parâmetros $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \phi, \nu)^\top$, enquanto que em Fonseca *et al.* (2008) considera-se o vetor $\boldsymbol{\xi}^* = (\boldsymbol{\beta}^\top, \sqrt{\phi}, \nu)^\top$.

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

de Vallejos & Steel (2013) a distribuição *a priori* de Jeffreys e a distribuição *a priori* de Jeffreys independente para o vetor $\boldsymbol{\xi}$ são obtidas a partir da matriz (2.13) e dadas, respectivamente, por

$$\pi^J(\boldsymbol{\xi}) \propto \frac{1}{\phi^{1+\frac{p}{2}}} \left(\frac{\nu+1}{\nu+3} \right)^{\frac{p}{2}} \left(\frac{\nu}{\nu+3} \right)^{\frac{1}{2}} \left\{ \Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}, \quad (2.14)$$

$$\pi^I(\boldsymbol{\xi}) \propto \frac{1}{\phi} \left(\frac{\nu}{\nu+3} \right)^{\frac{1}{2}} \left\{ \Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}. \quad (2.15)$$

Baseados no Teorema 1 de Fernandez & Steel (1999, p.156), e no Corolário 1 de Fonseca *et al.* (2008, p.328), é possível garantir a existência de uma distribuição *a posteriori* própria para $\boldsymbol{\xi}$ ao se utilizar a distribuição *a priori* $\pi^I(\boldsymbol{\xi})$. O mesmo, no entanto, não pode ser feito para a distribuição *a priori* $\pi^J(\boldsymbol{\xi})$, uma vez que esta não satisfaz a condição necessária apontada por Vallejos & Steel (2013, p.4) para que a distribuição *a posteriori* gerada seja própria:

$$\pi(\nu) = 0, \forall \nu \in (0, a],$$

com $a = p/(n-p)$.

Tendo em vista os argumentos expostos, consideraremos nesta seção a distribuição *a priori* de Jeffreys independente para o vetor de parâmetros $\boldsymbol{\theta}$, associado ao modelo linear heteroscedástico t-Student. Para tanto, assumiremos independência entre $\boldsymbol{\beta}$ e $(\boldsymbol{\gamma}^\top, \nu)^\top$, de maneira que a distribuição *a priori* resultante será da forma $\pi^I(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma}, \nu)$.

Assim, temos que

$$\pi(\boldsymbol{\beta}) \propto \left(\frac{\nu+1}{\nu+3} \right)^p \sqrt{\det \sum_{i=1}^n \frac{x_i x_i'}{\exp(z_i' \boldsymbol{\gamma})}} \propto 1,$$

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

e portanto, a distribuição *a priori* de Jeffreys independente é dada por $\pi(\boldsymbol{\theta}) \propto \sqrt{\det \mathcal{I}_2}$, em que

$$\mathcal{I}_2 = \begin{pmatrix} \frac{1}{2} \frac{\nu}{(\nu+3)} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top & -\frac{1}{(\nu+1)(\nu+3)} \sum_{i=1}^n \mathbf{z}_i \\ -\frac{1}{(\nu+1)(\nu+3)} \sum_{i=1}^n \mathbf{z}_i^\top & \frac{n}{4} \left\{ \psi' \left(\frac{\nu}{2} \right) - \psi' \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)} \right\} \end{pmatrix}.$$

Observe que a especificação da distribuição *a priori* proposta se completará após o cálculo do determinante da matriz quadrada \mathcal{I}_2 , cuja dimensão $k + 1$ depende da dimensão dos vetores \mathbf{z}'_i s.

Proposição 2.3.1. *As distribuições a priori de Jeffreys independentes para o vetor de parâmetros $\boldsymbol{\theta}$, considerando $k=2$ e $k=3$ são dadas respectivamente por*

$$\pi^I(\boldsymbol{\theta}) \propto \left(\frac{\nu}{\nu+3} \right) \left\{ \Psi \left(\frac{\nu}{2} \right) - \Psi \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}, \quad (2.16)$$

$$\pi^I(\boldsymbol{\theta}) \propto \left(\frac{\nu}{\nu+3} \right)^{\frac{3}{2}} \left\{ \Psi \left(\frac{\nu}{2} \right) - \Psi \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}. \quad (2.17)$$

Demonstração. Consideremos para o cálculo dos determinantes desta demonstração as notações: $c_1 \doteq \frac{\nu}{2(\nu+3)}$, $c_2 \doteq \frac{1}{(\nu+1)(\nu+3)}$ e $c_3 \doteq \frac{n}{4} \left\{ \Psi \left(\frac{\nu}{2} \right) - \Psi \left(\frac{\nu+1}{2} \right) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)} \right\}$.

Quando a dimensão das variáveis explicativas da regressão relacionada aos parâmetros de dispersão for igual a 2, temos que,

$$\sum_{i=1}^n \mathbf{z}_i = \begin{pmatrix} \kappa_{11} \\ \kappa_{12} \end{pmatrix} \quad \text{e} \quad \sum_{i=1}^n \mathbf{z}_i \mathbf{z}'_i = \begin{pmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{pmatrix},$$

em que $\kappa_{11} = n$, $\kappa_{12} = \kappa_{21} = \sum_{i=1}^n z_{i1}$ e $\kappa_{22} = \sum_{i=1}^n z_{i1}^2$.

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

Desta forma, obtemos a seguinte representação da matriz \mathcal{I}_2 ,

$$\mathcal{I}_2 = \begin{pmatrix} c_1\kappa_{11} & c_1\kappa_{12} & c_2\kappa_{11} \\ c_1\kappa_{21} & c_1\kappa_{22} & c_2\kappa_{12} \\ c_2\kappa_{11} & c_2\kappa_{12} & c_3 \end{pmatrix},$$

e seu determinante é dado por

$$\begin{aligned} \det \mathcal{I}_2 &= c_1^2 c_3 \kappa_{11} \kappa_{22} + c_1 c_2^2 \kappa_{11} \kappa_{21} \kappa_{12} + c_1 c_2^2 \kappa_{11} \kappa_{12} \kappa_{12} \\ &- c_1^2 c_2^2 \kappa_{11} \kappa_{11} \kappa_{22} - c_1 c_2^2 \kappa_{11} \kappa_{12} \kappa_{12} - c_1 c_3 \kappa_{21} \kappa_{12} \\ &= c_1^2 c_3 (\kappa_{11} \kappa_{22} - \kappa_{12} \kappa_{21}) + c_1 c_2^2 (\kappa_{11} \kappa_{12} \kappa_{21} - \kappa_{11} \kappa_{11} \kappa_{22}) \\ &= c_1^2 c_3 (\kappa_{11} \kappa_{22} - \kappa_{12} \kappa_{21}) - \kappa_{11} c_1 c_2^2 (\kappa_{11} \kappa_{22} - \kappa_{12} \kappa_{21}) \\ &= \{c_1^2 c_3 - \kappa_{11} c_1 c_2^2\} \det \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top. \end{aligned} \quad (2.18)$$

Logo, a distribuição *a priori* de Jeffreys independente para o vetor de parâmetros $\boldsymbol{\theta}$ é calculada como

$$\begin{aligned} \pi^I(\boldsymbol{\theta}) &\propto \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left\{ \left(\frac{\nu}{\nu+3}\right) \left[\frac{\Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right)}{8} - \frac{2(\nu+5)}{8\nu(\nu+1)(\nu+3)} \right] - \frac{1}{(\nu+1)^2(\nu+3)^2} \right\}^{\frac{1}{2}} \\ &\propto \left(\frac{\nu}{\nu+3}\right) \left\{ \Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}. \end{aligned}$$

Quando a dimensão das variáveis explicativas da regressão relacionada aos parâmetros de dispersão for igual a 3, temos que

$$\sum_{i=1}^n \mathbf{z}_i = \begin{pmatrix} \kappa_{11} \\ \kappa_{12} \\ \kappa_{13} \end{pmatrix} \quad \text{e} \quad \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^\top = \begin{pmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{pmatrix},$$

em que $\kappa_{11} = n$, $\kappa_{12} = \kappa_{21} = \sum_{i=1}^n z_{i1}$, $\kappa_{13} = \kappa_{31} = \sum_{i=1}^n z_{i2}$, $\kappa_{22} = \sum_{i=1}^n z_{i1}^2$ e

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

$\kappa_{33} = \sum_{i=1}^n z_{i2}^2$. Assim, obtemos a seguinte representação da matriz \mathcal{I}_2 ,

$$\mathcal{I}_2 = \begin{pmatrix} c_1 \kappa_{11} & c_1 \kappa_{12} & c_1 \kappa_{13} & c_2 \kappa_{11} \\ c_1 \kappa_{21} & c_1 \kappa_{22} & c_1 \kappa_{23} & c_2 \kappa_{12} \\ c_1 \kappa_{31} & c_1 \kappa_{32} & c_1 \kappa_{33} & c_2 \kappa_{13} \\ c_2 \kappa_{11} & c_2 \kappa_{12} & c_2 \kappa_{13} & c_3 \end{pmatrix},$$

e seu determinante é dado por

$$\begin{aligned} \det \mathcal{I}_2 &= (-1)^{1+1} (\mathcal{I}_2)_{11} \det(\mathcal{I}_2)_{-1,-1} + (-1)^{1+2} (\mathcal{I}_2)_{12} \det(\mathcal{I}_2)_{-1,-2} \\ &+ (-1)^{1+3} (\mathcal{I}_2)_{13} \det(\mathcal{I}_2)_{-1,-3} + (-1)^{1+4} (\mathcal{I}_2)_{14} \det(\mathcal{I}_2)_{-1,-4}, \end{aligned} \quad (2.19)$$

onde $(\mathcal{I}_2)_{-i,-j}$ denota a matriz \mathcal{I}_2 sem a i -ésima linha e sem a j -ésima coluna.

O cálculo dos determinantes das submatrizes $(\mathcal{I}_2)_{-i,-j}$, fornece os seguintes resultados

$$\begin{aligned} \det(\mathcal{I}_2)_{-1,-1} &= c_1 \kappa_{22} (c_1 c_2 \kappa_{33} - c_2^2 \kappa_{13}) \\ &- c_1 \kappa_{13} (c_1 c_3 \kappa_{23} - c_2^2 \kappa_{12} \kappa_{13}) \\ &+ c_2 \kappa_{12} \left(c_1 c_2 \kappa_{23} \kappa_{13} - \frac{1}{2} c_1 c_2 \kappa_{12} \kappa_{33} \right), \end{aligned} \quad (2.20)$$

$$\begin{aligned} \det(\mathcal{I}_2)_{-1,-2} &= c_1 \kappa_{12} (c_1 c_3 \kappa_{33} - c_2^2 \kappa_{13} \kappa_{33}) \\ &- c_1 \kappa_{23} (c_1 c_3 - c_2^2 \kappa_{11} \kappa_{13}) \\ &+ c_2 \kappa_{12} (c_1 c_2 \kappa_{13}^2 - c_1 c_2 \kappa_{11} \kappa_{33}), \end{aligned} \quad (2.21)$$

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

$$\begin{aligned}
\det(\mathcal{I}_2)_{-1,-3} &= c_1 \kappa_{12} (c_1 c_3 - c_2^2 \kappa_{12} \kappa_{13}) \\
&- c_1 \kappa_{22} (c_1 c_3 \kappa_{13} - c_2^2 \kappa_{12} \kappa_{13}) \\
&+ c_2 \kappa_{12} (c_1 c_2 \kappa_{13} \kappa_{12} - c_1 c_2 \kappa_{11} \kappa_{23}), \tag{2.22}
\end{aligned}$$

$$\begin{aligned}
\det(\mathcal{I}_2)_{-1,-4} &= c_1 \kappa_{12} (c_1 c_2 \kappa_{23} \kappa_{13} - (c_1 c_2 \kappa_{12} \kappa_{33}) \\
&- c_1 \kappa_{22} (c_1 c_2 \kappa_{13}^2 - c_1 c_2 \kappa_{11} \kappa_{33}) \\
&+ c_1 \kappa_{23} (c_1 c_2 \kappa_{13} \kappa_{12} - c_1 c_2 \kappa_{23} \kappa_{11}), \tag{2.23}
\end{aligned}$$

Substituindo as expressões (2.20) a (2.23) na Fórmula (2.19), obtemos a seguinte distribuição *a priori* de Jeffreys independente para o vetor de parâmetros $\boldsymbol{\theta}$

$$\begin{aligned}
\pi^I(\boldsymbol{\theta}) &\propto \left(\frac{\nu}{\nu+3} \right) \left\{ \left(\frac{\nu}{\nu+3} \right) \left[\frac{\Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right)}{8} - \frac{2(\nu+5)}{8\nu(\nu+1)(\nu+3)} \right] - \frac{1}{(\nu+1)^2(\nu+3)^2} \right\}^{\frac{1}{2}} \\
&\propto \left(\frac{\nu}{\nu+3} \right)^{\frac{3}{2}} \left\{ \Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{\frac{1}{2}}.
\end{aligned}$$

□

As distribuições *a priori* de Jeffreys obtidas desempenham um papel importante em análises com pouco ou nenhum conhecimento prévio a respeito dos parâmetros envolvidos. No entanto, é importante verificar se estas geram distribuições *a posteriori* próprias, tarefa que nem sempre é realizada com facilidade.

Considerando os modelos de regressão linear homocedásticos e com erros t-Student, em Fernandez & Steel (1999) verifica-se que, para $n \geq p + 1$, as distribuições *a priori* para o vetor de parâmetros $\boldsymbol{\xi}$ que são da forma

$$\pi(\boldsymbol{\xi}) \propto \frac{1}{\phi} \pi(\nu),$$

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

em que $\pi(\nu)$ é uma função de ν própria, geram distribuições *a posteriori* próprias.

Como veremos, este resultado será fundamental quando consideramos heterocedasticidade no modelo. Nas Figuras 2.4 (a) e (b) observamos que as distribuições *a priori* de Jeffreys propostas possuem esta propriedade, o que é confirmado através da proposição a seguir.

Proposição 2.3.2. *A função $\pi(\nu) = \left(\frac{\nu}{\nu+3}\right)^{\frac{k}{2}} \left\{ \Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\}^{1/2}$ é tal que,*

(i) *para $k = 2$, $\pi(\nu) = O(1)$ quando $\nu \rightarrow 0$ e $\pi(\nu) = O(\nu^{-2})$ quando $\nu \rightarrow \infty$ e*

(ii) *para $k = 3$, $\pi(\nu) = O(\nu^{-\frac{1}{2}})$ quando $\nu \rightarrow 0$ e $\pi(\nu) = O(\nu^{-2})$ quando $\nu \rightarrow \infty$.*

Demonstração. Observe que $\pi(\nu)$ é uma composição de funções contínuas em $(0, \infty)$, o que garante sua continuidade neste intervalo. Sendo assim, para mostrar que

$$\int_0^{\infty} \pi(\nu) < \infty, \quad (2.24)$$

basta mostrar que $\pi(\nu)$ tem caudas que convergem para zero. Pelo Corolário 1 de Fonseca *et al.* (2008, p. 328) temos que

$$\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left[\Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right]^{\frac{1}{2}} = O(\nu^{-\frac{1}{2}}), \quad \text{quando } \nu \rightarrow 0.$$

Como $\left(\frac{\nu}{\nu+3}\right)^{\frac{k-1}{2}} = O(\nu^{\frac{k-1}{2}})$, para valores suficientemente pequenos de ν , temos que $\pi(\nu) = O(1)$, para $k = 2$ e, $\pi(\nu) = O(\nu^{\frac{1}{2}})$, para $k = 3$, quando $\nu \rightarrow 0$.

Por outro lado, do Corolário 1 de Fonseca *et al.* (2008) temos também que,

$$\left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left[\Psi\left(\frac{\nu}{2}\right) - \Psi\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right]^{\frac{1}{2}} = O(\nu^{-2}), \quad \text{quando } \nu \rightarrow \infty.$$

Além disso, como $\left(\frac{\nu}{\nu+3}\right)^{\frac{k-1}{2}} = O(1)$, para valores suficientemente grandes de ν , e para $k = 2$ e $k = 3$, temos que $\pi(\nu) = O(\nu^{-2})$.

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

Da continuidade de $\pi(\nu)$ em $(0, \infty)$, juntamente com os fatos mostrados, temos que a integral (2.24) é própria. \square

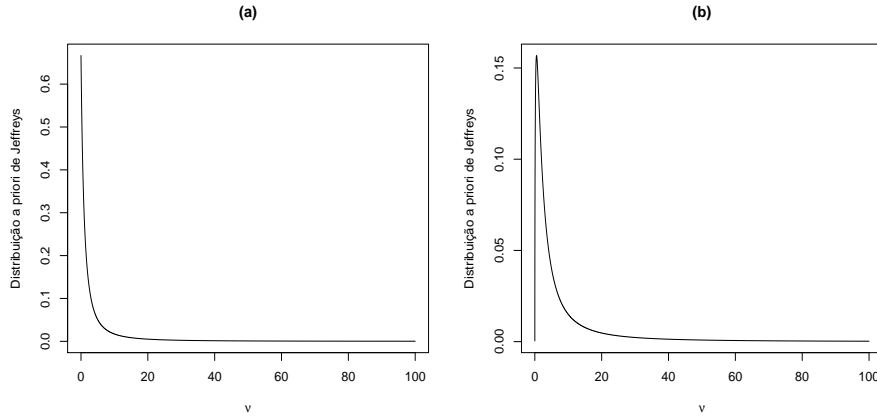


Figura 2.4: Distribuições *a priori* de Jeffreys para o parâmetro grau de liberdade do modelo linear heteroscedástico t-Student, considerando (a) $k=2$ e (b) $k=3$.

Pelo Teorema de Bayes, combinando a função de verossimilhança do modelo linear heteroscedástico t-Student com alguma das distribuições *a priori* obtidas na Proposição 2.3.1, temos que a distribuição *a posteriori* conjunta para $\boldsymbol{\theta}$, dada as observações $\mathbf{y} = (y_1, \dots, y_n)$, é

$$\pi(\boldsymbol{\theta}) \propto \prod_{i=1}^n \frac{\Gamma(\frac{\nu+1}{2})\nu^{\nu/2}}{\Gamma(\frac{\nu}{2})\sqrt{\pi}\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-\frac{(\nu+1)}{2}} \pi^I(\boldsymbol{\theta}). \quad (2.25)$$

Os próximos lemas são resultados auxiliares, usados para mostrar que a distribuição *a posteriori* (2.25) é própria.

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

Lema 2.3.3. *Seja $c_\mu(\sigma) = \frac{1}{\sigma} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}}$, com $\sigma > 0$, $\nu > 0$ e $\mu \in \mathbb{R}$. Temos que $c_\mu(\sigma)$ é crescente para todo $\sigma < \mu$.*

Demonstração. De fato, derivando c_μ com respeito a σ , temos

$$\begin{aligned}
 \dot{c}_\mu(\sigma) &= -\frac{1}{\sigma^2} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}} + \frac{2\mu^2(\nu+1)}{\sigma^4} \frac{1}{2} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}-1} > 0 \\
 &\Leftrightarrow \frac{2\mu^2(\nu+1)}{\sigma^4} \frac{1}{2} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}-1} > \frac{1}{\sigma^2} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}} \\
 &\Leftrightarrow \frac{\mu^2}{\sigma^2}(\nu+1) > \frac{1}{\sigma^2} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right] \\
 &\Leftrightarrow \mu^2(\nu+1) > \frac{1}{\sigma^2} \left[\nu + \left(\frac{\mu}{\sigma} \right)^2 \right] \\
 &\Leftrightarrow \mu^2(\nu+1) > \sigma^2\nu + \mu^2 \\
 &\Leftrightarrow \sigma^2 < \mu^2.
 \end{aligned}$$

□

Lema 2.3.4. *Seja $C_{\mu_1, \dots, \mu_n}(\sigma_1, \dots, \sigma_n) = \prod_{i=1}^n c_{\mu_i}(\sigma_i)$, com c_μ como especificada no Lema 2.3.3 e $\sigma_i^2 = \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})$, em que $\mathbf{z}_i, \boldsymbol{\gamma} \in \mathbb{R}^k$, para todo $i = 1, \dots, n$. Nestas condições, existem $m \in \mathbb{N}$, $m < n$, e $\tilde{\mathbf{z}} \in \mathbb{R}^k$ tais que $\tilde{\sigma}^2 = \exp(\tilde{\mathbf{z}}^\top \boldsymbol{\gamma})$ e*

$$\begin{aligned}
 C_{\mu_1, \dots, \mu_n}(\sigma_1, \dots, \sigma_n) &\leq \frac{1}{(\tilde{\sigma})^{m+2}} \prod_{j=1}^m \left[\nu + \left(\frac{\mu_j}{\tilde{\sigma}} \right)^2 \right]^{-\frac{\nu+1}{2}} \\
 &= \frac{1}{\exp\left(\left(\frac{m}{2} + 1\right)\tilde{\mathbf{z}}^\top \boldsymbol{\gamma}\right)} \prod_{j=1}^m \left[\nu + \frac{\mu_j^2}{\exp(\tilde{\mathbf{z}}^\top \boldsymbol{\gamma})} \right]^{-\frac{\nu+1}{2}}.
 \end{aligned}$$

Demonstração. Sejam $\sigma_{min}^2 = \exp(\mathbf{z}_{min}^\top \boldsymbol{\gamma})$ e $\sigma_{max}^2 = \exp(\mathbf{z}_{max}^\top \boldsymbol{\gamma})$ o mínimo e o máximo de $\sigma_1, \dots, \sigma_n$, respectivamente. Pelo Lema 2.3.3, para cada $i = 1, \dots, n$ temos que

(i) Se $\sigma_{min}, \sigma_{max} < \mu_i$, então c_{μ_i} é crescente no intervalo $[\sigma_{min}, \sigma_{max}]$, e portanto $c_{\mu_i}(\sigma) \leq c_{\mu_i}(\sigma_{max}), \forall \sigma \in [\sigma_{min}, \sigma_{max}]$;

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

(ii) Se $\sigma_{min}, \sigma_{max} > \mu_i$, então c_{μ_i} é decrescente no intervalo $[\sigma_{min}, \sigma_{max}]$, e portanto $c_{\mu_i}(\sigma) \leq c_{\mu_i}(\sigma_{min}), \forall \sigma \in [\sigma_{min}, \sigma_{max}]$ e

(iii) Se $0 < \sigma_{min} < \mu_i$ e $\sigma_{max} > \mu_i$, então $c_{\mu_i}(\sigma) \leq c_{\mu_i}(\mu_i), \forall \sigma \in [\sigma_{min}, \sigma_{max}]$.

Dados μ_1, \dots, μ_n , denotemos por n_1, n_2 e n_3 o número de elementos $\mu \in \{\mu_1, \dots, \mu_n\}$ que satisfazem, respectivamente, as hipóteses (i), (ii) e (iii), de forma que $n_1 + n_2 + n_3 = n$. Então,

$$\begin{aligned} C_{\mu_1, \dots, \mu_n}(\sigma_1, \dots, \sigma_n) &\leq \prod_{i=1}^{n_1} \frac{1}{\sigma_{max}} \left[\nu + \left(\frac{\mu_i}{\sigma_{max}} \right)^2 \right]^{-\frac{\nu+1}{2}} \\ &\times \prod_{j=1}^{n_2} \frac{1}{\sigma_{min}} \left[\nu + \left(\frac{\mu_j}{\sigma_{min}} \right)^2 \right]^{-\frac{\nu+1}{2}} \\ &\times \prod_{l=1}^{n_3} \frac{1}{\mu_l} [\nu + 1]^{-\frac{\nu+1}{2}} \\ &\leq \frac{1}{(\sigma_{min})^{n_2+2}} \prod_{j=1}^{n_2} \left[\nu + \left(\frac{\mu_j}{\sigma_{min}} \right)^2 \right]^{-\frac{\nu+1}{2}} \\ &= \frac{1}{\exp\left(\left(\frac{n_2}{2} + 1\right) \mathbf{z}_{min}^\top \boldsymbol{\gamma}\right)} \prod_{j=1}^{n_2} \left[\nu + \frac{\mu_j^2}{\exp(\mathbf{z}_{min}^\top \boldsymbol{\gamma})} \right]^{-\frac{\nu+1}{2}}. \end{aligned}$$

Fazendo $m = n_2$ e $\tilde{\mathbf{z}} = \mathbf{z}_{min}$, provamos o Lema. \square

Proposição 2.3.5. *A distribuição a posteriori conjunta para $\boldsymbol{\theta}$ dada em (2.25) é própria para $n \geq p + 1$.*

Demonstração. Pelo Lema 2.3.4 existem $m \in \mathbb{N}$, $m < n$, e $\tilde{\mathbf{z}} \in \mathbb{R}^k$ tais que $\tilde{\sigma}^2 = \exp(\tilde{\mathbf{z}}^\top \boldsymbol{\gamma})$ e

$$\begin{aligned} \prod_{i=1}^n \frac{\Gamma(\frac{\nu+1}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \sqrt{\pi \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})}} \left[\nu + \left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right)^2 \right]^{-\frac{\nu+1}{2}} \pi(\nu) &\leq \frac{1}{\tilde{\sigma}^2} \prod_{j=1}^m \frac{\Gamma(\frac{\nu+1}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \sqrt{\pi \tilde{\sigma}^2}} \left[\nu + \left(\frac{y_j - \mathbf{x}_j^\top \boldsymbol{\beta}}{\tilde{\sigma}} \right)^2 \right]^{-\frac{\nu+1}{2}} \pi(\nu) \\ &= \frac{1}{\exp(\tilde{\mathbf{z}}^\top \boldsymbol{\gamma})} \prod_{j=1}^m \frac{\Gamma(\frac{\nu+1}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \sqrt{\pi \exp(\frac{1}{2} \tilde{\mathbf{z}}^\top \boldsymbol{\gamma})}} \left[\nu + \frac{(y_j - \mathbf{x}_j^\top \boldsymbol{\beta})^2}{\exp(\tilde{\mathbf{z}}^\top \boldsymbol{\gamma})} \right]^{-\frac{\nu+1}{2}} \pi(\nu). \end{aligned}$$

Note que, o segundo termo da inequação é a distribuição *a posteriori* do modelo linear homocedástico, com distribuições *a priori* $\pi(\boldsymbol{\beta}) \propto 1$ para $\boldsymbol{\beta}$ e $\pi(\tilde{\sigma}, \nu) \propto \frac{\pi(\nu)}{\tilde{\sigma}}$ para $(\tilde{\sigma}, \nu)$. Pelo Teorema 1 de Fernandez & Steel (1999, p. 156), como $\pi(\nu)$ é própria, temos que a distribuição *a posteriori* do segundo termo da inequação também será própria para $n \geq p + 1$. Além disso, como a inequação é válida em todo o espaço paramétrico, pelo Teste da Comparação (no caso multidimensional), temos que $\pi(\boldsymbol{\theta}|\mathbf{y})$ também será própria desde que $n \geq p + 1$. \square

2.3.1 Geração de amostras MCMC da distribuição *a posteriori*

Ao mostrar que a distribuição *a posteriori* (2.25) é própria, estamos aptos a fazer inferência sobre os parâmetros. No entanto, devido sua complexidade em termos de integração, recorreremos ao uso de ferramentas computacionais para obtenção dos estimadores bayesianos.

Utilizamos nos estudos de simulação e na aplicação do capítulo seguinte o algoritmo Metropolis-Hastings para obter amostras MCMC da distribuição *a posteriori*.

Dado que na i -ésima iteração do algoritmo a cadeia de Markov dos parâmetros se encontra no estado $\boldsymbol{\theta}^{(i)} = (\boldsymbol{\beta}^{(i)\top}, \boldsymbol{\gamma}^{(i)\top}, \nu^{(i)\top})^\top$, geramos os candidatos a transição $\boldsymbol{\theta}^{(prop)} = (\boldsymbol{\beta}^{(prop)\top}, \boldsymbol{\gamma}^{(prop)\top}, \nu^{(prop)\top})^\top$ da seguinte maneira

$$\boldsymbol{\theta}^{(prop)} = \boldsymbol{\theta}^{(i)} + \mathbf{C}^\top \mathbf{Z}, \quad (2.26)$$

em que \mathbf{Z} é o vetor aleatório com distribuição normal multivariada, de dimensão $(p + k + 1)$, média $\mathbf{0}$ e matriz de covariâncias I_{p+k+1} , com I_{p+k+1} a matriz identidade de dimensão $(p + k + 1)$, e \mathbf{C} é tal que $\Sigma = \mathbf{C}^\top \mathbf{C}$, com Σ denotando a matriz de covariância da distribuição *a posteriori* de $\boldsymbol{\theta}|\mathbf{y}$.

2.3. DISTRIBUIÇÕES A PRIORI E A POSTERIORI

Para poder utilizar a distribuição proposta (2.26) de forma coerente, com valores de candidatos pertencentes ao espaço paramétrico, aplicamos a reparametrização $\phi(\nu) = \ln \nu$ ao grau de liberdade e geramos uma cadeia para $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \ln \nu)^\top$. Observe que devido a invariância da distribuição *a priori* de Jeffreys sob transformações 1-1, a distribuição *a posteriori* de $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \ln \nu)^\top$ dado \mathbf{y} é obtida como $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \ln \nu | \mathbf{y}) \propto L(\boldsymbol{\theta}, \boldsymbol{\gamma}, \ln \nu | \mathbf{y}) \pi(\boldsymbol{\theta}, \boldsymbol{\gamma}, \ln \nu)$. Ao final do processo, recuperamos os valores gerados de ν via função inversa.

As cadeias geradas foram de tamanho 100.000 e os descartes das amostras de aquecimento foram feitos com base nas suas médias ergódicas ao longo das iterações. Os saltos dados foram de tamanho 20, sendo este valor aumentado caso a cadeia gerada estivesse muito auto-correlacionada. A convergência foi verificada graficamente por meio dos histogramas, gráficos das cadeias e de autocorrelação. Utilizamos também a estatística de Geweke, a qual indica convergência quando está contida no intervalo $(-1.96, 1.96)$.

Capítulo 3

Estudo de simulação e aplicação

Em modelos de regressão com erros t-Student a estimação do grau de liberdade pode ser muito complexa. Em abordagens bayesianas, o uso de distribuições *a priori* impróprias para este parâmetro podem levar a obtenção de distribuições *a posteriori* impróprias (Fernandez & Steel, 1999), o que dificulta, por exemplo, o uso de distribuições *a priori* não informativas. Sendo assim, é comum que se fixe ou que se utilize distribuições informativas nestes casos, como pode ser observado em trabalhos como os de Zellner (1976), Fernandez & Steel (1999) e Geweke (1993). Neste trabalho as distribuições *a priori* de Jeffreys são derivadas para dados heteroscedásticos e, este capítulo tem como objetivo verificar o desempenho dessas distribuições, principalmente na estimação do grau de liberdade.

Apresentamos um estudo de simulação considerando o modelo linear heteroscedástico t-Student e utilizando as distribuições *a priori* de Jeffreys, propostas na Seção 2.3. Para fins de comparação, outras distribuições *a priori* foram usadas, assumindo que os vetores de parâmetros β e γ têm a distribuição *a priori* $\pi(\beta, \gamma) \propto 1$ e, como sugerido por Geweke (1993), consideramos que o grau de liberdade tem distribuição *a priori* exponencial, ou seja, $\pi(\nu) \propto \lambda \exp(-\lambda\nu), \lambda > 0$. Note que as distribuições *a priori*

de Jeffreys obtidas neste trabalho são funções que dependem apenas de ν . E, como pode ser observado na Figura 2.3, também têm comportamento exponencial. Apesar da similaridade de comportamento entre estas distribuições, as distribuições *a priori* de Jeffreys apresentam a vantagem de serem não informativas.

Na Seção 3.1, realizamos uma análise das propriedades frequentistas dos estimadores bayesianos obtidos. São considerados a raiz do erro quadrático médio (\sqrt{EQM}), o viés, o desvio padrão das estimativas (D_p) e a probabilidade de cobertura dos intervalos de credibilidade interquantil de 95%. Com exceção da probabilidade de cobertura, utilizaremos, devido a escala dos dados, os valores relativos destas medidas, obtidos através da divisão das mesmas pelo valor absoluto do parâmetro correspondente.

A raiz do erro quadrático médio simulada é dada pela expressão

$$\sqrt{EQM} = \left[\sum_{i=1}^n \frac{(\hat{\theta}_i - \theta)^2}{n} \right]^{1/2}.$$

O viés simulado é obtido fazendo a diferença entre as médias das estimativas e o verdadeiro valor do parâmetro.

O desvio padrão das estimativas é calculado como

$$\sqrt{D_p} = \left[\sum_{i=1}^n \frac{(\hat{\theta}_i - \bar{\hat{\theta}})^2}{n-1} \right]^{1/2}.$$

A estimativa da probabilidade de cobertura dos intervalos de credibilidade é encontrada calculando dos intervalos interquantil de credibilidade 95% das estimativas em cada amostra gerada, e então verificando a proporção de intervalos que contêm o verdadeiro valor do parâmetro.

O uso de distribuições *a priori* informativas pode ser bastante conveniente quando se há, de fato, algum conhecimento *a priori* a respeito do caso a ser analisado. A

elicitação de distribuições *a priori* deve no entanto ser feita com cautela, uma vez que considerações equivocadas, pode levar a resultados igualmente equivocados. Uma alternativa a este problema é o uso de distribuições *a priori* não informativas, como a de Jeffreys.

Na Seção 3.2 verificamos a flexibilidade da distribuição *a priori* de Jeffreys na obtenção de estimativas, através do ajuste do modelo linear heteroscedástico t-Student a conjuntos de dados com e sem observações atípicas. O desempenho da distribuição *a priori* de Jeffreys foi comparado ao desempenho das distribuições exponenciais, com pequenas e grandes médias.

Finalmente, na Seção 3.3 ajustamos o modelo linear heteroscedástico t-Student a um conjunto de dados reais, utilizando a distribuição *a priori* de Jeffreys proposta e comparamos os resultados obtidos através do ajuste do modelo linear heteroscedástico, cujos erros têm distribuição normal.

3.1 Propriedades Frequentistas

Apresentamos nesta seção um estudo com dados simulados, cuja finalidade é verificar o comportamento dos estimadores bayesianos dos parâmetros do modelo linear heteroscedástico t-Student. Consideramos nesta análise as distribuições *a priori* de Jeffreys e as distribuições *a priori* $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \nu) \propto \lambda \exp(-\lambda \nu)$, com parâmetros $\lambda = 1$, $\lambda = 0.1$, e $\lambda = 0.005$.

Para geração dos dados artificiais, tentamos reproduzir o cenário do conjunto de dados de pesquisa salarial, apresentado Capítulo 2. Desta forma, fixamos os valores $\beta_0 = -1400.18$, $\beta_1 = 7.8$, $\beta_2 = 31.96$, $\gamma_0 = 12.16$ e $\gamma_1 = -0.006$. As variáveis explicativas foram obtidas da distribuição Gama(170, 0.4) e Gama(1.9, 0.2), respectivamente. Para o parâmetro grau de liberdade, consideramos os valores da seguinte malha numérica $\nu = \{1, 2, 3, 5, 7, 10, 20\}$.

Para cada configuração, 500 réplicas foram geradas considerando amostras de tamanho 30 e 100. Para obtenção das estimativas, geramos as amostras MCMC da distribuição *a posteriori* através do algoritmo Metropolis-Hastings, e em cada réplica a convergência do processo foi monitorada verificando se as estatísticas de Geweke se encontravam no intervalo (-1.96, 1.96). Cadeias que não satisfiziam este a este critério, eram descartadas, e um novo conjunto de dados era gerado.

Nas Tabelas 3.1 a 3.4, apresentamos a raiz do erro quadrático médio relativo e o desvio padrão relativo, considerando as distribuições *a priori* de Jeffreys, e as distribuições exponenciais com parâmetros $\lambda = 1$, $\lambda = 0.1$ e $\lambda = 0.05$ para amostras de tamanho 30 e 100.

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.1: $\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição *a priori* de Jeffreys para os parâmetros do modelo linear heteroscedástico t-Student.

		$\sqrt{\text{EQM}}$ relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,3091	0,5414	0,6533	0,0513	3,1364	0,2036
	100	0,5626	0,2357	0,3145	0,0271	1,7619	0,0377
2	30	1,0951	0,4558	0,5693	0,0401	2,5564	0,2367
	100	0,5106	0,2117	0,2433	0,0189	1,2047	0,0593
3	30	0,9763	0,4046	0,4825	0,0387	2,3736	0,2292
	100	0,4507	0,1878	0,2329	0,0173	1,0845	0,0765
5	30	0,9284	0,3837	0,4399	0,0333	1,9536	0,2564
	100	0,4799	0,1973	0,2191	0,0156	0,9629	0,0892
7	30	0,8908	0,3637	0,4469	0,0329	1,9189	0,2405
	100	0,4788	0,1989	0,2069	0,0144	0,9094	0,0882
10	30	0,9018	0,3735	0,4201	0,0305	2,0735	0,2634
	100	0,4185	0,1733	0,1966	0,0134	0,9037	0,1058
20	30	0,8416	0,3496	0,4426	0,0255	1,8018	0,2515
	100	0,4043	0,1648	0,2087	0,0131	0,8037	0,1032
		D_p relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,3088	0,5414	0,6526	0,0526	3,132	0,1992
	100	0,5649	0,2365	0,3162	0,0272	1,7717	0,0379
2	30	1,0962	0,4563	0,5698	0,0399	2,5584	0,2357
	100	0,5105	0,2109	0,2434	0,0189	1,2059	0,0593
3	30	0,9761	0,4047	0,4829	0,0383	2,3759	0,2295
	100	0,4505	0,1877	0,2331	0,0173	1,0829	0,0761
5	30	0,9276	0,3833	0,4404	0,0337	1,9553	0,2559
	100	0,4772	0,1961	0,2192	0,0156	0,9639	0,0887
7	30	0,8917	0,3639	0,4472	0,0324	1,9206	0,2361
	100	0,4789	0,1989	0,2069	0,0144	0,9091	0,0875
10	30	0,9023	0,3736	0,4205	0,0301	2,0732	0,2584
	100	0,4188	0,1734	0,1968	0,0133	0,9046	0,1057
20	30	0,8425	0,3499	0,4429	0,0253	1,8029	0,2383
	100	0,4037	0,1647	0,2082	0,0131	0,8044	0,1035

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.2: $\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição *a priori* exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 1$.

		$\sqrt{\text{EQM}}$ relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,2305	0,5144	0,7008	0,0523	3,2135	0,1777
	100	0,6013	0,2484	0,2936	0,0239	1,5421	0,0349
2	30	1,1247	0,4566	0,4725	0,0554	2,5218	0,1659
	100	0,5183	0,2143	0,2572	0,0198	1,2712	0,0489
3	30	1,0526	0,4332	0,5203	0,0365	2,3838	0,1639
	100	0,4986	0,2054	0,2444	0,0173	1,1217	0,0579
5	30	0,9644	0,3994	0,4622	0,0319	2,2165	0,1726
	100	0,4766	0,1978	0,2294	0,0152	0,9696	0,0588
7	30	0,9599	0,3965	0,4862	0,0299	1,8745	0,1774
	100	0,4204	0,1744	0,2114	0,0148	0,8936	0,0532
10	30	0,9538	0,3961	0,4611	0,0279	1,9371	0,2173
	100	0,4343	0,1806	0,2192	0,0138	0,8354	0,0595
20	30	1,3526	0,5634	0,5839	0,0265	1,7705	0,3124
	100	0,4827	0,1982	0,2282	0,0129	0,8637	0,0694
		D_p relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,2313	0,5147	0,7003	0,0522	3,2152	0,1746
	100	0,6014	0,2485	0,2937	0,0239	1,5388	0,0357
2	30	1,1455	0,4645	0,4767	0,0565	2,5776	0,1629
	100	0,5187	0,2145	0,2575	0,0189	1,2725	0,0488
3	30	1,0513	0,4336	0,5198	0,0357	2,3859	0,1577
	100	0,4988	0,2055	0,2447	0,0173	1,1206	0,0574
5	30	0,9653	0,3998	0,4627	0,0311	2,2182	0,1528
	100	0,4766	0,1979	0,2288	0,0152	0,9702	0,0573
7	30	0,9607	0,3968	0,4862	0,0298	1,8878	0,1468
	100	0,4208	0,1745	0,2114	0,0148	0,8945	0,0513
10	30	0,9496	0,3943	0,4616	0,0276	1,9368	0,1586
	100	0,4347	0,1808	0,2194	0,0138	0,8362	0,0556
20	30	1,3537	0,5638	0,5846	0,0261	1,7427	0,1974
	100	0,4832	0,1984	0,2284	0,0129	0,8628	0,0554

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.3: $\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição *a priori* exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.1$.

		$\sqrt{\text{EQM}}$ relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,3866	0,5708	0,7212	0,052	3,0736	0,5828
	100	0,5879	0,2448	0,2941	0,0255	1,5123	0,0393
2	30	1,0781	0,4419	0,5354	0,0393	2,4038	0,5744
	100	0,5257	0,2168	0,2481	0,0206	1,1616	0,0916
3	30	1,0095	0,4188	0,4537	0,0368	2,3191	0,4509
	100	0,4648	0,1931	0,2433	0,0176	1,0996	0,1114
5	30	0,9088	0,3772	0,4631	0,0312	2,0483	0,3657
	100	0,4598	0,1904	0,2201	0,0148	0,9678	0,1148
7	30	0,9366	0,3898	0,4121	0,0386	2,1298	0,2564
	100	0,4348	0,1785	0,2217	0,0151	0,9364	0,1349
10	30	0,8617	0,3562	0,4258	0,0281	1,8262	0,1994
	100	0,4337	0,1785	0,2181	0,0138	0,9253	0,1263
20	30	0,8022	0,3352	0,4156	0,0269	1,7869	0,1893
	100	0,4005	0,1658	0,1914	0,0129	0,8763	0,1057
		D_p relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,3854	0,5702	0,7218	0,0527	3,0767	0,5645
	100	0,5885	0,2451	0,2944	0,0253	1,5133	0,0394
2	30	1,0779	0,4427	0,5334	0,0393	2,4046	0,5374
	100	0,5261	0,2174	0,2483	0,0206	1,1627	0,0911
3	30	1,0085	0,4182	0,4522	0,0368	2,3215	0,4277
	100	0,4653	0,1933	0,2435	0,0176	1,0995	0,1095
5	30	0,9097	0,3776	0,4614	0,0361	2,0499	0,3574
	100	0,4599	0,1904	0,2202	0,0149	0,9684	0,1142
7	30	0,9367	0,3897	0,4111	0,0297	2,1309	0,2559
	100	0,4331	0,1778	0,2218	0,0151	0,9368	0,1342
10	30	0,8609	0,3559	0,4253	0,0282	1,8263	0,1996
	100	0,4341	0,1786	0,2182	0,0138	0,9209	0,1265
20	30	0,8019	0,3352	0,4158	0,0269	1,7886	0,1869
	100	0,4009	0,1663	0,1916	0,0129	0,8769	0,1053

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.4: $\sqrt{\text{EQM}}$ e Desvio padrão amostral relativos, considerando distribuição *a priori* exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.005$.

		$\sqrt{\text{EQM}}$ relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,2654	0,5274	0,7118	0,0494	3,3209	0,2445
	100	0,6027	0,2488	0,2885	0,0238	1,5571	0,0387
2	30	1,0772	0,4477	0,5823	0,0414	2,6943	2,3882
	100	0,5057	0,2099	0,2564	0,0202	1,1841	0,0891
3	30	1,0009	0,4106	0,4804	0,0337	2,3615	0,8306
	100	0,5054	0,2109	0,2359	0,0176	1,1315	0,2168
5	30	0,9427	0,3881	0,4391	0,0292	2,0151	0,2808
	100	0,4575	0,1883	0,2156	0,0151	0,9746	1,8794
7	30	0,9409	0,3897	0,4584	0,0282	1,9251	0,2259
	100	0,4751	0,1947	0,2066	0,0142	0,9539	0,8953
10	30	0,8853	0,3672	0,4314	0,0259	1,8537	0,1122
	100	0,4138	0,1722	0,2046	0,0143	0,8909	0,1025
20	30	0,8319	0,3463	0,3748	0,0251	1,8169	0,0226
	100	0,4007	0,1653	0,1996	0,0132	0,8316	0,0381
		D_p relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	1,2664	0,5279	0,7087	0,0494	3,2892	0,2312
	100	0,6033	0,2497	0,2876	0,0236	1,5585	0,0386
2	30	1,0779	0,4483	0,5829	0,0414	2,6951	2,3716
	100	0,5062	0,2101	0,2566	0,0202	1,1761	0,0887
3	30	1,0004	0,4103	0,4802	0,0335	2,3623	0,8271
	100	0,5056	0,2109	0,2362	0,0176	1,1289	0,2147
5	30	0,9426	0,3881	0,4394	0,0292	2,0162	0,2803
	100	0,4579	0,1885	0,2156	0,0151	0,9753	1,8779
7	30	0,9415	0,3958	0,4589	0,0282	1,9259	0,2258
	100	0,4751	0,1946	0,2066	0,0142	0,9504	0,8941
10	30	0,8857	0,3673	0,4318	0,0258	1,8536	0,1121
	100	0,4135	0,1719	0,2049	0,0142	0,8892	0,1026
20	30	0,8303	0,3457	0,3747	0,0251	1,8187	0,0226
	100	0,4012	0,1654	0,1996	0,0132	0,8314	0,0381

3.1. PROPRIEDADES FREQUENTISTAS

Analisando as Tabelas 3.1, 3.2, 3.3 e 3.4, observa-se que para todas as distribuições *a priori* consideradas os valores da \sqrt{EQM} e do D_p são próximos, mesmo quando o tamanho da amostra é pequena. Além disso, nota-se uma diminuição destas medidas com o aumento amostral, como esperado.

As Figuras 3.1 e 3.2 sintetizam as informações das Tabelas 3.1, 3.2, 3.3 e 3.4 e mostram a variação da \sqrt{EQM} relativo para amostras de tamanho 30 e 100, respectivamente. Os D_p relativos foram omitidos por apresentarem comportamento semelhante.

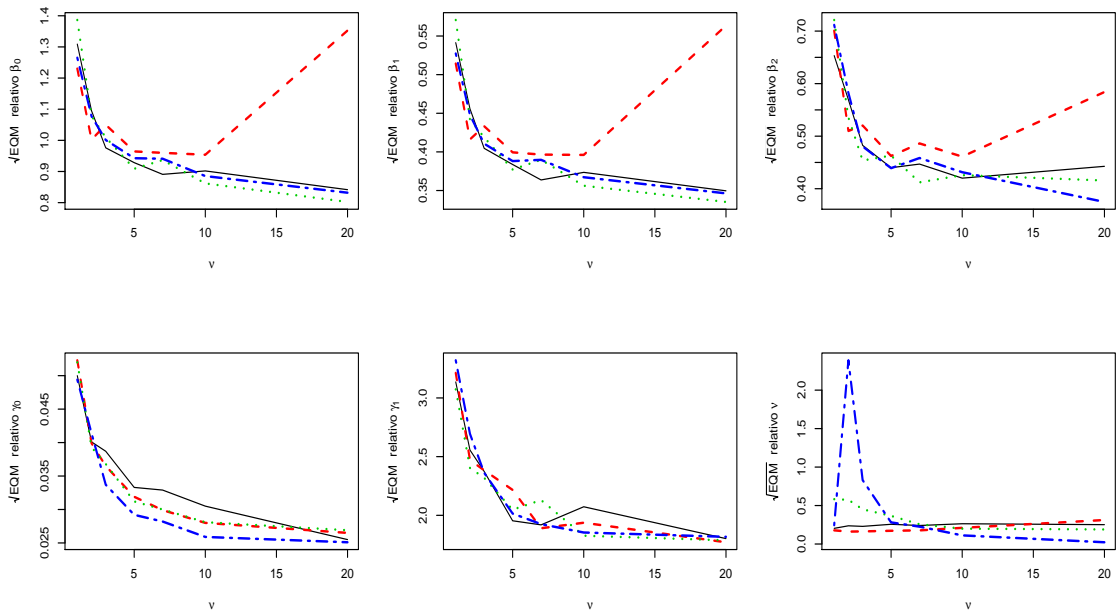


Figura 3.1: \sqrt{EQM} relativo simulado com amostras de tamanho $n = 30$, obtido utilizando as distribuições *a priori* de Jeffreys (linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul).

Percebe-se também maiores valores da \sqrt{EQM} e do D_p para graus de liberdade pequenos. Em geral, as distribuições *a priori* consideradas apresentaram comportamento parecidos. No entanto, a distribuição *a priori* exponencial de parâmetro $\lambda = 1$ apresentou maiores valores da \sqrt{EQM} e do D_p com relação as outras distribuições e um aumento quando $\nu < 5$ e $\nu > 10$.

3.1. PROPRIEDADES FREQUENTISTAS

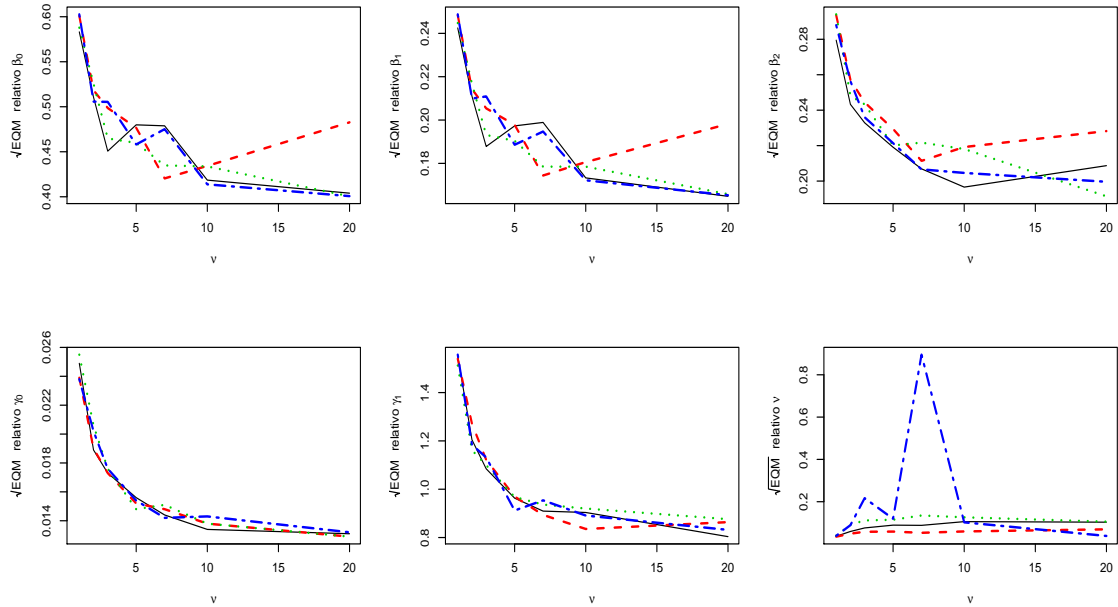


Figura 3.2: \sqrt{EQM} relativo simulado com amostras de tamanho $n = 100$, obtido utilizando as distribuições *a priori* de Jeffreys (linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul).

Nas Tabelas 3.5, 3.6, 3.7 e 3.8 apresentamos o viés relativo e a probabilidade de cobertura, considerando as distribuições *a priori* de Jeffreys, e as distribuições exponenciais com parâmetros $\lambda = 1$, $\lambda = 0.1$ e $\lambda = 0.05$ para amostras de tamanho 30 e 100. Indicaremos por 0 valores de viés relativo menor do que 10^{-4} .

Nota-se que mesmo para as amostras de tamanho 30, os valores do viés relativo são baixos e diminuem com o aumento do tamanho amostral. No geral as distribuições *a priori* consideradas apresentaram comportamento semelhante, no entanto algumas diferenças são notadas.

Para as distribuições *a priori* de Jeffreys os valores de viés relativo são bastante pequenos, característica que se mantém na medida em que se varia o grau de liberdade. Valores relativamente mais altos do que os demais, diminuem com o aumento da amostra.

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.5: Viés relativo e probabilidade de cobertura, considerando distribuição a priori de Jeffreys para os parâmetros do modelo linear heteroscedástico t-Student.

Viés relativo							
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	-0,0653	0,0228	0,0415	-0,0000	0,2161	0,0434
	100	0,0453	-0,0184	-0,0092	-0,0012	0,0114	0,0015
2	30	0,0000	-0,0031	0,0313	-0,0044	0,0522	0,0241
	100	-0,0265	0,0107	0,0095	0,0000	0,0000	0,0000
3	30	0,0423	-0,0163	0,0025	-0,0061	0,0167	0,0000
	100	-0,0243	0,0124	0,0016	0,0000	-0,0766	0,0012
5	30	0,0583	-0,0243	-0,0047	-0,0049	0,0275	-0,0193
	100	-0,0555	0,0232	-0,0083	-0,0012	0,0081	-0,0156
7	30	0,0085	-0,0021	-0,0102	-0,0056	-0,0308	-0,0469
	100	0,0193	-0,0072	-0,0058	-0,0000	0,0475	-0,0123
10	30	-0,0297	0,0131	-0,0036	-0,0049	-0,0893	-0,0524
	100	-0,0046	0,0011	0,0013	-0,0019	-0,0049	-0,0072
20	30	0,0015	-0,0038	0,0000	-0,0036	-0,0493	-0,0811
	100	0,0236	-0,0084	-0,0168	0,0000	0,0113	-0,0087
Probabilidade de Cobertura							
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	0,946	0,954	0,940	0,956	0,956	0,948
	100	0,952	0,954	0,942	0,946	0,948	0,940
2	30	0,942	0,946	0,938	0,942	0,940	0,952
	100	0,954	0,948	0,952	0,960	0,940	0,932
3	30	0,932	0,934	0,938	0,934	0,946	0,940
	100	0,952	0,948	0,946	0,948	0,958	0,936
5	30	0,942	0,944	0,944	0,938	0,972	0,928
	100	0,922	0,924	0,962	0,962	0,942	0,962
7	30	0,932	0,938	0,942	0,932	0,952	0,920
	100	0,920	0,912	0,942	0,952	0,948	0,940
10	30	0,930	0,926	0,922	0,938	0,940	0,904
	100	0,966	0,956	0,954	0,956	0,948	0,946
20	30	0,936	0,938	0,918	0,938	0,946	0,896
	100	0,952	0,960	0,942	0,936	0,952	0,946

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.6: Viés relativo e probabilidade de cobertura, considerando a distribuição *a priori* exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 1$.

Viés relativo							
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	0,0356	-0,0127	-0,0409	-0,0037	0,0997	0,0336
	100	0,0153	-0,0077	0,0111	-0,0000	0,0917	0,0000
2	30	0,0055	-0,0018	-0,0223	-0,0067	-0,1404	-0,0172
	100	-0,0085	0,0038	-0,0041	-0,0015	-0,0103	-0,0037
3	30	0,0077	-0,0037	-0,0325	-0,0079	-0,0346	-0,0452
	100	0,0183	-0,0082	-0,0029	-0,0000	0,0703	-0,0079
5	30	0,0038	-0,0021	0,0054	-0,0076	-0,0464	-0,0792
	100	-0,0208	0,0064	0,0192	0,0000	-0,0264	-0,0136
7	30	0,0212	-0,0053	-0,0273	-0,0027	-0,1251	-0,0997
	100	-0,0071	0,0044	-0,0094	-0,0000	0,0034	-0,0136
10	30	0,0993	-0,0411	0,0029	-0,0052	-0,0942	-0,1378
	100	-0,0027	0,0000	0,0000	-0,0000	0,0033	-0,0235
20	30	0,0289	-0,0149	-0,001	0,0047	0,3221	-0,2428
	100	0,0000	0,0000	-0,0000	-0,0000	0,0559	-0,0423
Probabilidade de Cobertura							
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	0,938	0,938	0,928	0,940	0,938	0,924
	100	0,946	0,950	0,942	0,962	0,936	0,950
2	30	0,948	0,946	0,944	0,954	0,958	0,940
	100	0,936	0,938	0,948	0,946	0,930	0,948
3	30	0,912	0,912	0,932	0,936	0,950	0,914
	100	0,938	0,938	0,924	0,934	0,940	0,916
5	30	0,916	0,910	0,914	0,932	0,930	0,904
	100	0,920	0,912	0,924	0,948	0,958	0,910
7	30	0,914	0,916	0,898	0,932	0,956	0,838
	100	0,948	0,944	0,944	0,948	0,958	0,946
10	300	0,880	0,884	0,872	0,940	0,936	0,740
	100	0,936	0,934	0,928	0,954	0,968	0,918
20	300	0,708	0,708	0,796	0,942	0,964	0,422
	100	0,886	0,894	0,902	0,952	0,930	0,834

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.7: Viés relativo e probabilidade de cobertura, considerando a distribuição *a priori* exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.1$.

Viés relativo							
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	-0,0837	0,0357	-0,0138	0,0031	0,0115	0,1472
	100	0,0029	-0,0012	0,0000	-0,0026	0,0393	0,0000
2	30	-0,0497	0,0169	0,0551	0,0017	0,0878	0,2044
	100	-0,0107	0,0048	-0,0039	-0,0000	-0,0094	0,0106
3	30	-0,0623	0,0286	-0,0412	-0,0000	-0,0056	0,1442
	100	-0,0015	0,0000	0,0013	-0,0000	-0,0521	0,0211
5	30	0,0099	-0,0000	-0,0446	-0,0035	-0,0405	0,0792
	100	-0,0162	0,0068	0,0071	-0,0000	0,0259	0,0128
7	30	-0,0403	0,0189	-0,0344	-0,0045	0,0659	0,0195
	100	0,0434	-0,0179	-0,0054	-0,0000	-0,0318	0,0147
10	30	0,0516	-0,0193	-0,0289	-0,0027	0,0784	0,0057
	100	-0,0066	0,0026	0,0066	-0,0000	-0,0063	-0,0000
20	30	-0,0387	0,0149	0,0103	-0,0016	0,0269	-0,0314
	100	0,0015	-0,0000	0,0016	-0,0000	0,0084	-0,0107
Probabilidade de Cobertura							
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	0,946	0,948	0,936	0,952	0,950	0,910
	100	0,930	0,934	0,948	0,950	0,962	0,942
2	30	0,944	0,946	0,936	0,944	0,972	0,950
	100	0,946	0,942	0,956	0,940	0,952	0,946
3	30	0,938	0,944	0,960	0,942	0,962	0,966
	100	0,962	0,954	0,944	0,934	0,940	0,964
5	30	0,944	0,936	0,942	0,936	0,954	0,944
	100	0,948	0,944	0,944	0,972	0,940	0,940
7	30	0,916	0,920	0,944	0,952	0,948	0,920
	100	0,956	0,962	0,920	0,934	0,950	0,946
10	30	0,948	0,940	0,928	0,946	0,948	0,944
	100	0,930	0,928	0,932	0,942	0,940	0,936
20	30	0,936	0,940	0,928	0,930	0,938	0,920
	100	0,948	0,950	0,958	0,952	0,940	0,954

3.1. PROPRIEDADES FREQUENTISTAS

Tabela 3.8: a distribuição *a priori* exponencial para os parâmetros do modelo linear heteroscedástico t-Student, com $\lambda = 0.05$.

		Viés relativo					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	-0,0223	-0,0000	0,0729	0,0015	0,4814	0,0823
	100	0,0182	-0,0056	-0,0245	-0,0028	-0,0268	-0,0028
2	30	0,0277	-0,0111	0,0059	-0,0025	0,1004	0,3009
	100	-0,0062	0,0031	0,0021	0,0015	0,1467	0,0096
3	30	-0,0568	0,0232	0,0256	-0,0035	-0,0859	0,0848
	100	0,0191	-0,0089	0,0046	-0,0000	0,0922	0,0315
5	30	-0,0427	0,0171	0,0133	-0,0013	0,0599	0,0211
	100	-0,0024	-0,0000	0,0093	-0,0000	0,0259	0,1121
7	30	0,0273	-0,0126	-0,0053	0,0000	-0,0654	0,0125
	100	-0,0204	0,0099	-0,0089	-0,0000	-0,0931	0,0609
10	30	0,0294	-0,0116	0,0012	-0,0024	-0,0961	0,0064
	100	-0,0249	0,0104	0,0000	0,0000	0,0689	0,0031
20	30	0,0641	-0,0248	-0,0183	0,0000	-0,0198	0,0000
	100	0,0074	-0,0039	-0,0069	-0,0000	0,0407	0,0017
		Probabilidade de Cobertura					
ν real	n	β_0	β_1	β_2	γ_0	γ_1	ν
1	30	0,958	0,966	0,924	0,944	0,934	0,942
	100	0,940	0,948	0,942	0,958	0,944	0,946
2	30	0,954	0,942	0,916	0,944	0,946	0,912
	100	0,946	0,946	0,936	0,948	0,960	0,950
3	30	0,940	0,938	0,930	0,948	0,924	0,920
	100	0,932	0,936	0,946	0,940	0,944	0,942
5	30	0,936	0,938	0,934	0,950	0,942	0,944
	100	0,942	0,942	0,948	0,952	0,942	0,934
7	30	0,928	0,920	0,926	0,938	0,938	0,910
	100	0,908	0,914	0,968	0,948	0,938	0,934
10	30	0,930	0,924	0,922	0,936	0,940	0,948
	100	0,950	0,946	0,948	0,930	0,936	0,950
20	30	0,950	0,948	0,942	0,938	0,934	0,962
	100	0,932	0,928	0,938	0,930	0,928	0,946

Com relação as distribuições exponenciais, percebe-se que para $\lambda = 1$, o viés relativo é maior comparado à distribuição *a priori* de Jeffreys quando o grau de liberdade é grande e o tamanho de amostra é pequeno. Ambas distribuições subestimam o parâmetro grau de liberdade conforme este aumenta.

Para as distribuições exponenciais com parâmetros $\lambda = 0.1$ e $\lambda = 0.005$, os valores de viés relativos são maiores quando temos graus de liberdade e amostras pequenos. Além disso, tais distribuições sobrestimam ν , principalmente quando $\lambda = 0.005$. Isto pode ser um indicativo de que distribuições deste tipo com médias elevadas, podem afetar na robustez do modelo.

As probabilidades de cobertura apresentadas nas Tabelas 3.5, 3.6, 3.7 e 3.8, se encontram resumidas nas Figuras 3.3 e 3.4. Observa-se que a distribuição exponencial de parâmetro $\lambda = 1$ possui baixa cobertura, quando o tamanho da amostra é pequeno. O aumento amostral melhora esta característica mas não resolve o problema para o grau de liberdade quando este assume valores altos.

Estes resultados mostram que as distribuições *a priori* utilizadas apresentam boas propriedades frequentistas. As distribuições exponenciais se desempenharam bem, mas com algumas ressalvas, como mencionamos. Percebemos que no caso destas distribuições, os resultados obtidos poderiam ser melhorados especificando-se convenientemente o valor do parâmetro λ . Entre os escolhidos para o estudo de simulação, o que apresentou melhor comportamento foi $\lambda = 0.1$.

As distribuições *a priori* de Jeffreys obtiveram um bom desempenho, e estabilidade ao longo da variação do grau de liberdade. Além disso, possui a vantagem de ser não informativa, não havendo a necessidade de se especificar parâmetro.

3.1. PROPRIEDADES FREQUENTISTAS

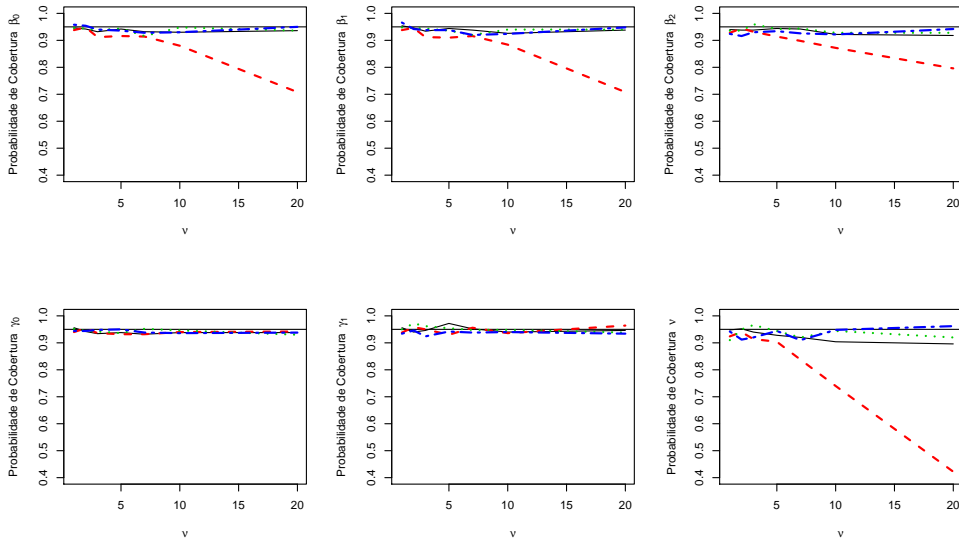


Figura 3.3: Probabilidade de cobertura com amostras de tamanho $n = 30$, obtida utilizando as distribuições *a priori* de Jeffreys (linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul).

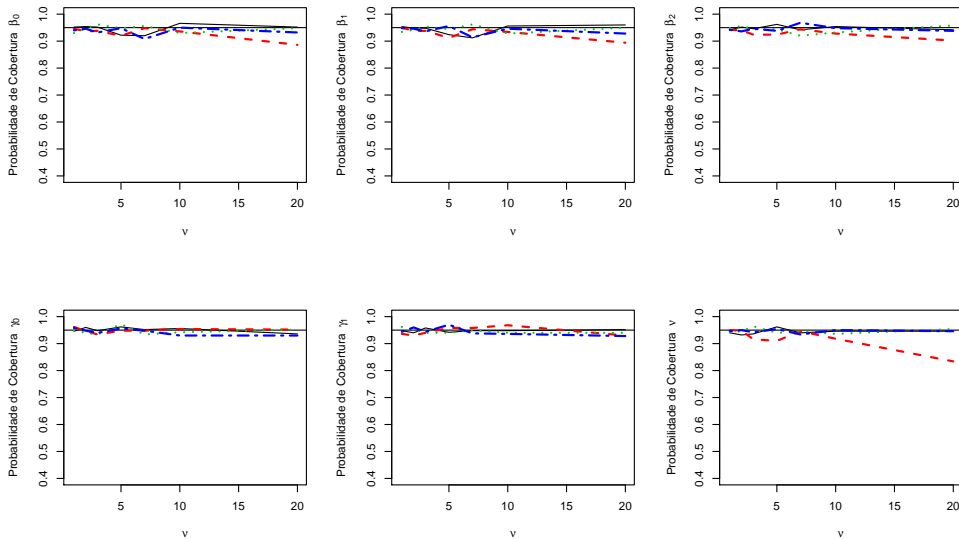


Figura 3.4: Probabilidade de cobertura com amostras de tamanho $n = 100$, obtida utilizando as distribuições *a priori* de Jeffreys (linha sólida preta), e as distribuições exponenciais com parâmetros $\lambda = 1$ (linha tracejada vermelha), $\lambda = 0.1$ (linha pontilhada verde) e $\lambda = 0.005$ (linha tracejada e pontilhada azul).

3.2 Robustez das estimativas em relação à perturbações na variável resposta

Nesta seção temos como objetivo avaliar o desempenho das distribuições *a priori* consideradas na Seção 2.1, na detecção de pontos influentes. Utilizamos nesta análise as distribuições *a priori* de Jeffreys e as distribuições *a priori* $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \nu) \propto \lambda \exp(-\lambda \nu)$, com parâmetros $\lambda = 1$, $\lambda = 0.1$, e $\lambda = 0.005$.

Para tanto, consideramos o ajuste do modelo linear heteroscedástico t-Student ao conjunto de dados artificiais obtido a partir do modelo linear heteroscedástico, cujos erros têm distribuição normal. Foram simuladas 100 observações, utilizando os seguintes valores para os parâmetros: $\beta_0 = 3$, $\beta_1 = 5$, $\beta_2 = 1.5$, $\gamma_0 = 1$ e $\gamma_1 = -0.006$. As variáveis explicativas foram obtidas através da distribuição Normal(0,1).

A convergência das cadeias geradas foi verificada através da estatística de Geweke, e os histogramas, gráficos das cadeias e de autocorrelação se encontram expostos nas Figuras 3.5, 3.6, 3.7 e 3.8.

Na Tabela 3.9 apresentamos as estimativas para os parâmetros do modelo ajustado, obtidas através da média *a posteriori*, os desvios padrão *a posteriori* e os intervalos interquantil de credibilidade 95%, para cada uma das distribuições *a priori* consideradas.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

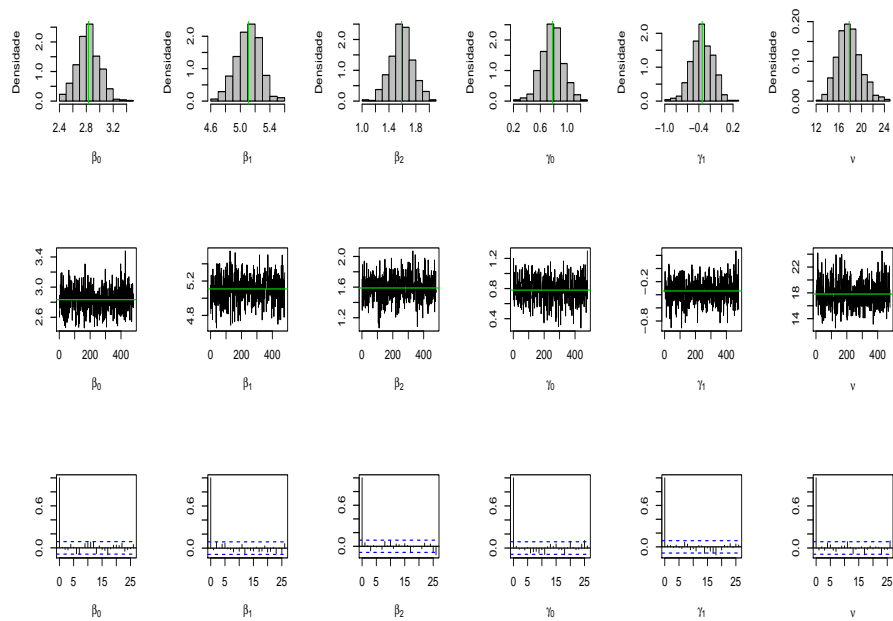


Figura 3.5: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Jeffreys, considerando o conjunto de dados na gerado Seção 3.2, sem perturbações nas observações.

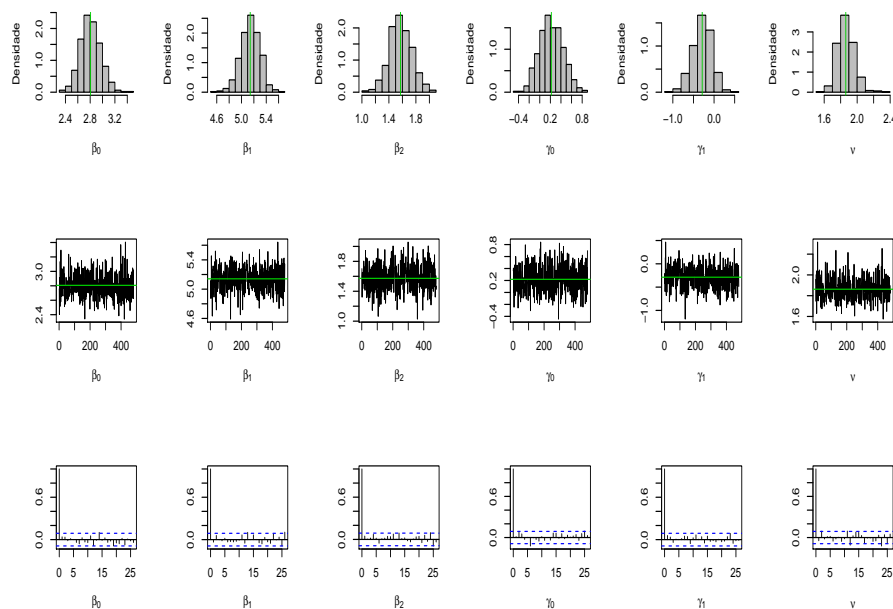


Figura 3.6: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Geweke com $\lambda = 1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

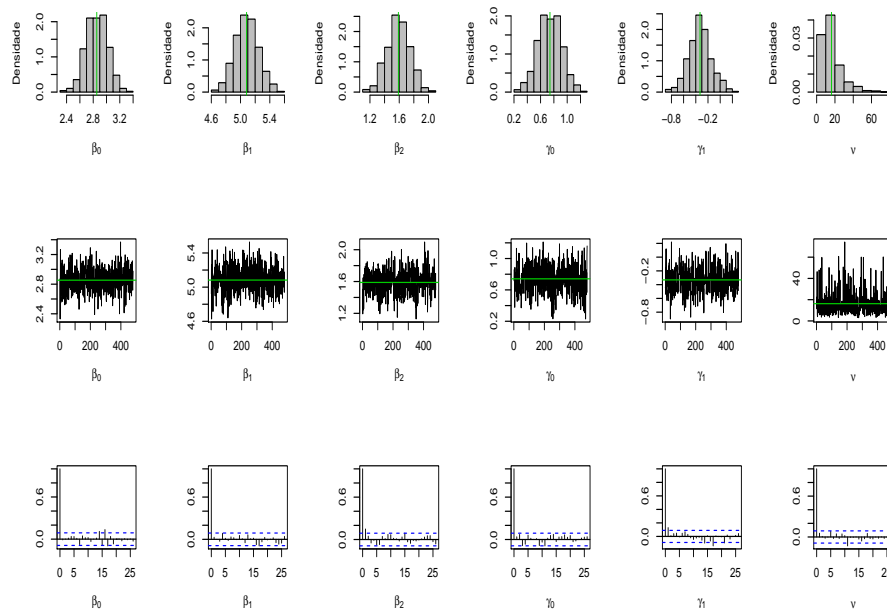


Figura 3.7: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Geweke com $\lambda = 0.1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.

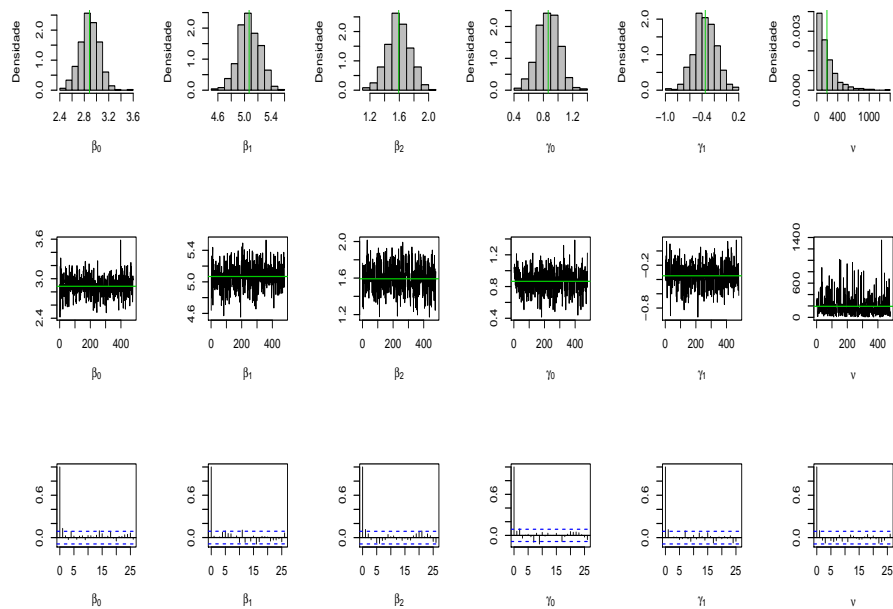


Figura 3.8: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Geweke com $\lambda = 0.005$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

Tabela 3.9: Estimativas pontuais, desvios padrão *a posteriori* e intervalos de credibilidade 95% para os parâmetros dos modelo linear heteroscedástico, considerando diferentes distribuições *a priori*.

Distribuição <i>a priori</i>	Parâmetros	Média <i>a posteriori</i>	Desvio Padrão <i>a posteriori</i>	Intervalo interquantil de credibilidade 95%
Jeffreys	β_0	2,8331	0,1642	(2,5211; 3,1491)
	β_1	5,1080	0,1625	(4,7693; 5,3963)
	β_2	1,5880	0,1612	(1,2594; 1,9253)
	γ_0	0,7793	0,1618	(0,4415; 1,099)
	γ_1	-0,3421	0,1858	(-0,7156; -0,0102)
	ν	17,8160	2,0691	(14,1892; 22,1469)
Geweke $\lambda = 1$	β_0	2,8064	0,1635	(2,5079; 3,1568)
	β_1	5,1410	0,1561	(4,8277; 5,4160)
	β_2	1,5724	0,1671	(1,2509; 1,9136)
	γ_0	0,2189	0,2209	(-0,1838; 0,6432)
	γ_1	-0,2877	0,2310	(-0,7256; 0,1315)
	ν	1,8620	0,1013	(1,6854; 2,0648)
Geweke $\lambda = 0, 1$	β_0	2,8537	0,1638	(2,5447; 3,1789)
	β_1	5,0825	0,1583	(4,7870; 5,4061)
	β_2	1,5885	0,1572	(1,2813; 1,8856)
	γ_0	0,7411	0,1770	(0,3816; 1,0968)
	γ_1	-0,3308	0,1854	(-0,6947; 0,0625)
	ν	16,2505	10,9703	(4,3515; 47,6964)
Geweke $\lambda = 0, 005$	β_0	2,8848	0,1569	(2,5763; 3,1900)
	β_1	5,0679	0,1609	(4,7559; 5,3746)
	β_2	1,5920	0,1551	(1,2908; 1,8908)
	γ_0	0,8658	0,1493	(0,5523; 1,1412)
	γ_1	-0,3480	0,1785	(-0,6721; -0,0016)
	ν	193,6025	187,2228	(12,7770; 730,1142)

Observamos estimativas semelhantes para o vetor de parâmetros β , com valores estimados próximos dos utilizados para geração dos dados e intervalos de credibilidade de amplitude pequena, para todas as distribuições *a priori* consideradas. Estimativas similares também podem ser notadas em relação ao vetor γ , com exceção às obtidas utilizando a distribuição *a priori* exponencial com $\lambda = 1$.

A diferença entre as distribuições *a priori* utilizadas é acentuada quando analisamos as estimativas obtidas para o grau de liberdade. Para as distribuições exponenciais com parâmetros $\lambda = 0.1$ e $\lambda = 0.005$ os valores estimados de ν são maiores quando comparados ao uso de $\lambda = 1$, ou seja, dessas distribuições a que melhor consegue

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

detectar a normalidade dos dados é a de maior média.

Este fato pode ser verificado na Tabela 3.10, na qual apresentamos as estimativas bayesianas dos critérios EAIC, EBIC, DIC, e LPML. O pior desempenho, segundo estas medidas, é atribuído à distribuição *a priori* exponencial com $\lambda = 1$, a qual fornece a menor estimativa para o grau de liberdade.

Tabela 3.10: Estimativas dos critérios de seleção bayesianos para os modelos ajustados.

Distribuição <i>a priori</i>	EAIC	EBIC	DIC	LPML
Jeffreys	386,0727	401,7037	379,3583	-189,7108
Exponencial(1)	401,2116	416,8426	394,3685	-197,4253
Exponencial(0, 1)	386,4532	402,0842	380,1749	-189,7568
Exponencial(0, 005)	385,6782	401,3092	378,7197	-189,3590

Ainda de acordo com estes critérios, as distribuições restantes não exibem diferenças significativas. A distribuição *a priori* de Jeffreys no entanto, apresenta a vantagem de ser não informativa, não havendo a necessidade em se ter conhecimento da normalidade dos dados para se obter boas estimativas, como nos outros casos.

Para verificar o desempenho das distribuições *a priori* na presença de observações atípicas, geramos um outlier no conjunto de dados simulado, perturbando a observação 1 da seguinte maneira,

$$y_1 = \max(y) + 3S_y.$$

A convergência das cadeias geradas foi verificada através da estatística de Geweke, e os gráficos dos histogramas, das cadeias e de autocorrelação se encontram expostos nas Figuras 3.9, 3.10, 3.11 e 3.12. Os resultados obtidos estão sintetizados na Tabela 3.11. Novamente, observamos estimativas semelhantes para o vetor de parâmetros β quando comparamos as quatro distribuições *a priori* utilizadas.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

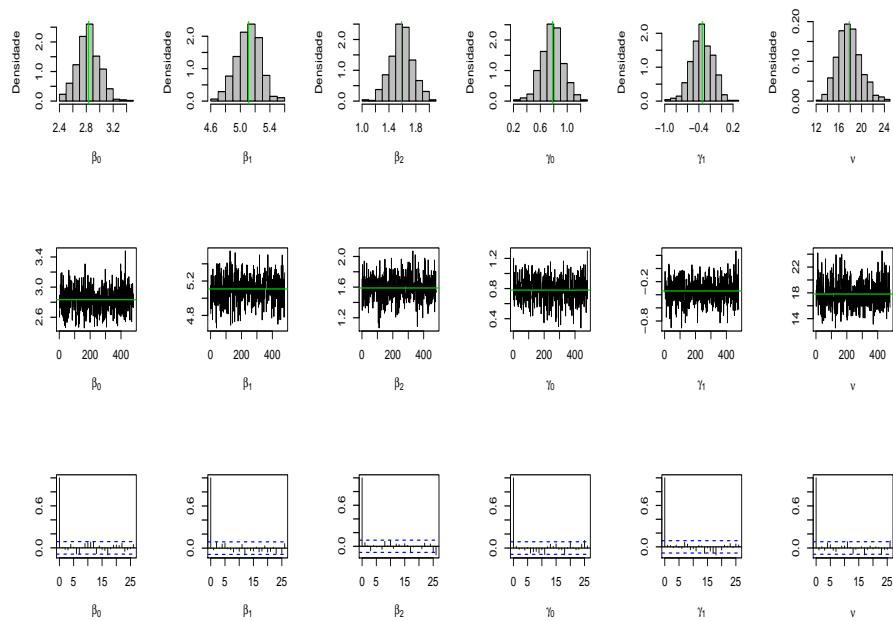


Figura 3.9: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Jeffreys, considerando o conjunto de dados na gerado Seção 3.2.

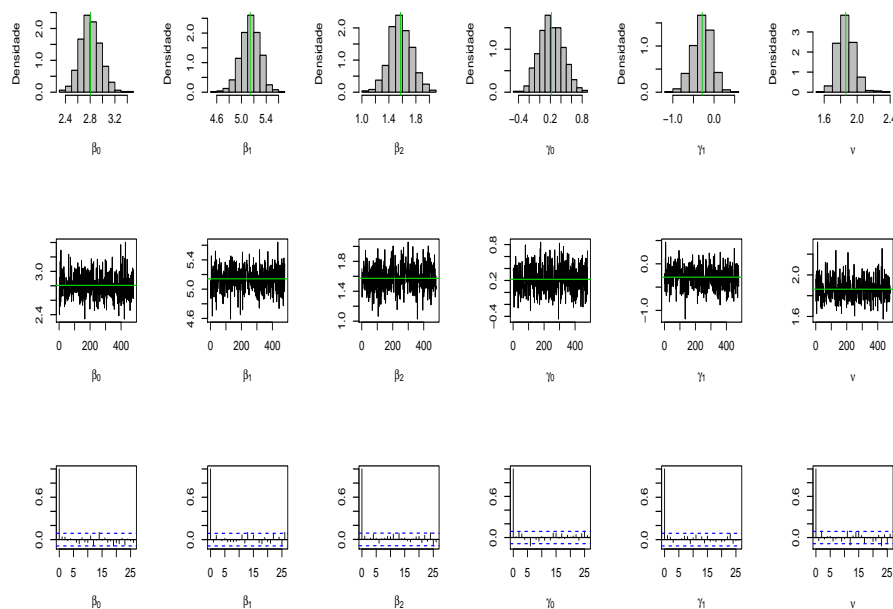


Figura 3.10: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Geweke com $\lambda = 1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

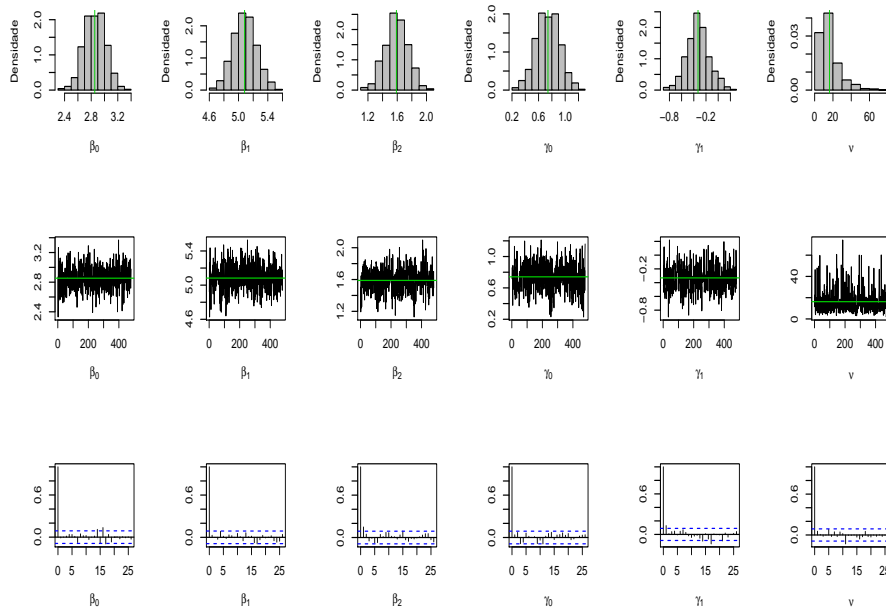


Figura 3.11: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Geweke com $\lambda = 0.1$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.

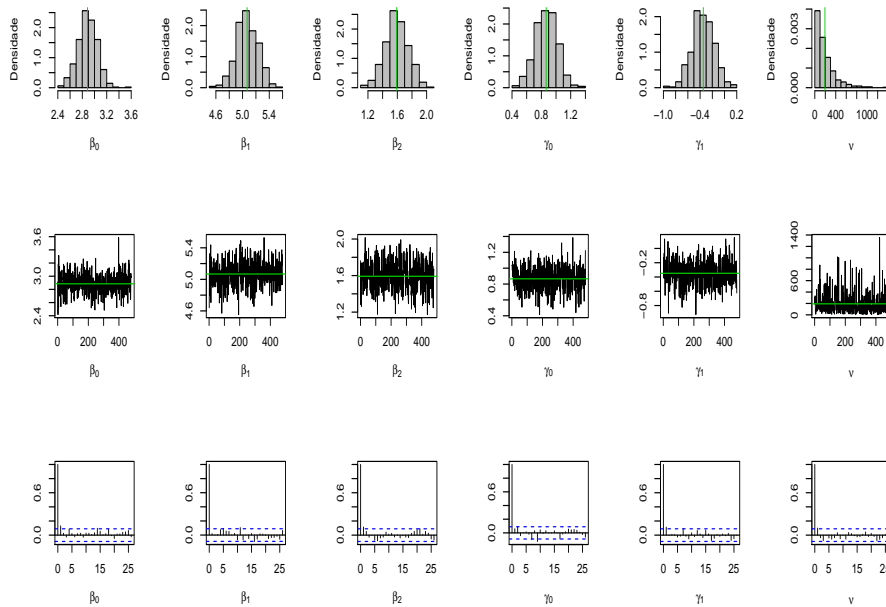


Figura 3.12: Histogramas, gráficos das cadeias geradas e de autocorrelação para o modelo linear heteroscedástico t-Student com distribuição *a priori* exponencial com $\lambda = 0.005$, considerando o conjunto de dados gerado na Seção 3.2, sem perturbações nas observações.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

Tabela 3.11: Estimativas pontuais, desvios padrão *a posteriori* e intervalos de credibilidade 95% para os parâmetros dos modelo linear heteroscedástico, considerando diferentes distribuições *a priori*.

Distribuição <i>a priori</i>	Parâmetros	Média <i>a posteriori</i>	Desvio Padrão <i>a posteriori</i>	Intervalo interquartil de credibilidade 95%
Jeffreys	β_0	2,7994	0,1653	(2,5037; 3,1289)
	β_1	5,1801	0,1564	(4,8458; 5,4434)
	β_2	1,5765	0,1722	(1,2167; 1,9046)
	γ_0	0,635	0,1931	(0,2569; 1,0112)
	γ_1	-0,2239	0,2059	(-0,6265; 0,1773)
	ν	4,6097	0,2642	(4,1599; 5,1716)
Geweke $\lambda = 1$	β_0	2,8195	0,1702	(2,4873; 3,1772)
	β_1	5,1764	0,1713	(4,8139; 5,4993)
	β_2	1,5643	0,1947	(1,1833; 1,9459)
	γ_0	-0,1169	0,2708	(-0,666; 0,3925)
	γ_1	-0,2173	0,2774	(-0,7414; 0,3066)
	ν	0,9354	0,0515	(0,8468; 1,0532)
Geweke $\lambda = 0, 1$	β_0	2,8023	0,1674	(2,4779; 3,1509)
	β_1	5,1639	0,1597	(4,8609; 5,4744)
	β_2	1,5816	0,1731	(1,2475; 1,9254)
	γ_0	0,6353	0,1952	(0,2469; 1,0166)
	γ_1	-0,2333	0,2119	(-0,6407; 0,1298)
	ν	4,637	0,2666	(4,1709; 5,1951)
Geweke $\lambda = 0, 005$	β_0	2,8163	0,1812	(2,4814; 3,1774)
	β_1	5,2055	0,1741	(4,8502; 5,5334)
	β_2	1,5998	0,1907	(1,2077; 1,9762)
	γ_0	1,0609	0,1712	(0,6970; 1,3798)
	γ_1	-0,0427	0,2177	(-0,4776; 0,3795)
	ν	18,3721	1,0931	(16,3692; 20,6609)

O mesmo acontece com o vetor γ , com exceção da distribuição exponencial de parâmetro $\lambda = 0.005$. Note que esta distribuição *a priori* possui média bastante elevada, o que contribui para o aumento da média *a posteriori* do grau de liberdade e portanto, da estimativa deste parâmetro. Por consequência, temos neste caso um modelo menos robusto, como veremos a seguir. Analogamente, ao se considerar o parâmetro $\lambda = 1$ obtemos uma estimativa menor para ν .

A distribuições *a priori* que fornecem os melhores ajustes, segundo as estimativas dos critérios de seleção bayesianos EAIC, EBIC, DIC e LPML, apresentados na Tabela 3.12, são as de Jeffreys e exponencial com parâmetro $\lambda = 0.1$.

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

Tabela 3.12: Estimativas dos critérios de seleção bayesianos para os modelos ajustados.

Distribuição <i>a priori</i>	EAIC	EBIC	DIC	LPML
Jeffreys	415,8046	431,4356	409,0372	-204,6254
Geweke $\lambda = 1$	436,9806	452,6117	429,8739	-216,1938
Geweke $\lambda = 0,1$	415,9170	431,5480	409,2907	-204,7633
Geweke $\lambda = 0,005$	444,9208	460,5519	438,0024	-221,5525

Para analisar a influência da observação perturbada sob o ajuste do modelo, utilizamos a estimativa Monte Carlo da medida de divergência de K-L, bem como o logaritmo das ordenadas preditivas condicionais.

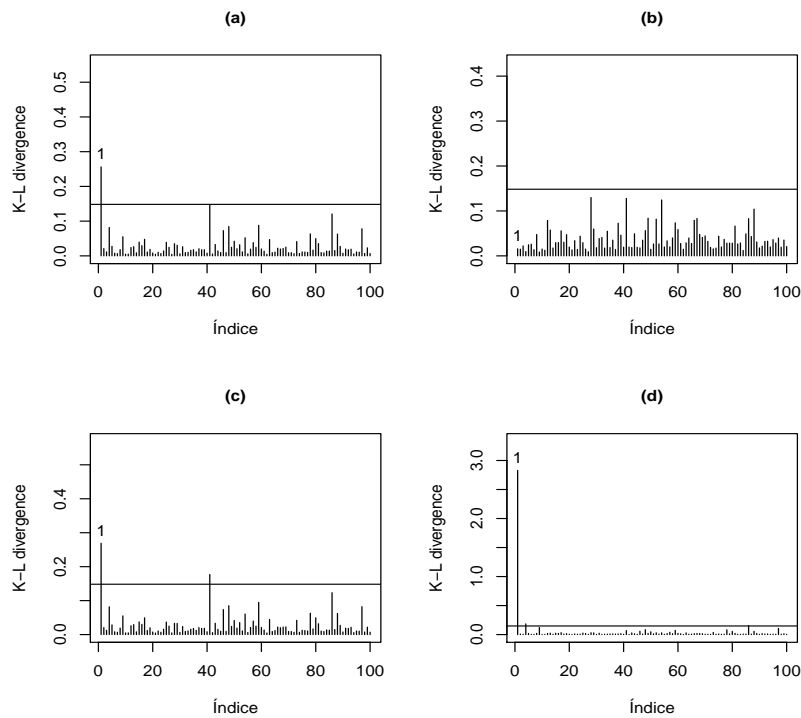


Figura 3.13: Estimativa Monte Carlo da medida de divergência de K-L, considerando as distribuições *a priori* de Jeffreys (a) e exponenciais com parâmetros $\lambda = 1$ (b), $\lambda = 0.1$ (c) e $\lambda = 0.005$ (d).

Verifica-se nas Figura 3.13 (a), (c) e (d) que a observação perturbada é detectada como possivelmente influente quando utilizamos as distribuições *a priori* de Jeffreys e exponenciais com parâmetros $\lambda = 0.1$ e $\lambda = 0.005$, respectivamente. No entanto, tal

3.2. ROBUSTEZ DAS ESTIMATIVAS EM RELAÇÃO À PERTURBAÇÕES NA VARIÁVEL RESPOSTA

medida é consideravelmente maior para o ultimo caso. Por estimar um grau de liberdade de menor valor, a distribuição *a priori* exponencial com $\lambda = 1$ possui observações com medida de K-L abaixo do ponto de corte, como vemos na Figura 3.13 (b).

Nota-se também na Figura 3.14 que a observação 1 tem ordenada preditiva condicional muito baixa em comparação com as restantes, sendo ainda menor para a distribuição exponencial de parâmetro $\lambda = 0.005$, como mostra a Figura 3.14 (e).

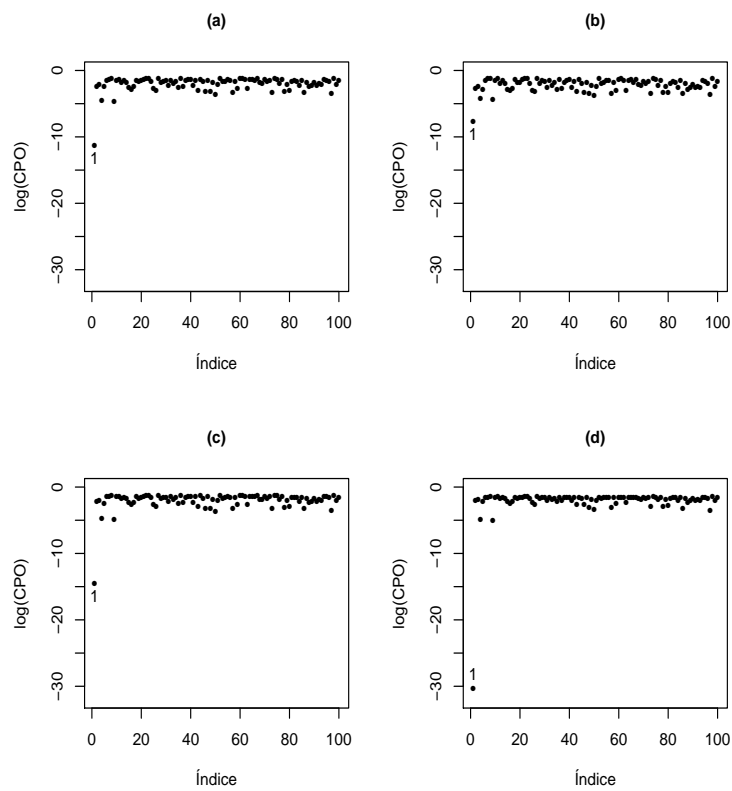


Figura 3.14: Logaritmo das ordenadas preditivas condicionais para o conjunto de dados simulado, considerando as distribuições *a priori* de Jeffreys (a) e exponenciais com parâmetros $\lambda = 1$ (b), $\lambda = 0.1$ (c) e $\lambda = 0.005$ (d).

Para os dois conjuntos de dados simulados, constatamos que a distribuição *a priori* de Jeffreys apresentou boa performance, com resultados equivalentes aos obtidos quando se utilizavam distribuições *a priori* informativas.

3.3 Aplicação dados de pesquisa salarial

Em modelos de regressão, o uso da distribuição normal na modelagem dos erros pode não ser adequado, podendo apresentar distorções nos resultados das análises quando o conjunto de dados contém alguma observação atípica. Ao acrescentarmos a estes termos a suposição de heteroscedasticidade, ainda percebemos sensibilidade do modelo em relação aos outliers, como pode ser visto nas Figuras 2.3 (a) e (b).

Para mostrar a superioridade do modelo linear heteroscedástico t-Student em situações semelhantes, apresentamos o seu ajuste ao conjunto de dados de pesquisa salarial descrito na Seção 2.1. Escolhemos uma observação para ser perturbada com o objetivo de verificar a influência que este ponto causará nas estimativas. Na prática, este cenário pode ocorrer devido a natureza dos dados (como por exemplo, um funcionário cujo salário é significativamente maior do que o dos demais), à erros de medição durante a coleta dos dados ou ainda, por erros humanos como a digitação de dados incorretos. Desta forma, consideramos a seguinte perturbação da observação 13: $y_{13} = \max(y) + S_y$.

Apresentamos também nesta aplicação, o ajuste do modelo linear heteroscedástico com erros normal e utilizamos a distribuição *a priori* de Jeffreys para os dois modelos considerados.

A convergência das cadeias geradas foi verificada através da estatística de Geweke. Os histogramas, os gráficos das cadeias e de autocorrelação se encontram expostos nas Figuras 3.15 e 3.16.

Na Tabela 3.13 apresentamos as estimativas para os parâmetros do modelo ajustado, obtidas através da média *a posteriori*, os desvios padrão *a posteriori* e os intervalos interquantil de credibilidade 95%, para cada um dos modelos considerados.

3.3. APLICAÇÃO DADOS DE PESQUISA SALARIAL

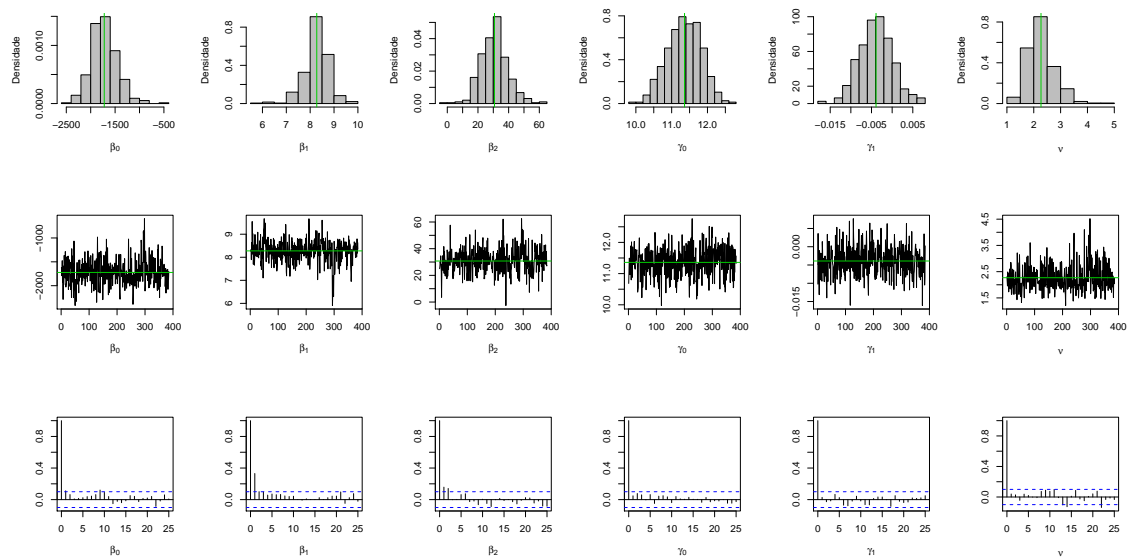


Figura 3.15: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico t-Student e distribuição *a priori* de Jeffreys, considerando o conjunto de dados da Seção 2.3.

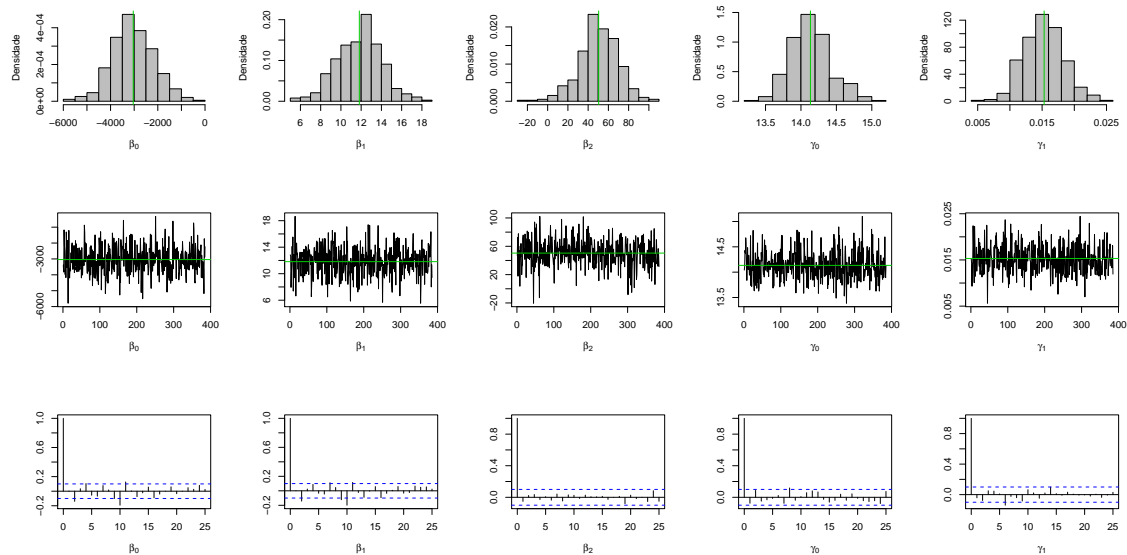


Figura 3.16: Histogramas, gráficos da cadeia e de autocorrelação para o modelo linear heteroscedástico com erros normal e distribuição *a priori* de Jeffreys, considerando o conjunto de dados da Seção 2.3.

3.3. APLICAÇÃO DADOS DE PESQUISA SALARIAL

Tabela 3.13: Estimativas pontuais, desvios padrão *a posteriori* e intervalos de credibilidade 95% para os parâmetros dos modelo linear heteroscedástico, considerando erros com distribuição t-Student e distribuição normal.

Distribuição dos Erros	Parâmetros	Média <i>a posteriori</i>	Desvio Padrão <i>a posteriori</i>	Intervalo interquantil de credibilidade 95%
t-Student	β_0	-1722,3354	269,9059	(-2249,7994; -1130,1691)
	β_1	8,2767	0,5065	(7,2266; 9,1782)
	β_2	30,7617	8,8939	(15,5295; 49,1652)
	γ_0	11,3595	0,4716	(10,4415; 12,2365)
	γ_1	-0,0039	0,0047	(-0,0113; 0,0044)
	ν	2,2697	0,4805	(1,4959; 3,3241)
Normal	β_0	-3039,7693	891,2922	(-4637,9496; -1297,5895)
	β_1	11,8188	2,1922	(7,5276; 16,1763)
	β_2	50,3971	19,1372	(9,6921; 86,5828)
	γ_0	14,1337	0,2821	(13,6288; 14,7441)
	γ_1	0,0153	0,0032	(0,0097; 0,0214)

Observamos que na presença do outlier, os dois modelos utilizados fornecem análises diferentes. Quando se considera a distribuição normal para os erros, o valor médio estimado para o salário dos funcionários é $\widehat{y}_{med} = 3515.515$. Esta quantidade é relativamente maior do que a obtida através do modelo t-Student, o qual fornece $\widehat{y}_{med} = 2820.032$.

Para a seleção dos modelos empregados neste estudo, utilizamos as estimativas dos critérios bayesianos EAIC, EBIC, DIC e LPML, expostas na Tabela 3.14. Podemos verificar que segundo todas estas medidas, o melhor ajuste aos dados é alcançado através do uso da distribuição t-Student.

Tabela 3.14: Estimativas dos critérios de seleção bayesianos para os modelos ajustados

Distribuição dos erros	EAIC	EBIC	DIC	LPML
t-Student	498,3884	506,9923	491,3877	-246,5252
Normal	536,8764	544,0464	531,9024	-270,3020

3.3. APLICAÇÃO DADOS DE PESQUISA SALARIAL

Para a analisar a adequabilidade dos modelos considerados calculamos, através da amostragem da distribuição preditiva condicional, os intervalos preditivos de probabilidade 50%, 75% e 95% para cada uma das 31 observações y_i . Se o modelo for adequado, é de se esperar que $(1 - \alpha) \times 100\%$ dos y_i caiam em seus respectivos intervalos preditivos de probabilidade $(1 - \alpha) \times 100\%$.

Tabela 3.15: Porcentagem das observações contidas em seus respectivos intervalos preditivos.

Distribuição dos erros	Intervalo preditivo de probabilidade 50%	Intervalo preditivo de probabilidade 75%	Intervalo preditivo de probabilidade 95%
t-Student	64	74	93
Normal	83	87	96

Neste sentido, pela Tabela 3.15 concluímos que modelo é mais adequado quando seus erros seguem distribuição t-Student, uma vez que as porcentagens das observações contidas nos intervalos preditivos estão mais próximas das porcentagens de definição dos mesmos. No caso dos erros normal, o que notamos são intervalos preditivos maiores devido a alta porcentagem de observações que estes contêm, o que indica que as distribuições preditivas neste caso, estão bastante dispersas.

Consideramos também o diagnóstico de influência, utilizando a estimativa Monte Carlo da medida de divergência de K-L, bem como o logaritmo das ordenadas preditivas condicionais. Vemos pela Figura 3.17 que a observação 13 é apontada como possível ponto influente nos dois modelos ajustados. No entanto, para o modelo com erros de distribuição normal, há um afastamento maior da medida de K-L do seu ponto de corte para y_{13} e a detecção da possível influência da observação 4. Além disso, o logaritmo da CPO em y_{13} é menor do que o obtido considerando erros t-Student.

3.3. APLICAÇÃO DADOS DE PESQUISA SALARIAL

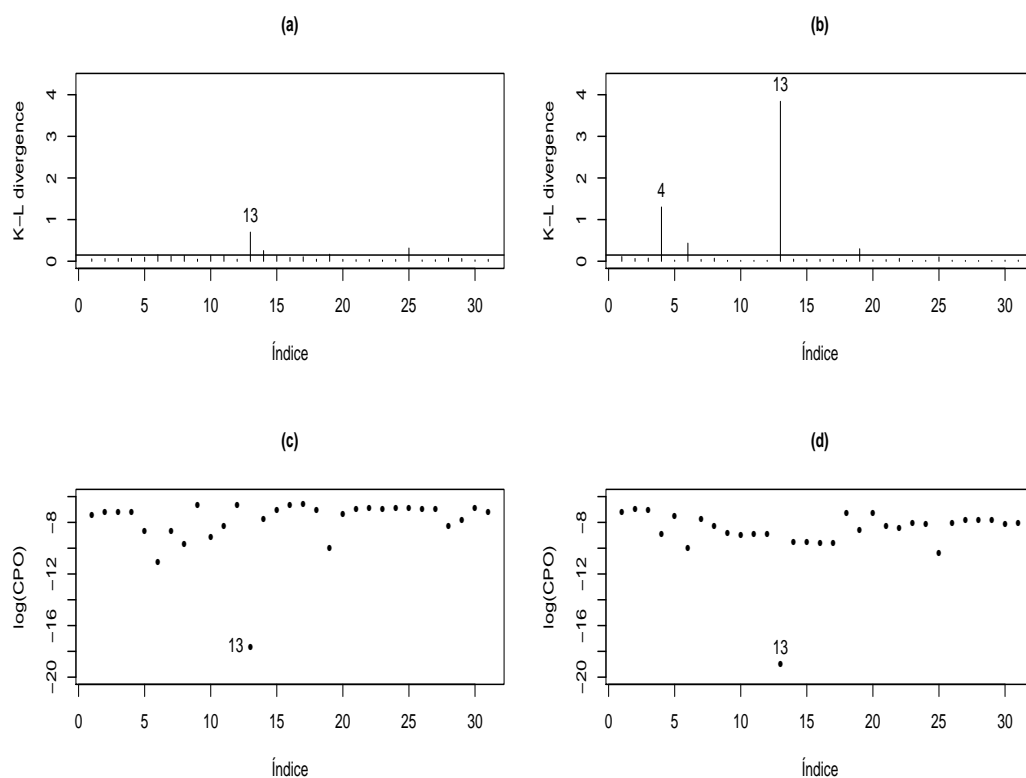


Figura 3.17: Estimativa Monte Carlo da medida de divergência de K-L e logaritmo da ordenada preditiva condicional, considerando os modelos heteroscedásticos com erros de distribuição t-Student (a) e (c) e de distribuição normal (b) e (d).

Sendo assim, analisamos o impacto que as observações detectadas causam nos ajustes através da remoção destes pontos do conjunto de dados. Ajustamos os modelos sem as observações apontadas como influentes e calculamos as variações relativas das estimativas da seguinte maneira:

$$VR = |(\hat{\theta} - \hat{\theta}^*)/\hat{\theta}^*|,$$

com $\hat{\theta}$ representando o vetor de estimativas sem a observação influente e $\hat{\theta}^*$ o vetor das estimativas para o conjunto de dados completo.

3.3. APLICAÇÃO DADOS DE PESQUISA SALARIAL

Tabela 3.16: Mudança percentual nos valores das estimativas.

Erros Normal						
Observação excluída	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\gamma}_0$	$\widehat{\gamma}_1$	$\widehat{\nu}$
{4}	37,16%	27,09%	55,36%	3,39%	7,8309%	-
{13}	63,62%	38,80%	16,46%	13,44%	134,09%	-
{4,13}	61,76%	39,94%	13,41%	13,66%	133,42%	-
Erros t-Student						
Observação excluída	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\gamma}_0$	$\widehat{\gamma}_1$	$\widehat{\nu}$
{13}	10,97%	4,15%	9,29%	1,53	42,98%	17,06%

Como vemos na Tabela 3.16, as maiores mudanças percentuais das estimativas ocorrem para o modelo normal, sendo notado maior impacto na exclusão da observação 13. Quando os erros são t-Student este ponto exerce menos alterações nos ajustes, e portanto, obtemos mais estabilidade nas análises na presença de observações atípicas.

Capítulo 4

Considerações finais e propostas futuras

Neste trabalho propomos o uso das distribuições *a priori* de Jeffreys para os parâmetros do modelo linear heteroscedástico t-Student, especificado na Seção 2.2. E verificamos que as distribuições *a posteriori* geradas, para os casos estudados ($k = 2$ e $k = 3$), são próprias.

Uma análise das propriedades frequentistas dos estimadores bayesianos obtidos, baseados na distribuição *a priori* de Jeffreys propostas, foi apresentada na Seção 3.1, e mostrou, para o cenário estudado, um bom comportamento de medidas como o erro quadrático médio, o desvio padrão amostral, o viés e a probabilidade de cobertura. Comparável ao de distribuições *a priori* que foram definidas utilizando informações a respeito do conjunto de dados.

A relevância destes resultados se deve ao fato de que, sob uma perspectiva bayesiana, o uso de distribuições *a priori* não informativas é o primeiro passo a ser dado em direção à automatização dos ajustes de modelos. Não havendo a necessidade de se especificar parâmetros ou hiperparâmetros de distribuições *a priori*, menos subjetiva

fica a análise, e mais confiáveis os resultados.

Outro aspecto estudado foi a robustez das estimativas na presença de observações atípicas. Verificamos, na Seção 3.2, um bom desempenho dos estimadores obtidos baseados na distribuição *a priori* de Jeffreys. E na Seção 3.3 vimos a superioridade do uso da distribuição t-Student para os erros, quando comparado ao uso da distribuição normal.

Diante dos resultados apresentados, listamos a seguir algumas extensões deste trabalho que podem ser consideradas futuramente.

4.1 Propostas futuras

- Análise bayesiana objetiva para os modelos de regressão não linear com erros t-Student, utilizando distribuição *a priori* de Jeffreys para os parâmetros do modelo;
- Verificação analítica da distribuição *a posteriori* gerada, com demonstrações que comprovem que esta é própria;
- Realização de simulações e aplicação envolvendo um conjunto de dados reais, para verificar o desempenho dos estimadores bayesianos obtidos através da distribuição *a posteriori* calculada e
- Desenvolvimento das análises anteriores, para o modelo não linear t-Student, considerando a hipótese de heteroscedasticidade dos erros.

Apêndice A

Derivadas parciais de primeira e segunda ordem da Seção 2.2

Neste apêndice apresentamos as derivadas parciais de primeira e segunda ordem da função log-verossimilhança do modelo linear heteroscedástico t-Student, apresentado na Seção 2.2 desta dissertação.

A.1 Derivadas parciais de primeira ordem

As derivadas parciais de primeira ordem de $l(\boldsymbol{\theta})$ com respeito aos parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ e ν são expressas através das seguintes equações:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = (\nu + 1) \sum_{i=1}^n \left\{ \mathbf{x}_i \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\}, \quad (\text{A.1})$$

em que $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}$ é um vetor de dimensão p , cujo elemento de índice j é dado por

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} = & (\nu + 1) \sum_{i=1}^n \left\{ x_{ij} \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \left. \times \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \right\}, \end{aligned}$$

com $j = 0, \dots, p - 1$;

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{\sum_{i=1}^n \mathbf{z}_i}{2} + \frac{(\nu + 1)}{2} \sum_{i=1}^n \left\{ \mathbf{z}_i \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\}, \quad (\text{A.2})$$

em que $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}$ é um vetor de dimensão k , cujo elemento de índice l é dado por

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \gamma_l} = & -\frac{\sum_{i=1}^n z_{il}}{2} + \frac{(\nu + 1)}{2} \sum_{i=1}^n \left\{ z_{il} \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \left. \times \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \right\}, \end{aligned}$$

com $l = 0, \dots, k - 1$. E, por fim,

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \nu} = & n \left\{ \frac{1}{2} \psi \left(\frac{\nu + 1}{2} \right) - \frac{1}{2} \psi \left(\frac{\nu}{2} \right) + \frac{1}{2} \log \frac{\nu}{2} + \frac{1}{2} \right\} \\ & - \frac{1}{2} \sum_{i=1}^n \left\{ \log \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right] \right\} \\ & - \frac{(\nu + 1)}{2} \sum_{i=1}^n \left\{ \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\}, \quad (\text{A.3}) \end{aligned}$$

em que $\psi(z) = \frac{d}{dz} \log \Gamma(z)$ é a função digama.

A.2 Derivadas parciais de segunda ordem

Considerando agora as derivadas parciais de segunda ordem de $l(\boldsymbol{\theta})$ com respeito aos parâmetros $\boldsymbol{\beta}, \boldsymbol{\gamma}$ e ν , obtemos as seguintes componentes:

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = & - (\nu + 1) \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i \mathbf{x}_i^\top}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\} \\ & + 2(\nu + 1) \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i \mathbf{x}_i^\top (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(2\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-2} \right\}, \end{aligned} \quad (\text{A.4})$$

em que $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}$ é a matriz de dimensão $p \times p$ cujo elemento (j_1, j_2) é dado por

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_{j_1} \partial \beta_{j_2}} = & - (\nu + 1) \sum_{i=1}^n \left\{ \frac{x_{ij_1} x_{ij_2}}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \right\} \\ & + 2(\nu + 1) \sum_{i=1}^n \left\{ \frac{x_{ij_1} x_{ij_2} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-2} \right\}, \end{aligned}$$

com $j_1, j_2 = 0, \dots, p-1$;

$$\begin{aligned} \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^\top}(\boldsymbol{\theta}) = & - (\nu + 1) \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i \mathbf{z}_i^\top (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\} \\ & + (\nu + 1) \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i \mathbf{z}_i^\top (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^3}{\exp(2\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-2} \right\} \end{aligned} \quad (\text{A.5})$$

em que $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^\top}$ é a matriz de dimensão $p \times k$, cujo elemento (j, l) é dado por

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_j \partial \gamma_l}(\boldsymbol{\theta}) = & - (\nu + 1) \sum_{i=1}^n \left\{ \frac{x_{ij} z_{il} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \right\} \\ & + (\nu + 1) \sum_{i=1}^n \left\{ \frac{x_{ij} z_{il} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^3}{\exp(2\gamma_0 + 2z_{i1}\gamma_1 + \dots + 2z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-2} \right\} \end{aligned}$$

A.2. DERIVADAS PARCIAIS DE SEGUNDA ORDEM

com $j = 0, \dots, p-1$ e $l = 0, \dots, k-1$;

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \nu}(\boldsymbol{\theta}) = & \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\} \\ & - (\nu + 1) \sum_{i=1}^n \left\{ \frac{\mathbf{x}_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-2} \right\} \end{aligned} \quad (\text{A.6})$$

em que $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta \partial \nu}$ é o vetor de dimensão p , cujo elemento de índice j é dado por

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_j \partial \nu}(\boldsymbol{\theta}) = & \sum_{i=1}^n \left\{ \frac{x_{ij}(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \right\} \\ & - (\nu + 1) \sum_{i=1}^n \left\{ \frac{x_{ij}(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-2} \right\} \end{aligned} \quad (\text{A.7})$$

com $j = 0, \dots, p-1$;

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} = & - \frac{(\nu + 1)}{2} \sum_{i=1}^n \left\{ \frac{\mathbf{z}_i \mathbf{z}_i^\top (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\} \\ & + \frac{(\nu + 1)}{2} \sum_{i=1}^n \left\{ \frac{\mathbf{z}_i \mathbf{z}_i^\top (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^4}{\exp(2\mathbf{z}_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma})} \right]^{-2} \right\} \end{aligned} \quad (\text{A.8})$$

em que $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top}$ é a matriz de dimensão $k \times k$ cujo elemento (l_1, l_2) é dado por

$$\begin{aligned} \frac{\partial^2 l}{\partial \gamma_{l_1} \partial \gamma_{l_2}}(\boldsymbol{\theta}) = & - \frac{(\nu + 1)}{2} \sum_{i=1}^n \frac{z_{il_1} z_{il_2} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \\ & \times \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \\ & + \frac{(\nu + 1)}{2} \sum_{i=1}^n \frac{z_{il_1} z_{il_2} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^4}{\exp(2\gamma_0 + 2z_{i1}\gamma_1 + \dots + 2z_{ik}\gamma_{k-1})} \\ & \times \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-2} \end{aligned} \quad (\text{A.9})$$

A.2. DERIVADAS PARCIAIS DE SEGUNDA ORDEM

com $l_1, l_2 = 0, \dots, k-1$;

$$\begin{aligned} \frac{\partial^2 l}{\partial \gamma \partial \nu}(\boldsymbol{\theta}) = & - \frac{(\nu+1)}{2} \sum_{i=1}^n \left\{ \frac{z_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(z_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(z_i^\top \boldsymbol{\gamma})} \right]^{-2} \right\} \\ & + \frac{1}{2} \sum_{i=1}^n \left\{ \frac{z_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(z_i^\top \boldsymbol{\gamma})} \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(z_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\} \end{aligned} \quad (\text{A.10})$$

em que $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \gamma \partial \nu}$ é o vetor de dimensão k , cujo elemento de índice l é dado por

$$\begin{aligned} \frac{\partial^2 l}{\partial \gamma_l \partial \nu}(\boldsymbol{\theta}) = & - \frac{(\nu+1)}{2} \sum_{i=1}^n \left\{ \frac{z_{il} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-2} \right\} \\ & + \frac{1}{2} \sum_{i=1}^n \left\{ \frac{z_{il} (y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right. \\ & \times \left. \left[\nu + \frac{(y_i - \beta_0 - x_{i1}\beta_1 - \dots - x_{ip-1}\beta_{p-1})^2}{\exp(\gamma_0 + z_{i1}\gamma_1 + \dots + z_{ik}\gamma_{k-1})} \right]^{-1} \right\} \end{aligned} \quad (\text{A.11})$$

com $l = 0, \dots, k-1$. E, por fim,

$$\begin{aligned} \frac{\partial l^2(\boldsymbol{\theta})}{\partial \nu^2} = & \frac{n}{4} \left\{ \psi' \left(\frac{\nu+1}{2} \right) - \psi' \left(\frac{\nu}{2} \right) + \frac{2}{\nu} \right\} - \sum_{i=1}^n \left\{ \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(z_i^\top \boldsymbol{\gamma})} \right]^{-1} \right\} \\ & + \frac{(\nu+1)}{2} \sum_{i=1}^n \left\{ \left[\nu + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\exp(z_i^\top \boldsymbol{\gamma})} \right]^{-2} \right\}. \end{aligned} \quad (\text{A.12})$$

As derivadas mistas restantes são obtidas fazendo $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \gamma \partial \boldsymbol{\beta}^\top} = \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^\top} \right]^\top$, $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \nu \partial \boldsymbol{\beta}} = \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \nu} \right]^\top$ e $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \nu \partial \boldsymbol{\gamma}} = \left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \nu} \right]^\top$.

Referências Bibliográficas

- Achcar, J. A., de Cássia Favoretti, A. & Mazucheli, J. (1996). Approximate bayesian analysis for the michaelis-menten model. *Journal of the Chilean Statistical Society*, **13**(1-2), 1–23.
- Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 99–102.
- Bernardo, J. M. (2009). *Modern Bayesian inference: Foundations and Objective Methods*, volume 200. Elsevier.
- Branco, M., Bolfarine, H., Iglesias, P. & Arellano-Valle, R. B. (2000). Bayesian analysis of the calibration problem under elliptical distributions. *Journal of Statistical Planning and Inference*, **90**(1), 69–85.
- Branco, M. D., Genton, M. G. & Liseo, B. (2013). Objective bayesian analysis of skew-t distributions. *Scandinavian Journal of Statistics*, **40**(1), 63–85.
- Cabral, C. R. B., Lachos, V. H. & Madruga, M. R. (2012). Bayesian

- analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *Journal of Statistical Planning and Inference*, **142**(1), 181–200.
- Cancho, V. G., Ortega, E. M. & Paula, G. A. (2010). On estimation and influence diagnostics for log-birnbaum–saunders student-t regression models: Full bayesian analysis. *Journal of Statistical Planning and Inference*, **140**(9), 2486–2496.
- Carlin, B. P. & Louis, T. A. (1997). Bayes and empirical bayes methods for data analysis. *Statistics and Computing*, **7**(2), 153–154.
- Chatterjee, S. & Hadi, A. S. (2009). *Sensitivity analysis in linear regression*, volume 327. John Wiley & Sons.
- Chen, M.-H., Shao, Q.-M. & Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.
- Cho, H., Ibrahim, J. G., Sinha, D. & Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, **65**(1), 116–124.
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 133–169.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275.

- Cysneiros, F. J. A., Paula, G. A. & Galea, M. (2007). Heteroscedastic symmetrical linear models. *Statistics & probability letters*, **77**(11), 1084–1090.
- Fernandez, C. & Steel, M. F. (1999). Multivariate student-t regression models: Pitfalls and inference. *Biometrika*, **86**(1), 153–167.
- Fonseca, T. C., Ferreira, M. A. & Migon, H. S. (2008). Objective bayesian analysis for the student-t regression model. *Biometrika*, **95**(2), 325–333.
- Garay, A. W. M. (2014). *Modelos de regressao para dados censurados sob Distribuicoes Simétricas*. Ph.D. thesis, Universidade de Sao Paulo.
- Gelfand, A. E. (1996). Model determination using sampling-based methods. *Markov chain Monte Carlo in practice*, pages 145–161.
- Gelman, A., Meng, X.-L. & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**(4), 733–760.
- Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, **8**(S1), S19–S40.
- Ghosh, M. (2011). Objective priors: An introduction for frequentists. *Statistical Science*, pages 187–202.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

- Lehmann, E. L. & Casella, G. (1998). *Theory of point estimation*, volume 31. Springer Science & Business Media.
- Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.
- McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, **84**, 473–478.
- Osiewalski, J. & Steel, M. F. (1993). Robust bayesian inference in elliptical regression models. *Journal of Econometrics*, **57**(1), 345–363.
- Paulino, C., Turkman, M. & Murteira, B. (2003). Estatística bayesiana, fundação clouste gulbenkian lisboa.
- Peng, F. & Dey, D. K. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian Journal of Statistics*, **23**(2), 199–213.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Vallejos, C. A. & Steel, M. F. (2013). On posterior propriety for the student-*t* linear regression model under jeffreys priors. *arXiv preprint arXiv:1311.1454*.
- Yiqi, B. (2012). *Estimação e diagnóstico na distribuição Weibull-*

REFERÊNCIAS BIBLIOGRÁFICAS

Binomial-Negativa em análise de sobrevivência. Ph.D. thesis, Universidade de São Paulo.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, **71**(354), 400–405.