UNICAMP

Carlos Alberto Huaira Contreras

Modelo de Regressão Linear Mistura de Escala Normal com Ponto de Mudança: Estimação e Diagnóstico

CAMPINAS
2014

Universidade Estadual de Campinas

Instituto de Matemática, Estatística
e Computação Científica

Carlos Alberto Huaira Contreras

Modelo de Regressão Linear Mistura de Escala Normal com Ponto de Mudança: Estimação e Diagnóstico

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Estatística.

Orientador Prof. Dr. Filidor Edilfonso Vilca Labra

Este exemplar corresponde á versão final da dissertação defendida pelo aluno Carlos Alberto Huaira Contreras, e orientada pelo Prof. Dr. Filidor Edilfonso Vilca Labra.

Assinatura do Orientador

Campinas
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

H86m

Huaira Contreras, Carlos Alberto, 1971-

Modelo de regressão linear mistura de escala normal com ponto de mudança

: estimação e diagnóstico / Carlos Alberto Huaira Contreras. – Campinas, SP : [s.n.], 2014.

Orientador: Filidor Edilfonso Vilca Labra.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Algoritmos de expectativa de maximização. 2. Misturas de escala. 3. Problemas de ponto de mudança. 4. Influência local. 5. Modelos lineares (Estatística). I. Vilca Labra, Filidor Edilfonso,1964-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Scale mixture of normal regression linear regression model with change point : estimation and diagnostics
Palavras-chave em inglês:
EM algorithms
Scale mixture
Change-point problems
Local influence
Linear models (Statistics)
Área de concentração: Estatística
Titulação: Mestre em Estatística
Banca examinadora:
Filidor Edilfonso Vilca Labra [Orientador]
Víctor Hugo Lachos Dávila
Camila Borelli Zeller
Data de defesa: 14-08-2014
Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 14 de agosto de 2014 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA

Prof(a). Dr(a). VÍCTOR HUGO LACHOS DÁVILA

Camila Borelli Zeller

Prof(a). Dr(a). CAMILA BORELLI ZELLER

Abstract

Linear models are widely used in statistics to describe the relationship between a response variable and one or more explanatory variables, where usually it is assumed the errors are normally distributed. Moreover, in linear regression model is assumed that the same linear model holds for the whole data set, but this is not always valid. The model may change after a specific point, and so a linear model with a change point would be appropriate for data set.
The main objective of work is to study some aspect of estimation and analysis of diagnostics in the regression linear with change point model under scale mixture of normal distributions. The analysis of diagnostics is based on the works of Cook (1986) and Zhu & Lee (2001). The results obtained represent a extension of some results obtained in the literature; see for example Chen (1998) and Osorio & Galea (2006). Finally, simulation studies are investigated through Monte Carlo simulations and numerical examples are presented to illustrate the proposed results.

Keywords: EM Algorithm; Mahalanobis Distance; Scale Mixtures of Normal Distributions; Local Influence; Linear Models; Change Point.

Resumo

Modelos lineares são frequentemente usados em estatística para descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas, onde geralmente os erros são assumidos como normalmente distribuídos. Além disso, em modelos de regressão linear assume-se que o mesmo modelo linear é válido para todo o conjunto de dados. O modelo pode mudar após um ponto específico e assim um modelo linear com um ponto de mudança poderá ser apropriado para o conjunto de dados.
O principal objetivo deste trabalho é estudar alguns aspectos de estimação e análise de diagnóstico em modelos de regressão linear com ponto de mudança sob distribuições de mistura de escala normal. A análise de diagnóstico é baseada nos trabalhos de Cook (1986) e Zhu & Lee (2001). Os resultados obtidos representam uma extensão de alguns resultados apresentados na literatura, ver por exemplo Chen (1998) e Osorio & Galea (2006). Finalmente, estudos de simulação através de simulações Monte Carlo são realizados e exemplos numéricos são apresentados para ilustrar os resultados propostos.

Palavras-chave: Algoritmo EM; Distância de Mahalanobis; Distribuições Mistura de Escala Normal; Influência Local; Modelos Lineares; Ponto de Mudança.

Sumário

Aos meus pais Ana Maria e Teodosio.
Aos meus irmãos Ana Patricia e Julio.
Á minha esposa Rosália e ao meu filho Gael.
Á inesquecivel Albertina (in memorian).

Agradecimentos

- Ao professor Filidor Vilca Labra pela confiança, sugestões, apoio e orientação na elaboração deste trabalho.

- Á professora Camila Borelli Zeller pela sua constante colaboração e sugestões para a obtenção dos resultados obtidos.

- Ao Departamento de Estatística IMECC-UNICAMP pela oportunidade oferecida.

- Ao IBOPE Midia pela disponibilização dos dados utilizados neste trabalho.

- A minha familia, tanto brasileira quanto peruana, que de diversas maneiras me encorajaram para completar esta etapa de vida.

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Os modelos estatísticos são construídos para explicar de forma aproximada a estrutura de um conjunto de dados mediante uma relação funcional que melhor se ajuste ao padrão das características medidas. Estes modelos estatísticos deverão satisfazer certas suposições para que seja considerada uma boa aproximação. De forma clássica, modelos de regressão linear - como todos os modelos lineares - são baseados na denominada teoria normal. Assim, uma suposição rotineira do modelo é que tanto os erros como os efeitos aleatórios são normalmente distribuídos. Sob esta suposição, a inferência estatística para os parâmetros é amplamente conhecida e encontra-se disponível en uma ampla variedade de programas computacionais de análise estatística, veja por exemplo, Cohen & Cohen (2008) e Everitt & Hothorn (2010) para aplicações no sistema R.

Em muitas áreas de pesquisa, é de interesse explicar como um conjunto de variáveis denominadas independentes (definidas a partir de um experimento controlado ou por uma base teórica previamente definida) afeta - conjunta ou individualmente - uma variável denominada dependente. Os chamados modelos de regressão (Draper & Smith, 1998) são ferramentas estatísticas utilizadas muito frequentemente para este fim. A literatura inclui exemplos de regressão linear e não linear, calibração, regressão inversa, dentre outros e as suas aplicações podem ser vistas em diferentes áreas de investigação tais como Medicina, Ciências Sociais, Econometría, Geologia, Química, Engenharia, dentre outras.

Uma situação particular ocorre quando o conjunto de observações pode ser indexado por alguma variável (tempo, espaço ou outra variável que indique uma ordem), e é possível observar que a partir de certo ponto desta indexação a distribuição destas observações muda. Esta situação caracteriza um problema de ponto de mudança (Gomes de Souza, 2004). Logo, pode-se especificar modelos de regressão linear com mudança nos parâmetros a partir de uma k-ésima observação de um conjunto de observações indexada.

Quando o ponto de mudança é desconhecido e o interesse é detectar e estimar este, Chen (1998) sugere uma metodologia que utiliza o Critério de Informação de Schwarz (SIC) de forma recursiva sobre todos os possíveis pontos do conjunto de dados passíveis de ser o ponto de mudança. Chen &

Gupta (2001) apresentam uma metodologia geral para deteção de pontos de mudança em modelos com diversas estruturas, assim, quando um ponto de mudança está presente e é desconhecido, a estrutura do modelo muda a partir desta k-ésima observação considerada, e três formas de mudança nas estruturas dos modelos podem ocorrer: (1) A variância do modelo considerado é constante no conjunto total de dados e a mudança se produz somente nos parâmetros do modelo de regressão,
(2) A mudança se produz na variância do modelo e os parâmetros do modelo de regressão são constantes no conjunto total de dados, e
(3) A mudança afeta a variância e os parâmetros do modelo de regressão.

Como foi dito anteriormente, a teoria de inferência estatística sobre modelos de regressão foi amplamente desenvolvida em torno da distribuição normal, isto justificado por algumas boas propriedades estatísticas. No entanto, a inferência baseada no modelo normal sofre de falta de robustez no sentido de ser muito sensível quando existem observações aberrantes (*outliers*) ou quando os dados provém de uma distribuição com caudas mais pesadas que a distribuição normal. Embora seja frequente o uso de transformações sobre a variável resposta com o fim de aproximar a uma distribuição normal (ou ao menos procurar simetria), sugere-se buscar um modelo teórico mais conveniente quando isto é possível. Assim, relaxar a suposição de normalidade para estimar os parâmetros do modelo tornou-se uma alternativa interessante.

O uso de distribuições simétricas com caudas mais pesadas que a distribuição normal tem-se mostrado úteis para reduzir a influência dos "*outliers*"sobre as estimativas de máxima verossimilhança. Uma classe de distribuições denominada de Mistura de Escala Normal (Andrews & Mallows, 1974) oferece alternativas neste sentido. Esta classe contém as distribuições normal, t de Student, slash, normal contaminada, dentre outras. e tem sido estudada com grande interesse nos últimos anos.

Assim, neste trabalho, consideramos o modelo de regressão linear, onde é assumido que os modelos considerados apresentam um ponto de mudança e as observações seguem uma distribuição na classe Mistura de Escala Normal. Considerando que a posição do ponto de mudança é desconhecida, serão encontrados os estimadores de máxima verossimilhança para as três formas de mudança nas estruturas do modelo descritas anteriormente. Adicionalmente, será apresentado um procedimento recursivo que considera o critério de informação de Schwarz para estimar um ponto de mudança desconhecido. As estimativas de máxima verossimilhança serão obtidas via o algoritmo EM, e a programação do algoritmo e do procedimento recursivo para determinação do ponto de mudança são implemetados no sistema R. Finalmente, é considerado um enfoque de análise de diagnóstico (Cook, 1986) baseado na metodologia de Zhu & Lee (2001).

1.1 Motivação

Esta dissertação é motivada por um problema real que está relcionado a procedimentos de medição de audiência. A questão geral é encontrar um modelo estatístico adequado que permita descrever a relação entre os resultados de audiência obtidos a partir do uso de instrumentos eletrônicos de medição e os obtidos a partir do método manual de preenchimento de questionário. Devido aos avanços tecnológicos ocorridos nos últimos anos, os procedimentos de medição são cada vez mais automatizados e incorporam novas utilidades que fazem que suas características sejam melhores que seus predecessores (que na maioria de casos empregavam muitos processos manuais). É natural então que estes novos procedimentos substituam os antigos.

Quando uma mudança nos procedimentos de medição é realizada, é importante conhecer qual é o efeito que esta mudança irá causar, e na grande maioria de vezes é preciso encontrar uma relação numérica entre valores de medição obtidos pelo novo procedimento e o antigo. A relação numérica em questão é imprescindível quando as medições são realizadas de forma contínua e por alguma razão torna-se necessário manter o histórico das medições, ou quando o procedimento será implantado em etapas e por algum tempo se conviverá com os dois procedimentos.

É preciso então, estimar valores de medição para novo procedimento obtido a partir dos valores de medição do antigo procedimento. Para isto, foi selecionado um período de tempo no qual os dois procedimentos de medição são utilizados para aferir em duas amostras representativas da área de cobertura estudada, e a partir destes dados se constrói um modelo de regressão que explique a medição pelo procedimento novo em função do procedimento antigo.

Para o caso específico que motiva esta dissertação, sabe-se que os valores das medições são realizadas de forma contínua e que cada medição possa ser identificada e ordenada por esta identificação, obtendo assim uma base de dados indexada. Adicionalmente, devido a certas condições conhecidas, é de esperar que a partir de um certo valor desta indexação os valores das medições sejam diferentes e produzam uma mudança na estrutura da relação do modelo de regressão, propondo assim a utilização de um modelo de regressão com um ponto de mudança. Por outro lado, a presença de dados aberrantes nas medições sugere a utilização de modelos robustos. Desta forma, serão consideradas distribuições simétricas com caudas mais pesadas que a distribuição normal, especificamente serão consideradas as distribuições da classe de Mistura de Escala Normal.

1.2 Objetivo do Trabalho

O objetivo do presente trabalho é apresentar um estudo de estimação e diagnóstico no Modelo de Regressão Linear (MRL) com um ponto de mudança, onde os erros seguem uma distribuição na classe Mistura de Escala Normal (MEN). A estimação dos parâmetros e a determinação do ponto de mudança são tratados com detalhes e são inspirados nos trabalhos de Chen (1998) e Osorio & Galea (2006). Quatro distribuições de probabilidade que formam parte da classe de Mistura de Escala Normal são utilizadas na obtenção de resultados específicos. Os resultados obtidos serão aplicados sobre dados simulados e reais.

Os objetivos específicos deste trabalho podem ser resumidos em:

i) Desenvolver a estimação por máxima verossimilhança para modelos de regressão linear mistura de escala normal que apresentam um ponto de mudança e apresentar a estimação de parâmetros mediante o algoritmo EM para cada um dos três casos de mudança considerados;

ii) Considerar um método para determinar a existência de um ponto de mudança e a posição deste no conjunto de dados a partir de Critério de Informação de Schwarz, para todos os modelos estudados;

iii) Abordar o problema de influência local nos modelos estudados considerando alguns esquemas de perturbação e seguindo a metodologia de Zhu & Lee (2001);

iv) Avaliar os resultados obtidos a partir de estudos de simulação de Monte Carlo;

v) Aplicar os resultados obtidos em conjuntos de dados reais. Um dos quais corresponde a um conjunto conhecido dentro da literatura estatística e para o qual existem alguns resultados em situações específicas (Holbert,1982).

1.3 Organização do Trabalho

A presente dissertação contêm seis capítulos e um apêndice que contêm os dados das aplicações. Os capítulos são organizados como segue abaixo.

No Capítulo 2, apresenta-se uma revisão dos principais conceitos que serão tratados nesta dissertação. O conceito de ponto de mudanca em modelos clássicos de regressão linear e a estimação de parâmetros via algoritmo EM são introduzidos para estes modelos. Finalmente, resultados importantes relacionados à classe de distribuições de Mistura de Escala Normal são apresentados.

No Capítulo 3, estuda-se os modelos de regressão linear (MRL) considerando um ponto de mudança e com erros distribuídos por distribuições da classe de Mistura de Escala Normal (MEN). A especificação do modelo, a estimação de máxima verossimilhança dos parâmetros, o algoritmo EM correspondente e a forma de determinação do ponto de mudança a partir de Critério de Informação de Schwarz (SIC) são apresentados para os diversos modelos considerados. Estudos de simulação e discussões sobre os resultados são apresentados em todas as situações.

O Capítulo 4 é dedicado ao estudo de diagnóstico no MRL com ponto de mudança baseado na metodologia de Zhu & Lee (2001). Resultados e discussões para alguns esquemas de perturbação são apresentados.

No Capítulo 5, apresenta-se aplicações sobre três conjuntos de dados reais.

Finalmente, no Capítulo 6 apresentamos as considerações finais e conclusões deste trabalho, assim como uma discussão para futuros trabalhos.

Capítulo 2

Principais Conceitos

Neste capítulo, apresentamos os principais conceitos que serão tratados nesta dissertação. Primeiramente, descrevemos o modelo de regressão linear com a presença de um ponto de mudança e com erros distribuídos normalmente e posteriormente erros com distribuição t de Student. Seguidamente, apresenta-se uma descrição do algoritmo EM usado na estimação dos parâmetros dos modelos estudados no presente trabalho. Finalmente, uma revisão das distribuições de Mistura de Escala Normal para o caso univariado é considerada.

2.1 Modelo de Regressão com Ponto de Mudança

Nesta seção, descrevemos o problema de presença de um ponto de mudança nos modelos de regressão. Consideramos duas situações, para os modelos de regressão, estudadas por Chen (1998) e Osorio & Galea (2006); quando os erros seguem uma distribuição normal com média zero e variância constante e uma extensão que considera uma distribuição t de Student com média zero e variância constante, respectivamente. Em ambas situações a mudança ocorre nos parâmetros de regressão.

2.1.1 O Problema de Ponto de Mudança

Intuitivamente pode-se considerar que qualquer conjunto de observações necessariamente ordenadas por alguma característica conhecida (de tempo, de espaço ou outra qualquer), apresenta um problema de ponto de mudança se a distribuição das observações muda após um ponto determinado e desconhecido deste ordenamento.

Uma definição formal para este problema é encontrado em Gomes de Souza (2004): dada uma sequência de variáveis aleatórias (ou vetores aleatórios) $\mathbf{x_1}$, $\mathbf{x_2}$,... com distribuições de probabilidade dadas por f_1,f_2... respectivamente onde os índices 1,2,... indicam uma ordem (por exemplo, instantes de tempo). Se diz que o ponto k (correspondente ao índice k) é o ponto de mudança desconhecido dessa sequência se $f_1 = f_2 = ... = f_k \neq f_{k+1} = f_{k+2} = ...$

Deve-se ter claro que a sequência definida não necessariamente está referida a intervalos de tempo que estejam distribuídos igualmente e sim a qualquer forma de ordem dos dados. Por outro lado, é possível generalizar a definição para mais de um ponto de mudança.

As várias características pelas quais os problemas de ponto de mudança podem ser clasificados são apresentadas por Brodsky & Darkhovsky (1993), estas estão referidas logo abaixo.

(1) Ao método de obtenção de dados, que podem ser separados em duas situações: sequenciais e de tamanho de amostra fixado. Nos procedimentos sequenciais, por não existir um tamanho de amostra determinado, a verificação de homogeniedade de f deve ser feita quando cada nova observação é realizada, os dados deverão ser examinados a cada nova observação quanto à hipótese de homogeneidade. Quando temos uma amostra fixa o processo de obtenção de dados é completado antes da hipótese de homogeneidade de f ser testada.

(2) À informação a priori sobre f_i, de onde os problemas de ponto de mudança podem ser considerados paramétricos, semi-paramétricos ou não paramétricos.

(3) Às características dos dados, determinando modelos com tempo contínuo ou discreto, uni ou multidimensionais ou com observações dependentes ou independentes.

(4) Ao tipo de mudança, que podem ser de dois tipos: Os de mudança abrupta como por exemplo um modelo da forma $Y_i = \alpha_1 + \beta_1 X_i$, para $i = 1, \ldots, k$ e $Y_i = \alpha_2 + \beta_2 X_i$, para $i = k+1, \ldots, n$ (onde X_i necessariamente mantém uma ordem respeito ao indexador); e os modelos com mudança gradual ou sem descontinuidade onde dado $X_1 \leq X_2 \leq ... \leq X_n$ existe λ ($X_k < \lambda < X_{k+1}$) tal que $Y_i = \alpha_1 + \beta_1 \lambda = \alpha_2 + \beta_2 \lambda$, estes modelos são conhecidos também como regressão segmentada.

(5) Ao número de mudanças, onde podemos ter modelos com um único ponto de mudança ou múltiplos pontos de mudança.

A respeito da inferência sobre os modelos com ponto de mudança, devem ser considerados os seguintes aspectos: Determinar a existência do ponto (ou pontos) de mudança, localizar a posição deste ponto (ou pontos), estimar todos os parâmetros de interesse do modelo e realizar análises preditivas.

Dentro da literatura estatística, o problema de ponto de mudança tem sido estudado com muito interesse ao longo dos anos. Estudos sobre problemas de ponto de mudança na média sobre uma sequência de variáveis aleatórias normais são tratados por Chernoff & Zacks (1964), Gardner (1969), Srivastava (1975), Worsley (1979) e Srivastava & Worsley (1986). Horváth (1993) e Chen & Gupta (1995) estendem este estudo para mudança simultânea na média e variância, sempre sobre distribuições normais da variável aleatória. Mais recentemente Bhatti (2000) utiliza alguns testes conhecidos no problema de ponto de mudança na variância de sequência de variáveis aleatórias normais com média não conhecida.

No que se refere aos modelos de regressão linear, muitos estudos na literatura estatística foram desenvolvidos considerando o problema de ponto de mudança sobre o modelo clássico que asssume normalidade dos erros aleatórios. Quandt (1958, 1960) intoduz o método de máxima verossimilhança para estimar e testar parâmetros de modelos de regressão segmentada. Ferreira (1975), Chin Choy & Broemeling (1980) e Holbert (1982) realizam estudos desde o ponto de vista bayesiano. Brown et al. (1975) e Hofrichter (2007) usam o método de residuais recursivos para a deteção de pontos de mudança em modelos de regressão linear múltipla. Hawkins (1989) utiliza o criterio de união-interseção. Kim & Siegmund (1989) e Kim (1994) utilizam o teste de razão de verossimilhança para deteção de pontos de mudança. Csörgő & Horváth (1997) apresentam propriedades assintóticas de métodos de deteção de ponto demudança. O uso de Criterio de Informação de Schwarz (Schwarz, 1978) para determinação de ponto de mudança em modelos de regressão linear é apresentado por Chen (1998) e Chen & Gupta (1997, 1999, 2001). Osorio & Galea (2006) utilizam o critério de Informação de Schwarz (SIC), na determinação do ponto de mudança num modelo de regressão linear onde os erros são distribuídos de acordo a uma distribuição t de Student.

Nas seguintes subseções, apresentamos a metodologia descrita por Chen (1998) e a extensão apresentada por Osorio & Galea (2006) para o modelo robusto t de Student. Em ambos casos considera-se a mudança nos parâmetros de regressão e variância constante.

2.1.2 Modelo de Regressão Linear Normal com Ponto de Mudança

Considerando o modelo de regressão:

$$\mathbf{Y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{2.1}$$

onde $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{p-1})^\top$ é um vetor e parâmetros desconhecidos de dimensão p, $\mathbf{x}_i = (1, x_{1i}, ..., x_{1(p-1)})^\top$ é a i-ésima linha da matriz de desenho $\mathbf{X}$ de dimensão $n \times p, (n > p)$ e os erros aleatórios ϵ_1,...,ϵ_n são independentes e identicamente distribuídos como $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $i = 1, \ldots, n$, onde σ^2 é um parâmetro desconhecido maior que zero.

Considerando que o ponto de mudança encontra-se na posição k, não conhecida, tem-se que $Y_i \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}_1, \sigma^2)$, para $i = 1, \ldots, k$ e $Y_i \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}_2, \sigma^2)$, para $i = k+1, \ldots, n$, ou seja, ponto de mudança na média da distribuição da variável resposta.

Chen (1998) aborda o problema de verificar a existência de um ponto de mudança no modelo de regressão (2.1) e determinar a posição deste. Para isto, considera-se um teste de hipótese da forma

$$\begin{aligned} H_0 &: \quad Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \ldots, n, \\ H_1 &: \quad \begin{cases} Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_1 + \epsilon_i, & i = 1, \ldots, k, \\ Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_2 + \epsilon_i, & i = k+1, \ldots, n, \end{cases} \end{aligned} \tag{2.2}$$

isto é, a hipótese nula indica que não existe ponto de mudança no modelo de regressão contra a hipótese alternativa que identifica um modelo de regressão com um ponto de mudança na observação k.

Quando a posição do ponto de mudança não é conhecida, H_1 deverá considerar uma coleção de modelos com ponto de mudança que considera as posições $p, ..., n-p$ e o objetivo neste caso será selecionar um modelo desta coleção.

A metodologia para a seleção proposta por Chen (1998) usa o Critério de Informação de Schwarz (SIC) definido como

$$SIC \;=\; -2\,\ell(\widehat{\boldsymbol{\theta}}) \;+\; s\;\log\;n, \tag{2.3}$$

onde $\ell(\widehat{\boldsymbol{\theta}})$ corresponde à função de log-verossimilhança avaliada na estimativa de máxima verossimilhança $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^{\top}, \widehat{\sigma^2})^{\top}$, s é o número de parâmetros do modelo e n é o tamanho de amostra.

Os critérios de decisão adotados são os seguintes:

- O modelo considerado apresenta um ponto de mudança, que equivale a rejeitar a hipótese nula, se

$$SIC(n) \;>\; min\{SIC(k),\, para\, k = p, ..., n-p\}. \tag{2.4}$$

- Quando a hipótese nula é rejeitada, a posição estimada do ponto de mudança via máxima verossimilhança corresponde ao valor $\widehat{k}$ que satisfaz

$$SIC(\widehat{k}) \;=\; min\{SIC(k) :\; k = p, ..., n-p\}. \tag{2.5}$$

Como mencionado em Osorio & Galea (2006), a maximização da função de log-verossimilhança equivale à minimização do Critério de Informação de Schwarz.

Sob a hipótese nula os estimadores de máxima verossimilhança dos parâmetros são

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{Y} \qquad e \qquad \widehat{\sigma^2} = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^{\top}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}).$$

O critério de informação de Schwarz, $SIC(n)$, pode ser expresso como

$$\begin{aligned} SIC(n) \;&=\; -2\,\ell_0(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2}) \;+\; (p+1)\;\log\;n \\ &=\; n\;\log[(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^{\top}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})] + n\,(\log 2\pi + 1) + (p+1-n)\log\,n. \end{aligned}$$

Sob a hipótese alternativa os estimadores de máxima verossimilhança dos parâmetros considerados podem ser expressados como

$$\widehat{\boldsymbol{\beta}_1} = (\mathbf{X_1}^{\top}\mathbf{X_1})^{-1}\mathbf{X_1}^{\top}\mathbf{Y_1}, \qquad \widehat{\boldsymbol{\beta}_2} = (\mathbf{X_2}^{\top}\mathbf{X_2})^{-1}\mathbf{X_2}^{\top}\mathbf{Y_2} \qquad e$$

$$\widehat{\sigma^2} = \frac{1}{n}[(\mathbf{Y_1} - \mathbf{X_1}\widehat{\boldsymbol{\beta}_1})^{\top}(\mathbf{Y_1} - \mathbf{X_1}\widehat{\boldsymbol{\beta}_1}) + (\mathbf{Y_2} - \mathbf{X_2}\widehat{\boldsymbol{\beta}_2})^{\top}(\mathbf{Y_2} - \mathbf{X_2}\widehat{\boldsymbol{\beta}_2})].$$

Neste caso, o critério de informação de Schwarz, $SIC(k)$, é dado por

$$\begin{aligned}SIC(k) &= -2\,\ell_k(\widehat{\boldsymbol{\beta}_1},\widehat{\boldsymbol{\beta}_2},\widehat{\sigma^2}) + (2p+1)\ \log\ n\\ &= n\ \log[(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}_1})^\top(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}_1}) + (\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}_2})^\top(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}_2})]\\ &\quad +n\,(\log 2\pi+1)+(2p+1-n)\log n,\end{aligned}$$

onde $k=p,...,n-p$, $\mathbf{X}_1$ é uma partição da matriz $\mathbf{X}$ que considera as k primeiras linhas, $\mathbf{Y}_1$ é uma partição do vetor $\mathbf{Y}$ que considera as k primeiras observações, $\mathbf{X}_2$ é uma partição da matriz $\mathbf{X}$ que considera as $n-k$ últimas linhas e $\mathbf{Y}_2$ é uma partição do vetor $\mathbf{Y}$ que considera as $n-k$ últimas observações.

2.1.3 Modelo de Regressão Linear t de Student com Ponto de Mudança

Osorio & Galea (2006) propõem um modelo de regressão linear t de Student, onde os erros aleatórios ϵ_1,...,ϵ_n do modelo considerado em (2.1) são independentes e identicamente distribuídas como $\epsilon_i \overset{\text{iid}}{\sim} t(0,\sigma^2,\nu)$, para $i=1,\ldots,n$, onde σ^2 é um parâmetro desconhecido e ν são os graus de liberdade da distribuição t de Student.Assim, quando o ponto de mudança enconta-se na posição k não conhecida tem-se que $Y_i \sim\ t(\mathbf{x}_i^\top\boldsymbol{\beta}_1,\sigma^2,\nu)$, para $i=1,\ldots,k$ e $Y_i \sim\ t(\mathbf{x}_i^\top\boldsymbol{\beta}_2,\sigma^2,\nu)$, para $i=k+1,\ldots,n$.

Considerando o comentário de Fernández & Steel (1999) sobre os graus de liberdade ν, estes serão considerados conhecidos e como sugerido por Lange et al. (1989) uma avaliação de vários possíveis valores de ν deverá ser feita para escolher o que maximize a função de verossimilhança. Quando $\nu\to\infty$ a distribuição t de Student converge à normal e os resultados apresentados por Chen (1998) podem ser obtidos.

A estimação de máxima verossimilhança sob hipótese nula considera a função de log-verossimilhança dada por

$$\ell_0(\boldsymbol{\beta},\sigma^2) = n\ \log \mathrm{G}(\nu) - \frac{n}{2}\ \log\sigma^2 - \frac{\nu+1}{2}\sum_{i=1}^{n}\log\{1+d_i/\nu\},$$

onde $$d_i = \frac{(Y_i-\mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma^2},\quad i=1,\ldots,n \quad \text{e} \quad \mathrm{G}(\nu)=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})}.$$

As funções escore são

$$U(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\sum_{i=1}^{n} q_i(Y_i-\mathbf{x_i}^\top\boldsymbol{\beta})\mathbf{x_i} = \frac{1}{\sigma^2}\mathbf{X}^\top\mathbf{Q}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \qquad e$$

$$U(\sigma^2) = -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(Y_i-\mathbf{x_i}^\top\boldsymbol{\beta})^2 = -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\mathbf{V_Q}(\boldsymbol{\beta}),$$

onde $$\mathbf{V_Q}(\boldsymbol{\beta}) = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{Q}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}), \qquad com$$

$$\mathbf{Q} = diag(q_1, ..., q_n), \qquad e \qquad q_i = \frac{\nu+1}{\nu+d_i}, \quad i = 1, \ldots, n.$$

Neste caso, o critério de informação de Schwarz, $SIC(n)$, pode ser escrito como

$$SIC(n) \ = -2\,n\ \log \mathrm{G}(\nu) + n\ \log \widehat{\sigma^2} + (\nu+1)\sum_{i=1}^{n} \log\{1 + d_i/\nu\} + (p+1)\ \log\ n.$$

Os resultados equivalentes ao considerar a hipótese alternativa são:

$$\ell_k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2) = n\ \log \mathrm{G}(\nu) - \frac{n}{2}\ \log \sigma^2 - \frac{\nu+1}{2}[\sum_{i=1}^{k} \log\{1 + d_{1i}/\nu\} + \sum_{i=k+1}^{n} \log\{1 + d_{2i}/\nu\}],$$

$$U(\boldsymbol{\beta}_1) = \frac{1}{\sigma^2}\mathbf{X_1}^\top \mathbf{V_1}(\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1), \qquad U(\boldsymbol{\beta}_2) = \frac{1}{\sigma^2}\mathbf{X_2}^\top \mathbf{V_2}(\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2), \qquad e$$

$$U(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}[\mathbf{V_{Q_1}}(\boldsymbol{\beta_1}) + \mathbf{V_{Q_2}}(\boldsymbol{\beta_2})],$$

onde $\mathbf{V_{Q_1}}(\boldsymbol{\beta_1}) = (\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1)^\top \mathbf{Q_1}(\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1), \mathbf{V_{Q_2}}(\boldsymbol{\beta_2}) = (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2)^\top \mathbf{Q_2}(\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2)$,
com $\mathbf{Q_1} = diag(q_1, ..., q_k)$ *e* $\mathbf{Q_2} = diag(q_{k+1}, ..., q_n)$

$$SIC(k) \ = -2\,n\ \log \mathrm{G}(\nu) + n\ \log \widehat{\sigma^2} + (\nu+1)\ [\sum_{i=1}^{k} \log\{1 + d_{1i}/\nu\}$$

$$+ \sum_{i=k+1}^{n} \log\{1 + d_{2i}/\nu\}\ ] + (2p+1)\ \log\ n,$$

onde $d_{1i} = \dfrac{(Y_i - \mathbf{x_i}^\top \boldsymbol{\beta}_1)^2}{\sigma^2}, i = 1, \ldots, k, \qquad d_{2i} = \dfrac{(Y_i - \mathbf{x_i}^\top \boldsymbol{\beta}_2)^2}{\sigma^2}, i = k+1, \ldots, n,$

Sob ambas hipóteses, observa-se que as equações de verossimilhança correspondem a um sistema de equações não linear e pode ser resolvido via métodos numéricos. Na próxima seção, descrevemos o algoritmo EM e a aplicação deste para estimação do modelo de regressão linear t de Student.

2.2 O Algoritmo EM

Nas últimas décadas o desenvolvimento de algoritmos computacionais orientados à realização de estimações de parâmetros tem sido um tema recorrente na bibliografia e desde então, muitas aplicações em diversas áreas tem sido apresentadas. Um grande número destas estão relacionadas a inferência e modelagem robusta sobre conjuntos de dados simétricos. O algoritmo EM (Dempster, Laird & Rubin, 1977) é um processo iterativo muito utilizado e eficiente na estimação de parâmetros de modelos com dados incompletos a partir da maximização das funções de máxima verossimilhança. Em muitos casos, a obtenção de estimativas de máxima verossimilhança pode ser simplificada ao utilizar a formulação de dados aumentados, também chamado de dados completos. Estes dados são a união dos dados denominados incompletos que são observados e os dados denominados perdidos.

Uma especificação do algoritmo é a seguinte: Seja $\mathbf{y}_o$ e $\mathbf{y}_f$ que denotam os dados observados e faltantes, respectivamente. Denota-se como $\mathbf{y}_c = (\mathbf{y}_o, \mathbf{y}_f)$ os dados completos que comportam os dados observados e faltantes conjuntamente. Sejam $f(\mathbf{y}_c|\boldsymbol{\theta})$ a função de verossimilhança dos dados completos, $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \log(f(\mathbf{y}_c|\boldsymbol{\theta}))$, $\boldsymbol{\theta} \in \Theta$, a função de log-verossimilhança dos dados completos e $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ o valor esperado da log-verossimilhança dos dados completos condicionado aos dados observados $\mathbf{y}_o$ e os parâmetros estimados atuais. Temos que:

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}_o, \widehat{\boldsymbol{\theta}}], \tag{2.6}$$

onde $\widehat{\boldsymbol{\theta}}$ são os parâmetros estimados usados na avaliação da esperança condicional e $\boldsymbol{\theta}$ são as estimativas atualizadas, obtidas pela maximização de Q.

Cada iteração do algoritmo EM é composta por dois passos: O cálculo da esperança condicional (E), e a maximização (M). Assim, para a r-ésima iteração temos:

Passo E: Calcular $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r-1)})$ como uma função de $\widehat{\boldsymbol{\theta}}^{(r-1)}$ e os dados observados;

Passo M: Encontrar $\widehat{\boldsymbol{\theta}}^{(r)}$, tal que, $Q(\widehat{\boldsymbol{\theta}}^{(r)}|\widehat{\boldsymbol{\theta}}^{(r-1)}) = Max_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r-1)})$.

Os dois passos são repetidos quantas vezes seja necessário. A convergência é assegurada uma vez que o algoritmo garante o aumento da verossimilhança em cada iteração, isto é, as funções de verossimilhança observada $\ell(\boldsymbol{\theta}|\mathbf{y}_o)$ obtidas via o algoritmo EM nas iterações (r) e $(r+1)$ guardam a seguinte relação $\ell(\boldsymbol{\theta}^{(r)}|\mathbf{y}_o) \leq \ell(\boldsymbol{\theta}^{(r+1)}|\mathbf{y}_o)$, o que permite afirmar que o algoritmo geralmente converge a um máximo local ou global da função de log-verossimilhança. A verificação que o máximo verdadeiro é alcançado sempre deve ser realizada, ao que se recomenda rodar várias vezes as iterações do algoritmo EM com diferentes valores iniciais.

O passo M no algoritmo EM pode ser substituído por um processo de maximização condicional (CM) de alguma função dos parâmetros que estão sendo estimados. Este algoritmo foi proposto por Meng & Rubin (1993) e é denominado algoritmo de maximização condicional de esperança (ECM). Neste caso, maximiza-se a função Q sujeita a restrições em $\boldsymbol{\theta}$, tornando o algoritmo computacionalmente mais simples.

A seguir apresentamos os algoritmos EM sugeridos por Osorio & Galea (2006) para a estimação de parâmetros do modelo de regressão linear t de Student.

- Quando o modelo não apresenta ponto de mudança, a função de log-verossimilhança observada pode ser escrita como

$$\ell_0(\mathbf{Y}|\nu;\boldsymbol{\theta}) = -\frac{n}{2}\,\log(2\,\pi\,\sigma^2) + \frac{1}{2}\log\,|\mathbf{Q}| - \frac{1}{2\sigma^2}\mathbf{V}_{\mathbf{Q}}(\boldsymbol{\beta}).$$

O algoritmo EM maximiza a função de log-verossimilhança anterior de forma iterativa. A seguir são descritos os dois passos para a r-ésima iteração:

Passo E: Partindo de uma estimativa inicial $\widehat{\boldsymbol{\theta}}^{(r-1)} = (\widehat{\boldsymbol{\beta}^{\top}}^{(r-1)}, \widehat{\sigma^2}^{(r-1)})$, calcula-se os pesos $q_i{}^{(r)}$ a partir da esperança condicional

$$E(U_i|Y_i;\widehat{\boldsymbol{\theta}}^{(r-1)}) = q_i^{(r)} = \frac{\nu+1}{\nu+d_i^{(r-1)}},$$

onde $\quad d_i^{(r-1)} = \dfrac{(Y_i - \mathbf{x_i}^{\top}\widehat{\boldsymbol{\beta}^{\top}}^{(r-1)})^2}{\widehat{\sigma^2}^{(r-1)}} \quad i = 1,\ldots,n.$

Passo M: Usando os pesos obtidos no passo anterior, as estimativas de máxima verossimilhança podem ser escritas como

$$\widehat{\boldsymbol{\beta}}^{(r)} = (\mathbf{X}^{\top}\mathbf{Q}^{(r)}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{Q}^{(r)}\mathbf{Y}, \qquad \widehat{\sigma^2}^{(r)} = \frac{1}{n}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}^{(r)})^{\top}\mathbf{Q}^{(r)}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}^{(r)}),$$

onde $\quad \mathbf{Q}^{(r)} = diag(q_1^{(r)},...,q_n^{(r)}).$

- Quando o modelo em questão apresenta ponto de mudança na posição k, a função de log-verossimilhança observada é

$$\ell_k(\mathbf{Y}|\nu;\boldsymbol{\theta}) = -\frac{n}{2}\,\log\,2\,\pi\,\sigma^2 + \frac{1}{2}\log\,|\mathbf{Q_1}| - \frac{1}{2\sigma^2}\mathbf{V}_{\mathbf{Q_1}}(\boldsymbol{\beta_1}) + \frac{1}{2}\log\,|\mathbf{Q_2}| - \frac{1}{2\sigma^2}\mathbf{V}_{\mathbf{Q_2}}(\boldsymbol{\beta_2}).$$

Neste caso, os dois passos do algoritmo EM na r-ésima iteração são descritos como:

Passo E: Partindo de uma estimativa inicial $\widehat{\boldsymbol{\theta}}^{(r-1)} = (\widehat{\boldsymbol{\beta}_1^{\top}}^{(r-1)}, \widehat{\boldsymbol{\beta}_2^{\top}}^{(r-1)}, \widehat{\sigma^2}^{(r-1)})$, calcula-se os pesos $q_i^{(r-1)}$ a partir da esperança condicional

$$E(U_i|Y_i;\boldsymbol{\theta}^{(r-1)}) = q_i^{(r-1)} = \frac{\nu+1}{\nu+d_i^{(r-1)}},$$

onde

$$d_i = \begin{cases} \dfrac{(Y_i - \mathbf{x_i}^{\top}\widehat{\boldsymbol{\beta}_1^{\top}}^{(r-1)})^2}{\widehat{\sigma^2}^{(r-1)}} & i = 1,\ldots,k, \\ \dfrac{(Y_i - \mathbf{x_i}^{\top}\widehat{\boldsymbol{\beta}_2^{\top}}^{(r-1)})^2}{\widehat{\sigma^2}^{(r-1)}} & i = k+1,\ldots,n. \end{cases}$$

Passo M: Usando os pesos obtidos no passo anterior, as estimativas, nesta etapa, são obtidas por

$$\widehat{\boldsymbol{\beta}_1}^{(r)} = (\mathbf{X_1}^\top \mathbf{Q_1}^{(r)} \mathbf{X_1})^{-1} \mathbf{X_1}^\top \mathbf{Q_1}^{(r)} \mathbf{Y_1},$$

$$\widehat{\boldsymbol{\beta}_2}^{(r)} = (\mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{X_2})^{-1} \mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{Y_2} \quad e$$

$$\widehat{\sigma^2}^{(r)} = \frac{1}{n}[(\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1^{(r)})^\top \mathbf{Q_1}^{(r)} (\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1^{(r)}) + (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2^{(r)})^\top \mathbf{Q_2}^{(r)} (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2^{(r)})],$$

com $\mathbf{Q_1}^{(r)} = diag(q_1^{(r)}, ..., q_k^{(r)})$, $\quad \mathbf{Q_2}^{(r)} = diag(q_{k+1}^{(r)}, ..., q_n^{(r)})$ e considerando que $\mathbf{X}_1$ é uma partição da matriz $\mathbf{X}$ que considera as k primeiras linhas, $\mathbf{Y}_1$ é uma partição do vetor $\mathbf{Y}$ que considera as k primeiras observações, $\mathbf{X}_2$ é uma partição da matriz $\mathbf{X}$ que considera as $n-k$ últimas linhas e $\mathbf{Y}_2$ é uma partição do vetor $\mathbf{Y}$ que considera as $n-k$ últimas observações.

Em ambas situações, os passos E e M são repetidos até o convergência do algoritmo.

2.3 Um Exemplo em Regressão Linear Simples

A seguir apresentamos a aplicação dos resultados mostrados nas duas seções anteriores sobre um conjunto de dados simulados. Os diversos resultados são comparados.

1) Conjunto de dados simulados

É considerada uma amostra de tamanho 20 do modelo de regressão $y_i = \beta_0 + \beta_1 x_i + e_i$, onde os erros e_i seguem uma distribuição t de Student com 2 graus de liberdade para $i = 1, \ldots, n$. Considera-se adicionalmente que o conjunto de dados apresenta um ponto de mudança na posição 10. Assim, os modelos que geram as observações apresentadas na Tabela 2.1 foram definidos como

$$y_i = 2 + 0.5\, x_i + e_i, \qquad i = 1, ..., 10,$$

$$y_i = 4 + 1,5\, x_i + e_i, \qquad i = 11, ..., 20.$$

2) Ajuste dos modelos

Primeiramente, assume-se que a distribuição dos erros do modelo segue uma normal com média zero e variância desconhecida para $i = 1, \ldots, n$. O conjunto de dados é ajustado a partir de um modelo de regressão linear simples utilizando os resultados apresentados por Chen (1998). Quando não se considera a existência de um ponto de mudança os resultados sob hipótese nula são utilizados. A reta obtida por este ajuste é mostrada na parte (a) da Figura 2.1.

Mantendo a suposição que a distribuição dos erros dos modelos ajustados segue uma normal com média zero e variância desconhecida para $i = 1, \ldots, n$ e considerando desta vez que existe uma ponto de mudança conhecido na posição 10, o conjunto de dados deve ser ajustado utilizando os resultados sob hipótese alternativa, obtendo duas retas. Levando em consideração a indexação e determinando k=10, a primeira reta é ajustada utilizando os dez primeiros dados e a

Tabela 2.1: Dados simulados para modelo de regressão linear simples proposto.

Indicador de Indexação	Variável Y	Variável X
1	7.68	9.84
2	8.25	9.68
3	4.16	7.87
4	20.53	5.53
5	5.48	4.47
6	4.57	3.25
7	7.34	9.26
8	0.97	9.78
9	3.28	5.64
10	11.60	8.61
11	32.99	18.95
12	31.63	18.09
13	21.42	11.36
14	31.04	17.18
15	24.47	14.32
16	32.19	19.04
17	28.94	15.90
18	21.46	10.39
19	19.80	10.43
20	27.95	16.41

segunda utilizando os últimos dez dados. As retas obtidas são mostradas na parte (b) da Figura 2.1.

Observe que nas duas situações anteriores as estimações de máxima verossimilhança dos parâmetros de regressão são os mesmos que os obtidos ao utilizar o método de mínimos quadrados. Para o caso do modelo com ponto de mudança devem considerar as partições adequadas de X e Y.

Ao considerar que os erros seguem uma distribuição t de Student com dois graus de liberdade para $i = 1, \ldots, n$, os resultados para os modelos de regressão robustos apresentados por Osorio & Galea (2006) serão utilizados. Novamente, os resultados sob hipótese nula são utilizados quando se assume que não existe um ponto de mudança e os resultados sob hipótese alternativa com k=10 são considerados ao assumir a existência de um ponto de mudança nessa posição. O algoritmo EM é utilizado para a estimação dos parâmetros em ambas situações.

As partes (c) e (d) da Figura 2.1 apresentam as retas ajustadas ao considerar os modelos robustos. O primeiro corresponde ao ajuste para um modelo de regressão t de Student sem ponto de mudança e o último para um modelo de regressão t de Student com ponto de mudança na posição 10.

3) Aplicação da metodologia para determinação de ponto de mudança

A Figura 2.2 apresenta os resultados da aplicação da metodologia para determinação de ponto de mudança que usa o critério de informação de Schwarz, (SIC), proposto por Chen (1998) para os modelos de regressão normal e t de Student (Osorio & Galea, 2006). Observa-se que em ambas

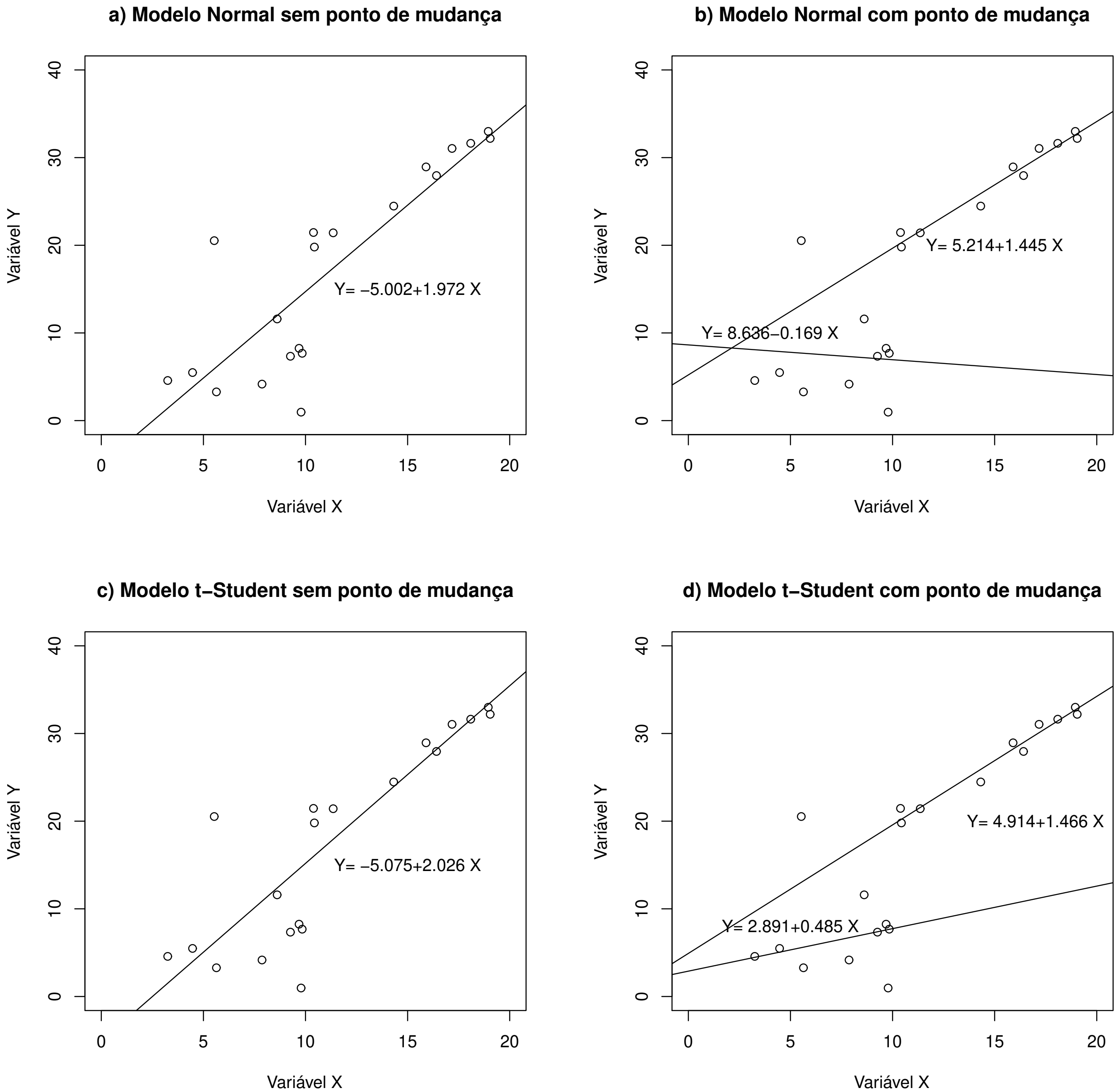


Figura 2.1: Ajuste de modelos de regressão Normal e t de Student para dados simulados considerando diversas suposições.

situações identifica-se o ponto de mudança na posição 10. O menor valor apresenta-se no $SIC(10)$ do modelo de regressão t de Student. A aplicação da metologia proposta leva a escolha do modelo que reflete as características do conjunto de dados simulado.

4) Comparativos dos modelos sugeridos

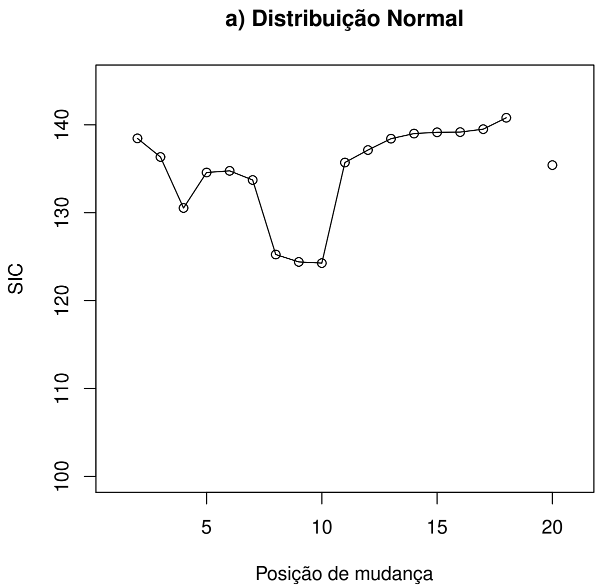


Figura 2.2: Valores SIC dos modelos de regressão Normal e t de Student ajustados sobre dados simulados.

Os modelos de regressão linear que não consideram um ponto de mudança captam a tendência da relação, no entanto, os ajustes para os dados com valores extremos apresentam maiores diferenças com os valores ajustados. É intuitivo que isso aconteça, pois ao considerar os dados como um único conjunto (o que não condiz com a simulação realizada), a única reta que ajustará os dados será menos precisa nos dados mais extremos, assim a avaliação sob a presença de um ponto de mudança torna-se importante.

Ao observar os ajustes do modelo de regressão normal com ponto de mudança na posição 10, a primeira reta mostra uma relação negativa (o valor de β_{12} é negativo, o que difere da estrutura da simulação), isto devido principalmente à observação número 10 do conjunto que causa um efeito de alavanca no ajuste. Assim, considerar um modelo de regressão normal não é a melhor opção quando dados aberrantes estão presentes em um conjunto de dados. Esta situação não ocorre quando é considerado o modelo de regressão t de Student. De fato, considerar modelos de regressão robustos torna-se uma opção para descrever melhor a relação entre variáveis de um conjunto de dados.

Finalmente, a metodologia para deteção de um ponto de mudança que considera o SIC é uma boa alternativa para este fim.

2.4 Distribuições de Mistura de Escala Normal

O interesse pelo estudo teórico de distribuições simétricas que possuem caudas mais pesadas que a distribuição normal tem sido crescente desde a década de 70. Neste contexto, Andrews & Mallows (1974) apresentam uma classe de distribuções simétricas denominadas de Mistura de Escala Normal que agrupa distribuições simétricas, entre elas as distribuições normal e t de Student. Nesta seção, apresentam-se as principais características desta classe de distribuições.

2.4.1 Definição e Notação

Uma variável aleatória Y tem distribuição de Mistura de Escala Normal com parâmetro de locação $\mu \in \mathbb{R}$ e parâmetro de dispersão σ^2 (maior que zero) se a sua função de densidade de probabilidade (f.d.p.) é dada por

$$f(y) = \int_0^\infty \phi(y; \mu, \kappa(u)\sigma^2) dH(u), \tag{2.7}$$

onde $\phi(.; \mu, \sigma^2)$ denota a f.d.p. de uma distribuição normal univariada com média μ e variância σ^2, $\kappa(.)$ é uma função de ponderação positiva, e U é uma variável aleatória positiva com função de distribuição acumulada (f.d.a.) $H(u; \boldsymbol{\nu})$, $\boldsymbol{\nu}$ é um escalar ou vetor de parâmetros da distribuição de U que controla as caudas da distribuição.

Esta distribuição é denotada por $MEN(\mu, \sigma^2, H)$.

2.4.2 Representação Estocástica

Seja Y uma variável aleatória distribuída como $MEN(\mu, \sigma^2, H)$. Este pode ser representado estocasticamente como

$$Y = \mu + \kappa^{1/2}(U)Z, \tag{2.8}$$

onde μ é um parâmetro de locação ($\mu \in \mathbb{R}$), Z é uma variável aleatória Normal unidimensional com média zero e variância σ^2, $\kappa(.)$ é uma função de ponderação positiva e U é uma variável aleatória positiva com (f.d.a.) H(u;$\boldsymbol{\nu}$) e função de densidade de probabilidade (f.d.p.) h(u;$\boldsymbol{\nu}$), independente de Z, onde $\boldsymbol{\nu}$ é um escalar ou vetor de parâmetros da distribuição de U.

O seguinte resultado é importante pois será utilizado em próximos desenvolvimentos.

Proposição 2.4.1. *Sabendo que Y tem distribuição de Mistura de Escala Normal unidimensional $MEN(\mu, \sigma^2, H)$, a distribuição de Y condicionada a U tem distribuição normal unidimensional da forma:*

$$Y \mid U = u \sim N(\mu, k(u)\sigma^2) \tag{2.9}$$

Demonstração. A prova é realizada diretamente a partir da representação estocástica dada em (2.8). □

As distribuções de Mistura de Escala Normal são constituídas por famílias paramétricas de distribuições probabilísticas que preservam a estrutura simétrica das distribuições normais.

Comparadas com outras familias de distribuições com estas características, a classe de distribuições de Mistura de Escala Normal é mais geral que a classe de distribuições Normal/Independente (Lange & Sinsheimer, 1993), porém mais restrita do que a classe de distribuições elípticas (Arellano-Valle, 1994), isto é, ela pode ser considerada uma sub classe desta.

A distribuções normal é um membro particular desta sub classe. Outras distribuições conhecidas que compõem esta classe de distribuições são: t de Student, Slash e Normal Contaminada. Em Andrews & Mallows (1974), encontra-se um estudo mais amplo das distribuições de Mistura de Escala Normal, incluindo a apresentação de outras distribuições que compõem esta sub classe.

Na Figura 2.3, apresentamos um diagrama que contextualiza as diversas classes de distribuições simétricas.

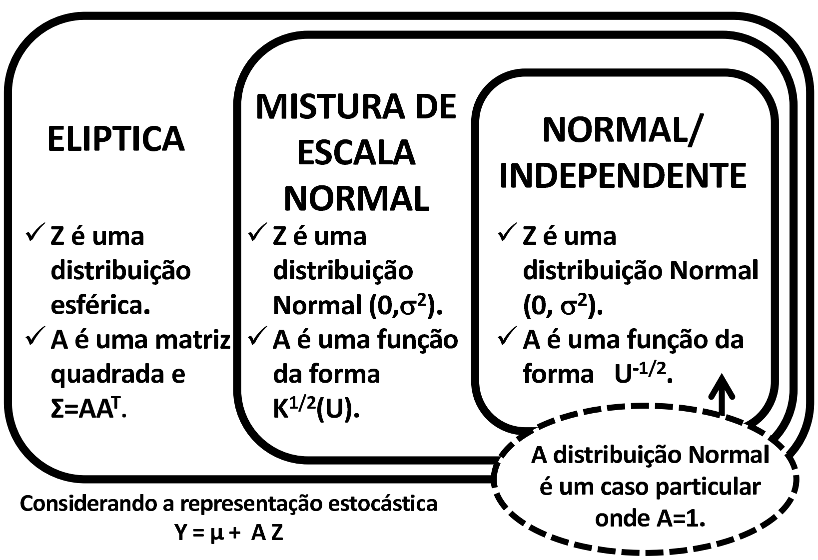


Figura 2.3: Diagrama para classes de distribuições de probabilidade simétricas

2.4.3 Propriedades

Considerando os supostos apresentados anteriormente, algumas propriedades interessantes da distribução de Mistura de Escala Normal são:

(P1) Se a esperança existe, então $E[Y] = \mu$, se $E[\kappa^{1/2}(U)] < \infty$.

(P2) Se a variância existe, então $V[Y] = E[\kappa(U)]\sigma^2$, se $E[\kappa(U)] < \infty$.

(P3) A f.d.p. definida em (2.7) é infinitamente diferenciável em y, μ e σ^2 exceto possivelmente em $y = \mu$. Esta é diferenciável na ordem j se nestes pontos $E[\kappa^{(1/2)+j}(U)] < \infty$.

(P4) Qualquer subvetor de Y tem densidade marginal com a mesma forma geral de Y.

(P5) A regressão de qualquer subvetor de Y sobre seu subvetor complementar é linear.

(P6) A função característica de Y é:

$$E[e^{itY}] = e^{it\mu} \int_0^\infty e^{(-\frac{1}{2}\kappa(u)t^2\sigma^2)} dH(u).$$

A verificação destas propriedades podem ser encontradas nos trabalhos e Box & Tiao (1973), Andrews & Mallows (1974), Dempster et al. (1980) e Fang, Kost & Ng (1990).

2.4.4 Algumas Distribuições Especificas

Quatro distribuições de probabilidade conhecidas e que fazem parte da classe de Mistura de Escala Normal são apresentadas na Tabela 2.2. Para cada uma delas se indica as funções $\kappa(.)$ e U que as caracterizam.

Tabela 2.2: Quatro distribuições de Mistura de Escala Normal univariadas.

Distribuição	Notação	$\kappa(.)$	U	
Normal	$N(\mu, \sigma^2)$	1	*Degenerada*	
t de Student	$t(\mu, \sigma^2, \nu)$	$1/u$	$Gamma(\frac{\nu}{2}, \frac{\nu}{2})$	(1)
Slash	$SL(\mu, \sigma^2, \nu)$	$1/u$	$Beta(\nu, 1)$	
Normal Contaminada	$CN(\mu, \sigma^2, \nu, \gamma)$	$1/u$	*Discreta*	(2)

(1) Considerando $Gamma(a, b)$ com média $\frac{a}{b}$.

(2) *Discreta com f.d.p.* $h(u; \boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)},\ 0 \leq \nu \leq 1,\ 0 < \gamma \leq 1.$

Levando em conta que para as aplicações que incluem um desenvolvimento computacional é importante que a distribuição de U seja conhecida e possível de ser trabalhada, os resultados específicos desta dissertação serão obtidos considerando estas quatro distribuições.

A Figura 2.4 mostra que para distribuições normais com diversos valores de média e variância é possível encontrar várias alternativas de distribuições de caudas pesadas dentro da classe da Mistura de Escala Normal. Por outo lado, observa-se que a simetria e mantida, porém as formas, especialmente nas caudas, apresentam diferenças que dependem dos parâmetros próprios destas distribuições. A convergência para a distribuição normal para o caso das distribuições t de Student e Slash ocorre quando $\nu \to \infty$. Para a distribuição normal contaminada, esta convergência ocorre quando $\nu \to 0$ ou $\gamma \to 1$.

Finalmente, um resultado importante para a implementação dos algoritmos computacionais EM é a obtenção dos momentos condicionais definidos por $q(d) = E[k^{-1}(U)|Y = y]$, onde $d = \frac{(y-\mu)^2}{\sigma^2}$ é a distância de Mahalanobis.

Considerando a Proposição 2.4.1 e usando adequadamente as informações da Tabela 2.2 na relação conhecida $f_U(u|Y = y) = \dfrac{f_Y(y|U = u)h(u;\boldsymbol{\nu})}{f_Y(y)}$, os valores $q(d)$ para as quatro distribuições estudadas são:

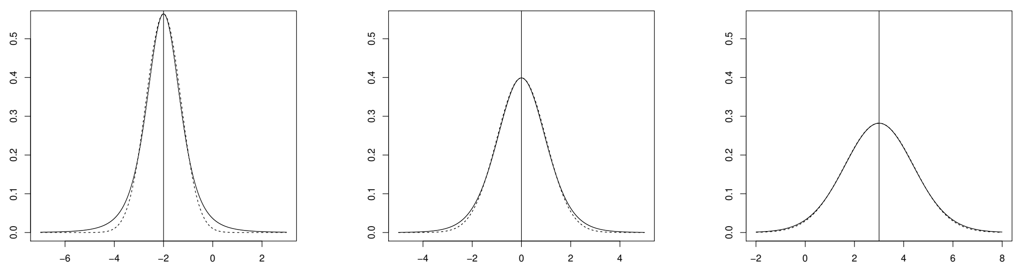
- Distribuição normal
$$q(d) = 1, \tag{2.10}$$
onde $f_Y(y) = \phi(y;\mu,\sigma^2), \qquad y \in \mathbb{R}$.

- Distribuição t de Student
$$q(d) = \frac{\nu+1}{\nu+d}, \tag{2.11}$$
onde $f_Y(y) = \dfrac{\Gamma(\frac{1+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{1/2}}\nu^{-1/2}\sigma^{-1}\left(1+\frac{1}{\nu}\right)^{-(1+\nu)/2}, \qquad y \in \mathbb{R}$.

- Distribuição slash
$$q(d) = \left(\frac{1+2\nu}{d}\right)\frac{P_1(1/2+\nu+1, d/2)}{P_1(1/2+\nu, d/2)}, \tag{2.12}$$
onde $f_Y(y) = \displaystyle\int_0^1 u^{\nu-1}\phi(y;\mu,u^{-1}\sigma^2)du, \qquad y \in \mathbb{R}$.
$P_x(a,b)$ denota a função de distribuição acumulada da $Gamma(a,b)$ com média $\frac{a}{b}$ avaliada no valor x.

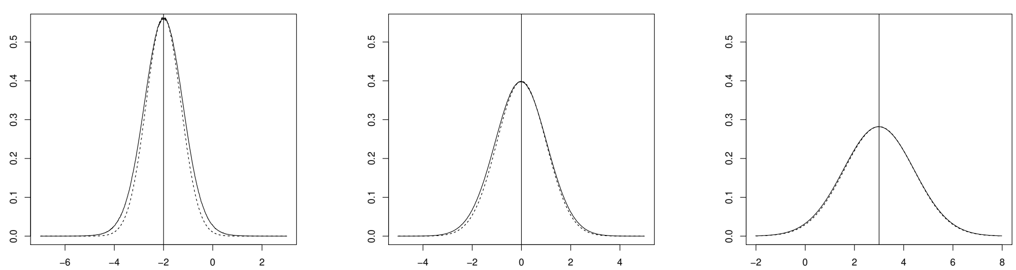
- Distribuição normal contaminada
$$q(d) = \frac{1-\nu+\nu\gamma^{3/2}\exp\{(1-\gamma)d/2\}}{1-\nu+\nu\gamma^{1/2}\exp\{(1-\gamma)d/2\}}, \tag{2.13}$$
onde $f_Y(y) = \nu\,\phi(y;\mu,\gamma^{-1}\sigma^2) + (1-\nu)\,\phi(y;\mu,\sigma^2), \qquad \mathbf{y} \in \mathbb{R}$.
O parâmetro ν pode ser interpretado como a proporção de pontos aberrantes e γ como um fator de escala.

ν=4 ν=10 ν=30



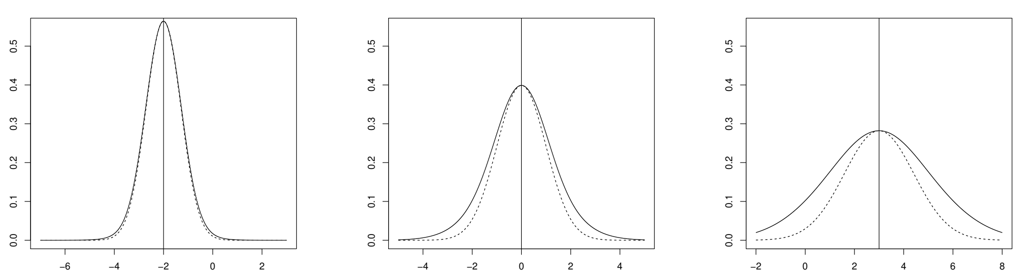
(b) Distribuição Slash

ν=4 ν=10 ν=30



(c) Distribuição Normal Contaminada

ν=0.1, γ=0.4 ν=0.5,γ=0.4 ν=0.9, γ=0.4



ν=0.4, γ=0.1 ν=0.4,γ=0.5 ν=0.4, γ=0.9

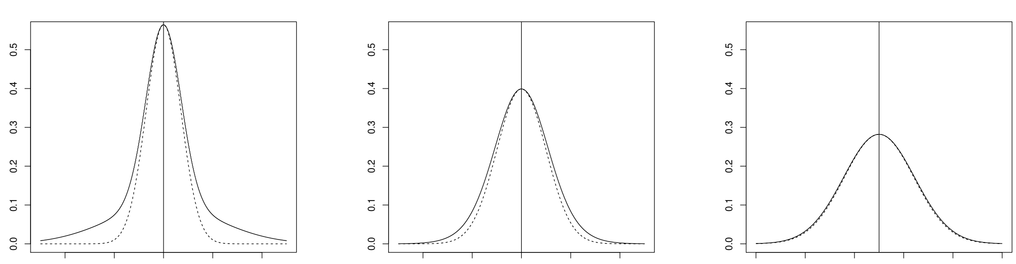


Figura 2.4: Algumas distribuições que compõem a classe de Mistura de Escala Normal (linha cheia) como alternativas para $N(-2, 0.5)$, $N(0, 1)$ e $N(3, 2)$ (linha pontillada).

(a) Distribuição t de Student

Capítulo 3

Modelos de Regressão Linear Mistura de Escala Normal com Ponto de Mudança

Diversos modelos estatísticos robustos considerando classes de distribuições simétricas tem sido apresentados na literatura estatística nos ultimos anos, veja por exemplo, Fernandez & Steel (2000) e Rosa et al. (2003). Estes modelos são apresentados como uma alternativa aos modelos que assumem normalidade e que se tornam restritivos em algumas situações práticas de modelagem estatística.

Neste capítulo estudam-se modelos de regressão linear robustos com erros distribuídos de acordo a uma distribuição da familia de mistura de escala normal e que apresentam um ponto de mudança. Assim considera-se que os erros aleatórios apresentam uma distribuição de caudas mais pessadas que a distribuição normal. O modelo de regressão linear clássico com ponto de mudança que considera os erros normalmente distribuídos é um caso particular, pois a distribuição normal faz parte da classe de distribuições considerada.

Primeiramente, apresenta-se de forma geral a especificação do modelo proposto e a função de verossimilhança associada ao modelo, logo é realizado um estudo de inferência que inclui a estimação de parâmetros e uma metodologia para identificação do ponto de mudança quando este não é conhecido para três situações específicas em modelos com ponto de mudança. Em todos os casos apresenta-se estudos de simulação.

As três configurações estudadas com respeito a mudança produzida em um modelo de regressão linear a partir da observação k são:

(1) Modelo de regressão linear com mudança nos coeficientes de regressão a partir da observação k.

(2) Modelo de regressão linear com mudança na variância dos erros a partir da observação k.

(3) Modelo de regressão linear com mudança nos coeficientes de regressão e na variância dos erros a partir da observação k.

3.1 Modelo de Regressão Linear

Considerando uma variável aleatória Y_i, que satisfaz a seguinte relação:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_{(p-1)i} x_{(p-1)i} + \epsilon_i, \tag{3.1}$$

onde i representa a i-ésima observação, β_0, β_1, β_2 ... $\beta_{(p-1)}$ são p coeficientes de regressão desconhecidos e ϵ_i é um erro aleatório para $i = 1, \ldots, n$. Vamos supor que os erros ϵ_i são variáveis aleatórias independentes, tais que $\epsilon_i \sim MEN(0, \sigma_i^2, H)$. O modelo de regressão linear Mistura de Escala Normal (MRL-MEN) considera que Y_i segue uma distribuição na classe Mistura de Escala Normal, isto é, $Y_i \sim MEN(\mu_i, \sigma_i^2, H)$, onde $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_{(p-1)} x_{(p-1)i}$, $i = 1, \ldots, n$.

Quando existe um ponto de mudança na observação k, o modelo geral de regressão linear Mistura de Escala Normal com ponto de mudança (MRL-MEN-PM) considera Y_i´s independentes com distribuição $MEN^{(1)}$, para $i = 1, \ldots, k$ e $MEN^{(2)}$, para $i = k+1, \ldots, n$, onde $MEN^{(1)}$ e $MEN^{(2)}$ é definida conforme a configuração estudada. Considerando as três configurações que serão tratadas nesta dissertação, os modelos de regressão estudados apresentarão mudanças na média de Y_i (isto é, coeficientes de regressão do modelo), na variância ou na média e variância conjuntamente. As distribuições consideradas ao especificar o modelo com ponto de mudança devem refletir cada uma destas estruturas. Assim, quando consideramos uma observação Y_i distribuída como $MEN(\mu_i, \sigma_i^2, H)$, as mudanças poderam ocorrer em μ_i, ou em σ_i^2 ou em ambos. O fato de ter H inalterada significa que a função de ponderação positiva e a variável aleatória positiva U não sofrem mudança.

3.2 Função de Verossimilhança

Considerando uma amostra aleatória observada $\mathbf{y} = (y_1, ..., y_n)^\top$ a função de log-verossimilhança para $\boldsymbol{\theta}$ é da forma

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \qquad i = 1, \ldots, n. \tag{3.2}$$

Sabendo que a função de verossimilhança para uma observação i é

$$L_i(\theta) = \int_0^\infty \frac{1}{(2\pi\, \kappa(u_i)\, \sigma_i^2)^{1/2}} \exp\left\{ -\frac{1}{2} (k(u_i)\, \sigma_i^2)^{-1}\, (y_i - \mathbf{x}_i^\top \beta)^2 \right\} dH(u_i),$$

e ao considerar a distância de Mahalanobis $d_i = \frac{(y_i - \mathbf{x}_i^\top \beta)^2}{\sigma_i^2}$, a expressão anterior pode ser escrita como

$$L_i(\theta) = \frac{1}{(2\pi\, \sigma_i^2)^{1/2}} \int_0^\infty \kappa(u_i)^{-1/2} \exp\left\{-\frac{1}{2}k(u_i)^{-1}\, d_i\right\} dH(u_i).$$

Por outro lado, considerando

$$\mathbb{I}_i(w) = \int_0^\infty \kappa^{-w}(u_i) \exp\left\{-\frac{1}{2}\kappa^{-1}(u_i) d_i\right\} dH(u_i), \tag{3.3}$$

temos que $\quad L_i(\theta) = \dfrac{1}{(2\pi\, \sigma_i^2)^{1/2}}\, \mathbb{K}_i, \quad$ onde $\quad \mathbb{K}_i = \mathbb{I}_i(1/2), \qquad i = 1, \ldots, n$. Logo,

$$\ell_i(\boldsymbol{\theta}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma_i^2 + \log \mathbb{K}_i, \qquad i = 1, \ldots, n, \tag{3.4}$$

e de (3.2)

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^n \log \sigma_i^2 + \sum_{i=1}^n \log \mathbb{K}_i, \qquad i = 1, \ldots, n. \tag{3.5}$$

Quando a mudança no posição k é indicada e o modelo particular que depende da configuração da mudança é especificado, a expressão (3.5) deverá ser trabalhada adequadamente. No entanto, esta expressão é o ponto de partida para a obtenção da função de log-verossimilhança de qualquer um dos modelos com ponto de mudança propostos anteriormente.

3.3 Matriz de Informação Observada

Para a obtenção da matriz de informação observada é preciso calcular as derivadas de primeira e segunda ordem de $\ell(\boldsymbol{\theta})$ com respeito à $\boldsymbol{\theta}$. De (3.2), temos que

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \qquad \text{e} \qquad \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}.$$

a) Derivada de primeira ordem de $\ell(\boldsymbol{\theta})$

Considerando (3.4) a primeira derivada de $\ell_i(\boldsymbol{\theta})$ $(i = 1, \ldots, n)$ com respeito à $\boldsymbol{\theta}$ é

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{\partial \log \sigma_i^2}{\partial \boldsymbol{\theta}} + \frac{\partial \log \mathbb{K}_i}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{\partial \log \sigma_i^2}{\partial \boldsymbol{\theta}} + \frac{1}{\mathbb{K}_i}\frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}}.$$

Logo, temos

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\sum_{i=1}^n \frac{\partial \log \sigma_i^2}{\partial \boldsymbol{\theta}} + \sum_{i=1}^n \frac{1}{\mathbb{K}_i}\frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}}. \tag{3.6}$$

b) Derivada de segunda ordem de $\ell(\boldsymbol{\theta})$

A segunda derivada de $\ell_i(\boldsymbol{\theta})$ $(i = 1, \ldots, n)$ com respeito à $\boldsymbol{\theta}$ é

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{1}{2} \frac{\partial^2 \log \sigma_i^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + \frac{1}{\mathbb{K}_i} \frac{\partial^2 \mathbb{K}_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{1}{\mathbb{K}_i^2} \frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}} \frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}^\top}.$$

Logo, temos

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 \log \sigma_i^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + \sum_{i=1}^{n} \left\{ \frac{1}{\mathbb{K}_i} \frac{\partial^2 \mathbb{K}_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{1}{\mathbb{K}_i^2} \frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}} \frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}^\top} \right\}. \tag{3.7}$$

Observa-se novamente que quando a mudança no posição k é indicada e o modelo particular que depende da configuração da mudança é especificado, as expressões (3.6) e (3.7) devem ser trabalhados adequadamente.

c) A expressão $\mathbb{K}_i$

Considerando (3.3) e sabendo que $\mathbb{K}_i = \mathbb{I}_i(1/2)$ no modelo que está sendo apresentado, a expressão $\mathbb{K}_i$ para cada distribuição considerada nesta dissertação pode ser escrita como:

- Normal

$$\mathbb{K}_i = (2\pi)^{1/2} \phi({d_i}^{1/2}; 0, 1),$$

onde $\phi(x; 0, 1)$ é a função de distribuição de probabilidade de uma normal padrão avaliado no ponto x.

- t de Student

$$\mathbb{K}_i = \frac{2^{\frac{1}{2}} \nu^{\frac{\nu}{2}} \Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) (d_i + \nu)^{\frac{\nu+1}{2}}},$$

onde $\Gamma(a)$ é a função gamma avaliada em a.

- Slash

$$\mathbb{K}_i = \frac{\nu\, 2^{\nu + \frac{1}{2}} \Gamma(\nu + \frac{1}{2})}{{d_i}^{\nu + \frac{1}{2}}} \int_0^1 f(u_i) du_i,$$

onde $f(u_i)$ é a função de distribuição de probabilidade $Gamma(\nu + 1/2, d_i/2)$ com média $\frac{2\nu+1}{d_i}$.

- Normal Contaminada

$$\mathbb{K}_i = (2\pi)^{1/2} [\nu\, \phi({d_i}^{1/2}; 0, 1/\gamma) + (1 - \nu)\, \phi({d_i}^{1/2}; 0, 1)],$$

onde $\phi(x; 0, b)$ é a função de distribuição de probabilidade de uma normal com média zero e variância b avaliado no ponto x.

Por outro lado, para a obtenção das derivadas da expressão $\mathbb{K}_i$ com respeito à $\boldsymbol{\theta}$ considera-se a seguinte transformação:

$$\mathbb{K}_i = h(g(d_i)) = \int_0^\infty c\, \exp\left(-g(d_i)\right) dH(u_i),$$

onde $z = g(d_i) = \frac{1}{2}\kappa^{-1}(u_i)d_i$ e $c = \kappa^{-1/2}(u_i)$.

Logo, a primeira e segunda derivadas de $\mathbb{K}_i$ com respeito à $\boldsymbol{\theta}$ podem ser expressadas como

$$\frac{\partial \mathbb{K}_i}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\mathbb{I}_i(3/2)\frac{\partial d_i}{\partial \boldsymbol{\theta}} \qquad e \qquad \frac{\partial^2 \mathbb{K}_i}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = \frac{1}{4}\mathbb{I}_i(5/2)\frac{\partial d_i}{\partial \boldsymbol{\theta}}\frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} - \frac{1}{2}\mathbb{I}_i(3/2)\frac{\partial^2 d_i}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}. \tag{3.8}$$

Os valores das expressões de $\mathbb{K}_i$ obtidas para as distribuições t de Student, Slash e Normal Contaminada são casos específicos da generalização de $\mathbb{I}_i(w)$ apresentada por Lachos et al. (2011), desta é possível obter $\mathbb{I}_i(3/2)$ e $\mathbb{I}_i(5/2)$ quando necessário. No caso da distribuição normal $\mathbb{I}_i(w) = \mathbb{K}_i$ para todo w.

Finalmente, a matriz de informação observada é obtida como $J = -\dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}$.

3.4 Determinação de Ponto de Mudança

Consideramos a metodologia proposta por Chen (1998) que sugere que o problema de ponto de mudança num modelo de regressão linear seja visto como um teste de hipótese da forma apresentada em (2.2). O Critério de Informação de Schwarz (SIC) definido em (2.3) é usado para verificar a existência de um ponto de mudança em um modelo de regressão linear.

Quando a posição k do ponto de mudança não é conhecida, H_1 deverá estabelecer uma coleção de modelos com ponto de mudança que considera as posições $k = p, ..., n - p$. Ao sugerir modelos de regressão linear Mistura de Escala Normal é necessário considerar também a suposição adotada para a distribuição dos erros aleatórios, o critério poderá sugerir também a distribuição a ser considerada.

A classe de Mistura de Escala Normal engloba diversas distribuições, neste trabalho consideramos as distribuições Normal, t de Student, Slash e Normal Contaminada. Qualquer uma delas pode ser usada na suposição sobre os erros aleatórios do modelo de regressão linear.

Critérios de decisão

- O modelo considerado apresenta um ponto de mudança, que equivale a rejeitar a hipótese nula, se:

$SIC(n) > SIC(k),\, para\, k = p, ..., n - p$

- Quando a hipótese nula é rejeitada, a posição estimada do ponto de mudança via máxima verossimilhança corresponde ao valor $\widehat{k}$ que satisfaz:

$SIC(\widehat{k}) = min\{SIC(k) : k = p, ..., n - p\}$

Chen & Gupta (2001) propõem abordar o problema geral de ponto de mudança como um teste de hipótese da forma

$$\begin{aligned} H_0 &: \quad \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = ... = \boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n = \boldsymbol{\theta}, \\ H_1 &: \quad \boldsymbol{\theta}_1 = ... = \boldsymbol{\theta}_k \neq \boldsymbol{\theta}_{k+1} = ... = \boldsymbol{\theta}_n, \end{aligned} \tag{3.9}$$

onde f_1,f_2,...,f_n são as funções de distribuição de probabilidade da sequência de variáveis aleatórias x_1,x_2,...,x_n que pertencem a uma familia paramétrica comum f($\boldsymbol{\theta}$). A hipótese nula indica que não existe ponto de mudança contra a hipótese alternativa que identifica um ponto de mudança na observação k.

O teste de hipótese sugerido por Chen (1998) é um caso particular desta generalização, assim é possível utilizar a metodologia proposta nas três configurações de mudança estudadas nesta dissertação propondo um teste de hipótese adequado para cada uma delas.

3.5 Mudança nos Coeficientes de Regressão

3.5.1 Especificação do Modelo

O modelo geral (3.1) pode ser escrito de forma matricial como:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{3.10}$$

onde $\boldsymbol{\beta}$=$(\beta_0, \beta_1, ..., \beta_{p-1})^\top$, $\mathbf{x}_i$=$(1, x_{1i}, ..., x_{1(p-1)})^\top$ e os erros ϵ_1,...,ϵ_n são independentes e identicamente distribuídos como $\epsilon_i \overset{\text{iid}}{\sim} MEN(0, \sigma^2, H)$ para $i = 1, \ldots, n$.

A considerar a existência de um ponto de mudança na posição k, temos que:

$$\begin{aligned} &Y_i \overset{\text{iid}}{\sim} MEN(\mathbf{x}_i^\top \boldsymbol{\beta}_1, \sigma^2; H), \qquad para \quad i = 1, \ldots, k \qquad e \\ &Y_i \overset{\text{iid}}{\sim} MEN(\mathbf{x}_i^\top \boldsymbol{\beta}_2, \sigma^2; H), \qquad para \quad i = k+1, \ldots, n. \end{aligned} \tag{3.11}$$

O teste de hipótese apresentado em (2.2) é o adequado para esta situação de mudança. Usando apropriadamente a expressão (3.5), as funções de log-verossimilhança para este modelo nas situações de hipótese nula e hipótese alternativa são, respetivamente:

$$\ell_0(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \sum_{i=1}^{n}\log\mathbb{K}_i, \tag{3.12}$$

$$\ell_k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \sum_{i=1}^{k}\log\mathbb{K}_i^{(1)} + \sum_{i=k+1}^{n}\log\mathbb{K}_i^{(2)}, \tag{3.13}$$

onde
$$d_i = \begin{cases} \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2}, & em \quad \mathbb{K}_i, \\ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\sigma^2}, & em \quad \mathbb{K}_i^{(1)}, \\ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2}{\sigma^2}, & em \quad \mathbb{K}_i^{(2)}. \end{cases}$$

As expressões (3.12) e (3.13) tornam-se importantes, pois serão utilizadas na aplicação do critério para determinação de ponto de mudança. Estas deverão ser avaliadas nas estimativas de máxima verossimilhança dos parâmetros. Assim, será possível encontrar os valores $\ell_0(\widehat{\boldsymbol{\theta}}) = \ell_0(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})$ e $\ell_k(\widehat{\boldsymbol{\theta}}) = \ell_k(\widehat{\boldsymbol{\beta}_1}, \widehat{\boldsymbol{\beta}_2}, \widehat{\sigma^2})$ que compõem o Critério de Informação Schwarz (SIC) utilizado como parte da metodologia proposta.

3.5.2 Derivadas Parciais de $\ell(\boldsymbol{\theta})$

Considerando os resultados em (3.8) apresentados para a expressão $\mathbb{K}_i$ e usando o resultado (3.6), os cálculos necessários para a obtenção das derivadas parciais de primeira ordem de $\ell(\boldsymbol{\theta})$ são

$$\frac{\partial \log \sigma^2}{\partial \boldsymbol{\beta}_1} = 0, \qquad \frac{\partial \log \sigma^2}{\partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial \log \sigma^2}{\partial \sigma^2} = \frac{1}{\sigma^2},$$

$$\frac{\partial d_i}{\partial \boldsymbol{\beta}_1} = \begin{cases} -2\dfrac{1}{\sigma^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial d_i}{\partial \boldsymbol{\beta}_2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ -2\dfrac{1}{\sigma^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial d_i}{\partial \sigma^2} = \begin{cases} -\dfrac{1}{\sigma^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2, & para \quad i = 1, \ldots, k, \\ -\dfrac{1}{\sigma^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2, & para \quad i = k+1, \ldots, n, \end{cases}$$

Assim,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{cases} \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{k} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)\mathbf{x}_i^\top \right\}, \\ \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_2} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=k+1}^{n} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)\mathbf{x}_i^\top \right\}, \\ \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4} \left[\displaystyle\sum_{i=1}^{k} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2 \right\} + \sum_{i=k+1}^{n} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2 \right\} \right]. \end{cases}$$

Da mesma forma, calculando as segundas derivadas parciais de $\ell(\boldsymbol{\theta})$ temos

$$\frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} = 0, \qquad \frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 \log \sigma^2}{\partial \sigma^2 \partial \sigma^2} = -\frac{1}{\sigma^4},$$

$$\frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\beta}_1 \partial \sigma^2} = 0, \qquad \frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\beta}_2 \partial \sigma^2} = 0, \qquad \frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} = 0,$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} = \begin{cases} 2\dfrac{1}{\sigma^2}\mathbf{x}_i\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma^2}\mathbf{x}_i\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial\sigma^2\partial\sigma^2} = \begin{cases} 2\dfrac{1}{\sigma^6}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2, & para \quad i = 1,\ldots,k, \\ 2\dfrac{1}{\sigma^6}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2, & para \quad i = k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial\boldsymbol{\beta}_1\partial\sigma^2} = \begin{cases} 2\dfrac{1}{\sigma^4}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)\mathbf{x}_i^\top, & para \quad i = 1,\ldots,k, \\ 0, & para \quad i = k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial\boldsymbol{\beta}_2\partial\sigma^2} = \begin{cases} 0, & para \quad i = 1,\ldots,k, \\ 2\dfrac{1}{\sigma^4}(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)\mathbf{x}_i^\top, & para \quad i = k+1,\ldots,n. \end{cases}$$

Observe-se que as primeiras derivadas parciais correpondem às funções escore para os parâmetros. Este conjunto de equações corresponde a um sistema não linear de equações, que adicionalmente inclui os termos $\mathbb{I}_i(1/2)$ e $\mathbb{I}_i(3/2)$ que por (3.3) são expressões matemáticas complexas que incluem uma integral e que sempre dependem dos parâmetros $\boldsymbol{\theta}$ presentes na expressão de d_i. Uma solução via métodos numéricos é possível.

3.5.3 O Algoritmo EM

Uma característica deste algoritmo é a formulação a partir de uma representação hierárquica que permite uma flexibilidade no tratamento das derivadas teóricas. Assim, considerando o modelo de regressão linear definido em (3.10)-(3.11) e da Proposição 2.4.1 temos que

$$Y_i \mid U = u_i \sim \begin{cases} N(\mathbf{x}_i^\top\boldsymbol{\beta}_1, \kappa(u_i)\sigma^2) & para \quad i = 1,\ldots,k, \\ N(\mathbf{x}_i^\top\boldsymbol{\beta}_2, \kappa(u_i)\sigma^2) & para \quad i = k+1,\ldots,n \end{cases}$$

$e \quad U_i \sim H(u_i;\boldsymbol{\nu})$, onde Y_i e U_i são independentes para todo $i = 1,\ldots,n$.

No caso da hipótese nula em (2.2), consideram-se os vetores $\mathbf{y} = (y_1,..,y_n)^\top$, $\mathbf{u} = (u_1,..,u_n)^\top$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top,\sigma^2)^\top$ que denota o vetor de parâmetros do modelo.

Para todo $i = 1,\ldots,n$, temos que

$$f(y_i,u_i|\boldsymbol{\theta}) = f(y_i|u_i,\boldsymbol{\theta})f(u_i) = \frac{1}{(2\pi\,\kappa(u_i)\,\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2}\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma^2}\right\}h(u_i;\boldsymbol{\nu}).$$

Assim, considerando que $y_1,..,y_n$ são independentes, a função de log-verossimilhança completa associada a $\mathbf{y_c} = (\mathbf{y}^\top,\mathbf{u}^\top)^\top$ é da forma:

$$\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma^2}\right\} + C$$

onde $C = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\kappa(u_i) + \sum_{i=1}^{n}h(u_i;\boldsymbol{\nu})$, que não depende de $\boldsymbol{\theta}$.

Para a hipótese alternativa apresentada em (2.2), considerando os vetores $\mathbf{y} = (y_1,..,y_n)^\top$, $\mathbf{u} = (u_1,..,u_n)^\top$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top,\boldsymbol{\beta}_2^\top,\sigma^2)^\top$ que denota o vetor de parâmetros do modelo, temos

$$f(y_i, u_i|\boldsymbol{\theta}) = f(y_i|u_i, \boldsymbol{\theta}) f(u_i) = \begin{cases} \dfrac{1}{(2\pi\, \kappa(u_i)\, \sigma^2)^{1/2}} \exp\left\{ -\dfrac{1}{2} \dfrac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\kappa(u_i)\, \sigma^2} \right\} h(u_i; \boldsymbol{\nu}), \\ \qquad\qquad\qquad\qquad \textit{para} \quad i = 1, \ldots, k, \quad e \\ \\ \dfrac{1}{(2\pi\, \kappa(u_i)\, \sigma^2)^{1/2}} \exp\left\{ -\dfrac{1}{2} \dfrac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\kappa(u_i)\, \sigma^2} \right\} h(u_i; \boldsymbol{\nu}), \\ \qquad\qquad\qquad\qquad \textit{para} \quad i = k+1, \ldots, n. \end{cases}$$

Também neste caso $y_1, ..., y_n$ são independentes e a função de log-verossimilhança completa associada a $\mathbf{y_c} = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ é da forma:

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u}) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \left[\sum_{i=1}^{k} \left\{ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\kappa(u_i)\, \sigma^2} \right\} + \sum_{i=k+1}^{n} \left\{ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2}{\kappa(u_i)\, \sigma^2} \right\} \right] + C$$

onde $C = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \kappa(u_i) + \sum_{i=1}^{n} h(u_i; \boldsymbol{\nu})$.

O cálculo da esperança de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ definida em (2.6) é dada por

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \left\{ q(d_i) \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2} \right\} + E[C|\mathbf{y}, \mathbf{u}],$$

para o caso da hipótese nula, e

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{k} \left\{ q(d_i) \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\sigma^2} \right\}$$

$$-\frac{1}{2} \sum_{i=k+1}^{n} \left\{ q(d_i) \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2}{\sigma^2} \right\} + E[C|\mathbf{y}, \mathbf{u}],$$

para o caso da hipótese alternativa, onde $q(d_i) = E[\kappa^{-1}(U_i)|y_i, \widehat{\boldsymbol{\theta}}]$.

Assim, para obter a esperança condicional $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ é preciso calcular $q(d_i)$ para todos os valores $i = 1, \ldots, n$. Para as quatro distribuições da classe de Mistura de Escala Normal estudadas nesta dissertação, estes resultados estão apresentados em (2.10)-(2.13).

Finalmente, a r-ésima iteração do algoritmo EM para a estimação de parâmetros deste modelo é descrita como segue abaixo

Passo E: Calcular as esperanças condicionais $q(d_i)^{(r)} = E[\kappa(U_i)|Y_i, \widehat{\boldsymbol{\theta}}^{(r-1)}]$ cujas formas são conhecidas e apresentadas na Tabela 3.1

Sob a hipótese nula, temos que $d_i^{(r)} = \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}^{(r-1)})^2}{\sigma^{2(r-1)}}, \quad \textit{para} \quad i = 1, \ldots, n.$

Sob a hipótese alternativa, vamos ter que

Tabela 3.1: Esperanças condicionais $q(d_i)$ das distribuições estudadas.

Distribuição	$q(d_i)$
Normal	1
t de Student	$\dfrac{\nu+1}{\nu+d_i}$
Slash	$\dfrac{(1+2\nu)\ P_1(3/2+\nu\,,\ d_i/2)}{d_i\ P_1(1/2+\nu\,,\ d_i/2)}$
Normal Contaminada	$\dfrac{1-\nu+\nu\gamma^{3/2}\exp\{(1-\gamma)d_i/2\}}{1-\nu+\nu\gamma^{1/2}\exp\{(1-\gamma)d_i/2\}}$

$P_x(a,b)$ é f.d.a. da $Gamma(a,b)$ com média $\frac{a}{b}$ é avaliada no valor x.

$$d_i^{(r)} = \begin{cases} \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}_1^{(r-1)})^2}{\sigma^{2(r-1)}}, & para \quad i = 1, \ldots, k, \\ \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}_2^{(r-1)})^2}{\sigma^{2(r-1)}}, & para \quad i = k+1, \ldots, n. \end{cases}$$

Sob quaisquer das hipóteses, $\widehat{\boldsymbol{\theta}}^{(0)}$ é a estimação inicial dos parâmetros obtida a partir do método de mínimos quadrados.

Passo M: As estimativas de máxima verossimilhança sob a hipótese nula são obtidas por

$$\widehat{\boldsymbol{\beta}}^{(r)} = (\mathbf{X}^\top \mathbf{Q}^{(r)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Q}^{(r)} \mathbf{Y},$$
$$\widehat{\sigma^2}^{(r)} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^{(r)})^\top \mathbf{Q}^{(r)} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^{(r)}),$$

onde $\mathbf{Q}^{(r)} = diag(q(d_1)^{(r)}, ..., q(d_n)^{(r)})$.

As estimativas de máxima verossimilhança sob a hipótese alternativa são obtidas por

$$\widehat{\boldsymbol{\beta}_1}^{(r)} = (\mathbf{X_1}^\top \mathbf{Q_1}^{(r)} \mathbf{X_1})^{-1} \mathbf{X_1}^\top \mathbf{Q_1}^{(r)} \mathbf{Y_1},$$
$$\widehat{\boldsymbol{\beta}_2}^{(r)} = (\mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{X_2})^{-1} \mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{Y_2},$$
$$\widehat{\sigma^2}^{(r)} = \frac{1}{n}[(\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1^{(r)})^\top \mathbf{Q_1}^{(r)} (\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1^{(r)}) + (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2^{(r)})^\top \mathbf{Q_2}^{(r)} (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2^{(r)})],$$

onde $\mathbf{Q_1^{(r)}} = diag(q(d_1)^{(r)}, ..., q(d_k)^{(r)})$, $\mathbf{Q_2^{(r)}} = diag(q(d_{k+1})^{(r)}, ..., q(d_n)^{(r)})$, $\mathbf{X}_1$ é uma partição da matriz $\mathbf{X}$ que considera os k primeiras linhas, $\mathbf{Y}_1$ é uma partição do vetor $\mathbf{Y}$ que considera os k primeiras observações, $\mathbf{X}_2$ é uma partição da matriz $\mathbf{X}$ que considera os $n-k$ últimas linhas e $\mathbf{Y}_2$ é uma partição do vetor $\mathbf{Y}$ que considera os $n-k$ últimas observações.

No fim de cada iteração, logo do passo M é realizado um processo de otimização que consiste em maximizar a expressão $\sum_{i=1}^n \log \mathbb{K}_i$ que tem o objetivo de encontrar o melhor conjunto de valores para $\boldsymbol{\nu}$ que é um escalar ou um vetor de parâmetros da distribuição de U.

3.5.4 Estudos de Simulação

Nesta subseção, apresenta-se alguns resultados de diversas simulações realizadas para avaliar o desempenho do algoritmo EM e a metodologia proposta. Serão avaliadas as estimativas dos parâmetros, a aplicação do critério para determinação do ponto de mudança e a otimização dos valores de $\boldsymbol{\nu}$. Todas as amostras simuladas são geradas desde o modelo de regressão com ponto de mudança nos coeficientes de regressão

$$\begin{aligned} y_i &= 0.8 + 1.2x_i + \epsilon_i, \quad para \quad i = 1, \ldots, k, \\ y_i &= 2.0 + 1.5x_i + \epsilon_i, \quad para \quad i = k+1, \ldots, n, \end{aligned}$$

considerando $\sigma^2 = 1$ e $x_i \sim U(0, 20)$ para $i = 1, \ldots, n$.

1) **Estimativa dos parâmetros**

Foram configuradas várias combinações de valores de n e k, e para cada uma delas, simulações Monte Carlo de 100 e 1000 amostras foram geradas. Os valores dos parâmetros estimados nas amostras geradas foram avaliados. Os erros aleatórios simulados seguem as seguintes distribuições: Normal, t de Student ($\nu = 2$), Slash ($\nu = 4$) e Normal Contaminada ($\nu = 0, 2, \gamma = 0.3$). Os resultados das simulações Monte Carlo encontram-se nas Tabelas 3.2 - 3.5.

Podemos observar nestas tabelas que os parâmetros são bem estimados em todas as combinações avaliadas e para todas as estruturas de erros aleatórios consideradas. Assim, o desempenho do algoritmo EM na estimação dos parâmetros é eficiente.

Nas amostras de tamanho pequeno, quando $n = 40$, os resultados das estimativas são tão eficientes quanto àquelas obtidas de amostras de tamanhos maiores. Esta observação torna-se importante, pois muitas das aplicações são realizadas sob amostras de tamanho pequeno, por exemplo, a aplicação de Holbert (1982) será realizada com um conjunto de dados com 35 observações.

Todas os resultados que consideram 100 amostras Monte Carlo apresentam desvios padrão similares aos obtidos considerando 1000 amostras Monte Carlo. Além disso, os valores das médias das estimativas para todos os parâmetros são próximos dos considerados para a simulação.

2) **Determinação da posição do ponto de mudança**

Foram simuladas 250 amostras Monte Carlo, cada uma delas de tamanho $n = 40$, com o ponto de mudança na posição $k = 20$ e considerando as especificações dadas no inicio desta seção e t de Student ($\nu = 4$), Slash ($\nu = 2$) e Normal Contaminada ($\nu = 0.2, \gamma = 0.3$). Para cada amostra gerada, a posição do ponto de mudança é determinada utilizando os critérios de avaliação do valor SIC descritos na Seção 3.4. Para a aplicação do critério devem ser calculados 38 modelos em cada amostra Monte Carlo, assim a simulação avalia 9500 modelos.

Tabela 3.2: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão Normal, considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ^2
Posição do ponto de mudança k=20							
100	n=40	Média	0.82	1.20	2.11	1.49	0.88
		d.p	0.51	0.05	0.46	0.04	0.20
1000	n=40	Média	0.80	1.20	2.00	1.50	0.89
		d.p	0.47	0.04	0.46	0.04	0.21
Posição do ponto de mudança k=200							
100	n=400	Média	0.78	1.20	1.97	1.50	1.00
		d.p	0.14	0.01	0.15	0.01	0.07
1000	n=400	Média	0.80	1.20	2.00	1.50	0.99
		d.p	0.14	0.01	0.14	0.01	0.07
Posição do ponto de mudança k=500							
100	n=1000	Média	0.80	1.20	2.00	1.50	0.99
		d.p	0.09	0.01	0.08	0.01	0.05
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.09	0.01	0.09	0.01	0.04
Posição do ponto de mudança k=10							
100	n=40	Média	0.85	1.19	1.98	1.50	0.90
		d.p	0.74	0.06	0.32	0.03	0.22
1000	n=40	Média	0.84	1.20	2.00	1.50	0.91
		d.p	0.73	0.06	0.38	0.03	0.21
Posição do ponto de mudança k=100							
100	n=400	Média	0.80	1.20	1.98	1.50	0.98
		d.p	0.19	0.02	0.11	0.01	0.07
1000	n=400	Média	0.80	1.20	2.00	1.50	0.99
		d.p	0.21	0.02	0.11	0.01	0.07
Posição do ponto de mudança k=250							
100	n=1000	Média	0.79	1.20	2.01	1.50	0.99
		d.p	0.13	0.01	0.07	0.01	0.04
1000	n=1000	Média	0.79	1.20	2.00	1.50	1.00
		d.p	0.12	0.01	0.07	0.01	0.04
Posição do ponto de mudança k=30							
100	n=40	Média	0.78	1.20	1.92	1.50	0.89
		d.p	0.38	0.03	0.64	0.06	0.23
1000	n=40	Média	0.81	1.20	1.98	1.50	0.90
		d.p	0.39	0.03	0.69	0.06	0.21
Posição do ponto de mudança k=300							
100	n=400	Média	0.79	1.20	2.03	1.50	0.98
		d.p	0.12	0.01	0.18	0.02	0.07
1000	n=400	Média	0.79	1.20	2.00	1.50	0.99
		d.p	0.11	0.01	0.21	0.02	0.07
Posição do ponto de mudança k=750							
100	n=1000	Média	0.80	1.20	1.99	1.50	0.99
		d.p	0.07	0.01	0.12	0.01	0.04
1000	n=1000	Média	0.80	1.20	2.01	1.50	1.00
		d.p	0.07	0.01	0.13	0.01	0.05

Tabela 3.3: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão t de Student ($\nu = 2$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ^2
Posição do ponto de mudança k=20							
100	n=40	Média	0.74	1.21	1.99	1.50	0.86
		d.p	0.72	0.05	0.61	0.05	0.31
1000	n=40	Média	0.81	1.20	1.98	1.50	0.89
		d.p	0.69	0.06	0.63	0.06	0.33
Posição do ponto de mudança k=200							
100	n=400	Média	0.78	1.20	2.02	1.50	0.98
		d.p	0.20	0.02	0.16	0.01	0.10
1000	n=400	Média	0.80	1.20	2.00	1.50	0.99
		d.p	0.18	0.02	0.19	0.02	0.11
Posição do ponto de mudança k=500							
100	n=1000	Média	0.80	1.20	1.99	1.50	0.99
		d.p	0.12	0.01	0.12	0.01	0.06
1000	n=1000	Média	0.80	1.20	1.99	1.50	1.00
		d.p	0.11	0.01	0.11	0.01	0.07
Posição do ponto de mudança k=10							
100	n=40	Média	0.78	1.20	2.02	1.50	0.90
		d.p	0.92	0.08	0.51	0.05	0.33
1000	n=40	Média	0.79	1.20	1.99	1.50	0.90
		d.p	1.09	0.10	0.50	0.04	0.33
Posição do ponto de mudança k=100							
100	n=400	Média	0.76	1.20	2.01	1.50	1.00
		d.p	0.27	0.02	0.15	0.01	0.12
1000	n=400	Média	0.80	1.20	2.00	1.50	0.99
		d.p	0.26	0.02	0.15	0.01	0.11
Posição do ponto de mudança k=250							
100	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.15	0.01	0.09	0.01	0.06
1000	n=1000	Média	0.79	1.20	2.00	1.50	1.00
		d.p	0.16	0.01	0.09	0.01	0.07
Posição do ponto de mudança k=30							
100	n=40	Média	0.73	1.21	1.87	1.51	0.85
		d.p	0.52	0.04	0.93	0.09	0.29
1000	n=40	Média	0.79	1.20	2.06	1.50	0.91
		d.p	0.51	0.04	1.07	0.10	0.37
Posição do ponto de mudança k=300							
100	n=400	Média	0.81	1.20	1.97	1.50	0.98
		d.p	0.16	0.01	0.30	0.03	0.11
1000	n=400	Média	0.81	1.20	1.99	1.50	0.99
		d.p	0.15	0.01	0.27	0.02	0.11
Posição do ponto de mudança k=750							
100	n=1000	Média	0.78	1.20	1.97	1.50	0.99
		d.p	0.09	0.01	0.17	0.02	0.06
1000	n=1000	Média	0.79	1.20	2.00	1.50	0.99
		d.p	0.09	0.01	0.17	0.01	0.07

Tabela 3.4: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão Slash ($\nu = 4$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ^2
Posição do ponto de mudança k=20							
100	n=40	Média	0.78	1.20	2.06	1.49	0.89
		d.p	0.58	0.05	0.56	0.05	0.21
1000	n=40	Média	0.79	1.20	2.01	1.50	0.90
		d.p	0.54	0.05	0.54	0.05	0.23
Posição do ponto de mudança k=200							
100	n=400	Média	0.83	1.20	2.02	1.50	0.99
		d.p	0.16	0.01	0.17	0.01	0.08
1000	n=400	Média	0.79	1.20	2.00	1.50	0.99
		d.p	0.17	0.01	0.16	0.01	0.08
Posição do ponto de mudança k=500							
100	n=1000	Média	0.79	1.20	2.01	1.50	1.00
		d.p	0.11	0.01	0.12	0.01	0.05
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.10	0.01	0.11	0.01	0.05
Posição do ponto de mudança k=10							
100	n=40	Média	0.80	1.20	1.98	1.50	0.88
		d.p	0.87	0.08	0.46	0.04	0.22
1000	n=40	Média	0.83	1.20	2.00	1.50	0.89
		d.p	0.78	0.07	0.42	0.04	0.23
Posição do ponto de mudança k=100							
100	n=400	Média	0.78	1.20	1.99	1.50	0.98
		d.p	0.26	0.02	0.11	0.01	0.07
1000	n=400	Média	0.79	1.20	2.00	1.50	0.99
		d.p	0.23	0.02	0.14	0.01	0.07
Posição do ponto de mudança k=250							
100	n=1000	Média	0.77	1.20	1.99	1.50	1.00
		d.p	0.15	0.01	0.08	0.01	0.04
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.14	0.01	0.09	0.01	0.05
Posição do ponto de mudança k=30							
100	n=40	Média	0.80	1.20	1.94	1.50	0.92
		d.p	0.45	0.04	0.93	0.08	0.26
1000	n=40	Média	0.80	1.20	1.97	1.50	0.91
		d.p	0.44	0.04	0.80	0.07	0.23
Posição do ponto de mudança k=300							
100	n=400	Média	0.80	1.20	2.04	1.50	1.00
		d.p	0.15	0.01	0.24	0.02	0.07
1000	n=400	Média	0.80	1.20	1.99	1.50	0.99
		d.p	0.13	0.01	0.23	0.02	0.07
Posição do ponto de mudança k=750							
100	n=1000	Média	0.80	1.20	1.98	1.50	1.00
		d.p	0.08	0.01	0.14	0.01	0.05
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.09	0.01	0.15	0.01	0.05

Tabela 3.5: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão Normal Contaminada ($\nu = 0.2, \gamma = 0.3$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ^2
Posição do ponto de mudança k=20							
100	n=40	Média	0.64	1.22	2.08	1.49	0.90
		d.p	0.55	0.05	0.51	0.05	0.27
1000	n=40	Média	0.80	1.20	2.03	1.50	0.91
		d.p	0.54	0.05	0.57	0.05	0.26
Posição do ponto de mudança k=200							
100	n=400	Média	0.83	1.20	1.98	1.50	0.99
		d.p	0.15	0.01	0.17	0.01	0.08
1000	n=400	Média	0.81	1.20	2.00	1.50	0.99
		d.p	0.16	0.01	0.17	0.01	0.08
Posição do ponto de mudança k=500							
100	n=1000	Média	0.80	1.20	2.02	1.50	1.00
		d.p	0.10	0.01	0.11	0.01	0.05
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.11	0.01	0.11	0.01	0.05
Posição do ponto de mudança k=10							
100	n=40	Média	0.70	1.20	2.07	1.50	0.87
		d.p	0.77	0.07	0.45	0.04	0.25
1000	n=40	Média	0.80	1.20	2.02	1.50	0.91
		d.p	0.83	0.07	0.45	0.04	0.25
Posição do ponto de mudança k=100							
100	n=400	Média	0.84	1.20	2.01	1.50	1.00
		d.p	0.26	0.02	0.13	0.01	0.08
1000	n=400	Média	0.81	1.20	1.99	1.50	0.99
		d.p	0.24	0.02	0.13	0.01	0.08
Posição do ponto de mudança k=250							
100	n=1000	Média	0.83	1.20	2.00	1.50	0.99
		d.p	0.16	0.01	0.09	0.01	0.05
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00
		d.p	0.15	0.01	0.08	0.01	0.05
Posição do ponto de mudança k=30							
100	n=40	Média	0.82	1.20	2.10	1.49	0.90
		d.p	0.42	0.04	0.91	0.07	0.26
1000	n=40	Média	0.79	1.20	1.96	1.50	0.91
		d.p	0.44	0.04	0.84	0.07	0.24
Posição do ponto de mudança k=300							
100	n=400	Média	0.81	1.20	1.99	1.50	0.99
		d.p	0.13	0.01	0.25	0.02	0.09
1000	n=400	Média	0.80	1.20	2.00	1.50	0.99
		d.p	0.13	0.01	0.22	0.02	0.08
Posição do ponto de mudança k=750							
100	n=1000	Média	0.80	1.20	2.01	1.50	1.00
		d.p	0.08	0.01	0.14	0.01	0.05
1000	n=1000	Média	0.80	1.20	2.01	1.50	0.99
		d.p	0.08	0.01	0.15	0.01	0.05

Da mesma forma que em Hofrichter (2007), o desempenho da metodologia de determinação do ponto de mudança que utiliza o critério SIC é realizado avaliando a distribuição de frequência das posições de ponto de mudança obtidas nas amostras Monte Carlo. A Figura 3.1 mostra que para todas as situações a frequência maior de ocorrências se encontra na posição 20. Além disso, valores próximos da posição 20 apresentam também frequências de ocorrência altas. Isto leva a concluir que a metodologia proposta consegue determinar adequadamente o ponto de mudança.

Finalmente, selecionou-se de cada conjunto de amostras Monte Carlo, uma amostra onde o ponto de mudança estimado se encontra na posição 20, e todos os valores SIC obtidos para esta amostra são avaliados. A Figura 3.2 mostra que para todos os casos o menor valor do SIC corresponde à posição 20. Desta forma, se garante a coerência do critério utilizado.

3) **Otimização dos valores de $\boldsymbol{\nu}$**

Considerando as mesmas simulações Monte Carlo do item anterior, os resultados otimizados de $\boldsymbol{\nu}$ na posição estimada do ponto de mudança foram avaliados. Neste caso, foram desconsiderados valores atípicos. A Figura 3.3 mostra que em média os valores otimizados de $\boldsymbol{\nu}$ estão próximos dos valores originais de simulação. Isto garante o bom desempenho do algoritmo na otimização dos valores avaliados. Os gráficos box-plot mostram que os valores otimizados se encontram na sua grande maioria perto do valor usado para a simulação.

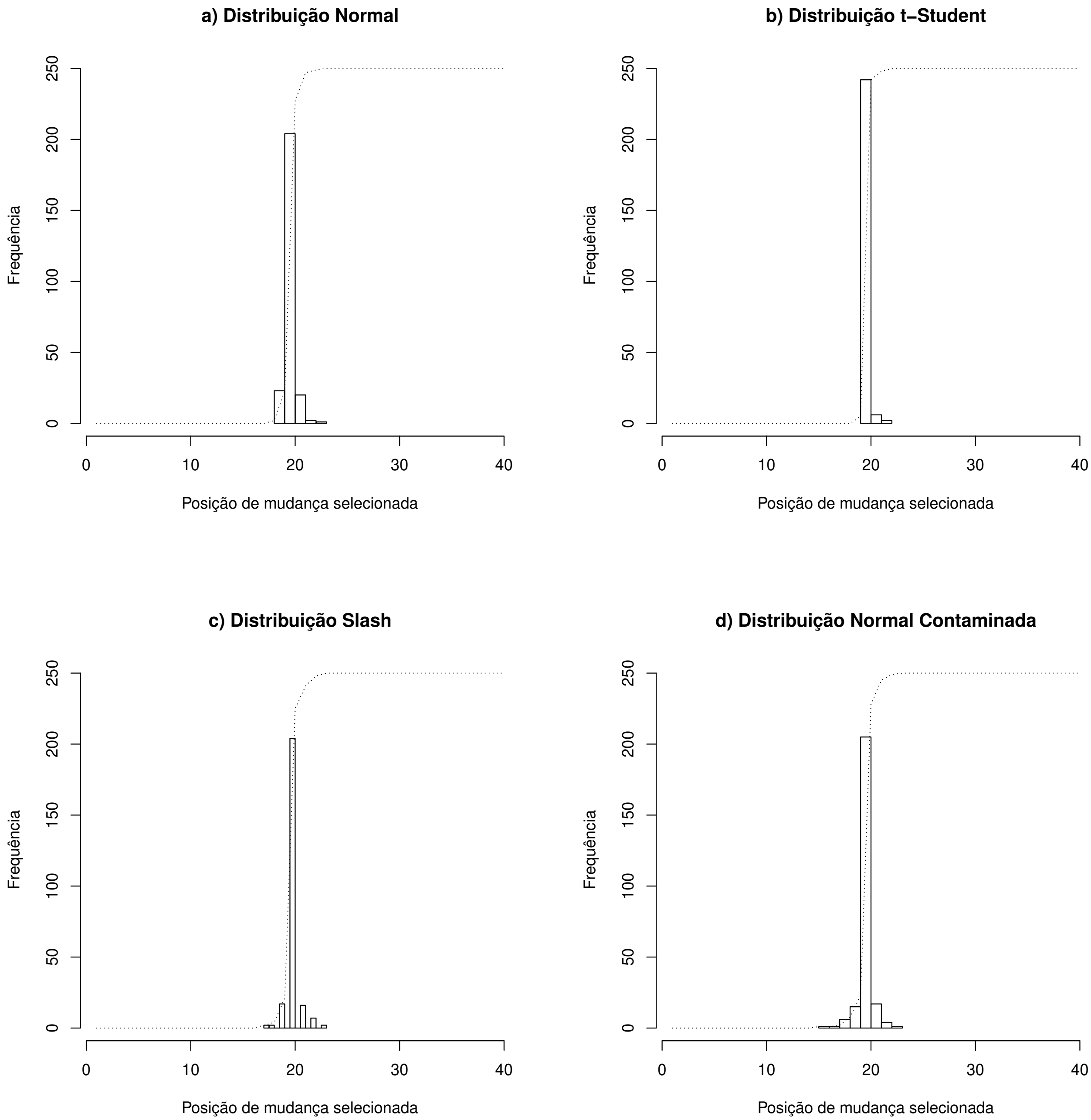


Figura 3.1: Frequências absolutas das posições dos pontos de mudança estimados nas amostras Monte Carlo simuladas no modelo de regressão linear com mudança nos coeficientes de regressão.

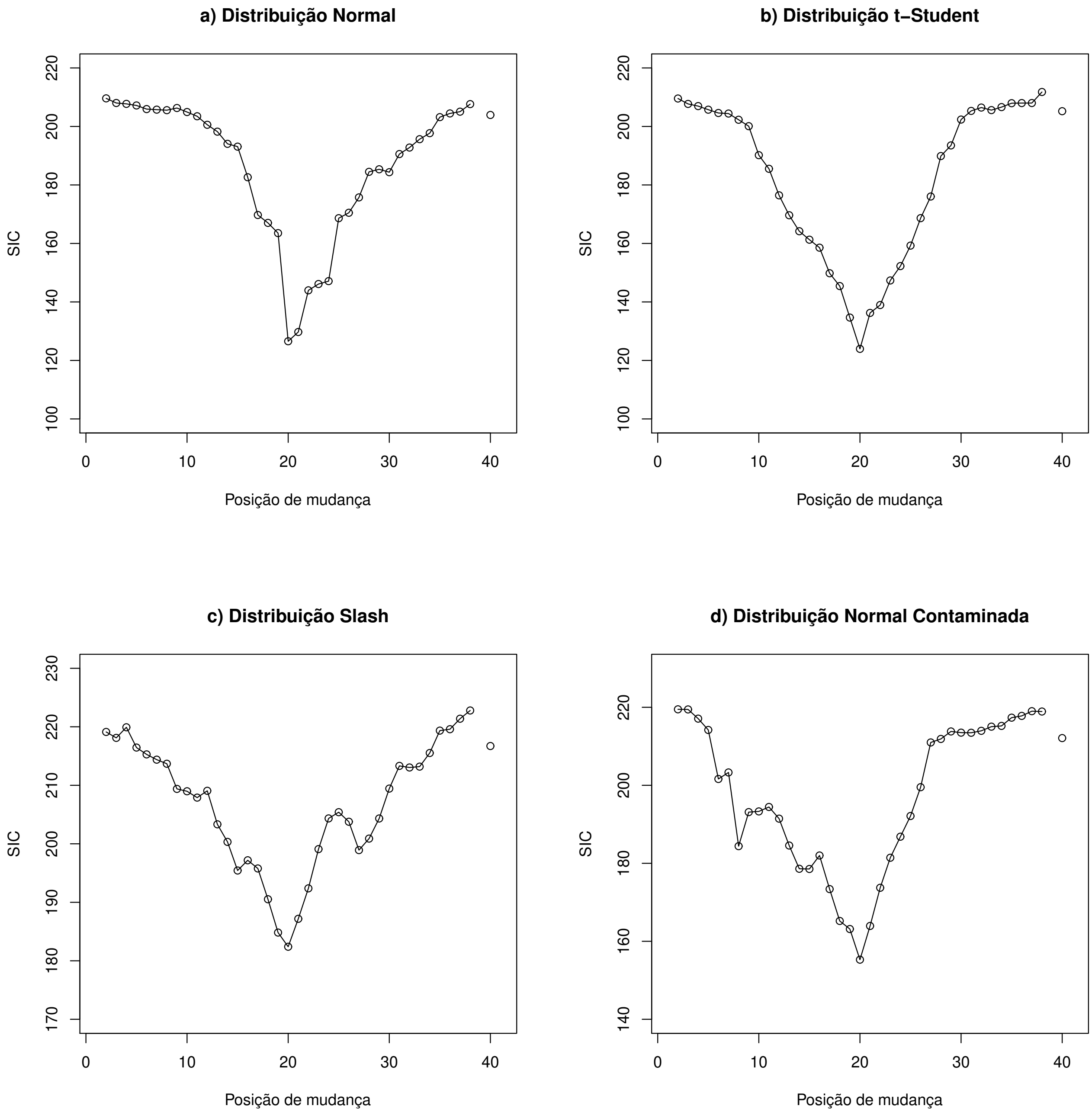


Figura 3.2: Aplicação do critério SIC sobre uma amostra Monte Carlo simulada, onde o ponto de mudança estimado é $k = 20$, no modelo de regressão linear com mudança nos coeficientes de regressão.

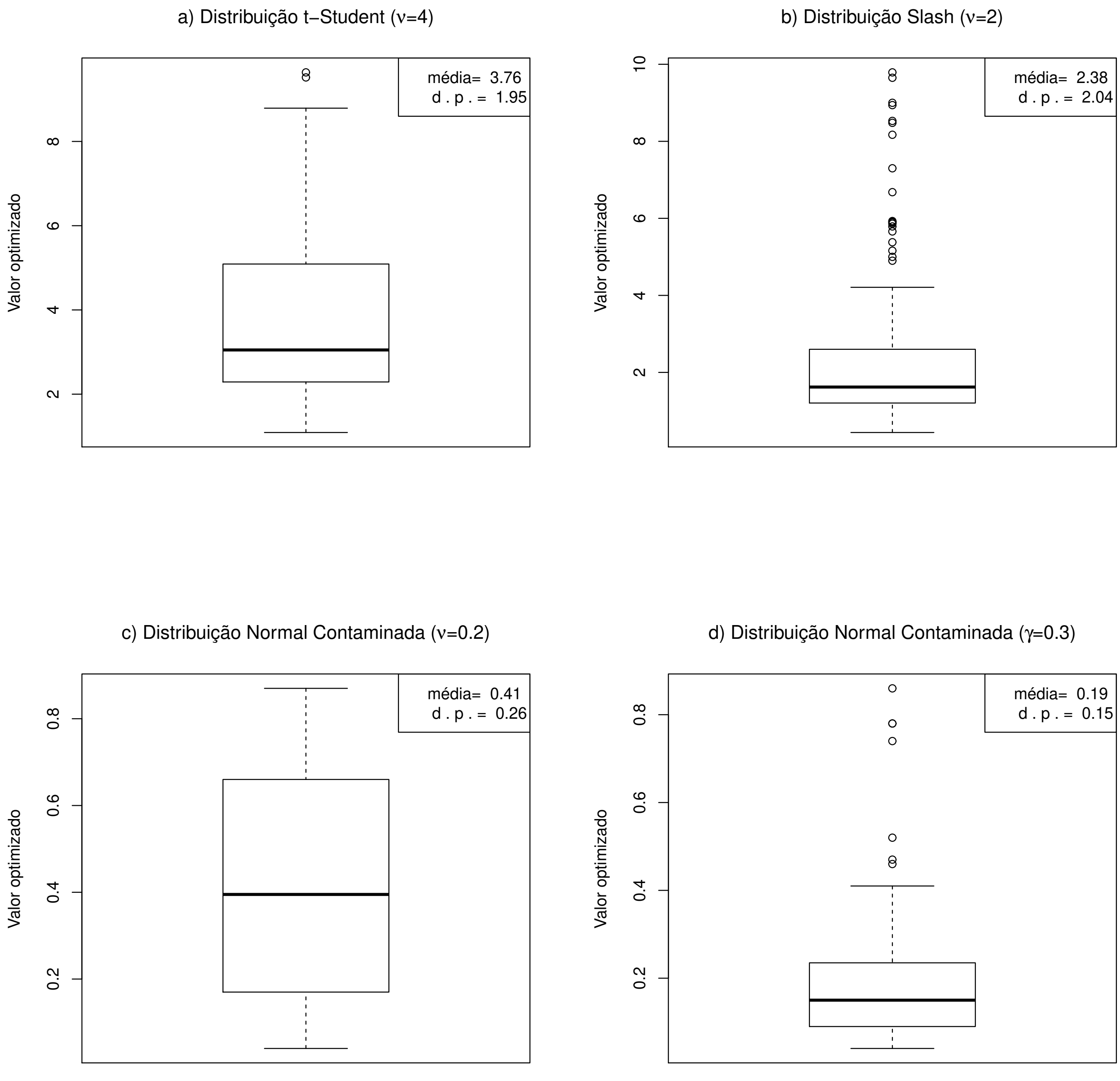


Figura 3.3: Resultado de simulações sobre otimização do parâmetro $\boldsymbol{\nu}$ das distribuições t de Student, Slash e Normal Contaminada no modelo de regressão linear com mudança nos coeficientes de regressão.

3.6 Mudança na Variância

3.6.1 Especificação do Modelo

O modelo geral (3.1) pode ser escrito de forma matricial como:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{3.14}$$

onde $\boldsymbol{\beta}=(\beta_0, \beta_1, ..., \beta_{p-1})^\top$ e $\mathbf{x}_i=(1, x_{1i}, ..., x_{1(p-1)})^\top$. Ao considerar a existência de um ponto de mudança a partir da posição k temos que os erros ϵ_1,...,ϵ_n são independentes e distribuídos como:

$$\epsilon_i \sim \begin{cases} MEN(0, \sigma_1^2, H), & para \quad i = 1, \ldots, k, \\ MEN(0, \sigma_2^2, H), & para \quad i = k+1, \ldots, n, \end{cases}$$

e

$$\begin{array}{llll} \mathbf{Y}_i \sim MEN(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma_1^2; H), & para & i = 1, \ldots, k & e \\ \mathbf{Y}_i \sim MEN(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma_2^2; H), & para & i = k+1, \ldots, n. & \end{array} \tag{3.15}$$

Para esta estrutura de mudança o teste de hipótese a considerar é:

$$\begin{array}{ll} H_0 : & Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \quad e \quad \sigma^2, \quad i = 1, \ldots, n, \\ H_1 : & \begin{cases} Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \quad e \quad \sigma_1^2, \quad i = 1, \ldots, k, \\ Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \quad e \quad \sigma_2^2, \quad i = k+1, \ldots, n. \end{cases} \end{array} \tag{3.16}$$

Usando apropriadamente a expressão (3.5), as funções de log-verossimilhança para este modelo nas situações de hipótese nula e hipótese alternativa apresentadas em (3.16) são, respetivamente:

$$\ell_0(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + \sum_{i=1}^{n} \log \mathbb{K}_i, \tag{3.17}$$

e

$$\ell_k(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2) = -\frac{n}{2} \log 2\pi - \frac{k}{2} \log \sigma_1^2 - \frac{(n-k)}{2} \log \sigma_2^2 + \sum_{i=1}^{k} \log \mathbb{K}_i^{(1)} + \sum_{i=k+1}^{n} \log \mathbb{K}_i^{(2)}, \tag{3.18}$$

onde $$d_i = \begin{cases} \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2}, & em \quad \mathbb{K}_i, \\ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma_1^2}, & em \quad \mathbb{K}_i^{(1)}, \\ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma_2^2}, & em \quad \mathbb{K}_i^{(2)}. \end{cases}$$

As expressões (3.17) e (3.18) tornam-se importantes, pois serão utilizadas na aplicação do critério para determinação de ponto de mudança. Estas deverão ser avaliadas nas estimativas de máxima verossimilhança dos parâmetros. Assim, será possível encontrar os valores $\ell_0(\widehat{\boldsymbol{\theta}}) = \ell_0(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})$ e $\ell_k(\widehat{\boldsymbol{\theta}}) = \ell_k(\widehat{\boldsymbol{\beta}}, \widehat{\sigma_1^2}, \widehat{\sigma_2^2})$ que compõem o Critério de Informação Schwarz (SIC) utilizado como parte da metodologia proposta.

3.6.2 Derivadas Parciais de $\ell(\boldsymbol{\theta})$

Considerando os resultados apresentados para a expressão $\mathbb{K}_i$ e usando o resultado (3.6), o cálculos necessários para a obtenção das derivadas parciais de primeira ordem de $\ell(\boldsymbol{\theta})$ são:

$$\frac{\partial \log \sigma_1^2}{\partial \boldsymbol{\beta}} = 0, \qquad \frac{\partial \log \sigma_1^2}{\partial \sigma_2^2} = 0, \qquad \frac{\partial \log \sigma_2^2}{\partial \boldsymbol{\beta}} = 0, \qquad \frac{\partial \log \sigma_2^2}{\partial \sigma_1^2} = 0,$$

$$\frac{\partial \log \sigma_1^2}{\partial \sigma_1^2} = \begin{cases} \dfrac{1}{\sigma_1^2}, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial \log \sigma_2^2}{\partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ \dfrac{1}{\sigma_2^2}, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial d_i}{\partial \boldsymbol{\beta}} = \begin{cases} -2\dfrac{1}{\sigma_1^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ -2\dfrac{1}{\sigma_2^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial d_i}{\partial \sigma_1^2} = \begin{cases} -\dfrac{1}{\sigma_1^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial d_i}{\partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ -\dfrac{1}{\sigma_2^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2, & para \quad i = k+1, \ldots, n, \end{cases}$$

Assim,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{cases} \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \dfrac{1}{\sigma_1^2}\displaystyle\sum_{i=1}^{k}\left\{\frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\mathbf{x}_i^\top\right\} + \dfrac{1}{\sigma_2^2}\displaystyle\sum_{i=1}^{k}\left\{\frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\mathbf{x}_i^\top\right\}, \\ \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_1^2} = -\dfrac{k}{2\sigma_1^2} + \dfrac{1}{2\sigma_1^4}\displaystyle\sum_{i=1}^{k}\left\{\frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2\right\}, \\ \dfrac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_2^2} = -\dfrac{(n-k)}{2\sigma_2^2} + \dfrac{1}{2\sigma_2^4}\displaystyle\sum_{i=k+1}^{n}\left\{\frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2\right\}. \end{cases}$$

Da mesma forma, calculando as segundas derivadas parciais de $\ell(\boldsymbol{\theta})$ temos:

$$\frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \sigma_2^2 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}\partial \sigma_1^2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}\partial \sigma_2^2} = 0,$$

$$\frac{\partial^2 \log \sigma_1^2}{\partial \sigma_1^2 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \sigma_1^2 \partial \sigma_1^2} = 0,$$

$$\frac{\partial^2 \log \sigma_1^2}{\partial \sigma_1^2 \partial \sigma_1^2} = \begin{cases} -\dfrac{1}{\sigma_1^4}, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 \log \sigma_2^2}{\partial \sigma_2^2 \partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ -\dfrac{1}{\sigma_2^4}, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta} \partial \sigma_1^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta} \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \sigma_1^2 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 d_i}{\partial \sigma_1^2 \partial \sigma_2^2} = 0.$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = \begin{cases} 2\dfrac{1}{\sigma_1^2}\mathbf{x}_i\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma_2^2}\mathbf{x}_i\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \sigma_1^2 \partial \sigma_1^2} = \begin{cases} 2\dfrac{1}{\sigma_1^6}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \sigma_2^2 \partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma_2^6}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \sigma_1^2} = \begin{cases} 2\dfrac{1}{\sigma_1^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma_2^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

Observe que as primeiras derivadas parciais correpondem às funções escore para os parâmetros. Este conjunto de equações corresponde a um sistema não linear de equações, que adicionalmente inclui os termos $\mathbb{I}_i(1/2)$ e $\mathbb{I}_i(3/2)$ que por (3.3) são expressões matemáticas complexas que incluem uma integral e que sempre dependem dos parâmetros $\boldsymbol{\theta}$ presentes em d_i. Uma solução via métodos numéricos é possível.

3.6.3 O Algoritmo EM

Uma característica deste algoritmo é a formulação a partir de uma representação hierárquica que permite uma flexibilidade no tratamento das derivadas teóricas. Assim, considerando o modelo de regressão linear definido em (3.14)-(3.20) e da Proposição 2.4.1 temos que:

$$Y_i \mid U = u_i \ \sim \ \begin{cases} N_1(\mathbf{x}_i^\top \boldsymbol{\beta}, \kappa(u_i)\sigma_1^2), & para \quad i = 1, \ldots, k, \\ N_1(\mathbf{x}_i^\top \boldsymbol{\beta}, \kappa(u_i)\sigma_2^2), & para \quad i = k+1, \ldots, n \end{cases}$$

$e \quad U_i \ \sim \ H(u_i; \boldsymbol{\nu})$, onde Y_i e U_i são independentes para todo $i = 1, \ldots, n$.

No caso da hipótese nula, apresentada em (3.16), consideram-se os vetores $\mathbf{y} = (y_1, .., y_n)^\top$, $\mathbf{u} = (u_1, .., u_n)^\top$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$ que denota o vetor de parâmetros do modelo.

Para todo $i = 1, \ldots, n$, temos que

$$f(y_i, u_i|\boldsymbol{\theta}) = f(y_i|u_i, \boldsymbol{\theta})f(u_i) = \frac{1}{(2\pi\ \kappa(u_i)\ \sigma^2)^{1/2}} \exp\left\{-\frac{1}{2}\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\kappa(u_i)\ \sigma^2}\right\} h(u_i; \boldsymbol{\nu}).$$

Assim, considerando que $y_1, .., y_n$ são independentes, a função de log-verossimilhança completa associada a $\mathbf{y_c} = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ é da forma:

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma^2}\right\} + C$$

onde $C = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\kappa(u_i) + \sum_{i=1}^{n} h(u_i;\boldsymbol{\nu})$.

Para a hipótese alternativa, apresentada em (3.16), considerando os vetores $\mathbf{y} = (y_1, .., y_n)^\top$, $\mathbf{u} = (u_1, .., u_n)^\top$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_1^2, \sigma_2^2)^\top$ que denota o vetor de parâmetros do modelo, temos

$$f(y_i, u_i|\boldsymbol{\theta}) = f(y_i|u_i, \boldsymbol{\theta})f(u_i) = \begin{cases} \dfrac{1}{(2\pi\,\kappa(u_i)\,\sigma_1^2)^{1/2}}\exp\left\{-\dfrac{1}{2}\dfrac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma_1^2}\right\}h(u_i;\boldsymbol{\nu}), \\ \qquad\qquad para \quad i = 1, \ldots, k, \quad e \\ \\ \dfrac{1}{(2\pi\,\kappa(u_i)\,\sigma_2^2)^{1/2}}\exp\left\{-\dfrac{1}{2}\dfrac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma_2^2}\right\}h(u_i;\boldsymbol{\nu}), \\ \qquad\qquad para \quad i = k+1, \ldots, n. \end{cases}$$

Também neste caso $y_1, .., y_n$ são independentes e a função de log-verossimilhança completa associada a $\mathbf{y_c} = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ é da forma:

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u}) = -\frac{k}{2}\log(\sigma_1^2) - \frac{(n-k)}{2}\log(\sigma_2^2) - \frac{1}{2}\left[\sum_{i=1}^{k}\left\{\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma_1^2}\right\} + \sum_{i=k+1}^{n}\left\{\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)\,\sigma_2^2}\right\}\right] + C,$$

onde $C = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\kappa(u_i) + \sum_{i=1}^{n} h(u_i;\boldsymbol{\nu})$.

O cálculo da esperança de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ definida em (2.6) é dada por

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\sigma^2}\right\} + E[C|\mathbf{y}, \mathbf{u}],$$

para o caso da hipótese nula, e

$$\begin{aligned} E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = -\frac{k}{2}\log(\sigma_1^2) - \frac{(n-k)}{2}\log(\sigma_2^2) - \frac{1}{2}\sum_{i=1}^{k}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\sigma_1^2}\right\} \\ -\frac{1}{2}\sum_{i=k+1}^{n}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\sigma_2^2}\right\} + E[C|\mathbf{y}, \mathbf{u}], \end{aligned}$$

para o caso da hipótese alternativa, onde $q(d_i) = E[\kappa^{-1}(U_i)|y_i, \widehat{\boldsymbol{\theta}}]$.

Assim, para obter a esperança condicional $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ é preciso calcular $q(d_i)$ para todos os valores $i = 1, \ldots, n$. Para as quatro distribuições da classe de Mistura de Escala Normal estudadas nesta dissertação, estes resultados estão apresentados em (2.10)-(2.13).

Finalmente, a r-ésima iteração do algoritmo EM para a estimação de parâmetros deste modelo é descrita como segue abaixo.

Passo E: Calcular as esperanças condicionais $q(d_i)^{(r)} = E(\kappa(U_i)|Y_i, \widehat{\boldsymbol{\theta}}^{(r-1)})$ cujas formas conhecidas estão apresentadas na Tabela 3.1.

Sob a hipótese nula, temos que $d_i^{(r)} = \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}^{(r-1)})^2}{\sigma^{2(r-1)}}, \quad para \quad i = 1, \ldots, n.$

Sob a hipótese alternativa, vamos ter que

$$d_i^{(r)} = \begin{cases} \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}^{(r-1)})^2}{\sigma_1^{2(r-1)}}, & para \quad\quad i = 1, \ldots, k, \\ \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}^{(r-1)})^2}{\sigma_2^{2(r-1)}}, & para \quad i = k+1, \ldots, n. \end{cases}$$

Sob quaisquer das hipóteses, $\widehat{\boldsymbol{\theta}}^{(0)}$ é a estimação inicial dos parâmetros obtida a partir do método de mínimos quadrados.

Passo M: As estimativas de máxima verossimilhança sob a hipótese nula são obtidas por

$$\widehat{\boldsymbol{\beta}}^{(r)} = (\mathbf{X}^\top \mathbf{Q}^{(r)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Q}^{(r)} \mathbf{Y},$$

$$\widehat{\sigma^2}^{(r)} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^{(r)})^\top \mathbf{Q}^{(r)} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^{(r)}),$$

onde $\mathbf{Q}^{(r)} = diag(q(d_1)^{(r)}, ..., q(d_n)^{(r)})$.

As estimativas de máxima verossimilhança sob a hipótese alternativa são obtidas por

$$\widehat{\boldsymbol{\beta}}^{(r)} = (\frac{1}{\sigma_1^{2(r-1)}} \mathbf{X_1}^\top \mathbf{Q_1}^{(r)} \mathbf{X_1} + \frac{1}{\sigma_2^{2(r-1)}} \mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{X_2})^{-1} (\frac{1}{\sigma_1^{2(r-1)}} \mathbf{X_1}^\top \mathbf{Q_1}^{(r)} \mathbf{Y_1} + \frac{1}{\sigma_2^{2(r-1)}} \mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{Y_2}),$$

$$\widehat{\sigma_1^2}^{(r)} = \frac{1}{k} [(\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}^{(r)})^\top \mathbf{Q_1}^{(r)} (\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}^{(r)})],$$

$$\widehat{\sigma_2^2}^{(r)} = \frac{1}{(n-k)} [(\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}^{(r)})^\top \mathbf{Q_2}^{(r)} (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}^{(r)})],$$

onde $\mathbf{Q_1^{(r)}} = diag(q(d_1)^{(r)}, ..., q(d_k)^{(r)})$, $\mathbf{Q_2^{(r)}} = diag(q(d_{k+1})^{(r)}, ..., q(d_n)^{(r)})$, $\mathbf{X}_1$ é uma partição da matriz $\mathbf{X}$ que considera os k primeiras linhas, $\mathbf{Y}_1$ é uma partição do vetor $\mathbf{Y}$ que considera os k primeiras observações, $\mathbf{X}_2$ é uma partição da matriz $\mathbf{X}$ que considera os $n-k$ últimas linhas e $\mathbf{Y}_2$ é uma partição do vetor $\mathbf{Y}$ que considera os $n-k$ últimas observações.

No fim de cada iteração, logo do passo M é realizado um processo de otimização que consiste em maximizar a expressão $\sum_{i=1}^{n} \log \mathbb{K}_i$ que tem o objetivo de encontrar o melhor conjunto de valores para $\boldsymbol{\nu}$ que é um escalar ou um vetor de parâmetros da distribuição de U.

3.6.4 Estudos de Simulação

Nesta subseção, apresenta-se alguns resultados de diversas simulações realizadas para avaliar o desempenho do algoritmo EM e a metodologia proposta. Serão avaliadas as estimativas dos parâmetros, a aplicação do critério para determinação do ponto de mudança e a otimização dos valores de $\boldsymbol{\nu}$. Todas as amostras simuladas são geradas desde o modelo de regressão com ponto de mudança na variância

$$y_i = 0.8 + 1.2x_i + \epsilon_i, \quad para \quad i = 1, \ldots, n,$$

considerando $\sigma_1^2 = 1$ para $i = 1, \ldots, k$, $\sigma_2^2 = 9$ para $i = k+1, \ldots, n$ e $x_i \sim U(0, 20)$ para $i = 1, \ldots, n$.

1) **Estimativa dos parâmetros**

Foram configuradas várias combinações de valores de n e k, e para cada uma delas, simulações Monte Carlo de 100 e 1000 amostras foram geradas. Os valores dos parâmetros estimados nas amostras geradas foram avaliados. Os erros aleatórios simulados seguem as seguintes distribuições: Normal, t de Student ($\nu = 2$), Slash ($\nu = 4$) e Normal Contaminada ($\nu = 0.2, \gamma = 0.3$). Os resultados das simulações Monte Carlo encontram-se nas Tabelas 3.6 - 3.9.

Podemos observar nestas tabelas que os parâmetros são bem estimados, em todas as combinações avaliadas e para todas as estruturas de erros aleatórios consideradas, Assim, o desempenho do algoritmo EM na estimação dos parâmetros é eficiente. Nas amostras de tamanho pequeno, quando $n = 40$, os resultados das estimativas são tão eficientes quanto àquelas obtidas de amostras de tamanhos maiores.

Todas os resultados que consideram 100 amostras Monte Carlo apresentam desvios padrão similares, no entanto os valores das médias das estimativas para todos os parâmetros são próximos dos considerados para a simulação.

2) **Determinação da posição do ponto de mudança**

Foram simuladas 250 amostras Monte Carlo, cada uma delas de tamanho $n = 40$, com o ponto de mudança na posição $k = 20$ e considerando as especificações dadas no inicio desta seção. Para cada amostra gerada, a posição do ponto de mudança é determinada utilizando os critérios de avaliação do valor SIC descritos na Seção 3.4. Os erros aleatórios simulados seguem as seguintes distribuições: Normal, t de Student ($\nu = 4$), Slash ($\nu = 2$) e Normal Contaminada ($\nu = 0.2, \gamma = 0.3$).

Da mesma forma que em Hofrichter (2007), o desempenho da metodologia de determinação do ponto de mudança que utiliza o critério SIC é realizado avaliando a distribuição de frequência das posições de ponto de mudança obtidas nas amostras Monte Carlo.

Tabela 3.6: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança na variância Normal, considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_0	β_1	σ_1^2	σ_2^2
Posição do ponto de mudança k=20						
100	n=40	Média	0.81	1.20	0.89	8.42
		d.p	0.45	0.04	0.32	2.60
1000	n=40	Média	0.81	1.20	0.91	8.91
		d.p	0.46	0.04	0.29	2.92
Posição do ponto de mudança k=200						
100	n=400	Média	0.80	1.20	1.00	9.15
		d.p	0.15	0.01	0.12	1.02
1000	n=400	Média	0.80	1.20	0.99	8.98
		d.p	0.14	0.01	0.10	0.88
Posição do ponto de mudança k=500						
100	n=1000	Média	0.81	1.20	1.00	8.97
		d.p	0.09	0.01	0.06	0.50
1000	n=1000	Média	0.80	1.20	1.00	9.00
		d.p	0.08	0.01	0.06	0.58
Posição do ponto de mudança k=10						
100	n=40	Média	0.80	1.20	0.88	8.83
		d.p	0.67	0.06	0.47	2.34
1000	n=40	Média	0.80	1.20	0.84	8.87
		d.p	0.60	0.05	0.43	2.34
Posição do ponto de mudança k=100						
100	n=400	Média	0.80	1.20	0.97	8.98
		d.p	0.18	0.02	0.15	0.75
1000	n=400	Média	0.81	1.20	0.99	8.96
		d.p	0.17	0.02	0.14	0.74
Posição do ponto de mudança k=250						
100	n=1000	Média	0.78	1.20	0.99	9.03
		d.p	0.11	0.01	0.09	0.49
1000	n=1000	Média	0.80	1.20	0.99	8.98
		d.p	0.11	0.01	0.09	0.45
Posição do ponto de mudança k=30						
100	n=40	Média	0.80	1.20	0.93	8.66
		d.p	0.36	0.03	0.24	3.85
1000	n=40	Média	0.79	1.20	0.95	9.09
		d.p	0.37	0.03	0.24	4.08
Posição do ponto de mudança k=300						
100	n=400	Média	0.80	1.20	0.99	8.93
		d.p	0.10	0.01	0.07	1.44
1000	nn=400	Média	0.80	1.20	0.99	8.93
		d.p	0.11	0.01	0.08	1.26
Posição do ponto de mudança k=750						
100	n=1000	Média	0.81	1.20	0.99	8.83
		d.p	0.08	0.01	0.05	0.81
1000	n=1000	Média	0.80	1.20	1.00	8.99
		d.p	0.07	0.01	0.05	0.80

A Figura 3.4 mostra que para todas as situações a frequência maior de ocorrências se encontra

Tabela 3.7: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança na variância t de Student ($\nu = 2$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_0	β_1	σ_1^2	σ_2^2
Posição do ponto de mudança k=20						
100	n=40	Média	0.85	1.20	0.99	10.07
		d.p	0.74	0.06	0.53	5.09
1000	n=40	Média	0.83	1.20	1.00	9.74
		d.p	0.61	0.05	0.59	5.38
Posição do ponto de mudança k=200						
100	n=400	Média	0.79	1.20	1.01	9.12
		d.p	0.19	0.01	0.16	1.55
1000	n=400	Média	0.81	1.20	1.00	9.14
		d.p	0.17	0.01	0.16	1.43
Posição do ponto de mudança k=500						
100	n=1000	Média	0.79	1.20	0.99	8.92
		d.p	0.11	0.01	0.10	0.87
1000	n=1000	Média	0.80	1.20	1.00	9.07
		d.p	0.11	0.01	0.10	0.92
Posição do ponto de mudança k=10						
100	n=40	Média	0.81	1.19	0.97	9.46
		d.p	0.81	0.08	0.85	4.06
1000	n=40	Média	0.79	1.20	1.03	9.45
		d.p	0.81	0.07	1.04	4.43
Posição do ponto de mudança k=100						
100	n=400	Média	0.81	1.20	0.98	9.04
		d.p	0.23	0.02	0.22	1.27
1000	n=400	Média	0.79	1.20	1.00	9.03
		d.p	0.22	0.02	0.23	1.18
Posição do ponto de mudança k=250						
100	n=1000	Média	0.79	1.20	1.00	9.02
		d.p	0.13	0.01	0.13	0.68
1000	n=1000	Média	0.80	1.20	1.00	9.02
		d.p	0.14	0.01	0.14	0.72
Posição do ponto de mudança k=30						
100	n=40	Média	0.81	1.20	1.01	11.27
		d.p	0.50	0.04	0.43	8.06
1000	n=40	Média	0.80	1.20	0.98	11.48
		d.p	0.51	0.04	0.43	9.81
Posição do ponto de mudança k=300						
100	n=400	Média	0.79	1.20	0.98	9.24
		d.p	0.14	0.01	0.14	2.06
1000	n=400	Média	0.80	1.20	1.00	9.07
		d.p	0.14	0.01	0.13	2.02
Posição do ponto de mudança k=750						
100	n=1000	Média	0.78	1.20	1.02	9.27
		d.p	0.09	0.01	0.09	1.48
1000	n=1000	Média	0.80	1.20	1.00	8.99
		d.p	0.10	0.01	0.08	1.26

Tabela 3.8: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança na variância Slash ($\nu = 4$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_0	β_1	σ_1^2	σ_2^2
Posição do ponto de mudança k=20						
100	n=40	Média	0.75	1.20	0.94	9.30
		d.p	0.54	0.05	0.33	3.46
1000	n=40	Média	0.80	1.20	0.91	9.00
		d.p	0.51	0.05	0.32	3.07
Posição do ponto de mudança k=200						
100	n=400	Média	0.82	1.20	0.99	8.98
		d.p	0.16	0.01	0.10	0.79
1000	n=400	Média	0.80	1.20	0.99	9.00
		d.p	0.15	0.01	0.11	0.98
Posição do ponto de mudança k=500						
100	n=1000	Média	0.79	1.20	0.99	9.00
		d.p	0.09	0.01	0.06	0.66
1000	n=1000	Média	0.80	1.20	1.00	8.99
		d.p	0.09	0.01	0.07	0.63
Posição do ponto de mudança k=10						
100	n=40	Média	0.81	1.20	0.83	9.01
		d.p	0.69	0.06	0.38	2.44
1000	n=40	Média	0.84	1.20	0.89	9.01
		d.p	0.71	0.06	0.52	2.41
Posição do ponto de mudança k=100						
100	n=400	Média	0.84	1.20	0.96	8.85
		d.p	0.20	0.02	0.15	0.85
1000	n=400	Média	0.80	1.20	0.98	9.00
		d.p	0.20	0.02	0.15	0.81
Posição do ponto de mudança k=250						
100	n=1000	Média	0.79	1.20	1.01	9.08
		d.p	0.13	0.01	0.09	0.52
1000	n=1000	Média	0.80	1.20	0.99	8.97
		d.p	0.13	0.01	0.10	0.49
Posição do ponto de mudança k=30						
100	n=40	Média	0.86	1.20	0.93	9.43
		d.p	0.43	0.04	0.25	5.40
1000	n=40	Média	0.81	1.20	0.94	8.93
		d.p	0.45	0.04	0.26	4.65
Posição do ponto de mudança k=300						
100	n=400	Média	0.80	1.20	0.98	9.09
		d.p	0.13	0.01	0.08	1.32
1000	n=400	Média	0.80	1.20	0.99	9.04
		d.p	0.13	0.01	0.08	1.29
Posição do ponto de mudança k=750						
100	n=1000	Média	0.79	1.20	1.00	9.03
		d.p	0.08	0.01	0.05	0.90
1000	n=1000	Média	0.80	1.20	1.00	9.05
		d.p	0.08	0.01	0.06	0.86

Tabela 3.9: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança na variância Normal Contaminada ($\nu = 0.2, \gamma = 0.3$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_0	β_1	σ_1^2	σ_2^2
Posição do ponto de mudança k=20						
100	n=40	Média	0.77	1.20	0.93	8.66
		d.p	0.54	0.04	0.36	3.10
1000	n=40	Média	0.81	1.20	0.93	9.30
		d.p	0.52	0.05	0.36	3.56
Posição do ponto de mudança k=200						
100	n=400	Média	0.81	1.20	0.99	8.88
		d.p	0.14	0.01	0.12	1.05
1000	n=400	Média	0.80	1.20	0.99	9.02
		d.p	0.15	0.01	0.12	1.03
Posição do ponto de mudança k=500						
100	n=1000	Média	0.80	1.20	1.00	8.89
		d.p	0.10	0.01	0.07	0.58
1000	n=1000	Média	0.80	1.20	1.00	9.00
		d.p	0.10	0.01	0.07	0.68
Posição do ponto de mudança k=10						
100	n=40	Média	0.84	1.19	0.90	8.94
		d.p	0.70	0.06	0.54	2.64
1000	n=40	Média	0.81	1.20	0.96	9.16
		d.p	0.72	0.06	0.60	2.86
Posição do ponto de mudança k=100						
100	n=400	Média	0.78	1.20	0.99	8.86
		d.p	0.19	0.02	0.18	0.88
1000	n=400	Média	0.80	1.20	0.99	9.01
		d.p	0.21	0.02	0.16	0.86
Posição do ponto de mudança k=250						
100	n=1000	Média	0.81	1.20	1.00	9.16
		d.p	0.13	0.01	0.09	0.54
1000	n=1000	Média	0.80	1.20	0.99	8.99
		d.p	0.13	0.01	0.10	0.54
Posição do ponto de mudança k=30						
100	n=40	Média	0.70	1.21	0.98	10.27
		d.p	0.35	0.03	0.27	5.20
1000	n=40	Média	0.81	1.20	0.94	9.32
		d.p	0.43	0.04	0.30	5.17
Posição do ponto de mudança k=300						
100	n=400	Média	0.80	1.20	1.00	9.13
		d.p	0.11	0.01	0.10	1.32
1000	n=400	Média	0.80	1.20	0.99	9.08
		d.p	0.13	0.01	0.10	1.50
Posição do ponto de mudança k=750						
100	n=1000	Média	0.80	1.20	1.00	8.89
		d.p	0.09	0.01	0.06	0.85
1000	n=1000	Média	0.80	1.20	1.00	9.08
		d.p	0.08	0.01	0.06	0.95

na posição 20. Além disso, valores próximos da posição 20 apresentam também frequências de ocorrência altas. Dessa forma, isto leva a concluir que a metodologia proposta consegue determinar adequadamente o ponto de mudança.

Finalmente, selecionou-se de cada conjunto de amostras Monte Carlo, uma amostra onde o ponto de mudança estimado se encontra na posição 20, e todos os valores SIC obtidos para esta amostra são avaliados. A Figura 3.5 mostra que para todos os casos o menor valor do SIC corresponde à posição 20. Desta forma, se garante a coerência do critério utilizado.

3) **Otimização dos valores de $\boldsymbol{\nu}$**

Considerando as mesmas simulações Monte Carlo do item anterior, foram selecionadas as amostras onde a posição estimada do ponto de mudança é $k = 20$. Para estas amostras, os valores de $\boldsymbol{\nu}$ foram avaliados. Neste caso, foram desconsiderados valores extremos.

A Figura 3.6 mostra que em média os valores otimizados de $\boldsymbol{\nu}$ estão próximos dos valores originais de simulação. Isto garante o bom desempenho do algoritmo na otimização dos valores avaliados.

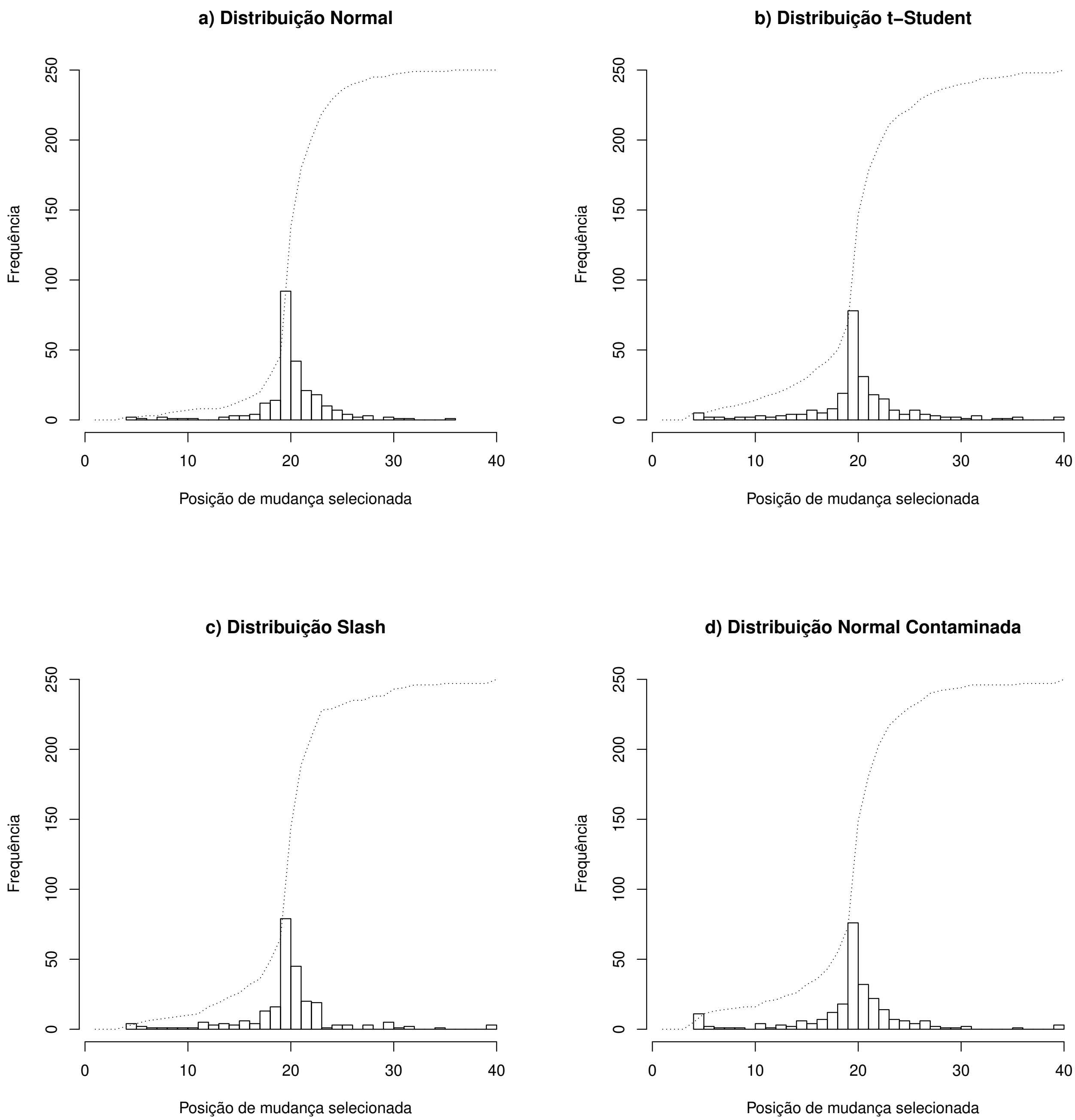


Figura 3.4: Frequências absolutas das posições dos pontos de mudança estimados nas amostras Monte Carlo simuladas no modelo de regressão linear com mudança na variância.

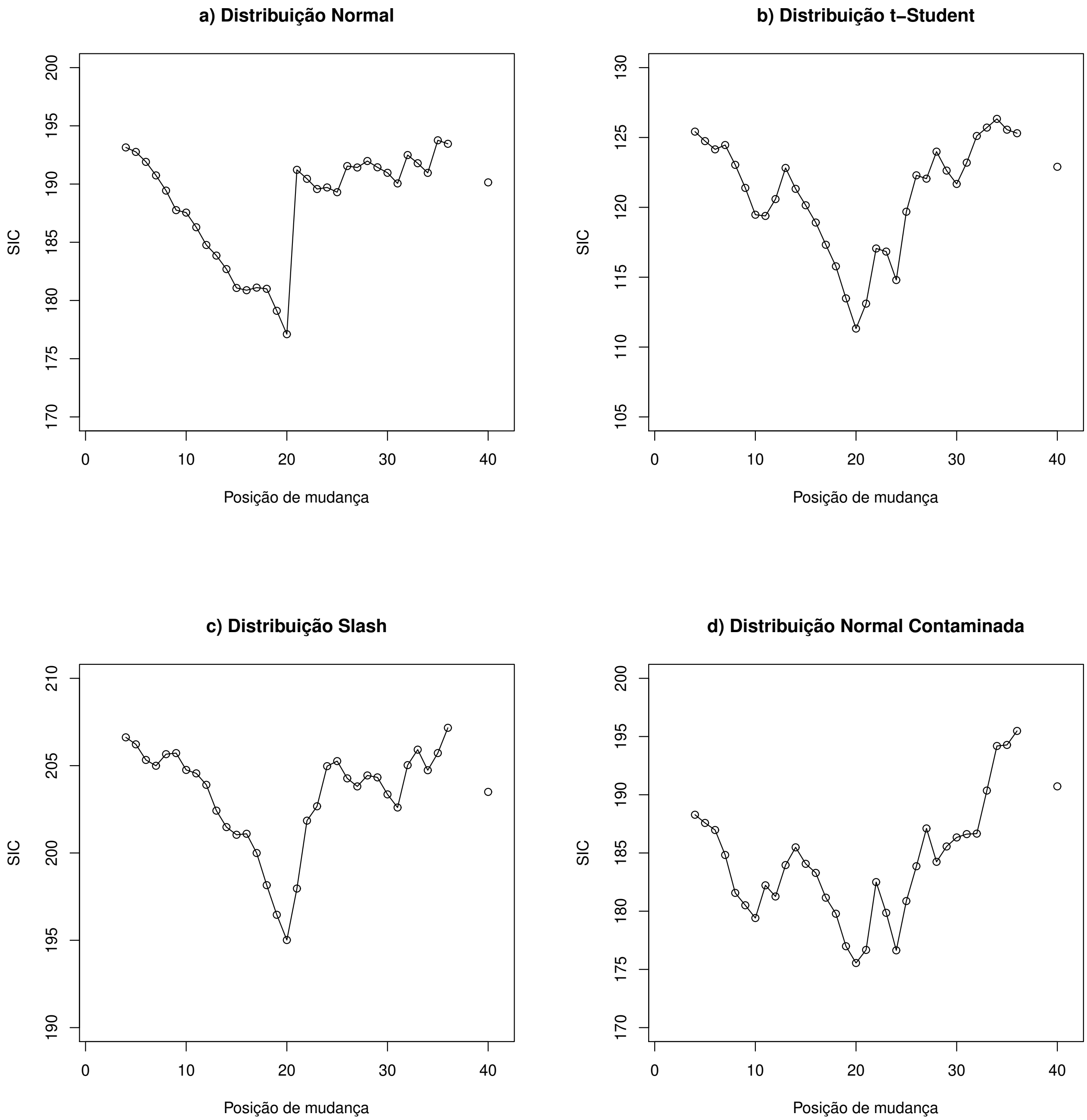


Figura 3.5: Aplicação do critério SIC sobre uma amostra Monte Carlo simulada, onde o ponto de mudança estimado é $k = 20$, no modelo de regressão linear com mudança na variância.

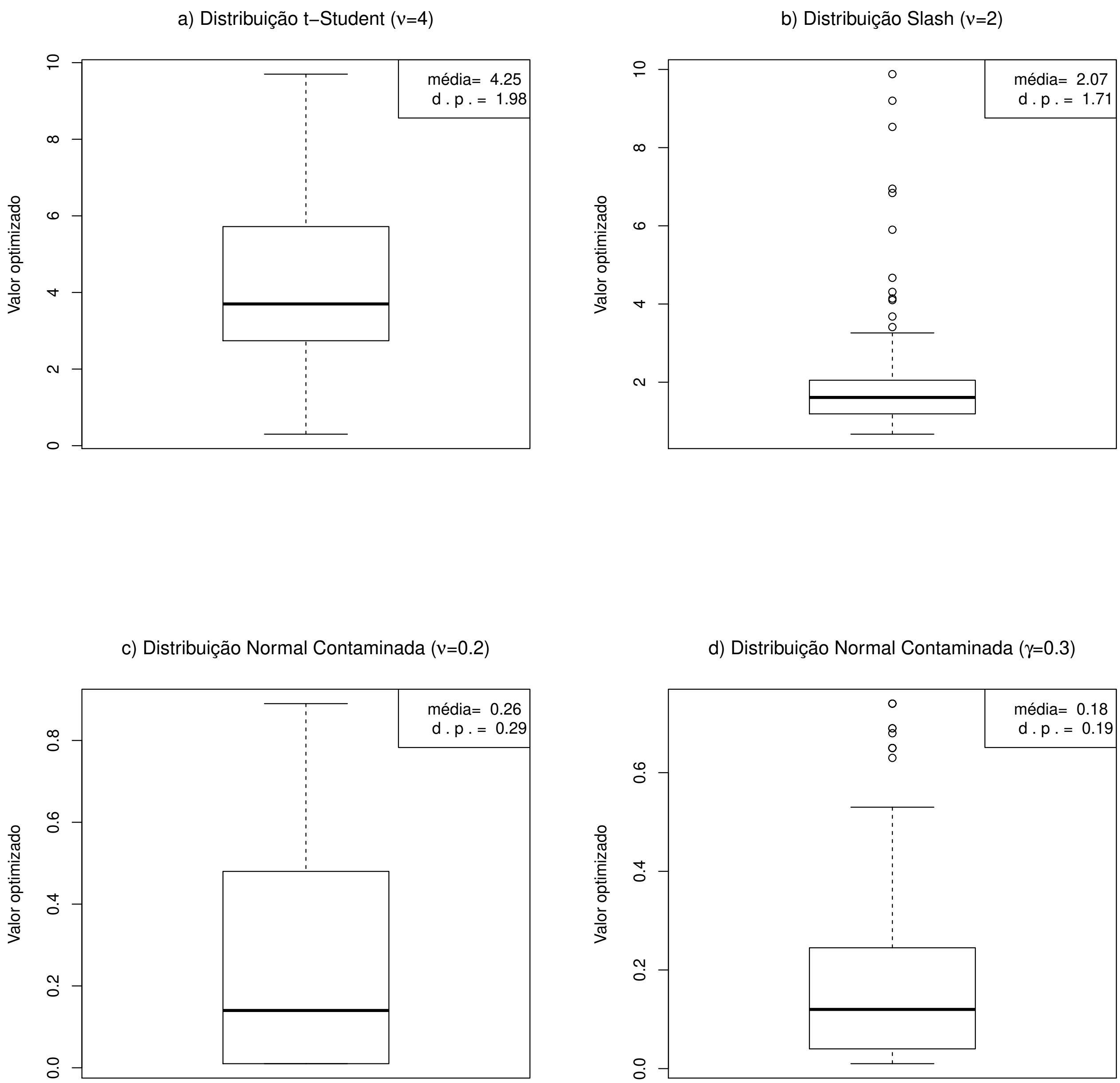


Figura 3.6: Resultado de simulações sobre otimização do parâmetro $\boldsymbol{\nu}$ das distribuições t de Student, Slash e Normal Contaminada no modelo de regressão linear com mudança na variância.

3.7 Mudança nos Coeficientes de Regressão e Variância

3.7.1 Especificação do Modelo

O modelo geral (3.1) pode ser escrito de forma matricial como:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \tag{3.19}$$

onde $\boldsymbol{\beta}$=$(\beta_0, \beta_1, ..., \beta_{p-1})^\top$ e $\mathbf{x}_i$=$(1, x_{1i}, ..., x_{1(p-1)})^\top$. Ao considerar a existência de mudança nos coeficientes de regressão e a variância a partir da posição k, temos que os erros ϵ_1,...,ϵ_n são independentes e distribuídas como:

$$e_i \sim \begin{cases} MEN(0, \sigma_1^2, H), & para \quad i = 1, \ldots, k, \\ MEN(0, \sigma_2^2, H), & para \quad i = k+1, \ldots, n, \end{cases}$$

e também:

$$\begin{array}{lll} \mathbf{Y}_i \sim MEN(\mathbf{x}_i^\top \boldsymbol{\beta}_1, \sigma_1^2; H), & para & i = 1, \ldots, k \quad e \\ \mathbf{Y}_i \sim MEN(\mathbf{x}_i^\top \boldsymbol{\beta}_2, \sigma_2^2; H), & para & i = k+1, \ldots, n. \end{array} \tag{3.20}$$

Para esta estrutura de mudança o teste de hipótese a considerar é:

$$\begin{array}{ll} H_0 : & Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i \quad e \quad \sigma^2, \qquad i = 1, \ldots, n, \\ \\ H_1 : & \begin{cases} Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_1 + \epsilon_i \quad e \quad \sigma_1^2, & i = 1, \ldots, k, \\ Y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_2 + \epsilon_i \quad e \quad \sigma_2^2, & i = k+1, \ldots, n. \end{cases} \end{array} \tag{3.21}$$

Usando apropriadamente a expressão (3.5), as funções de log-verossimilhança para este modelo nas situações de hipótese nula e hipótese alternativa apresentadas em (3.21) são, respetivamente:

$$\ell_0(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 + \sum_{i=1}^{n} \log \mathbb{K}_i, \tag{3.22}$$

para $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1^2, \sigma_2^2)$, temos que

$$\ell_k(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi - \frac{k}{2}\log \sigma_1^2 - \frac{(n-k)}{2}\log \sigma_2^2 + \sum_{i=1}^{k} \log \mathbb{K}_i^{(1)} + \sum_{i=k+1}^{n} \log \mathbb{K}_i^{(2)}, \tag{3.23}$$

onde $$d_i = \begin{cases} \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2}, & em \quad \mathbb{K}_i, \\ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\sigma_1^2}, & em \quad \mathbb{K}_i^{(1)}, \\ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2}{\sigma_2^2}, & em \quad \mathbb{K}_i^{(2)}. \end{cases}$$

As expressões (3.22) e (3.23) tornam-se importantes, pois serão utilizadas na aplicação do critério para determinação de ponto de mudança.Estas deverão ser avaliadas nas estimativas de máxima verossimilhança dos parâmetros. Assim, será possível encontrar os valores $\ell_0(\widehat{\boldsymbol{\theta}}) = \ell_0(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2})$ e $\ell_k(\widehat{\boldsymbol{\theta}}) = \ell_k(\widehat{\boldsymbol{\beta}_1}, \widehat{\boldsymbol{\beta}_2}, \widehat{\sigma_1^2}, \widehat{\sigma_2^2})$ que compõem o Critério de Informação Schwarz (SIC) utilizado como parte da metodologia proposta.

3.7.2 Derivadas Parciais de $\ell(\boldsymbol{\theta})$

Considerando os resultados apresentados para a expressão $\mathbb{K}_i$ e usando o resultado (3.6), o cálculos necessários para a obtenção das derivadas parciais de primeira ordem de $\ell(\boldsymbol{\theta})$são:

$$\frac{\partial \log \sigma_1^2}{\partial \boldsymbol{\beta}_1} = 0, \qquad \frac{\partial \log \sigma_1^2}{\partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial \log \sigma_1^2}{\partial \sigma_2^2} = 0,$$
$$\frac{\partial \log \sigma_2^2}{\partial \boldsymbol{\beta}_1} = 0, \qquad \frac{\partial \log \sigma_2^2}{\partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial \log \sigma_2^2}{\partial \sigma_1^2} = 0,$$
$$\frac{\partial \log \sigma_1^2}{\partial \sigma_1^2} = \begin{cases} \frac{1}{\sigma_1^2}, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial \log \sigma_2^2}{\partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ \frac{1}{\sigma_2^2}, & para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial d_i}{\partial \boldsymbol{\beta}_1} = \begin{cases} -2\frac{1}{\sigma_1^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)\mathbf{x}_i\top, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial d_i}{\partial \boldsymbol{\beta}_2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ -2\frac{1}{\sigma_2^2}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)\mathbf{x}_i\top, & para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial d_i}{\partial \sigma_1^2} = \begin{cases} -\frac{1}{\sigma_1^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial d_i}{\partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ -\frac{1}{\sigma_2^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2, & para \quad i = k+1, \ldots, n, \end{cases}$$

Assim,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{cases} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1} = \frac{1}{\sigma_1^2} \sum_{i=1}^{k} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)\mathbf{x}_i \right\}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_2} = \frac{1}{\sigma_2^2} \sum_{i=k+1}^{n} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)\mathbf{x}_i \right\}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_1^2} = -\frac{k}{2\sigma_1^2} + \frac{1}{2\sigma_1^4} \sum_{i=1}^{k} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2 \right\}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_2^2} = -\frac{(n-k)}{2\sigma_2^2} + \frac{1}{2\sigma_2^4} \sum_{i=k+1}^{n} \left\{ \frac{\mathbb{I}_i(3/2)}{\mathbb{I}_i(1/2)} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2 \right\}. \end{cases}$$

Da mesma forma, calculando as segundas derivadas parciais de $\ell(\boldsymbol{\theta})$ temos:

$$\frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \sigma_2^2 \partial \sigma_2^2} = 0,$$
$$\frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \sigma_1^2 \partial \sigma_1^2} = 0,$$

$$\frac{\partial^2 \log \sigma_1^2}{\partial \sigma_1^2 \partial \sigma_1^2} = \begin{cases} -\dfrac{1}{\sigma_1^4}, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 \log \sigma_2^2}{\partial \sigma_2^2 \partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ -\dfrac{1}{\sigma_2^4}, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_1 \partial \sigma_1^2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_1 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_2 \partial \sigma_1^2} = 0,$$

$$\frac{\partial^2 \log \sigma_1^2}{\partial \boldsymbol{\beta}_2 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_1^2}{\partial \sigma_1^2 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_1 \partial \sigma_1^2} = 0,$$

$$\frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_1 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_2 \partial \sigma_1^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \boldsymbol{\beta}_2 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 \log \sigma_2^2}{\partial \sigma_1^2 \partial \sigma_2^2} = 0,$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} = 0, \qquad \frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_1 \partial \sigma_2^2} = 0, \qquad \frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_2 \partial \sigma_1^2} = 0, \qquad \frac{\partial^2 d_i}{\partial \sigma_1^2 \partial \sigma_2^2} = 0.$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1} = \begin{cases} 2\dfrac{1}{\sigma_1^2}\mathbf{x}_i\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma_2^2}\mathbf{x}_i\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \sigma_1^2 \partial \sigma_1^2} = \begin{cases} 2\dfrac{1}{\sigma_1^6}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \sigma_2^2 \partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma_2^6}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_1 \partial \sigma_1^2} = \begin{cases} 2\dfrac{1}{\sigma_1^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)\mathbf{x}_i^\top, & para \quad i = 1, \ldots, k, \\ 0, & para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 d_i}{\partial \boldsymbol{\beta}_2 \partial \sigma_2^2} = \begin{cases} 0, & para \quad i = 1, \ldots, k, \\ 2\dfrac{1}{\sigma_2^4}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)\mathbf{x}_i^\top, & para \quad i = k+1, \ldots, n, \end{cases}$$

Observe que as primeiras derivadas parciais correpondem às funções escore para os parâmetros. Este conjunto de equações corresponde a um sistema não linear de equações, que adicionalmente inclui os termos $\mathbb{I}_i(1/2)$ e $\mathbb{I}_i(3/2)$ que por (3.3) são expressões matemáticas complexas que incluem uma integral e que sempre dependem dos parâmetros $\boldsymbol{\theta}$ presentes em d_i. Uma solução via métodos numéricos é possível.

3.7.3 O Algoritmo EM

Uma característica deste algoritmo é a formulação a partir de uma representação hierárquica que permite uma flexibilidade no tratamento das derivadas teóricas. Assim, considerando o modelo de regressão linear definido em (3.10)-(3.11) e da Proposição 2.4.1 temos que:

$$Y_i \mid U = u_i \sim \begin{cases} N_1(\mathbf{x}_i^\top \boldsymbol{\beta}_1, \kappa(u_i)\sigma_1^2) & para \quad i = 1, \ldots, k, \\ N_1(\mathbf{x}_i^\top \boldsymbol{\beta}_2, \kappa(u_i)\sigma_2^2) & para \quad i = k+1, \ldots, n, \end{cases}$$

$e \quad U_i \sim H(u_i; \boldsymbol{\nu})$, onde Y_i e U_i são independentes para todo $i = 1, \ldots, n$.

No caso da hipótese nula, apresentada em (3.21), consideram-se os vetores $\mathbf{y} = (y_1, .., y_n)^\top$, $\mathbf{u} = (u_1, .., u_n)^\top$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$ que denota o vetor de parâmetros do modelo.

Para todo $i = 1, \ldots, n$, temos que:

$$f(y_i, u_i|\boldsymbol{\theta}) = f(y_i|u_i, \boldsymbol{\theta}) f(u_i) = \frac{1}{(2\pi\, \kappa(u_i)\, \sigma^2)^{1/2}} \exp\left\{ -\frac{1}{2} \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\kappa(u_i)\, \sigma^2} \right\} h(u_i; \boldsymbol{\nu})$$

Assim, considerando que $y_1, .., y_n$ são independentes, a função de log-verossimilhança completa associada a $\mathbf{y_c} = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ é da forma:

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u}) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \left\{ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\kappa(u_i)\, \sigma^2} \right\} + C,$$

onde $C = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \kappa(u_i) + \sum_{i=1}^{n} h(u_i; \boldsymbol{\nu})$.

Para a hipótese alternativa, apresentada em (3.21), considerando os vetores $\mathbf{y} = (y_1, .., y_n)^\top$, $\mathbf{u} = (u_1, .., u_n)^\top$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \sigma_1^2, \sigma_2^2)^\top$ que denota o vetor de parâmetros do modelo, temos

$$f(y_i, u_i|\boldsymbol{\theta}) = f(y_i|u_i, \boldsymbol{\theta}) f(u_i) = \begin{cases} \dfrac{1}{(2\pi\, \kappa(u_i)\, \sigma_1^2)^{1/2}} \exp\left\{ -\dfrac{1}{2} \dfrac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\kappa(u_i)\, \sigma_1^2} \right\} h(u_i; \boldsymbol{\nu}), \\ \qquad para \quad i = 1, \ldots, k, \quad e \\ \\ \dfrac{1}{(2\pi\, \kappa(u_i)\, \sigma_2^2)^{1/2}} \exp\left\{ -\dfrac{1}{2} \dfrac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2}{\kappa(u_i)\, \sigma_2^2} \right\} h(u_i; \boldsymbol{\nu}), \\ \qquad para \quad i = k+1, \ldots, n. \end{cases}$$

Também neste caso $y_1, .., y_n$ são independentes e a função de log-verossimilhança completa associada a $\mathbf{y_c} = (\mathbf{y}^\top, \mathbf{u}^\top)^\top$ é da forma:

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u}) = -\frac{k}{2} \log(\sigma_1^2) - \frac{(n-k)}{2} \log(\sigma_2^2) - \frac{1}{2} \left[\sum_{i=1}^{k} \left\{ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2}{\kappa(u_i)\, \sigma_1^2} \right\} + \sum_{i=k+1}^{n} \left\{ \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2}{\kappa(u_i)\, \sigma_2^2} \right\} \right] + C,$$

onde $C = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \kappa(u_i) + \sum_{i=1}^{n} h(u_i; \boldsymbol{\nu})$.

O cálculo da esperança de $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ definida em (2.6) é dada por

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{u})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \left\{ q(d_i) \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2} \right\} + E[C|\mathbf{y}, \mathbf{u}],$$

para o caso da hipótese nula, e

$$E[\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})|\mathbf{y},\widehat{\boldsymbol{\theta}}] = -\frac{k}{2}\log(\sigma_1^2) - \frac{(n-k)}{2}\log(\sigma_2^2) - \frac{1}{2}\sum_{i=1}^{k}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2}{\sigma_1^2}\right\}$$
$$-\frac{1}{2}\sum_{i=k+1}^{n}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2}{\sigma_2^2}\right\} + E[C|\mathbf{y},\mathbf{u}],$$

para o caso da hipótese alternativa, onde $q(d_i) = E[\kappa^{-1}(U_i)|y_i,\widehat{\boldsymbol{\theta}}]$.

Assim, para obter a esperança condicional $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ é preciso calcular $q(d_i)$ para todos os valores $i = 1,\ldots,n$. Para as quatro distribuições da classe de Mistura de Escala Normal estudadas nesta dissertação, estes resultados estão apresentados em (2.10)-(2.13).

Finalmente, a r-ésima iteração do algoritmo EM para a estimação de parâmetros deste modelo é descrito como segue abaixo

Passo E: Calcular as esperanças condicionais $q(d_i)^{(r)} = E(\kappa(U_i)|Y_i,\widehat{\boldsymbol{\theta}}^{(r-1)})$ cujas formas conhecidas estão apresentadas na Tabela 3.1.

Sob a hipótese nula, temos que $d_i^{(r)} = \dfrac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}^{(r-1)})^2}{\sigma^{2(r-1)}}, \quad para \quad i = 1,\ldots,n.$

Sob a hipótese alternativa, temos que

$$d_i^{(r)} = \begin{cases} \dfrac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_1^{(r-1)})^2}{\sigma_1^{2(r-1)}}, & para \quad i = 1,\ldots,k, \\ \dfrac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_2^{(r-1)})^2}{\sigma_2^{2(r-1)}}, & para \quad i = k+1,\ldots,n. \end{cases}$$

Sob quaisquer das hipóteses, $\widehat{\boldsymbol{\theta}}^{(0)}$ é a estimação inicial dos parâmetros obtida a partir do método de mínimos quadrados.

Passo M: As estimativas de máxima verossimilhança sob a hipótese nula são obtidas por

$$\widehat{\boldsymbol{\beta}}^{(r)} = (\mathbf{X}^\top\mathbf{Q}^{(r)}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Q}^{(r)}\mathbf{Y},$$

$$\widehat{\sigma^2}^{(r)} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^{(r)})^\top\mathbf{Q}^{(r)}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^{(r)}),$$

onde $\mathbf{Q}^{(r)} = diag(q(d_1)^{(r)},...,q(d_n)^{(r)})$.

Os estimadores de máxima verossimilhança sob a hipótese alternativa são obtidos como:

$$\widehat{\boldsymbol{\beta}_1}^{(r)} = (\mathbf{X_1}^\top\mathbf{Q_1}^{(r)}\mathbf{X_1})^{-1}\mathbf{X_1}^\top\mathbf{Q_1}^{(r)}\mathbf{Y_1},$$

$$\widehat{\boldsymbol{\beta}_2}^{(r)} = (\mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{X_2})^{-1} \mathbf{X_2}^\top \mathbf{Q_2}^{(r)} \mathbf{Y_2},$$
$$\widehat{\sigma_1^2}^{(r)} = \frac{1}{k}[(\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1^{(r)})^\top \mathbf{Q_1}^{(r)} (\mathbf{Y_1} - \mathbf{X_1}\boldsymbol{\beta}_1^{(r)})],$$
$$\widehat{\sigma_2^2}^{(r)} = \frac{1}{(n-k)}[(\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2^{(r)})^\top \mathbf{Q_2}^{(r)} (\mathbf{Y_2} - \mathbf{X_2}\boldsymbol{\beta}_2^{(r)})],$$

onde $\mathbf{Q_1^{(r)}} = diag(q(d_1)^{(r)}, ..., q(d_k)^{(r)})$, $\mathbf{Q_2^{(r)}} = diag(q(d_{k+1})^{(r)}, ..., q(d_n)^{(r)})$, $\mathbf{X}_1$ é uma partição da matriz $\mathbf{X}$ que considera os k primeiras linhas, $\mathbf{Y}_1$ é uma partição do vetor $\mathbf{Y}$ que considera os k primeiras observações, $\mathbf{X}_2$ é uma partição da matriz $\mathbf{X}$ que considera os $n-k$ últimas linhas e $\mathbf{Y}_2$ é uma partição do vetor $\mathbf{Y}$ que considera os $n-k$ últimas observações.

No fim de cada iteração, logo do passo M é realizado um processo de otimização que consiste em maximizar a expressão $\sum_{i=1}^{n} \log \mathbb{K}_i$ que tem o objetivo de encontrar o melhor conjunto de valores para $\boldsymbol{\nu}$ que é um escalar ou um vetor de parâmetros da distribuição de U.

3.7.4 Estudos de Simulação

Nesta subseção, apresenta-se alguns resultados de diversas simulações realizadas para avaliar o desempenho do algoritmo EM e a metodologia proposta. Serão avaliadas as estimativas dos parâmetros, a aplicação do critério para determinação do ponto de mudança e a optimização dos valores de $\boldsymbol{\nu}$. Todas as amostras simuladas são geradas desde o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância

$$\begin{aligned} y_i &= 0.8 + 1.2x_i + \epsilon_i, \quad para \quad i = 1, \ldots, k, \\ y_i &= 2.0 + 1.5x_i + \epsilon_i, \quad para \quad i = k+1, \ldots, n, \end{aligned}$$

considerando $\sigma_1^2 = 1$ para $i = 1, \ldots, k$ e $\sigma_2^2 = 9$ para $i = k+1, \ldots, n$ e $x_i \sim U(0, 20)$ para $i = 1, \ldots, n$.

1) **Estimativa dos parâmetros**

Foram configuradas várias combinações de valores de n e k, e para cada uma delas, simulações Monte Carlo de 100 e 1000 amostras foram geradas. Os valores dos parâmetros estimados nas amostras geradas foram avaliados. Os erros aleatórios simulados seguem as seguintes distribuições: Normal, t de Student ($\nu = 2$), Slash ($\nu = 4$) e Normal Contaminada ($\nu = 0, 2, \gamma = 0.3$). Os resultados das simulações Monte Carlo encontram-se nas Tabelas 3.10 - 3.13.

Podemos observar nestas tabelas que os parâmetros são bem estimados em todas as combinações avaliadas e para todas as estruturas de erros aleatórios consideradas, Assim, o desempenho do algoritmo EM na estimação dos parâmetros é eficiente.

Nas amostras de tamanho pequeno, quando $n = 40$, os resultados das estimativas são tão eficientes quanto àquelas obtidas de amostras de tamanhos maiores. Esta observação torna-se importante, pois muitas das aplicações são realizadas sobre amostras de tamanho pequeno, por exemplo, a aplicação de Holbert (1982) é realizada sobre um conjunto de dados com 35 observações.

2) **Determinação da posição do ponto de mudança**

Foram simuladas 250 amostras Monte Carlo, cada uma delas de tamanho $n = 40$, com o ponto de mudança na posição $k = 20$ e considerando as especificações dadas no início desta seção. Para cada amostra gerada, a posição do ponto de mudança é determinada utilizando os critérios de avaliação do valor SIC descritos na Seção 3.4. Os erros aleatórios simulados seguem as seguintes distribuições: Normal, t de Student ($\nu = 4$), Slash ($\nu = 2$) e Normal Contaminada ($\nu = 0, 2, \gamma = 0.3$).

Da mesma forma que em Hofrichter (2007), o desempenho da metodologia de determinação do ponto de mudança que utiliza o critério SIC é realizado avaliando a distribuição de frequência das posições de ponto de mudança obtidas nas amostras Monte Carlo.

A Figura 3.7 mostra que para todas as situações a frequência maior de ocorrências se encontra na posição 20. Além disso, valores próximos da posição 20 apresentam também frequências de ocorrência altas. Isto leva a concluir que a metodologia proposta consegue determinar adequadamente o ponto de mudança.

Finalmente, selecionou-se de cada conjunto de amostras Monte Carlo, uma amostra onde o ponto de mudança estimado se encontra na posição 20, e todos os valores SIC obtidos para esta amostra são avaliados. A Figura 3.8 mostra que para todos os casos o menor valor do SIC corresponde à posição 20. Desta forma, se garante a coerência do critério utilizado.

3) **Otimização dos valores de $\boldsymbol{\nu}$**

Considerando as mesmas simulações Monte Carlo do item anterior, foram selecionadas as amostras onde a posição estimada do ponto de mudança é $k = 20$. Para estas amostras, os valores de $\boldsymbol{\nu}$ foram avaliados. Neste caso foram desconsiderados valores extremos.

A Figura 3.9 mostra que em média os valores otimizados de $\boldsymbol{\nu}$ estão próximos dos valores originais de simulação. Isto garante o bom desempenho do algoritmo na optimização dos valores avaliados.

Tabela 3.10: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão e variância Normal, considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ_1^2	σ_2^2
Posição do ponto de mudança k=20								
100	n=40	Média	0.84	1.20	1.93	1.50	0.89	8.07
		d.p	0.45	0.04	1.46	0.13	0.27	2.68
1000	n=40	Média	0.78	1.20	2.01	1.50	0.90	8.17
		d.p	0.47	0.04	1.37	0.12	0.30	2.66
Posição do ponto de mudança k=200								
100	n=400	Média	0.78	1.20	2.02	1.50	0.98	8.72
		d.p	0.14	0.01	0.41	0.04	0.11	0.90
1000	n=400	Média	0.80	1.20	1.95	1.50	0.99	8.94
		d.p	0.14	0.01	0.44	0.04	0.10	0.90
Posição do ponto de mudança k=500								
100	n=1000	Média	0.81	1.20	1.96	1.50	0.99	8.91
		d.p	0.09	0.01	0.27	0.02	0.07	0.60
1000	n=1000	Média	0.80	1.20	2.00	1.50	0.99	8.96
		d.p	0.09	0.01	0.26	0.02	0.06	0.56
Posição do ponto de mudança k=10								
100	n=40	Média	0.61	1.21	1.88	1.51	0.82	8.78
		d.p	0.68	0.06	1.05	0.09	0.41	2.59
1000	n=40	Média	0.80	1.20	1.98	1.50	0.80	8.46
		d.p	0.73	0.06	1.12	0.10	0.39	2.20
Posição do ponto de mudança k=100								
100	n=400	Média	0.83	1.20	1.98	1.50	0.97	8.79
		d.p	0.21	0.02	0.36	0.03	0.13	0.74
1000	n=400	Média	0.80	1.20	2.02	1.50	0.97	8.97
		d.p	0.20	0.02	0.34	0.03	0.14	0.72
Posição do ponto de mudança k=250								
100	n=1000	Média	0.79	1.20	1.99	1.50	1.00	8.95
		d.p	0.12	0.01	0.22	0.02	0.09	0.51
1000	n=1000	Média	0.79	1.20	2.00	1.50	0.99	8.97
		d.p	0.13	0.01	0.22	0.02	0.09	0.46
Posição do ponto de mudança k=30								
100	n=40	Média	0.85	1.20	1.85	1.52	0.93	7.18
		d.p	0.35	0.03	2.11	0.19	0.26	3.92
1000	n=40	Média	0.82	1.20	2.01	1.49	0.94	7.29
		d.p	0.39	0.03	2.07	0.18	0.25	3.67
Posição do ponto de mudança k=300								
100	n=400	Média	0.80	1.20	1.98	1.50	0.99	9.05
		d.p	0.12	0.01	0.62	0.05	0.08	1.33
1000	n=400	Média	0.80	1.20	1.99	1.50	0.99	8.81
		d.p	0.11	0.01	0.60	0.05	0.08	1.29
Posição do ponto de mudança k=750								
100	n=1000	Média	0.80	1.20	1.93	1.50	0.98	9.01
		d.p	0.07	0.01	0.42	0.03	0.05	0.83
1000	n=1000	Média	0.80	1.20	2.02	1.50	1.00	8.90
		d.p	0.07	0.01	0.39	0.03	0.05	0.78

Tabela 3.11: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão e variância t de Student ($\nu = 2$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ_1^2	σ_2^2
Posição do ponto de mudança k=20								
100	n=40	Média	0.82	1.20	2.31	1.46	0.93	8.67
		d.p	0.56	0.05	2.16	0.20	0.43	4.45
1000	n=40	Média	0.79	1.20	2.00	1.50	0.94	8.54
		d.p	0.65	0.06	2.01	0.17	0.53	4.73
Posição do ponto de mudança k=200								
100	n=400	Média	0.81	1.20	2.03	1.50	0.98	9.17
		d.p	0.19	0.02	0.59	0.05	0.16	1.69
1000	n=400	Média	0.79	1.20	2.02	1.50	0.99	9.04
		d.p	0.19	0.02	0.58	0.05	0.16	1.47
Posição do ponto de mudança k=500								
100	n=1000	Média	0.79	1.20	2.00	1.50	1.01	9.20
		d.p	0.11	0.01	0.30	0.03	0.10	0.89
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00	8.97
		d.p	0.11	0.01	0.35	0.03	0.10	0.90
Posição do ponto de mudança k=10								
100	n=40	Média	0.81	1.20	2.19	1.48	0.73	8.57
		d.p	0.88	0.07	1.66	0.15	0.49	3.60
1000	n=40	Média	0.80	1.20	1.99	1.50	0.90	8.61
		d.p	1.08	0.10	1.58	0.14	0.83	3.65
Posição do ponto de mudança k=100								
100	n=400	Média	0.78	1.20	1.93	1.51	0.97	8.75
		d.p	0.31	0.03	0.44	0.04	0.22	1.06
1000	n=400	Média	0.80	1.20	1.99	1.50	1.00	8.97
		d.p	0.26	0.02	0.46	0.04	0.23	1.22
Posição do ponto de mudança k=250								
100	n=1000	Média	0.77	1.20	2.01	1.50	1.00	8.89
		d.p	0.14	0.01	0.32	0.03	0.16	0.72
1000	n=1000	Média	0.80	1.20	1.99	1.50	1.01	9.00
		d.p	0.16	0.01	0.27	0.02	0.14	0.75
Posição do ponto de mudança k=30								
100	n=40	Média	0.81	1.20	1.74	1.52	1.02	8.84
		d.p	0.48	0.04	2.82	0.25	0.44	8.96
1000	n=40	Média	0.76	1.20	2.00	1.50	0.99	8.30
		d.p	0.49	0.04	3.26	0.30	0.45	8.05
Posição do ponto de mudança k=300								
100	n=400	Média	0.82	1.20	2.00	1.50	0.99	9.00
		d.p	0.16	0.01	0.79	0.07	0.13	2.14
1000	n=400	Média	0.79	1.20	1.98	1.50	1.00	8.99
		d.p	0.15	0.01	0.78	0.07	0.13	2.15
Posição do ponto de mudança k=750								
100	n=1000	Média	0.81	1.20	2.07	1.49	1.00	9.17
		d.p	0.10	0.01	0.47	0.04	0.08	1.33
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00	8.93
		d.p	0.10	0.01	0.49	0.04	0.08	1.26

Tabela 3.12: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão e variância Slash ($\nu = 4$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ_1^2	σ_2^2
Posição do ponto de mudança k=20								
100	n=40	Média	0.77	1.20	2.05	1.50	0.85	8.75
		d.p	0.53	0.04	1.68	0.14	0.30	3.49
1000	n=40	Média	0.78	1.20	2.07	1.49	0.90	8.27
		d.p	0.50	0.04	1.58	0.14	0.32	2.86
Posição do ponto de mudança k=200								
100	n=400	Média	0.79	1.20	1.92	1.51	0.99	8.91
		d.p	0.15	0.01	0.48	0.04	0.12	1.13
1000	n=400	Média	0.80	1.20	1.98	1.50	0.99	8.90
		d.p	0.16	0.01	0.49	0.04	0.10	0.94
Posição do ponto de mudança k=500								
100	n=1000	Média	0.79	1.20	2.06	1.50	0.99	9.04
		d.p	0.11	0.01	0.31	0.03	0.06	0.64
1000	n=1000	Média	0.80	1.20	2.01	1.50	1.00	8.98
		d.p	0.10	0.01	0.31	0.03	0.07	0.60
Posição do ponto de mudança k=10								
100	n=40	Média	0.78	1.20	1.97	1.50	0.84	8.81
		d.p	0.67	0.06	1.32	0.11	0.44	2.55
1000	n=40	Média	0.84	1.20	1.98	1.50	0.82	8.50
		d.p	0.84	0.07	1.22	0.11	0.43	2.37
Posição do ponto de mudança k=100								
100	n=400	Média	0.80	1.20	1.99	1.50	0.98	8.92
		d.p	0.20	0.02	0.40	0.04	0.16	0.83
1000	n=400	Média	0.80	1.20	1.99	1.50	0.99	8.94
		d.p	0.24	0.02	0.40	0.03	0.15	0.78
Posição do ponto de mudança k=250								
100	n=1000	Média	0.81	1.20	2.01	1.50	1.00	9.02
		d.p	0.16	0.01	0.23	0.02	0.09	0.51
1000	n=1000	Média	0.79	1.20	2.01	1.50	0.99	9.00
		d.p	0.14	0.01	0.25	0.02	0.10	0.47
Posição do ponto de mudança k=30								
100	n=40	Média	0.81	1.20	2.29	1.46	0.94	7.63
		d.p	0.45	0.04	2.55	0.23	0.26	3.96
1000	n=40	Média	0.79	1.20	1.95	1.50	0.95	7.59
		d.p	0.42	0.04	2.35	0.21	0.27	4.00
Posição do ponto de mudança k=300								
100	n=400	Média	0.79	1.20	2.06	1.50	1.00	8.99
		d.p	0.13	0.01	0.75	0.07	0.08	1.41
1000	n=400	Média	0.81	1.20	1.96	1.50	1.00	8.86
		d.p	0.13	0.01	0.72	0.06	0.08	1.33
Posição do ponto de mudança k=750								
100	n=1000	Média	0.80	1.20	1.99	1.50	1.00	8.92
		d.p	0.08	0.01	0.42	0.03	0.06	0.83
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00	8.98
		d.p	0.09	0.01	0.43	0.04	0.06	0.84

Tabela 3.13: Médias e desvios padrão (d.p) das estimativas dos parâmetros do modelo de regressão linear com mudança nos coeficientes de regressão e variância Normal Contaminada ($\nu = 0.2, \gamma = 0.3$), considerando 100 e 1000 amostras simuladas e diversos tamanhos de amostra (n) e posição do ponto de mudança (k).

Nro. de Simulações	Nro. de casos		β_{10}	β_{11}	β_{20}	β_{20}	σ_1^2	σ_2^2
Posição do ponto de mudança k=20								
100	n=40	Média	0.80	1.20	1.82	1.52	0.89	8.03
		d.p	0.58	0.05	1.95	0.16	0.35	3.52
1000	n=40	Média	0.80	1.20	2.00	1.50	0.93	8.43
		d.p	0.55	0.05	1.64	0.14	0.37	3.28
Posição do ponto de mudança k=200								
100	n=400	Média	0.80	1.20	1.96	1.50	1.00	8.80
		d.p	0.17	0.01	0.45	0.04	0.13	1.01
1000	n=400	Média	0.80	1.20	2.03	1.50	1.00	8.92
		d.p	0.17	0.01	0.49	0.04	0.12	1.05
Posição do ponto de mudança k=500								
100	n=1000	Média	0.79	1.20	1.95	1.50	0.99	9.07
		d.p	0.09	0.01	0.31	0.03	0.08	0.64
1000	n=1000	Média	0.80	1.20	2.00	1.50	1.00	8.98
		d.p	0.10	0.01	0.32	0.03	0.08	0.66
Posição do ponto de mudança k=10								
100	n=40	Média	0.77	1.20	2.14	1.49	0.78	8.14
		d.p	0.81	0.08	1.20	0.11	0.40	2.38
1000	n=40	Média	0.79	1.20	1.96	1.50	0.83	8.56
		d.p	0.82	0.07	1.34	0.12	0.48	2.69
Posição do ponto de mudança k=100								
100	n=400	Média	0.80	1.20	2.01	1.50	0.95	8.89
		d.p	0.22	0.02	0.43	0.04	0.14	0.83
1000	n=400	Média	0.79	1.20	1.99	1.50	0.98	8.90
		d.p	0.23	0.02	0.40	0.04	0.16	0.86
Posição do ponto de mudança k=250								
100	n=1000	Média	0.81	1.20	2.02	1.50	0.99	8.97
		d.p	0.13	0.01	0.25	0.02	0.10	0.57
1000	n=1000	Média	0.81	1.20	2.00	1.50	1.00	8.97
		d.p	0.15	0.01	0.26	0.02	0.10	0.53
Posição do ponto de mudança k=30								
100	n=40	Média	0.84	1.20	2.12	1.48	0.93	6.81
		d.p	0.36	0.03	2.90	0.24	0.30	4.03
1000	n=40	Média	0.81	1.20	2.07	1.49	0.96	7.72
		d.p	0.44	0.04	2.51	0.22	0.29	4.44
Posição do ponto de mudança k=300								
100	n=400	Média	0.79	1.20	1.92	1.51	1.00	8.86
		d.p	0.12	0.01	0.72	0.06	0.09	1.55
1000	n=400	Média	0.79	1.20	1.99	1.50	0.99	8.90
		d.p	0.13	0.01	0.71	0.06	0.09	1.56
Posição do ponto de mudança k=750								
100	n=1000	Média	0.80	1.20	2.05	1.50	1.00	8.98
		d.p	0.08	0.01	0.41	0.04	0.06	1.09
1000	n=1000	Média	0.80	1.20	2.02	1.50	1.00	8.90
		d.p	0.08	0.01	0.44	0.04	0.06	0.94

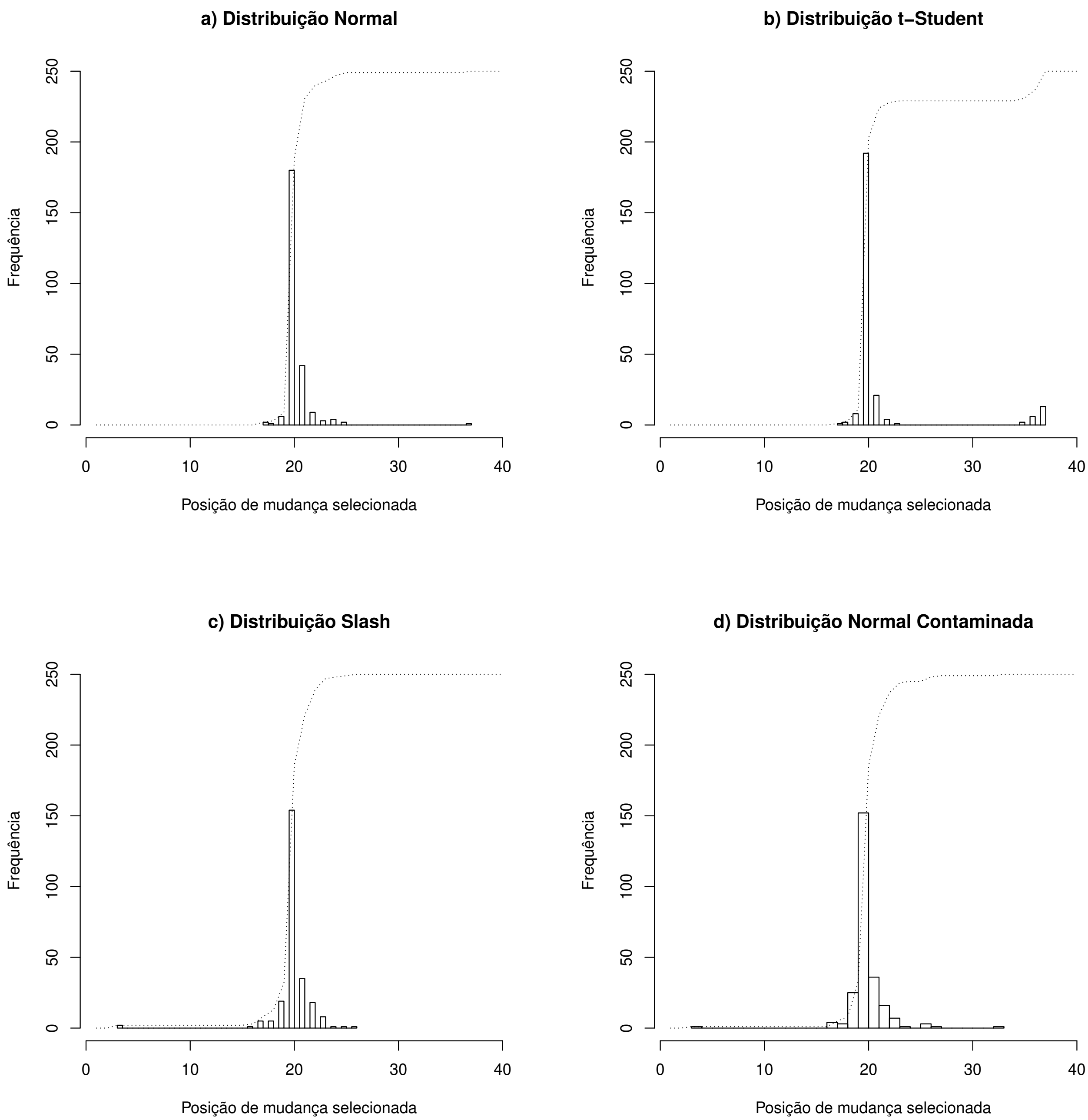


Figura 3.7: Frequências absolutas das posições dos pontos de mudança estimados nas amostras Monte Carlo simuladas no modelo de regressão linear com mudança nos coeficientes de regressão e variância.

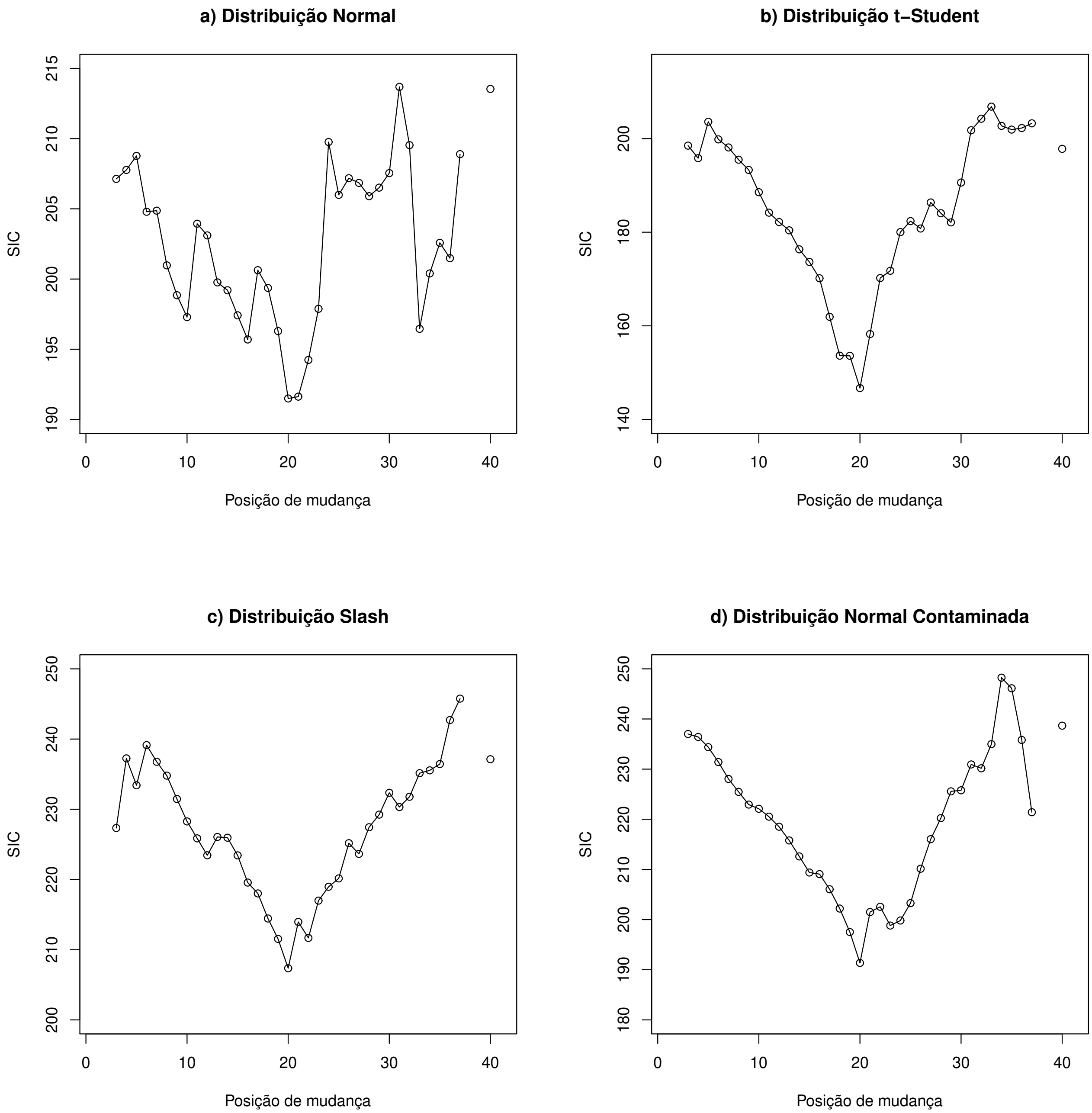


Figura 3.8: Aplicação do critério SIC sobre uma amostra Monte Carlo simulada, onde o ponto de mudança estimado é $k = 20$, no modelo de regressão linear com mudança nos coeficientes de regressão e variância.

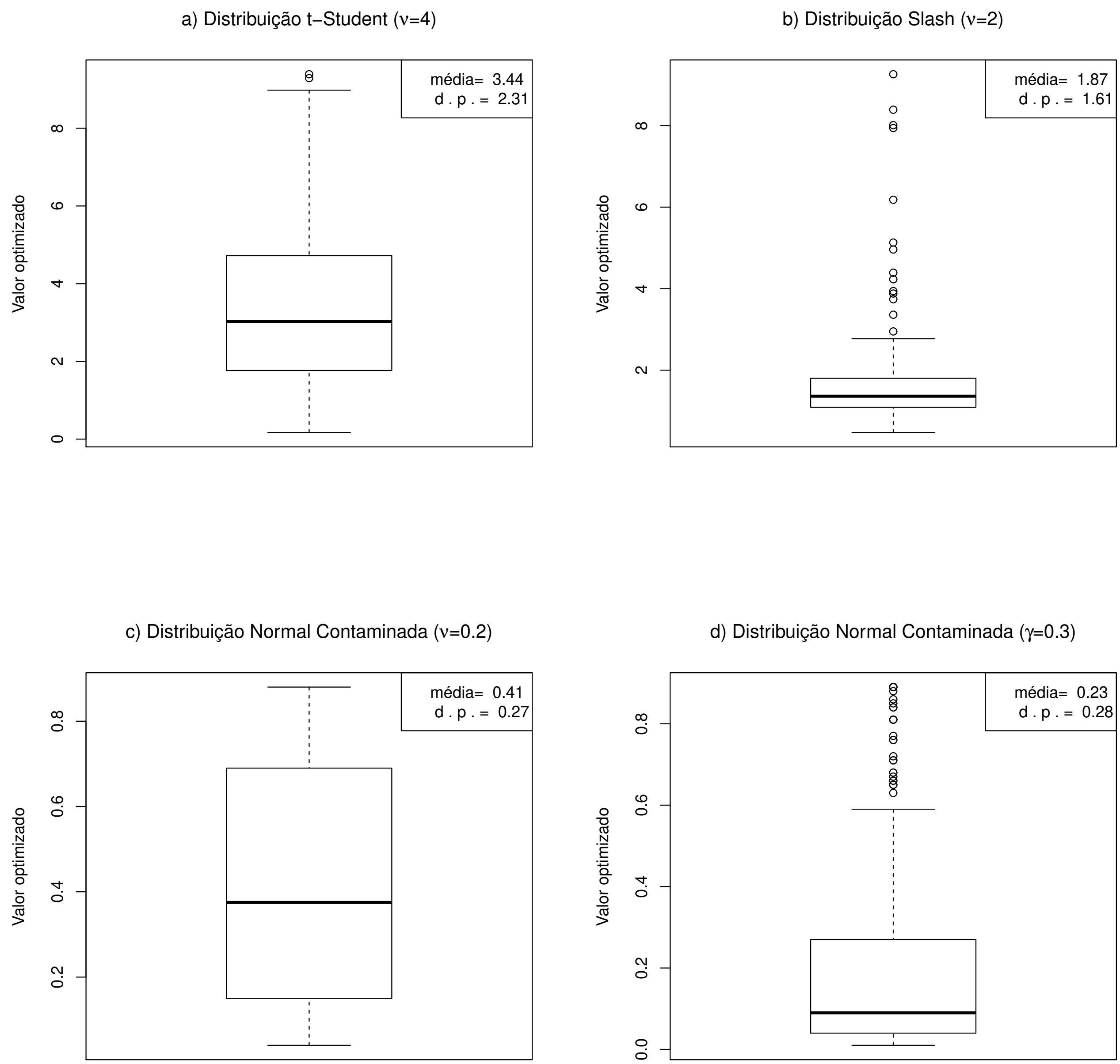


Figura 3.9: Resultado de simulações sobre otimização do parâmetro $\boldsymbol{\nu}$ das distribuições t de Student, Slash e Normal Contaminada no modelo de regressão linear com mudança nos coeficientes de regressão e variância.

Capítulo 4

Diagnóstico de Influência em modelos de regressão MEN com ponto de mudança

O método para avaliação de influência local para pequenas perturbações de um modelo estatístico proposto por Cook (1986) e a generalização deste para aplicação com dados incompletos proposto por Zhu & Lee (2001) têm sendo aplicados com muito sucesso em diversas áreas da Estatística, veja, por exemplo, Lachos (2002), Lachos et al. (2011) e Zeller (2009).

Neste capítulo, o método de influência local é estudado para os três modelos de regressão linear Mistura de Escala Normal com ponto de mudança apresentados no capítulo anterior. Três esquemas de perturbação serão apresentados.

(1) Perturbação por ponderação de casos, para detectar observações com influência no processo de estimação. Este esquema é equivalente a uma medida de influência por eliminação de casos (Osorio, 2006);

(2) Perturbação nas variáveis explicativas, para detectar possíveis maus condicionamentos entre algumas colunas da matriz $\mathbf{X}$;

(3) Perturbação na variável resposta, para detetar observações com influência sobre seus próprios valores previstos.

4.1 Diagnóstico de Influência

Um modelo estatístico procura sempre representar o fenômeno estudado da forma mais adequada. Desta forma, uma avaliação da estabilidade do modelo postulado deve ser realizado. Assim, um conjunto de técnicas estatísticas que identifiquem e avaliem observações que alterem substancialmente os resultados obtidos dentro do conjunto de dados analisado deve ser considerado dentro do processo de modelagem estatística. Estes são os denominados métodos de diagnóstico de influência.

Existem dois métodos desenvolvidos para a realização do diagnóstico de influência: o método de influência global, que permite identificar dados influentes mediante a eliminação de casos (Cook & Weisberg, 1982) e o método de influência local, que permite avaliar a influência quando pequenas perturbações no modelo estatístico ou nos dados usando uma medida de influência apropriada baseada no afastamento da verossimilhança (Cook, 1986).

Consideram-se dados influentes aqueles que controlam aspectos da análise estatística e produzem alterações relevantes nos resultados da análise quando a mesma for excluída (no caso da influência global) ou submetida a uma pequena perturbação (no caso da influência local). Finalmente, uma investigação completa de observações influentes é possível somente depois de que elas tenham sido identificadas.

4.2 Influência Local

Zhu & Lee (2001) propõem um método de avaliação de influência local para dados incompletos baseado no procedimento proposto por Cook (1986). Esta proposta utiliza a esperança condicional da função de log-verossimilhança dos dados completos obtidos para a implementação do algoritomo EM no lugar da função de log-verossimilhança dos dados observados. Os resultados analíticos apresentados são muito similares aos obtidos pelo procedimento clássico de influência local e permite avaliações de modelos mais complexos, para os quais a função de log-verossimilhança apresenta um tratamento algébraico complexo como é o caso dos modelos de regressão linear Mistura de Escala Normal com ponto de mudança. O método proposto por Zhu & Lee (2001) é descrito a seguir.

Considerando um vetor de perturbações $\boldsymbol{\omega} = (\omega_1, \omega_1, ..., \omega_g)^\top$ variando na região aberta $\Omega \in \mathbb{R}^g$. Seja $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|Y_c)$ a função de log-verossimilhança dos dados completos para o modelo perturbado. Assumindo que existe um $\boldsymbol{\omega}_0$, tal que $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0|Y_c) = \ell_c(\boldsymbol{\theta}|Y_c)$ para todo $\boldsymbol{\theta}$.

Considerando os resultados do algoritmo EM, propõe-se a função Q-afastamento

$$f_Q(\boldsymbol{\omega}) = 2[Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\widehat{\boldsymbol{\theta}})],$$

onde $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ é a estimativa de $\boldsymbol{\theta}$ que maximiza $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = E\{\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}|Y_o, \widehat{\boldsymbol{\theta}})\}$.

A função $f_Q(\boldsymbol{\omega})$ pode ser vista como uma medida de diferença entre $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ que é maior ou igual a zero e alcança seu mínimo global em $\boldsymbol{\omega}_0$. Quando nenhuma perturbação é introduzida $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)$ é igual a estimativa de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$.

Segundo Cook (1986) e Zhu & Lee (2001), o gráfico de influência é definido como $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$ e adicionalmente a curvatura normal $C_{f_Q,d}$ de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$ na direção do vetor unitário $\mathbf{d}$ pode ser usado para resumir o comportamento local de $f_Q(\boldsymbol{\omega})$. Esta é definida como

$$C_{f_Q,\mathbf{d}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{d} \quad \text{e} \quad -\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\omega}_0}^\top \left[-\ddot{Q}_\theta(\hat{\boldsymbol{\theta}})\right]^{-1} \Delta_{\boldsymbol{\omega}_0}, \tag{4.1}$$

$$\text{onde} \quad \ddot{Q}_{\theta}(\hat{\boldsymbol{\theta}}) = \left.\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{e} \quad \Delta_{\boldsymbol{\omega}} = \left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})}.$$

A curvatura normal baseada em $f_Q(\boldsymbol{\omega})$ pode ser entendida como uma generalização da curvatura normal proposta por Cook (1986) e também a expressão $-\ddot{Q}_{\boldsymbol{\omega}_0}$ torna-se fundamental para a deteção de observações localmente influentes. Por outro lado, Poon & Poon (1999) propõem um método de influência local baseado na curvatura normal conformalizada, que é uma medida padronizada da curvatura normal utilizada por Cook (1986). Logo, com base em (4.1), Zhu & Lee (2001) propõem a curvatura normal conformalizada $B_{f_Q,d}$ em $\boldsymbol{\omega}_0$ na direção do vetor unitário $\mathbf{d}$ e apresentam suas propriedades teóricas. Esta é definida como

$$B_{f_Q,\mathbf{d}} = \frac{-2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_0} \mathbf{d}}{\mathrm{tr}(-2\ddot{Q}_{\boldsymbol{\omega}_0})}. \tag{4.2}$$

A avaliação de casos influentes é baseada na inspeção visual do gráfico de $M(0)_l$ para $l = 1, ..., g$ versus o índice l. A obtenção de $M(0)_l$ é realizada a partir de $B_{f_Q,\mathbf{d}_l} = -2\mathbf{d}_l^\top \ddot{Q}_{\boldsymbol{\omega}_0} \mathbf{d}_l / \mathrm{tr}[-2\ddot{Q}_{\boldsymbol{\omega}_0}]$ onde $\mathbf{d}_l$ é um vetor de perturbação básica, tal que a l-ésima entrada é igual a 1 e as restantes iguais a 0.

Existem algumas alternativas para determinar os valores de referência ("*benchmark*") utilizados para avaliar se uma observação do conjunto de dados é potencialmente influente. Lee & Xu (2004) propõem o uso do valor $\overline{M(0)} + c^* SM(0)$ como o valor de referência, onde $\overline{M(0)}$ e $SM(0)$ são, respetivamente, a média e desvio padrão dos valores $M(0)_l$, para $l = 1, ..., g$ e c^* é uma constante positiva arbitrária.

Zhang & King (2005) e Zevallos et al. (2012) encontram o valor de referência usando simulações Monte Carlo. Assim, neste trabalho, são geradas s amostras simuladas com a mesma estrutura do modelo postulado. Para cada m-ésima amostra simulada, $m = 1, ..., s$, encontra-se o valor $2 \times \overline{M(0)}_m$ seguindo a sugestão de Verbeke & Molenberghs (1997), e finalmente o valor de referência é determinado pelo percentil 95 destes s valores.

4.3 Influência Local em MRL-MEN-PM nos Coeficientes de Regressão

4.3.1 Matriz Hessiana

Considerando os resultados apresentados na Subseção 3.5.3, temos que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = C - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\left[\sum_{i=1}^{k}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2}{\sigma^2}\right\} + \sum_{i=k+1}^{n}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2}{\sigma^2}\right\}\right],$$

onde $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \sigma^2)^\top$ e $q(d_i) = E[\kappa^{-1}(U_i)|y_i, \widehat{\boldsymbol{\theta}}]$. Logo, os elementos da matriz hessiana $\ddot{Q}_\theta(\hat{\boldsymbol{\theta}})$ são

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^\top} = -\frac{1}{\sigma^2} \sum_{i=1}^{k} \left\{ q(d_i)\mathbf{x}_i\mathbf{x}_i^\top \right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^\top} = -\frac{1}{\sigma^2} \sum_{i=k+1}^{n} \left\{ q(d_i)\mathbf{x}_i\mathbf{x}_i^\top \right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \left[\sum_{i=1}^{k} \left\{ q(d_i)(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)^2 \right\} + \sum_{i=k+1}^{n} \left\{ q(d_i)(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)^2 \right\} \right],$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^\top} = \mathbf{0},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_1 \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=1}^{k} \left\{ q(d_i)(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_1)\mathbf{x}_i \right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_2 \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=k+1}^{n} \left\{ q(d_i)(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_2)\mathbf{x}_i \right\}.$$

com $d_i = \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}_1)^2}{\sigma^2}$ para $i = 1, \ldots, k$ e $d_i = \dfrac{(y_i - \mathbf{x_i}^\top \boldsymbol{\beta}_2)^2}{\sigma^2}$ para $i = k+1, \ldots, n$.
A obtenção dos valores $q(d_i)$, $i = 1, \ldots, n$, para as quatro distribuições da classe de Mistura de Escala Normal estudadas nesta dissertação, estão apresentados em (2.10)-(2.13).

4.3.2 Esquemas de Perturbação

Considerando os três esquemas de perturbação estudados nesta dissertação, nesta subseção são apresentadas as matrizes $\Delta_{\boldsymbol{\omega}}$ necessárias para a avaliação da influência local. Para cada um dos esquemas de perturbação a matriz $\Delta_{\boldsymbol{\omega}_0}$ pode ser particionada como

$$\Delta_{\boldsymbol{\omega}_0} = (\Delta_{\boldsymbol{\beta}_1}^\top, \Delta_{\boldsymbol{\beta}_2}^\top, \Delta_{\sigma^2}^\top)^\top,$$

onde $\Delta_{\boldsymbol{\beta}_1} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{p\times g}$, $\Delta_{\boldsymbol{\beta}_2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{p\times g}$, e $\Delta_{\sigma^2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, g é a dimensão do vetor de perturbação $\boldsymbol{\omega}$, p e n o número de linhas e colunas da matriz $\mathbf{X}$, respetivamente.

1) **Ponderação de casos**

Atribuindo uma ponderação arbitrária para o valor esperado da função log-verossimilhança dos dados completos, a função Q perturbada, é possível capturar distorções em direções gerais, a qual é representada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}),$$

onde
$$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma^2 + q(d_i)\frac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_1)^2}{\sigma^2}], & para \quad i = 1, \dots, k, \\ -\frac{1}{2}[\log\sigma^2 + q(d_i)\frac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_2)^2}{\sigma^2}], & para \quad i = k+1, \dots, n. \end{cases}$$

Neste caso, o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (1, ..., 1)^\top$. Considerando o modelo com ponto de mudança na posição k, este esquema de perturbação pode ser reduzido a dois casos especiais

- $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{k} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + \sum_{i=k+1}^{n} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ e
- $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{k} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + \sum_{i=k+1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1 \partial\omega_i} = \begin{cases} \frac{1}{\sigma^2} q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)\mathbf{x}_i^\top & , para \quad i = 1, \dots, k, \\ 0 & , para \quad i = k+1, \dots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2 \partial\omega_i} = \begin{cases} 0 & , para \quad i = 1, \dots, k, \\ \frac{1}{\sigma^2} q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)\mathbf{x}_i^\top & , para \quad i = k+1, \dots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\sigma^2 \partial\omega_i} = \begin{cases} -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2 & , para \quad i = 1, \dots, k, \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2 & , para \quad i = k+1, \dots, n. \end{cases}$$

2) **Perturbação nas variáveis explicativas**

O interesse é perturbar uma variável explicativa contínua específica $\mathbf{x}_s, s = 1, ..., p$. Um esquema de perturbação aditivo é obtido fazendo $\mathbf{x}_{w_i s} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i = 1, \dots, n$ e onde S_{x_s} é um fator de escala que pode ser o desvio padrão de $\mathbf{x}_s$ e $\mathbf{1}_s$ é um vetor com 1 na posição s e 0 nas outras posições. Considerando o modelo com ponto de mudança na posição k dois casos especiais deste esquema de perturbação são de interesse

- $\mathbf{x}_{w_i s} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i = 1, \dots, k$ e
- $\mathbf{x}_{w_i s} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i = k+1, \dots, n$.

Neste esquema, a função Q perturbada é dada por
$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}}),$$

onde
$$Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma^2 + q(d_i)\frac{(y_i - \mathbf{x}_{w_i s}^\top\boldsymbol{\beta}_1)^2}{\sigma^2}], & para \quad i = 1, \dots, k, \\ -\frac{1}{2}[\log\sigma^2 + q(d_i)\frac{(y_i - \mathbf{x}_{w_i s}^\top\boldsymbol{\beta}_2)^2}{\sigma^2}], & para \quad i = k+1, \dots, n. \end{cases}$$
e o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (0, ..., 0)^\top$

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1 \partial\omega_i} = \begin{cases} \frac{S_{x_s}}{\sigma^2} q(d_i)(y_i - 2\mathbf{x}_{w_i s}^\top\boldsymbol{\beta}_1))\mathbf{1}_s^\top & , para \quad i = 1, \dots, k, \\ 0 & , para \quad i = k+1, \dots, n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\omega_i} = \begin{cases} 0 & ,para \quad i=1,\ldots,k, \\ \frac{S_{x_s}}{\sigma^2}q(d_i)(y_i - 2\mathbf{x}_{w_is}^\top\boldsymbol{\beta}_2))\mathbf{1}_s^\top & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = \begin{cases} -\frac{S_{x_s}}{\sigma^4}q(d_i)(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta}_1)\boldsymbol{\beta}_1^\top\mathbf{1}_s & ,para \quad i=1,\ldots,k, \\ -\frac{S_{x_s}}{\sigma^4}q(d_i)(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta}_2)\boldsymbol{\beta}_2^\top\mathbf{1}_s & ,para \quad i=k+1,\ldots,n. \end{cases}$$

3) **Ponderação na variável resposta**

O interesse é perturbar a variável resposta $y_i, i = 1, ..., n$. Um esquema de perturbação é obtido fazendo $y_{w_i} = y_i + w_i S_y$, para $i = 1, \ldots, n$, onde S_y é um fator de escala que pode ser o desvio padrão de y. Considerando o modelo com ponto de mudança na posição k dois casos especiais deste esquema de perturbação são de interesse:

- $y_{w_i} = y_i + w_i S_y$, para $i = 1, \ldots, k$ e
- $y_{w_i} = y_i + w_i S_y$, para $i = k+1, \ldots, n$.

Neste esquema, a função Q perturbada é dada por

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}}),$$

onde

$$Q_i(\boldsymbol{\theta},\boldsymbol{\omega}_i|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma^2 + q(d_i)\frac{(y_{w_i} - \mathbf{x_i}^\top\boldsymbol{\beta}_1)^2}{\sigma^2}], & para \quad i=1,\ldots,k, \\ -\frac{1}{2}[\log\sigma^2 + q(d_i)\frac{(y_{w_i} - \mathbf{x_i}^\top\boldsymbol{\beta}_2)^2}{\sigma^2}], & para \quad i=k+1,\ldots,n. \end{cases}$$

e o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (0, ..., 0)^\top$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\omega_i} = \begin{cases} \frac{S_y}{\sigma^2}q(d_i)\mathbf{x}_i^\top & ,para \quad i=1,\ldots,k, \\ 0 & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\omega_i} = \begin{cases} 0 & ,para \quad i=1,\ldots,k, \\ \frac{S_y}{\sigma^2}q(d_i)\mathbf{x}_i^\top & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = \begin{cases} \frac{S_y}{\sigma^4}q(d_i)(y_{w_i} - \mathbf{x}_i^\top\boldsymbol{\beta}_1) & ,para \quad i=1,\ldots,k, \\ \frac{S_y}{\sigma^4}q(d_i)(y_{w_i} - \mathbf{x}_i^\top\boldsymbol{\beta}_2) & ,para \quad i=k+1,\ldots,n. \end{cases}$$

4.3.3 Estudos de Simulação

Nesta subseção, apresenta-se alguns resultados de diversas simulações realizadas para avaliar o desempenho dos métodos de influência propostos e baseados em Zeller (2009). No primeiro estudo de simulação avalia-se a influência de "*outliers*"no método de estimação de $\boldsymbol{\theta}$. O segundo estudo de simulação avalia a capacidade da metodologia de detecção de dados atípicos e a sensibilidade das estimativas de máxima verossimilhança às observações extremas sob as diferentes distribuições de Mistura de Escala Normal estudadas nesta dissertação. Serão considerados para este estudo os esquemas de ponderação de casos, perturbação na variável explicativa e perturbação de variável resposta e as duas alternativas para determinação dos valores de referência ("*benchmark*") descritas na Seção 4.2. Todas as amostras simuladas são geradas desde o modelo de regressão com ponto de mudança nos coeficientes de regressão

$$\begin{aligned} y_i &= 2 + 5x_i + \epsilon_i, \quad \textit{para} \quad i = 1, \ldots, k, \\ y_i &= 4 + 10x_i + \epsilon_i, \quad \textit{para} \quad i = k+1, \ldots, n, \end{aligned}$$

considerando $\sigma^2 = 1$ e $x_i \sim U(0,1)$ para $i = 1, \ldots, n$.

1) **Influência de "*outliers*"no método de estimação**

Seguindo o procedimento sugerido por Pinheiro et al. (2001), a flexibilidade dos modelos de regressão Mistura de Escala Normal com ponto de mudança na observação k é estudada considerando duas situações. A primeira situação considera a influência de uma única observação aberrante situada antes da observação k na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, e a segunda situação considera duas observações aberrantes, uma antes do ponto k e a outra depois do ponto k na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$.

Sem perda de generalidade, é simulado um conjunto de dados de tamanho $n = 40$ com um ponto de mudança na observação $k = 20$ a partir do modelo de regressão com ponto de mudança nos coeficientes de regressão, proposto no início desta subseção, e considerando uma distribuição t de Student com $\nu = 3$. Para esta amostra se ajustam modelos de regressão Mistura de Escala Normal considerando as distribuições Normal, t de Student, Slash e Normal Contaminada.

Considerando que $\Im$ é um valor de contaminação, temos que uma observação i contaminada é obtida como $y_i(\Im) = y_i + \Im$. Assim, na primeira situação contaminamos a observação 10 e na segunda situação contaminamos simultaneamente as observações 10 e 30. Para valores $\Im$ entre -10 e 10 avaliam-se as mudanças relativas dos parâmetros do modelo, isto é, o percentual de mudança dos parâmetros para os diferentes valores de $\Im$. Nas Figuras 4.1 e 4.2 se apresentam os resultados das mudanças relativas das estimativas dos parâmetros do modelo proposto sob distribuição Normal, t de Student, Slash e Normal contaminada para os diferentes valores de contaminação $\Im$ considerados, para a simulação nas duas situações avaliadas.

O percentual de mudança é obtido como $100 \times ((\hat{\theta}_i(\Im) - \hat{\theta}_i)/\hat{\theta}_i)$ onde $\hat{\theta}_i$ denota a estimativa original do i-ésimo parâmetro do modelo considerado e $\hat{\theta}_i(\Im)$ denota a estimativa do i-ésimo parâmetro considerando a contaminação com o valor $\Im$.

Observa-se que em ambas situações as estimativas que consideram caudas mais pesadas (t de Student, Slash e Normal Contaminada) são menos afetadas pelas variações dos valores de contaminação $\Im$ em relação a modelo que considera a distribuição Normal. Por outro lado, na primeira situação a estimativa dos parâmetros β_{02} e β_{12} não é afetada pela presença da observação aberrante na posição 10, isto porque na estimação destes parâmetros somente são considerados

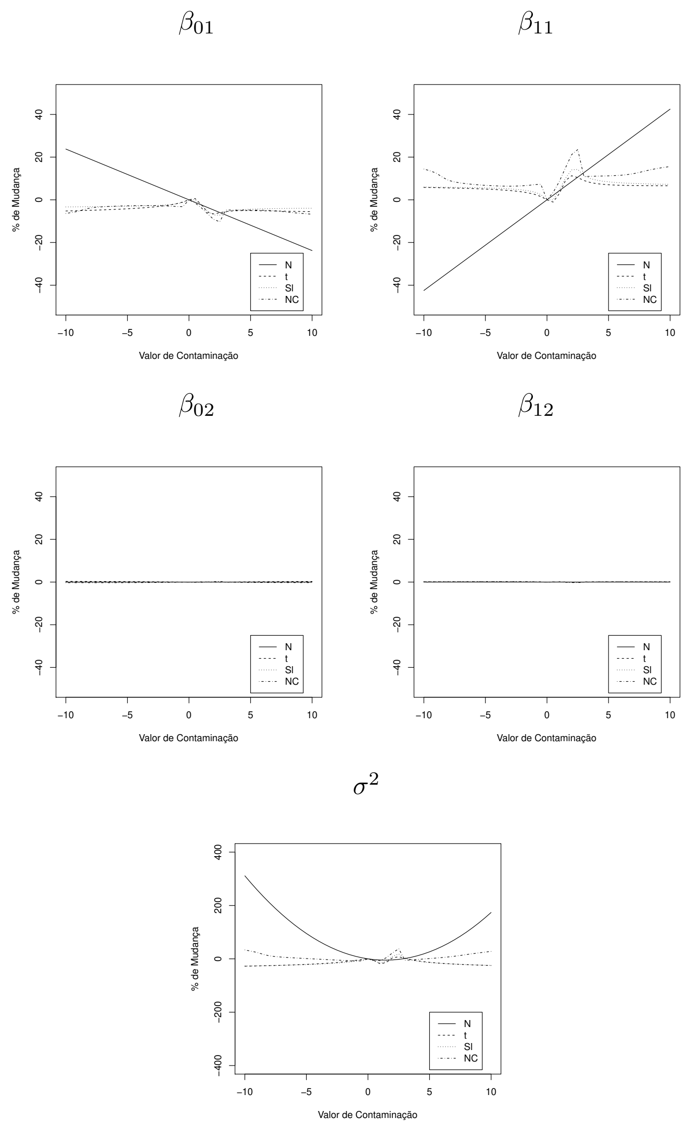


Figura 4.1: Mudanças relativas nas estimativas dos parâmetros do modelo de regressão linear Mistura de Escala normal com ponto de mudança nos coeficientes de regressão considerando distribuições Normal, t de Student, Slash e Normal Contaminada para diferentes valores de contaminação $\Im$ na observação 10.

os dados nas posições acima de $k = 20$, já na segunda situação considerada a estimativa destes parâmetros apresentam variação devido à presença da observação aberrante 30. Finalmente, quando o modelo considera a distribuição Normal, a variância σ^2 apresenta maior variação a medida que existam mais dados aberrantes.

2) **Esquemas de perturbações**

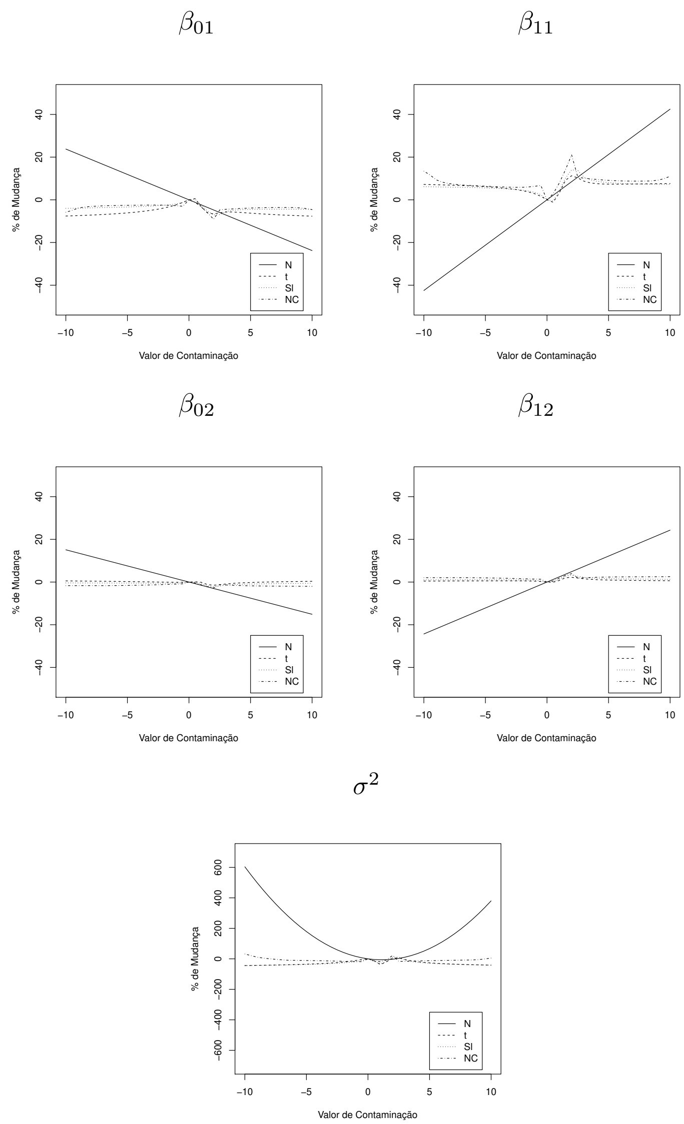


Figura 4.2: Mudanças relativas nas estimativas dos parâmetros do modelo de regressão linear Mistura de Escala normal com ponto de mudança nos coeficientes de regressão considerando distribuições Normal, t de Student, Slash e Normal Contaminada para diferentes valores de contaminação $\Im$ nas observações 10 e 30.

Seguindo a sugestão de Ortega et al. (2003), a perturbação da amostra simulada se realiza considerando $x_{max_1} \leftarrow x_{max_1} + cte \times (\mathbf{x}_1^\top \mathbf{x}_1)^{1/2}$ e $x_{max_2} \leftarrow x_{max_2} + cte \times (\mathbf{x}_2^\top \mathbf{x}_2)^{1/2}$ simultaneamente, onde $\mathbf{x}_1 = (x_1, ..., x_k)^\top$, $\mathbf{x}_2 = (x_{k+1}, ..., x_n)^\top$, x_{max_1} e x_{max_2} são os valores máximos dos vetores $\mathbf{x}_1$ e $\mathbf{x}_2$ respetivamente e cte é uma constante positiva.

Adicionalmente, sabendo que p_1 e p_2 são as posições das observações x_{max_1} e x_{max_2} respetivamente, considera-se uma segunda situação onde a variável resposta é perturbada nas posições

p_1 e p_2 por $y_{p_1} \leftarrow y_{p_1} + cte \times (\mathbf{y}_1^\top \mathbf{y}_1)^{1/2}$ e $y_{p_2} \leftarrow y_{p_2} + cte \times (\mathbf{y}_2^\top \mathbf{y}_2)^{1/2}$ simultaneamente, onde $\mathbf{y}_1 = (y_1, ..., x_k)^\top$ e $\mathbf{y}_2 = (y_{k+1}, ..., y_n)^\top$.

Primeiramente, são simuladas 400 amostras de tamanho $n = 40$ com um ponto de mudança na observação $k = 20$ a partir do modelo de regressão com ponto de mudança nos coeficientes de regressão, proposto no início desta subseção. As primeiras 100 amostras sob distribuição Normal, as seguintes 100 sob distribuição t de Student com $\nu = 3$, as seguintes 100 sob distribuição Slash com $\nu = 2$ e finalmente as últimas 100 amostras sob distribuição Normal Contaminada com $(\nu = 0.3, \gamma = 0.1)$.

Cada uma destas amostras simuladas é perturbada de acordo as duas situações (perturbação na variável explicativa e perturbação na variável resposta), descritas nos parágrafos anteriores, considerando $cte = 0.1$ e avalia-se a aplicação da metodologia de diagnóstico de influência para os esquemas de perturbação estudados. Em todas as situações, utilizamos os dois valores de referência ("*benchmark*") descritos na Seção 4.2. Para a proposta de Lee & Xu (2004), considera-se $c* = 3$ e para a obtenção do valor de referência via simulações Monte Carlo, considera-se $s = 1000$. Adicionalmente, para cada esquema de perturbação avalia-se o caso geral que calcula um único valor de referência para $i = 1, \ldots, n$ e o caso especial apresentado quando a mudança na posição k é conhecida que determina doia valores de referência, a primera considerando as observações até o ponto de mudança e a segunda considerando as observações após ponto de mudança.

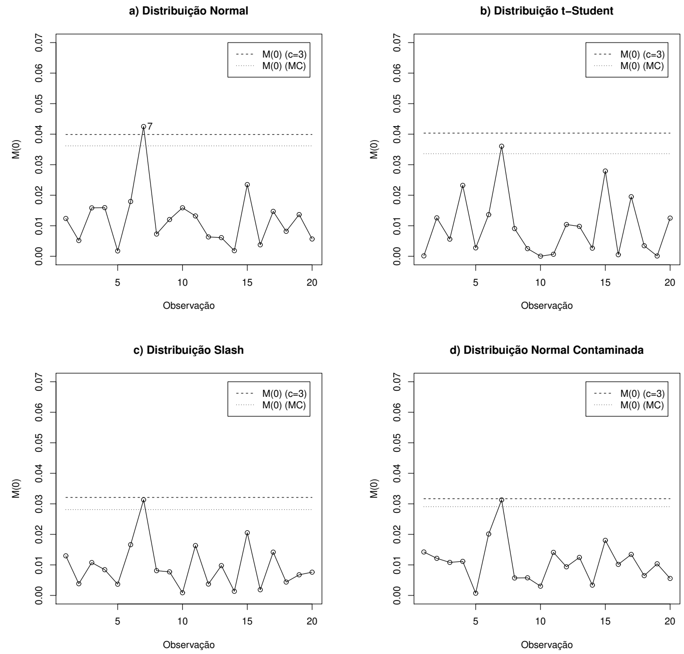
A Tabela 4.1 mostra o resultado obtidos para os diversos aspectos avaliados na aplicação da metodologia de diagóstico de influência. Primeiramente, a maior frequência observada quando se considera distribuição Normal mostra a sensibilidade das estimativas de máxima verossimilhança dos parâmetros na presença de dados atípicos neste caso enquanto que o uso de distribuições com caudas pesadas apresenta maior flexibilidade pois as frequências são menores em todos os casos. Por outro lado, ao observar as frequências considerando a distribuição Normal que o valor de referência obtido via simulações Monte Carlo é menos restrito que a proposta de Lee & Xu (2004), no sentido que este indica maior quantidade de observações como influentes e finalmente, o esquema especial quando a mudança no posição k é conhecida apresenta maior frequência de deteção de observações influentes. O valor de referência proposto por Lee & Xu (2004) parece ser uma alternativa computacional apropriada devido ao tempo computacional, assim adotaremos esta marca de referência para a avaliação dos conjuntos de dados reais.

Avaliações adicionais e as próprias formulações matemáticas consideradas para a avaliação de influência indicam que os valores dos parâmetros são importantes pelo qual sugire-se como uma boa prática para avaliação de modelos com ponto de mudança o uso do esquema especial de perturbação proposto.

Finalmente, considerando uma única amostra de tamanho $n = 40$ obtida desde $x_i \sim U(0, 1)$ para $i = 1, \ldots, n$ se geram quatro conjuntos de dados que seguem o modelo de regressão com ponto de mudança nos coeficientes de regressão proposto no inicio desta subseção e que consideram distribuição Normal, distribuição t de Student com $\nu = 3$, distribuição Slash com $\nu = 2$ e distribuição Normal Contaminada com $(\nu = 0.3, \gamma = 0.1)$, respetivamente. Estes quatro conjuntos de dados são perturbados de acordo as duas situações (perturbação de $\mathbf{x}$ e perturbação de $\mathbf{y}$) descritas considerando $cte = 0.1$ e avalia-se novamenete a aplicação da metodologia de diagóstico de influência.

As Figuras 4.3-4.4 mostram a influência das observações 7 e 29, que correspondem justamente às observações perturbadas. Observa-se novamente a sensibilidade das estimativas de máxima verossimilhança dos parâmetros na presença de dados atípicos quando é considerada a distribuição Normal enquanto que o uso de distribuições com caudas pesadas apresenta maior flexibilidade. Além disso, nota-se que o valor de referência proposto por Lee & Xu (2004) é mais restrito no sentido de identificação de observações aberrantes que o obtido via simulações Monte Carlo. Para a perturbação de casos, os resultados são similares. Eles não são mostrados aqui por brevidade.

Primeiro esquema especial de perturbação



Segundo esquema especial de perturbação

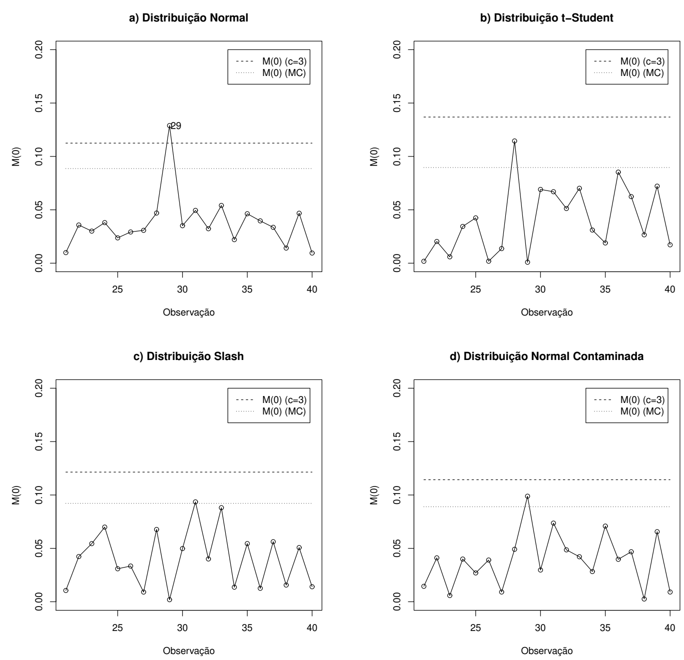
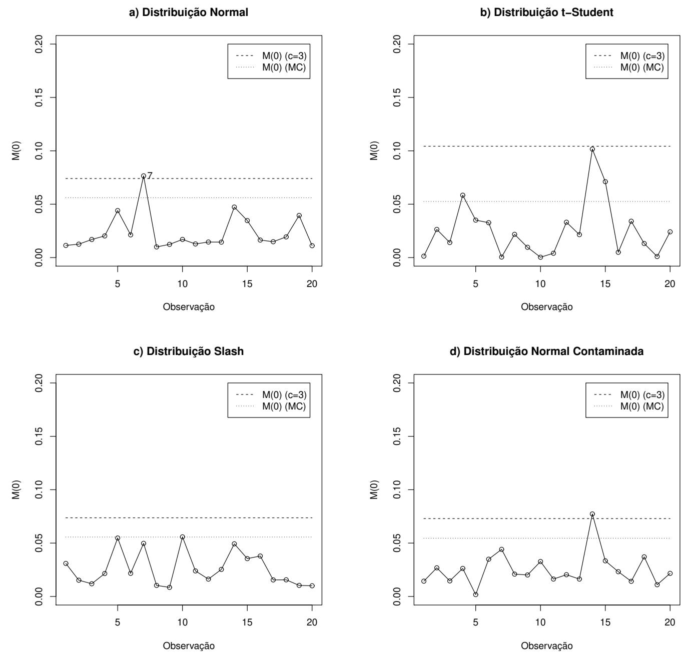


Figura 4.3: Gráfico de $M(0)$ para perturbação da variável explicativa no modelo de regressão linear com mudança nos coeficientes de regressão, considerando marcas de referência proposta por Lee & Xu (2004), onde $c^{*} = 3$ e via simulações Monte Carlo sob diversas distribuições que conformam a classe de Mistura de Escala Normal.

Primeiro esquema especial de perturbação



Segundo esquema especial de perturbação

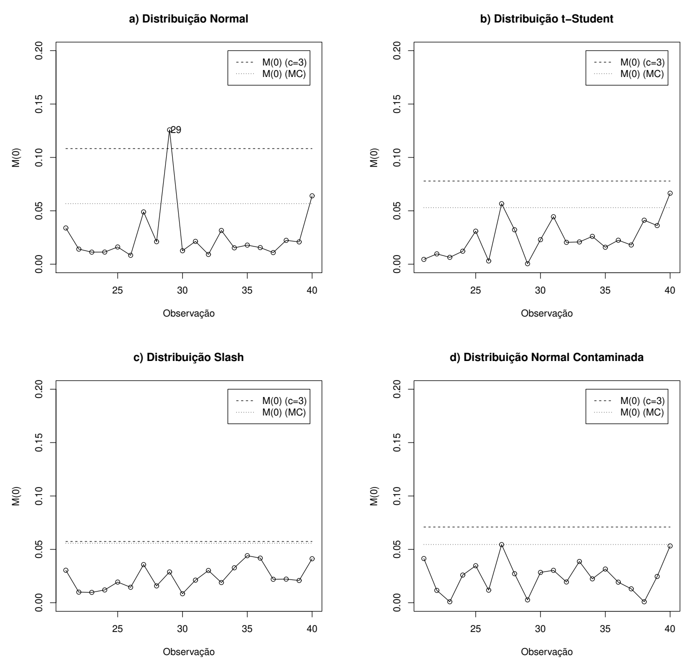


Figura 4.4: Gráfico de $M(0)$ para perturbação da variável resposta no modelo de regressão linear com mudança nos coeficientes de regressão, considerando marcas de referência proposta por Lee & Xu (2004) onde $c^* = 3$ e e via simulações Monte Carlo sob diversas distribuições que conformam a classe de Mistura de Escala Normal.

Tabela 4.1: Resultados para diagnóstico de influência no MRL com mudança nos coeficientes de regressão considerando 100 amostras simuladas para cada uma das distribuições de Mistura de Escala Normal estudadas: Normal (N), t de Student (t), Slash (Sl) e Normal Contaminada (NC).

Observação Perturbada	Esquema para M(0)	Esquema de Perturbação	Distribuições N	t	SL	NC
Valor de referência M(0) de Lee & Xu (2004), onde $c^* = 3$						
Perturbação da amostra sobre variável explicativa						
Até o ponto de mudança	Geral	Ponderação de casos	0	0	0	0
		Variável explicativa	0	0	0	0
	Especial	Ponderação de casos	99	57	89	76
		Variável explicativa	46	21	35	22
Após ponto de mudança	Geral	Ponderação de casos	100	40	87	69
		Variável explicativa	100	33	83	64
	Especial	Ponderação de casos	100	30	83	63
		Variável explicativa	77	17	48	24
Perturbação da amostra sobre variável resposta						
Até o ponto de mudança	Geral	Ponderação de casos	0	0	0	0
		Variável resposta	8	0	2	2
	Especial	Ponderação de casos	8	1	3	3
		Variável resposta	76	1	2	3
Após ponto de mudança	Geral	Ponderação de casos	65	0	2	1
		Variável resposta	100	0	2	1
	Especial	Ponderação de casos	10	0	1	0
		Variável resposta	100	0	2	0
Valor de referência M(0) obtida via simulação Monte Carlo						
Observação Perturbada	Esquema para M(0)	Esquema de Perturbação	Distribuições N	t	SL	NC
Perturbação da amostra sobre variável explicativa						
Até o ponto de mudança	Geral	Ponderação de casos	2	0	4	4
		Variável explicativa	0	0	4	3
	Especial	Ponderação de casos	71	1	40	26
		Variável explicativa	8	0	11	7
Após ponto de mudança	Geral	Ponderação de casos	100	0	4	2
		Variável explicativa	100	0	2	2
	Especial	Ponderação de casos	100	0	2	2
		Variável explicativa	100	0	2	1
Perturbação da amostra sobre variável resposta						
Até o ponto de mudança	Geral	Ponderação de casos	0	0	0	0
		Variável resposta	88	1	9	19
	Especial	Ponderação de casos	17	0	16	10
		Variável resposta	84	0	4	15
Após ponto de mudança	Geral	Ponderação de casos	100	0	2	2
		Variável resposta	100	0	2	1
	Especial	Ponderação de casos	99	0	2	1
		Variável resposta	100	0	2	1

4.4 Influência Local em MRL-MEN-PM na Variância

4.4.1 Matriz Hessiana

Considerando os resultados apresentados na Subseção 3.6.3 temos que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = C - \frac{k}{2}\log\sigma_1^2 - \frac{n-k}{2}\log\sigma_2^2 - \frac{1}{2}\left[\sum_{i=1}^{k}\left\{q(d_i)\frac{(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2}{\sigma_1^2}\right\} + \sum_{i=k+1}^{n}\left\{q(d_i)\frac{(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2}{\sigma_2^2}\right\}\right],$$

onde $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma_1^2, \sigma_2^2)^\top$ e $q(d_i) = E[\kappa^{-1}(U_i)|y_i, \widehat{\boldsymbol{\theta}}]$. Logo, os elementos da matriz hessiana $\ddot{Q}_\theta(\hat{\boldsymbol{\theta}})$ são

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\frac{1}{\sigma_1^2}\sum_{i=1}^{k}\left\{q(d_i)\mathbf{x}_i\mathbf{x}_i^\top\right\} - \frac{1}{\sigma_2^2}\sum_{i=k+1}^{n}\left\{q(d_i)\mathbf{x}_i\mathbf{x}_i^\top\right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\sigma_1^2} = \frac{k}{2\sigma_1^4} - \frac{1}{\sigma_1^6}\sum_{i=1}^{k}\left\{q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2\right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\sigma_2^2} = \frac{n-k}{2\sigma_2^4} - \frac{1}{\sigma_2^6}\sum_{i=k+1}^{n}\left\{q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2\right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma_1^2} = -\frac{1}{\sigma_1^4}\sum_{i=1}^{k}\left\{q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})\mathbf{x}_i\right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma_2^2} = -\frac{1}{\sigma_2^4}\sum_{i=k+1}^{n}\left\{q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})\mathbf{x}_i\right\},$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\sigma_2^2} = \mathbf{0}.$$

com $d_i = \dfrac{(y_i-\mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma_1^2}$, para $i = 1,\ldots,k$ e $d_i = \dfrac{(y_i-\mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma_2^2}$ para $i = k+1,\ldots,n$.
A obtenção dos valores $q(d_i)$, $i = 1,\ldots,n$, para as quatro distribuições da classe de Mistura de Escala Normal estudadas nesta dissertação, estão apresentados em (2.10)-(2.13).

4.4.2 Esquemas de Perturbação

Considerando os três esquemas de perturbação estudados nesta dissertação, nesta subseção são apresentadas as matrices $\Delta_{\boldsymbol{\omega}}$ necessárias para a avaliação da influência local. Para cada um dos esquemas de perturbação a matriz $\Delta_{\boldsymbol{\omega}_0}$ pode ser particionada como

$$\Delta_{\boldsymbol{\omega}_0} = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_{\sigma_1^2}, \Delta_{\sigma_2^2})^\top,$$

onde $\Delta_{\boldsymbol{\beta}} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{p\times g}$, $\Delta_{\sigma_1^2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, e $\Delta_{\sigma_2^2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, g é a dimensão do vetor de perturbação $\boldsymbol{\omega}$, p e n o número de linhas e colunas da matriz $\mathbf{X}$, respetivamente.

1) **Ponderação de casos**
Atribuindo uma ponderação arbitrária para o valor esperado da função log-verossimilhança dos dados completos, a função Q perturbada, é possível capturar distorções em direções gerais, a qual é representada por

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}),$$

onde $$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \begin{cases} -\dfrac{1}{2}[\log\sigma_1^2 + q(d_i)\dfrac{(y_i-\mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma^2}], & para \quad i=1,\dots,k, \\ -\dfrac{1}{2}[\log\sigma_2^2 + q(d_i)\dfrac{(y_i-\mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma^2}], & para \quad i=k+1,\dots,n. \end{cases}$$

Neste caso, o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (1,...,1)^\top$. Considerando o modelo com ponto de mudança na posição k, este esquema de perturbação se reduz a dois casos especiais

- $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{k}\omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + \sum_{i=k+1}^{n} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ e
- $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{k} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + \sum_{i=k+1}^{n}\omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \begin{cases} \dfrac{1}{\sigma_1^2} q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})\mathbf{x}_i^\top & ,para \quad i=1,\dots,k, \\ \dfrac{1}{\sigma_2^2} q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})\mathbf{x}_i^\top & ,para \quad i=k+1,\dots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\omega_i} = \begin{cases} -\dfrac{1}{2\sigma_1^2} + \dfrac{1}{2\sigma_1^4} q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2 & ,para \quad i=1,\dots,k, \\ 0 & ,para \quad i=k+1,\dots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\omega_i} = \begin{cases} 0 & ,para \quad i=1,\dots,k, \\ -\dfrac{1}{2\sigma_2^2} + \dfrac{1}{2\sigma_2^4} q(d_i)(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2 & ,para \quad i=k+1,\dots,n. \end{cases}$$

2) **Perturbação nas variáveis explicativas**
O interesse é perturbar uma variável explicativa continua especifica $\mathbf{x}_s, s=1,...,p$. Um esquema de perturbação aditivo é obtido fazendo $\mathbf{x}_{w_is} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i=1,\dots,n$ e onde S_{x_s} é um fator de escala que pode ser o desvio padrão de $\mathbf{x}_s$ e $\mathbf{1}_s$ é um vetor com 1 na posição s e 0 nas outras posições. Considerando o modelo com ponto de mudança na posição k dois casos especiais deste esquema de perturbação são de interesse

- $\mathbf{x}_{w_is} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i=1,\dots,k$ e
- $\mathbf{x}_{w_is} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i=k+1,\dots,n$.

Neste esquema, a função Q perturbada é dada por

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}}),$$

onde $$Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma_1^2 + q(d_i)\frac{(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta})^2}{\sigma_1^2}], & para \quad i = 1, \ldots, k, \\ -\frac{1}{2}[\log\sigma_2^2 + q(d_i)\frac{(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta})^2}{\sigma_2^2}], & para \quad i = k+1, \ldots, n. \end{cases}$$
e o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (0, ..., 0)^\top$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \begin{cases} \frac{S_{x_s}}{\sigma_1^2} q(d_i)(y_i - 2\mathbf{x}_{w_is}^\top\boldsymbol{\beta}))\mathbf{1}_s^\top & ,para \quad i = 1, \ldots, k, \\ \frac{S_{x_s}}{\sigma_2^2} q(d_i)(y_i - 2\mathbf{x}_{w_is}^\top\boldsymbol{\beta}))\mathbf{1}_s^\top & ,para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\omega_i} = \begin{cases} -\frac{S_{x_s}}{\sigma_1^4} q(d_i)(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta})\boldsymbol{\beta}^\top\mathbf{1}_s & ,para \quad i = 1, \ldots, k, \\ 0 & ,para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\omega_i} = \begin{cases} 0 & ,para \quad i = 1, \ldots, k, \\ -\frac{S_{x_s}}{\sigma_2^4} q(d_i)(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta})\boldsymbol{\beta}^\top\mathbf{1}_s & ,para \quad i = k+1, \ldots, n. \end{cases}$$

3) **Ponderação na variável resposta**

O interesse é perturbar a variável resposta $y_i, i = 1, ..., n$. Um esquema de perturbação é obtido fazendo $y_{w_i} = y_i + w_i S_y$, para $i = 1, \ldots, n$, onde S_y é um fator de escala que pode ser o desvio padrão de y. Considerando o modelo com ponto de mudança na posição k dois casos especiais deste esquema de perturbação são de interesse

- $y_{w_i} = y_i + w_i S_y$, para $i = 1, \ldots, k$ e
- $y_{w_i} = y_i + w_i S_y$, para $i = k+1, \ldots, n$.

Neste esquema, a função Q perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}}),$$

onde $$Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma_1^2 + q(d_i)\frac{(y_{w_i} - \mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma_1^2}], & para \quad i = 1, \ldots, k, \\ -\frac{1}{2}[\log\sigma_2^2 + q(d_i)\frac{(y_{w_i} - \mathbf{x_i}^\top\boldsymbol{\beta})^2}{\sigma_2^2}], & para \quad i = k+1, \ldots, n. \end{cases}$$
e o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (0, ..., 0)^\top$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \begin{cases} \frac{S_y}{\sigma_1^2} q(d_i)\mathbf{x}_i^\top & ,para \quad i = 1, \ldots, k, \\ \frac{S_y}{\sigma_2^2} q(d_i)\mathbf{x}_i^\top & ,para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\omega_i} = \begin{cases} \frac{S_y}{\sigma_1^4} q(d_i)(y_{w_i} - \mathbf{x}_i^\top\boldsymbol{\beta}) & ,para \quad i = 1, \ldots, k, \\ 0 & ,para \quad i = k+1, \ldots, n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\omega_i} = \begin{cases} 0 & ,para \quad i=1,\ldots,k, \\ \dfrac{S_y}{\sigma_2^4}q(d_i)(y_{w_i}-\mathbf{x}_i^\top\boldsymbol{\beta}) & ,para \quad i=k+1,\ldots,n. \end{cases}$$

4.4.3 Estudos de Simulação

Nesta subseção apresenta-se resultados das simulações realizadas para avaliar o desempenho dos métodos de influência em modelos de regressão com ponto de mudança na variância, os dois estudos de simulação apresentados realizam as mesmas avaliações descritas na Subseção 4.3.3. Neste caso, todas as amostras simuladas são geradas desde o modelo de regressão com ponto de mudança na variância

$$y_i = 4 + 10x_i + \epsilon_i, \quad para \quad i=1,\ldots,n,$$

e considerando $\sigma_1^2 = 1$ para $i=1,\ldots,k$, $\sigma_2^2 = 3$ para $i=k+1,\ldots,n$ e $x_i \sim U(0,1)$ para $i=1,\ldots,n$.

1) **Influência de "*outliers*"no método de estimação**

O mesmo procedimento descrito na Subseção 4.3.3 é utilizado nesta simulação, desta vez considerando o modelo de regressão com ponto de mudança na variância proposto no inicio desta subseção. Nas Figuras 4.5 e 4.6 se apresentam os resultados das mudanças relativas das estimativas dos parâmetros do modelo proposto sob distribuição Normal, t de Student, Slash e Normal Contaminada para os diferentes valores de contaminação $\Im$ considerados para a simulação nas duas situações avaliadas. Observa-se que em ambas situações as estimativas que consideram caudas mais pesadas (t de Student, Slash e Normal Contaminada) são menos afetadas pelas variações dos valores de contaminação $\Im$ em relação a modelo que considera a distribuição Normal, também que os parâmetros β_0 e β_1 são afetados pela presença das observações aberrantes, isto devido a que as estimação de estes parâmetros está em função das variâncias. Na primeira situação, as estimativas de σ_2^2 apresentam pequenas variações explicadas pelo método de estimação que inclui uma optimização nos graus de liberdade das distribuições com caudas pesadas. Na segunda situação, os parâmetros σ_1^2 e σ_2^2 apresentam variação devido à presença das observações aberrante 10 e 30.

2) **Esquemas de perturbação**

O procedimento descrito na Subseção 4.3.3 para este estudo de simulação é utilizado, desta vez considerando o modelo de regressão com ponto de mudança na variância proposto no inicio desta subseção. O número de amostras simuladas, os valores de n e k, os valores para os graus de liberdade das distribuições, assim como para cte, $c*$ e s são exatamente os mesmo que os considerados nas simulações correspondentes da Subseção 4.3.3.

Tanto a Tabela 4.2 quanto os as Figuras 4.7-4.8 mostram os resultados das simulações para

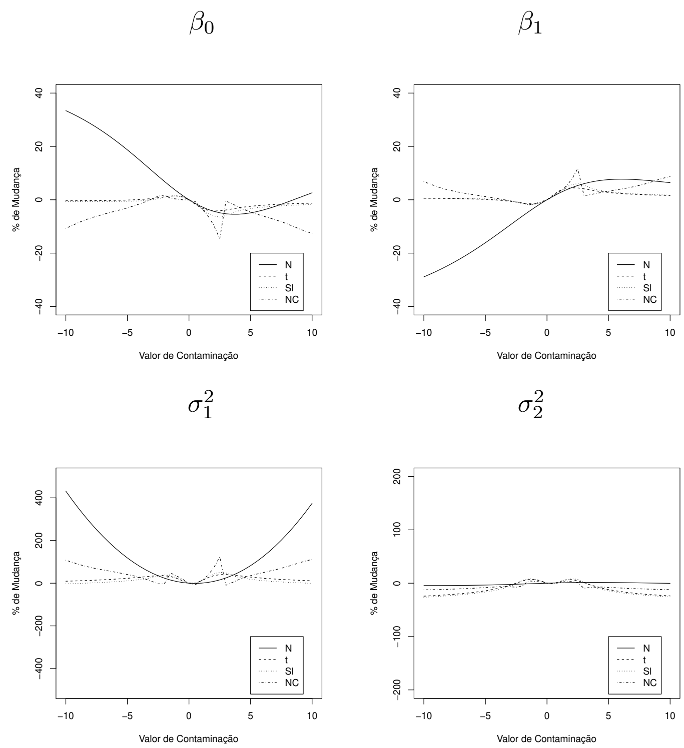


Figura 4.5: Mudanças relativas nas estimativas dos parâmetros do modelo de regressão linear Mistura de Escala normal com ponto de mudança na variância considerando distribuições Normal, t de Student, Slash e Normal Contaminada para diferentes valores de contaminação $\Im$ na observação 10.

o modelo de regressão com ponto de mudança na variância. Estes resultados são similares aos obtidos na Subseção 4.3.3, portanto as conclusões apresentadas anteriormente são válidas também para modelos de regressão com ponto de mudança na variância. As Figuras 4.7-4.8 mostram desta vez a influência das observações 18 e 34, que são às observações perturbadas nesta simulação.

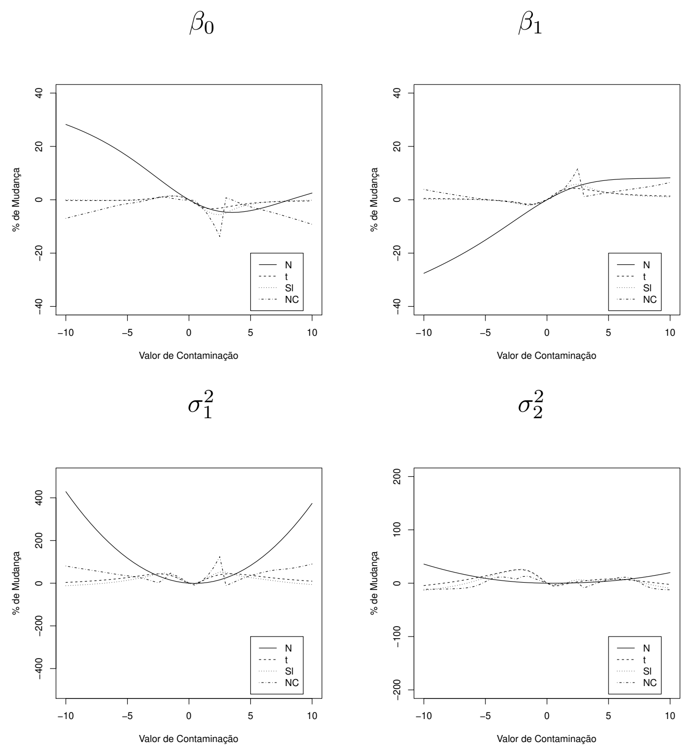


Figura 4.6: Mudanças relativas nas estimativas dos parâmetros do modelo de regressão linear Mistura de Escala normal com ponto de mudança na variância considerando distribuições Normal, t de Student, Slash e Normal Contaminada para diferentes valores de contaminação $\Im$ nas observações 10 e 30.

4.5 Influência Local em MRL-MEN-PM nos Coeficientes de Regressão e Variância

4.5.1 Matriz Hessiana

Considerando os resultados apresentados na Subseção 3.7.3 temos que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = C - \frac{k}{2}\log\sigma_1^2 - \frac{n-k}{2}\log\sigma_2^2 - \frac{1}{2}\left[\sum_{i=1}^{k}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2}{\sigma_1^2}\right\} + \sum_{i=k+1}^{n}\left\{q(d_i)\frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2}{\sigma_2^2}\right\}\right],$$

onde $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \sigma_1^2, \sigma_2^2)^\top$ e $q(d_i) = E[\kappa^{-1}(U_i)|y_i, \widehat{\boldsymbol{\theta}}]$. Logo, os elementos da matriz hessiana $\ddot{Q}_\theta(\hat{\boldsymbol{\theta}})$ são

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1^\top} = -\frac{1}{\sigma_1^2}\sum_{i=1}^{k}\left\{q(d_i)\mathbf{x}_i\mathbf{x}_i^\top\right\},$$

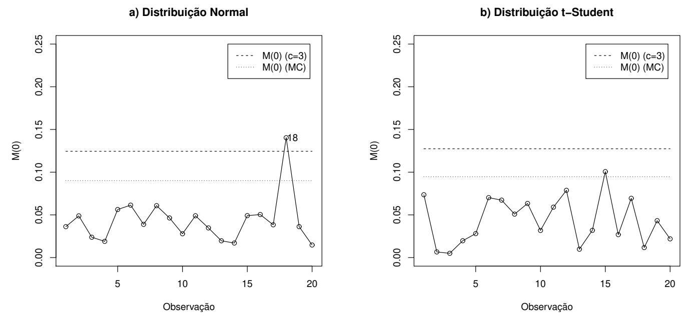
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2^\top} = -\frac{1}{\sigma_2^2}\sum_{i=k+1}^{n}\left\{q(d_i)\mathbf{x}_i\mathbf{x}_i^\top\right\},$$

Tabela 4.2: Resultados para diagnósticos de influência no modelo de regressão linear com mudança na variância considerando 100 amostras simuladas para cada uma das distribuições de Mistura de Escala Normal estudadas: Normal (N), t de Student (t), Slash (Sl) e Normal Contaminada (NC).

Valor de referência M(0) de Lee & Xu (2004) onde $c^* = 3$						
Observação Perturbada	Esquema para M(0)	Esquema de Perturbação	Distribuições N	t	SL	NC
Perturbação da amostra sobre variável explicativa						
Até o ponto de mudança	Geral	Ponderação de casos	100	41	94	57
		Variável explicativa	93	32	88	49
	Especial	Ponderação de casos	100	33	91	52
		Variável explicativa	72	21	68	17
Após ponto de mudança	Geral	Ponderação de casos	0	1	0	2
		Variável explicativa	0	1	0	0
	Especial	Ponderação de casos	99	61	97	75
		Variável explicativa	8	11	16	9
Perturbação da amostra sobre variável resposta						
Até o ponto de mudança	Geral	Ponderação de casos	51	0	28	4
		Variável resposta	100	0	35	6
	Especial	Ponderação de casos	5	0	7	0
		Variável resposta	100	0	29	4
Após ponto de mudança	Geral	Ponderação de casos	0	0	0	1
		Variável resposta	0	0	0	0
	Especial	Ponderação de casos	5	0	5	1
		Variável resposta	53	0	3	0
Valor de referência M(0) obtida vía simulação Monte Carlo						
Observação Perturbada	Esquema para M(0)	Esquema de Perturbação	Distribuições N	t	SL	NC
Perturbação da amostra sobre variável explicativa						
Até o ponto de mudança	Geral	Ponderação de casos	100	0	44	11
		Variável explicativa	100	0	44	10
	Especial	Ponderação de casos	100	0	39	10
		Variável explicativa	100	0	38	6
Após ponto de mudança	Geral	Ponderação de casos	35	0	11	6
		Variável explicativa	39	0	4	3
	Especial	Ponderação de casos	88	24	93	28
		Variável explicativa	76	6	75	23
Perturbação da amostra sobre variável resposta						
Até o ponto de mudança	Geral	Ponderação de casos	100	0	39	10
		Variável resposta	100	0	46	10
	Especial	Ponderação de casos	99	0	35	7
		Variável resposta	100	0	38	7
Após ponto de mudança	Geral	Ponderação de casos	11	0	5	4
		Variável resposta	40	0	3	1
	Especial	Ponderação de casos	51	18	77	21
		Variável resposta	86	3	77	21

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\sigma_1^2} = \frac{k}{2\sigma_1^4} - \frac{1}{\sigma_1^6}\sum_{i=1}^{k}\left\{q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2\right\},$$

Primeiro esquema especial de perturbação



Segundo esquema especial de perturbação

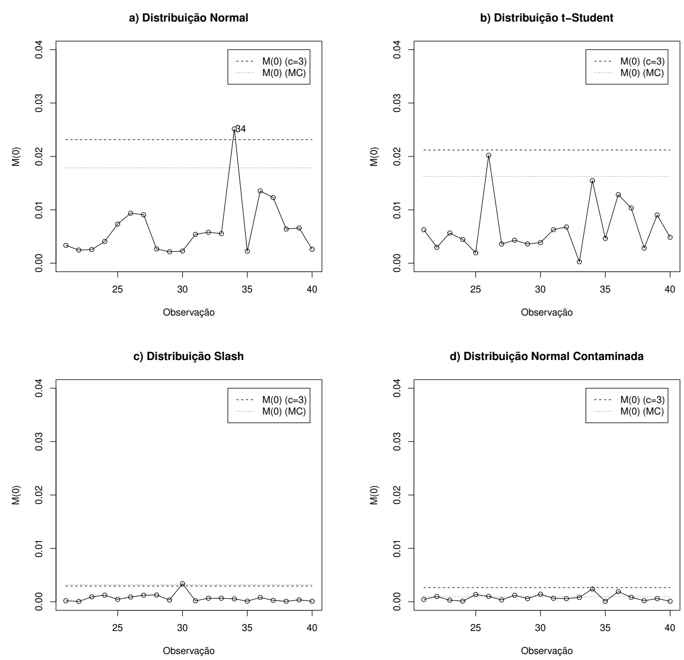


Figura 4.7: Gráfico de $M(0)$ para perturbação da variável explicativa no modelo de regressão linear com mudança na variância, considerando marcas de referência proposta por Lee & Xu (2004) onde $c^* = 3$ e via simulações Monte Carlo sob diversas distribuições que conformam a classe de Mistura de Escala Normal.

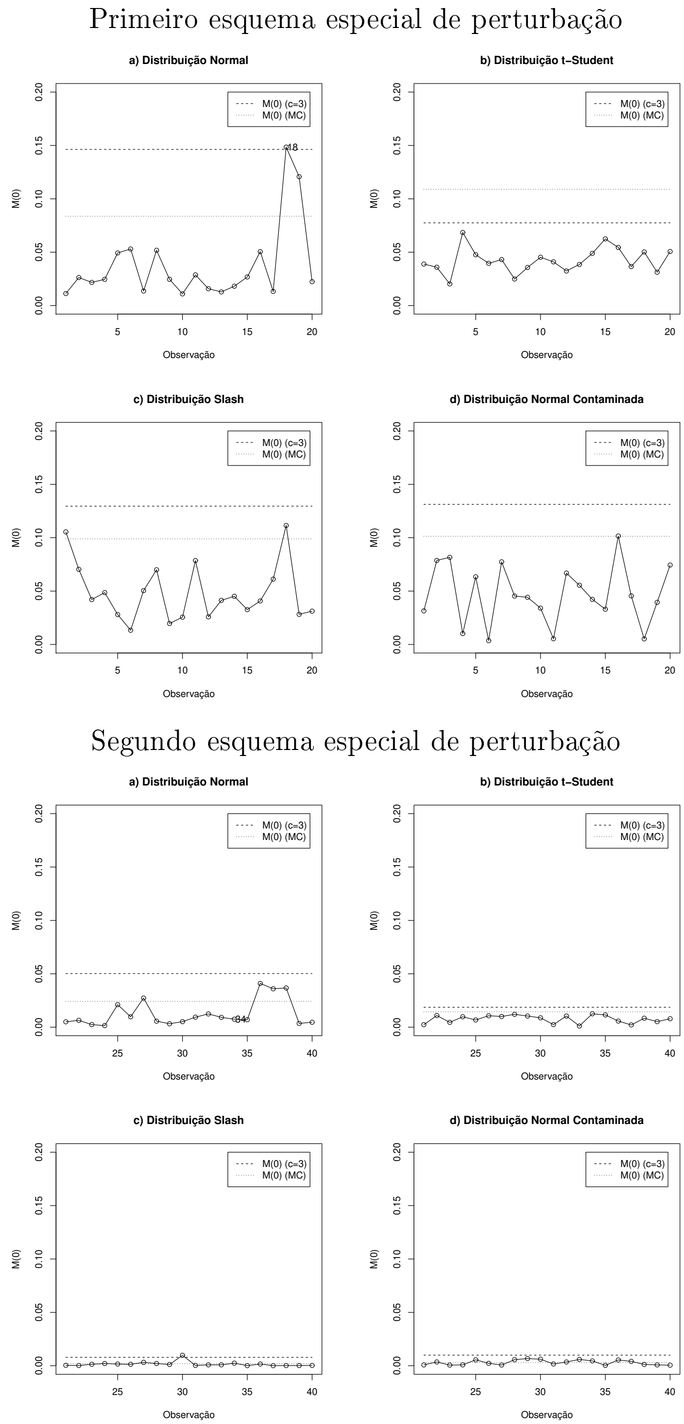


Figura 4.8: Gráfico de $M(0)$ para perturbação da variável resposta no modelo de regressão linear com mudança na variância, considerando marcas de referência proposta por Lee & Xu (2004) onde $c^* = 3$ e via simulações Monte Carlo sob diversas distribuições que conformam a classe de Mistura de Escala Normal.

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\sigma_2^2} = \frac{n-k}{2\sigma_2^4} - \frac{1}{\sigma_2^6}\sum_{i=k+1}^{n}\left\{q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2\right\},$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2^\top} = \mathbf{0},$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\sigma_1^2} = -\frac{1}{\sigma_1^4}\sum_{i=1}^{k}\left\{q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)\mathbf{x}_i\right\},$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\sigma_2^2} = \mathbf{0},$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\sigma_1^2} = \mathbf{0},$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\sigma_2^2} = -\frac{1}{\sigma_1^4}\sum_{i=k+1}^{n}\left\{q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)\mathbf{x}_i\right\},$$
$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\sigma_2^2} = \mathbf{0}.$$

com $d_i = \dfrac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_1)^2}{\sigma_1^2}$ para $i = 1, \ldots, k$ e $d_i = \dfrac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_2)^2}{\sigma_2^2}$ para $i = k+1, \ldots, n$.
A obtenção dos valores $q(d_i)$, $i = 1, \ldots, n$, para as quatro distribuições da classe de Mistura de Escala Normal estudadas nesta dissertação, são apresentadas em (2.10)-(2.13).

4.5.2 Esquemas de Perturbação

Considerando os três esquemas de perturbação estudados nesta dissertação, nesta subseção são apresentadas as matrices $\Delta_{\boldsymbol{\omega}}$ necessárias para a avaliação da influência local. Para cada um dos esquemas de perturbação a matriz $\Delta_{\boldsymbol{\omega}_0}$ pode ser particionada como

$$\Delta_{\boldsymbol{\omega}_0} = (\Delta_{\boldsymbol{\beta}_1}^\top, \Delta_{\boldsymbol{\beta}_2}^\top, \Delta_{\sigma_1^2}, \Delta_{\sigma_2^2})^\top,$$

onde $\Delta_{\boldsymbol{\beta}_1} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{p\times g}$, $\Delta_{\boldsymbol{\beta}_2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{p\times g}$, $\Delta_{\sigma_1^2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, e $\Delta_{\sigma_2^2} = \left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\omega}_0} \in \mathbb{R}^{1\times g}$, g é a dimensão do vetor de perturbação $\boldsymbol{\omega}$, p e n o número de linhas e colunas da matriz $\mathbf{X}$ respetivamente.

1) **Ponderação de casos**

Atribuindo uma ponderação arbitrária para o valor esperado da função log-verossimilhança dos dados completos, a função Q perturbada, é possível capturar distorções em direções gerais, as qual é representada por

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}),$$

onde
$$Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma_1^2 + q(d_i)\frac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_1)^2}{\sigma^2}], & para \quad i = 1, \ldots, k, \\ -\frac{1}{2}[\log\sigma_2^2 + q(d_i)\frac{(y_i - \mathbf{x_i}^\top\boldsymbol{\beta}_2)^2}{\sigma^2}], & para \quad i = k+1, \ldots, n. \end{cases}$$

Neste caso o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (1, ..., 1)^\top$. Considerando o modelo com ponto de mudança na posição k, este esquema de perturbação se reduz a dois casos especiais

- $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{k}\omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + \sum_{i=k+1}^{n} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ e
- $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{k} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) + \sum_{i=k+1}^{n}\omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são

$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\omega_i} = \begin{cases} \frac{1}{\sigma_1^2}q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)\mathbf{x}_i^\top & , para \quad i = 1, \ldots, k, \\ \mathbf{0} & , para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\omega_i} = \begin{cases} \mathbf{0} & , para \quad i = 1, \ldots, k, \\ \frac{1}{\sigma_2^2}q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)\mathbf{x}_i^\top & , para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\omega_i} = \begin{cases} -\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_1^4}q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_1)^2 & , para \quad i = 1, \ldots, k, \\ 0 & , para \quad i = k+1, \ldots, n, \end{cases}$$
$$\frac{\partial^2 Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\omega_i} = \begin{cases} 0 & , para \quad i = 1, \ldots, k, \\ -\frac{1}{2\sigma_2^2} + \frac{1}{2\sigma_2^4}q(d_i)(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2 & , para \quad i = k+1, \ldots, n. \end{cases}$$

2) **Perturbação nas variáveis explicativas**

O interesse é perturbar uma variável explicativa continua especifica $\mathbf{x}_s, s = 1, ..., p$. Um esquema de perturbação aditivo é obtido fazendo $\mathbf{x}_{w_is} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i = 1, \ldots, n$ e onde S_{x_s} é um fator de escala que pode ser o desvio padrão de $\mathbf{x}_s$ e $\mathbf{1}_s$ é um vetor com 1 na posição s e 0 nas outras posições. Considerando o modelo com ponto de mudança na posição k dois casos especiais deste esquema de perturbação são de interesse

- $\mathbf{x}_{w_is} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i = 1, \ldots, k$ e
- $\mathbf{x}_{w_is} = \mathbf{x}_i + w_i S_{x_s}\mathbf{1}_s$, para $i = k+1, \ldots, n$.

Neste esquema, a função Q perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}, \omega_i|\hat{\boldsymbol{\theta}}),$$

onde
$$Q_i(\boldsymbol{\theta}, \boldsymbol{\omega}_i|\hat{\boldsymbol{\theta}}) = \begin{cases} -\frac{1}{2}[\log\sigma_1^2 + q(d_i)\frac{(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta}_1)^2}{\sigma_1^2}], & para \quad i = 1, \ldots, k, \\ -\frac{1}{2}[\log\sigma_2^2 + q(d_i)\frac{(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta}_2)^2}{\sigma_2^2}], & para \quad i = k+1, \ldots, n. \end{cases}$$

e o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (0, ..., 0)^\top$.

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\omega_i} = \begin{cases} \dfrac{S_{x_s}}{\sigma_1^2}q(d_i)(y_i - 2\mathbf{x}_{w_is}^\top\boldsymbol{\beta}_1))\mathbf{1}_s^\top & ,para \quad i=1,\ldots,k, \\ \mathbf{0} & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\omega_i} = \begin{cases} \mathbf{0} & ,para \quad i=1,\ldots,k, \\ \dfrac{S_{x_s}}{\sigma_2^2}q(d_i)(y_i - 2\mathbf{x}_{w_is}^\top\boldsymbol{\beta}_2))\mathbf{1}_s^\top & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\omega_i} = \begin{cases} -\dfrac{S_{x_s}}{\sigma_1^4}q(d_i)(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta}_1)\boldsymbol{\beta}_1^\top\mathbf{1}_s & ,para \quad i=1,\ldots,k, \\ 0 & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\omega_i} = \begin{cases} 0 & ,para \quad i=1,\ldots,k, \\ -\dfrac{S_{x_s}}{\sigma_2^4}q(d_i)(y_i - \mathbf{x}_{w_is}^\top\boldsymbol{\beta}_2)\boldsymbol{\beta}_2^\top\mathbf{1}_s & ,para \quad i=k+1,\ldots,n. \end{cases}$$

3) **Ponderação na variável resposta**

O interesse é perturbar a variável resposta $y_i, i = 1, ..., n$. Um esquema de perturbação é obtido fazendo $y_{w_i} = y_i + w_i S_y$, para $i = 1, \ldots, n$ e onde S_y é um fator de escala que pode ser o desvio padrão de y. Considerando o modelo com ponto de mudança na posição k dois casos especiais deste esquema de perturbação são de interesse

- $y_{w_i} = y_i + w_i S_y$, para $i = 1, \ldots, k$ e
- $y_{w_i} = y_i + w_i S_y$, para $i = k+1, \ldots, n$.

Neste esquema, a função Q perturbada é dada por

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta},\omega_i|\hat{\boldsymbol{\theta}}),$$

onde
$$Q_i(\boldsymbol{\theta},\boldsymbol{\omega}_i|\hat{\boldsymbol{\theta}}) = \begin{cases} -\dfrac{1}{2}[\log\sigma_1^2 + q(d_i)\dfrac{(y_{w_i} - \mathbf{x_i}^\top\boldsymbol{\beta}_1)^2}{\sigma_1^2}], & para \quad i=1,\ldots,k, \\ -\dfrac{1}{2}[\log\sigma_2^2 + q(d_i)\dfrac{(y_{w_i} - \mathbf{x_i}^\top\boldsymbol{\beta}_2)^2}{\sigma_2^2}], & para \quad i=k+1,\ldots,n. \end{cases}$$

e o modelo não perturbado é obtido quando $\boldsymbol{\omega}_0 = (0, ..., 0)^\top$

Logo, as derivadas necessárias para a obtenção de $\Delta_{\boldsymbol{\omega}}$ e que deverão ser avaliadas em $\boldsymbol{\omega}_0$ são

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_1\partial\omega_i} = \begin{cases} \dfrac{S_y}{\sigma_1^2}q(d_i)\mathbf{x}_i^\top & ,para \quad i=1,\ldots,k, \\ \mathbf{0} & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_2\partial\omega_i} = \begin{cases} \mathbf{0} & ,para \quad i=1,\ldots,k, \\ \dfrac{S_y}{\sigma_2^2}q(d_i)\mathbf{x}_i^\top & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_1^2\partial\omega_i} = \begin{cases} \dfrac{S_y}{\sigma_1^4}q(d_i)(y_{w_i} - \mathbf{x}_i^\top\boldsymbol{\beta}_1) & ,para \quad i=1,\ldots,k, \\ 0 & ,para \quad i=k+1,\ldots,n, \end{cases}$$

$$\frac{\partial^2 Q_i(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma_2^2\partial\omega_i} = \begin{cases} 0 & ,para \quad i=1,\ldots,k, \\ \dfrac{S_y}{\sigma_2^4}q(d_i)(y_{w_i} - \mathbf{x}_i^\top\boldsymbol{\beta}_2) & ,para \quad i=k+1,\ldots,n. \end{cases}$$

4.5.3 Estudos de Simulação

Nesta subseção apresenta-se resultados das simulações realizadas para avaliar o desempenho dos métodos de influência em modelos de regressão com ponto de mudança nos coeficientes de regressão e variância, os dois estudos de simulação apresentados realizam as mesmas avaliações descritas na Subseção 4.3.3. Neste caso, todas as amostras simuladas são geradas desde o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância

$$\begin{aligned} y_i &= 2 + 5x_i + \epsilon_i, \quad para \quad i = 1, \ldots, k, \\ y_i &= 4 + 10x_i + \epsilon_i, \quad para \quad i = k+1, \ldots, n, \end{aligned}$$

e considerando $\sigma_1^2 = 1$ para $i = 1, \ldots, k$ e $\sigma_2^2 = 3$ para $i = k+1, \ldots, n$ e $x_i \sim U(0,1)$ para $i = 1, \ldots, n$.

1) **Influência de "*outliers*"no método de estimação**

O mesmo procedimento descrito na Subseção 4.3.3 é utilizado nesta simulação, desta vez considerando o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância proposto no inicio desta subseção. Nas Figuras 4.9 e 4.10 se apresentam os resultados das mudanças relativas das estimativas dos parâmetros do modelo proposto sob distribuição Normal, t de Student, Slash e Normal Contaminada para os diferentes valores de contaminação $\Im$ considerados para a simulação nas duas situações avaliadas.

Assim como para os outros modelos estudados, observa-se que em ambas situações as estimativas que consideram modelos com caudas mais pesadas (t de Student, Slash e Normal Contaminada) são menos afetadas pelas variações dos valores de contaminação $\Im$ em relação a modelo que considera a distribuição Normal.

Na primeira situação simulada, as variações nas estimativas de β_{02} e β_{12} e σ_2^2 que ocorrem para os distintos valores de contaminação $\Im$ são explicadas unicamente pelo método de estimação que inclui uma optimização nos graus de liberdade das distribuições com caudas pesadas, observe que o modelo sob distribuição Normal não apresenta nenhuma mudança nestes parâmetros.

Na segunda situação simulada, todos os parâmetros apresentam variação para os distintos valores de contaminação $\Im$ devido à presença das duas observações aberrante nas posições 10 e 30. Diferentemente do primeiro modelo estudado, a variância σ_1^2 não apresenta mudança na variação com a presencia de mais dados aberrantes.

2) **Esquemas de perturbação**

O procedimento descrito na Subseção 4.3.3 para este estudo de simulação é utilizado, desta

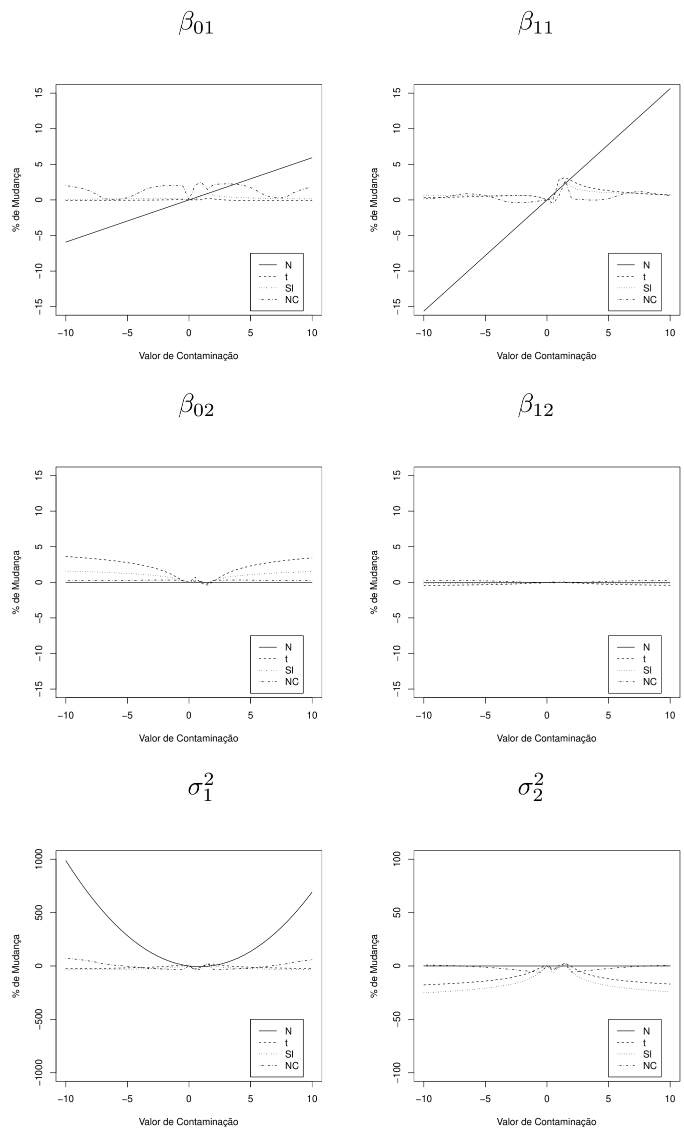


Figura 4.9: Mudanças relativas nas estimativas dos parâmetros do modelo de regressão linear Mistura de Escala normal com ponto de mudança nos coeficientes de regressão e variância considerando distribuições Normal, t de Student, Slash e Normal Contaminada para diferentes valores de contaminação $\Im$ na observação 10.

vez considerando o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância proposto no inicio desta subseção. São mantidas todas as condições das simulações exceto cte correspondentes nas subseções 4.3.3 e 4.4.3 exceto o valor de cte.

A Tabela 4.3, que resume as simulações usando $cte = 0.1$, e as Figuras 4.11-4.12 que considera $cte = 0.3$ mostram os resultados das simulações para o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância. Estes resultados são similares aos obtidos na Subseção

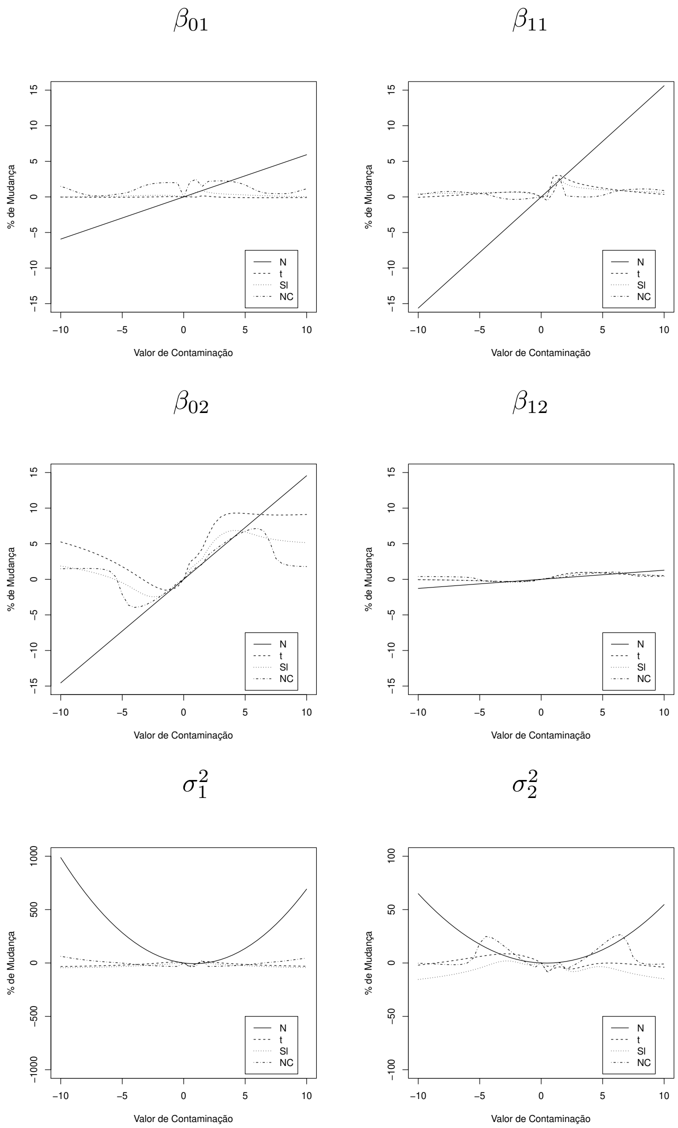


Figura 4.10: Mudanças relativas nas estimativas dos parâmetros do modelo de regressão linear Mistura de Escala normal com ponto de mudança nos coeficientes de regressão e variância considerando distribuições Normal, t de Student, Slash e Normal Contaminada para diferentes valores de contaminação $\Im$ nas observações 10 e 30.

4.3.3, portanto as conclusões apresentadas anteriormente são válidas também para modelos de regressão com ponto de mudança na variância. As Figuras 4.11-4.12 mostram desta vez a influência das observações 10 e 25, que são às observações perturbadas nesta simulação.

Tabela 4.3: Resultados para diagnósticos de influência no modelo de regressão linear com mudança nos coeficientes de regressão e variância considerando 100 amostras simuladas para cada uma das distribuições de Mistura de Escala Normal estudadas: Normal (N), t de Student (t), Slash (Sl) e Normal Contaminada (NC).

Valor de referência M(0) de Lee & Xu (2004) onde $c^* = 3$						
Observação Perturbada	Esquema para M(0)	Esquema de Perturbação	Distribuições N	t	SL	NC
Perturbação da amostra sobre variável explicativa						
Até o ponto de mudança	Geral	Ponderação de casos	95	60	89	76
		Variável explicativa	68	38	62	57
	Especial	Ponderação de casos	99	61	94	85
		Variável explicativa	44	32	40	38
Após ponto de mudança	Geral	Ponderação de casos	13	28	21	21
		Variável explicativa	5	15	8	9
	Especial	Ponderação de casos	94	73	86	79
		Variável explicativa	29	26	18	15
Perturbação da amostra sobre variável resposta						
Até o ponto de mudança	Geral	Ponderação de casos	31	1	27	16
		Variável resposta	99	3	55	19
	Especial	Ponderação de casos	2	2	0	3
		Variável resposta	90	1	26	7
Após ponto de mudança	Geral	Ponderação de casos	7	1	7	4
		Variável resposta	0	0	0	0
	Especial	Ponderação de casos	6	2	9	3
		Variável resposta	51	0	12	7
Valor de referência M(0) obtida vía simulação Monte Carlo						
Observação Perturbada	Esquema para M(0)	Esquema de Perturbação	Distribuições N	t	SL	NC
Perturbação da amostra sobre variável explicativa						
Até o ponto de mudança	Geral	Ponderação de casos	100	13	86	35
		Variável explicativa	98	7	77	29
	Especial	Ponderação de casos	100	7	84	32
		Variável explicativa	96	4	63	23
Após ponto de mudança	Geral	Ponderação de casos	85	11	64	33
		Variável explicativa	54	4	23	15
	Especial	Ponderação de casos	94	30	81	45
		Variável explicativa	25	4	13	11
Perturbação da amostra sobre variável resposta						
Até o ponto de mudança	Geral	Ponderação de casos	92	7	80	29
		Variável resposta	100	13	90	34
	Especial	Ponderação de casos	74	4	67	23
		Variável resposta	99	3	63	21
Após ponto de mudança	Geral	Ponderação de casos	52	5	34	22
		Variável resposta	3	0	0	0
	Especial	Ponderação de casos	56	26	49	30
		Variável resposta	85	6	66	24

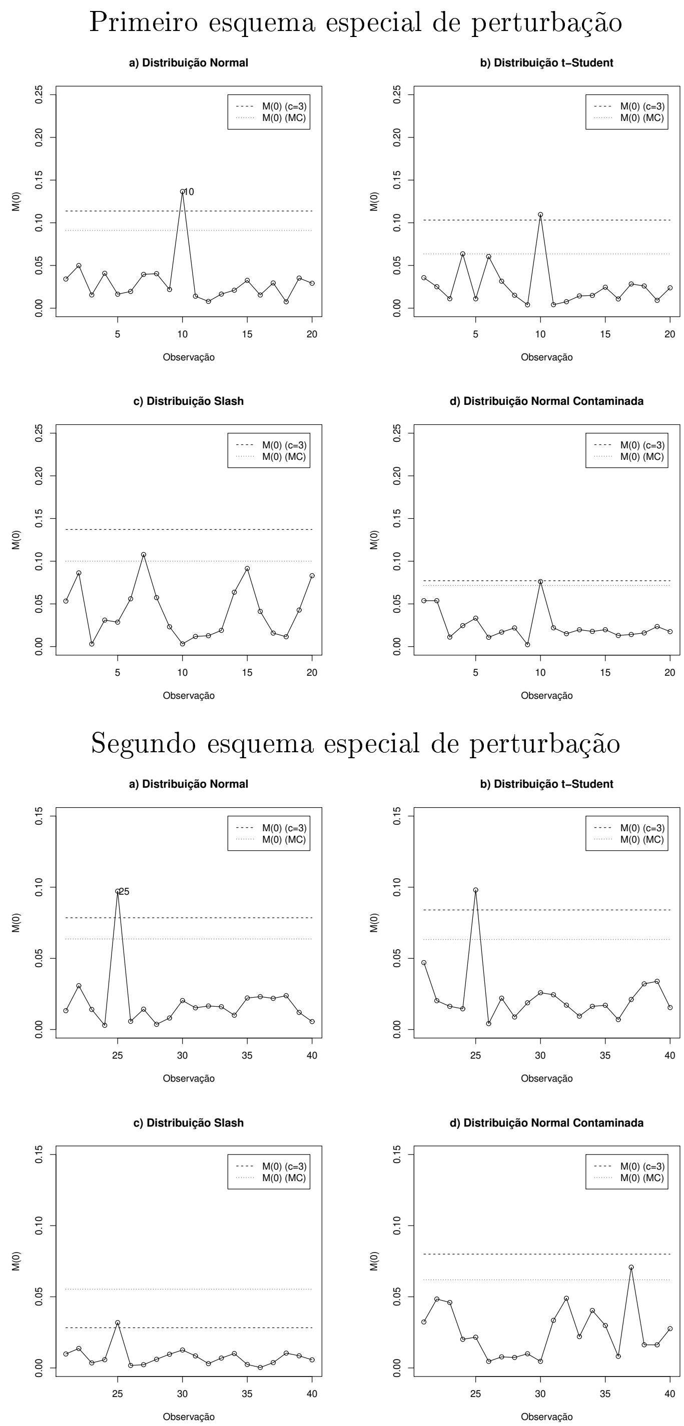
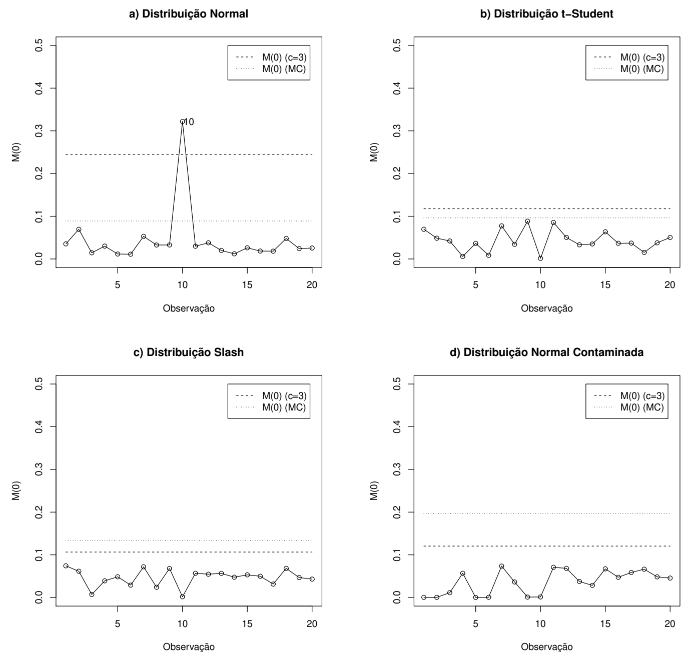


Figura 4.11: Gráfico de $M(0)$ para perturbação da variável explicativa no modelo de regressão linear com mudança nos coeficientes de regressão e variância, considerando marcas de referência proposta por Lee & Xu (2004) onde $c^* = 3$ e via simulações Monte Carlo sob diversas distribuições que conformam a classe de Mistura de Escala Normal.

Primeiro esquema especial de perturbação



Segundo esquema especial de perturbação

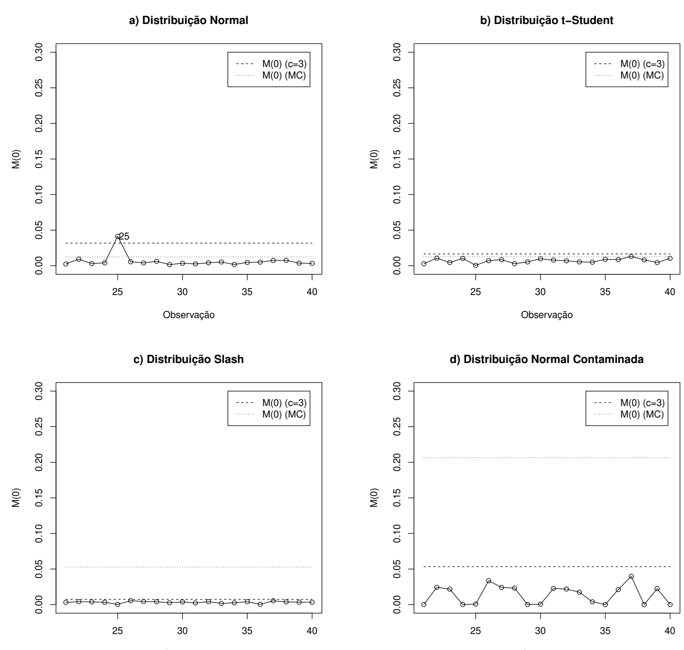


Figura 4.12: Gráfico de $M(0)$ para perturbação da variável resposta no modelo de regressão linear com mudança nos coeficientes de regressão e variância, considerando marcas de referência proposta por Lee & Xu (2004) onde $c^* = 3$ e via simulações Monte Carlo sob diversas distribuições que conformam a classe de Mistura de Escala Normal.

Capítulo 5

Aplicações

Neste capítulo, a metodologia desenvolvida neste trabalho é aplicada a conjuntos de dados reais. O primeiro corresponde a um conjunto de dados conhecido na literatura estatística que contém informações sobre volume de vendas de ações e estudado por Holbert (1982). Finalmente dois conjuntos de dados sobre medição de audiências de televisão correspondentes a dois dias de medição também serão considerados.

5.1 Conjunto de Dados Holbert

Um conjunto de dados mensais correspondentes ao período entre Janeiro de 1967 e Novembro de 1969 sobre volume de vendas (em milhões de dólares) dos mercados de ações de Boston (BSE) e de New York (NYAMSE). Inicialmente foi analisado via regressão segmentada por McGee & Carleton (1970) para validar a afirmação de um relatório do 03 de janeiro de 1970 na edição do Business Week sugerindo que as bolsas regionais foram prejudicadas pela abolição de "Give-ups"dada em dezembro de 1968.

Holbert (1982) analisa os mesmos dados usando regressão linear com ponto de mudança nos coeficientes de regressão desde o ponto de vista bayesiano. Posteriormente Chen (1998) analisa estes dados usando o método SIC para detectar mudança nos coeficientes de um modelo de regressão linear sob suposição de normalidade. Ainda Osorio & Galea (2006) consideram um modelo de regressão linear t de Student e utilizam o método SIC para detectar os pontos de mudança nos coeficientes de regressão.

Nesta seção, consideramos o modelo de regressão linear Mistura de Escala Normal e ilustramos a metodologia proposta para seleção de modelo e determinação do ponto de mudança que utiliza o procedimento SIC. Finalmente a metodologia para análise de influência para o modelo sugerido é aplicado. Os dados em questão, denominados Dados Holbert, estão no anexo.

Primeiramente, seguindo Chen (1998) e Osorio & Galea (2006), consideramos o modelo de regressão com mudança nos coeficientes de regressão. O procedimento SIC proposto é aplicado para a determinação do ponto de mudança, sob as quatro distribuições pertencentes à classe de Mistura de Escala Normal estudados nesta dissertação. O resumo dos resultados se apresentam na Tabela 5.1.

Tabela 5.1: Dados Holbert: Resultados para critério SIC para modelo de regressão linear com mudança nos coeficientes de regressão.

Modelo	$minSIC(k)$	$SIC(n)$	$\widehat{k}$
Normal	358.185	361.496	23
t de Student	357.996	361.462	23
	(ν=2.455)	(ν=2.939)	
Slash	358.021	361.591	23
	(ν=0.859)	(ν=0.980)	
Normal contaminada	354.934	362.812	9
	(ν=0.611,γ=0.012)	(ν=0.264,γ=0.097)	

Observa-se que ao considerar o modelo sob distribuição Normal o ponto de mudança encontra-se na posição $k = 23$, resultado que está em concordância com a conclusão obtida por Chen (1998). No entanto, utilizando o critério SIC observamos que os modelos sob distribuições com caudas pesadas são os mais adequados e destes, o modelo sob distribuição Normal Contaminada com ponto de mudança na posição $k = 9$ é o que melhor se ajusta aos dados. Este ponto de mudança na posição $k = 9$ já foi observado por Osorio & Galea (2006) quando utiliza o modelo t de Student com v = 1. Porém, note-se também que o valor $minSIC(k)$ do critério para o modelo sob distribuição Normal Contaminada apresenta o valor menor quando comparado com os mostrados no trabalho de Osorio & Galea (2006). A Figura 5.1 mostra os valores do critério de informação de Schwarz (SIC) contra a posição de mudança para todas as distribuições consideradas. Osorio & Galea (2006)

A Tabela 5.2 contém as estimativas de máxima verossimilhança para os parâmetros dos quatro modelos propostos, acompanhados de seus respectivos desvios padrão (entre parênteses) obtidos via matriz de informação observada.

A Figura 5.2 apresenta o gráfico de dispersão com o modelo de regressão com ponto de mudança nos coeficientes de regressão ajustado sob a distribuição Normal Contaminada. O gráfico mostra que a dispersão dos dados é diferente para os grupos associados à primeira reta e à segunda, Assim, o modelo de regressão com ponto de mudança nos coeficientes de regressão e variãncia será também avaliado.

A Tabela 5.3 apresenta o resumo de resultados da aplicação do procedimento SIC para a determinação do ponto de mudança sob as quatro distribuições da classe MEN estudadas quando o modelo de regressão com mudança nos coeficientes de regressão e variância é considerado. Neste

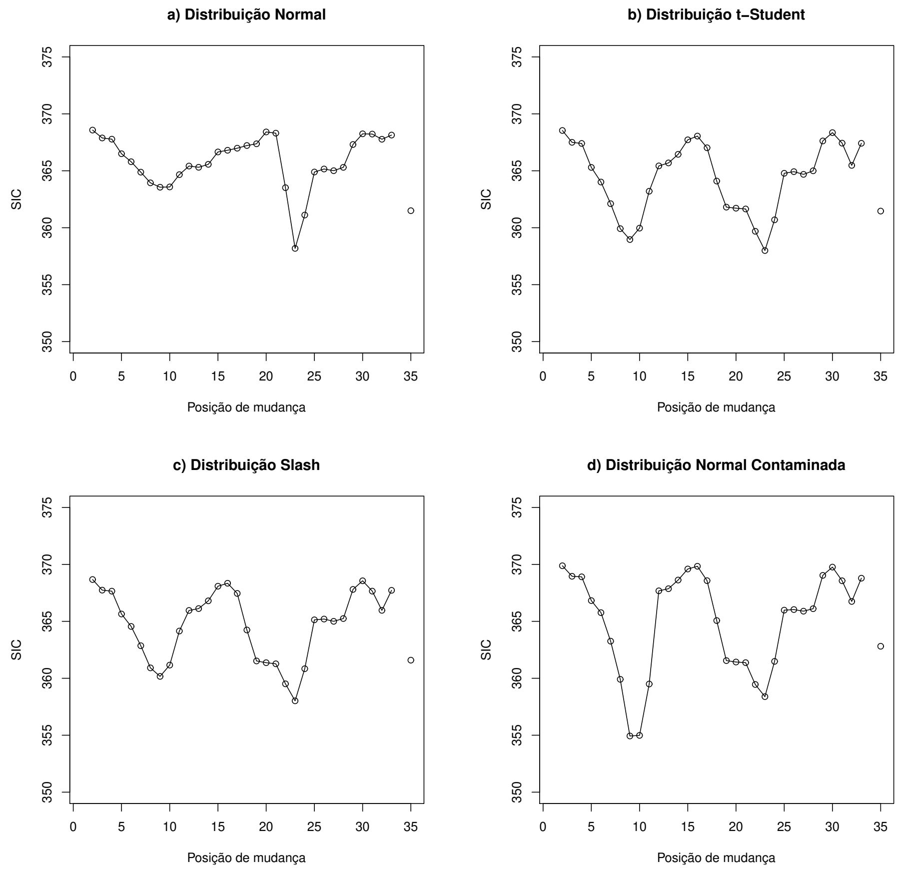


Figura 5.1: Dados Holbert: SIC para ajustes com diversas distribuições. Modelo de regressão com mudança nos coeficientes de regressão.

caso, todos os modelos avaliados indicam o ponto de mudança na posição $k = 9$. Novamente o modelo sob distribuição Normal Contaminada é o que se ajusta melhor aos dados. Além disso, comparando os resultados da tabelas 5.1 e 5.3, temos que o modelo de regressão com mudança nos coeficientes de regressão e variância sob distribuição Normal Contaminada (ν=0.736,γ=0.009) é o mais adequado para o conjunto de dados avaliado.

A Figura 5.3 mostra os valores do critério de informação de Schwarz (SIC) contra a posição de mudança para todas as distribuições estudadas, considerando o modelo de regressão com mudança nos coeficientes de regressão e variância. A Tabela 5.4 contém as estimativas de máxima verossimilhança para os parâmetros dos quatro modelos propostos, acompanhados dos respectivos desvios padrães obtidos via matriz de informação observada. Ao comparar os dois tipos de modelos avaliados sob distribuição Normal Contaminada observa-se que a maior diferença se produz na estimativa do coeficiente β_0 calculado para os dados até o ponto de mudança $k = 9$, as demais estimativas dos coeficientes de regressão mostram valores similares.

Tabela 5.2: Dados Holbert: Estimativa dos parâmetros para os modelos de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições.

Modelo	Casos	β_0	β_1	σ^2
Normal	1-23	-110.310	0.018	980.503
		(40.6577)	(0.0030)	(234.3851)
	24-35	11.075	0.007	
		(57.6394)	(0.0042)	
t de Student	1-23	-92.834	0.016	367.871
(ν=2.455)		(26.6999)	(0.0020)	(131.8429)
	24-35	15.585	0.006	
		(42.8789)	(0.0030)	
Slash	1-23	-94.529	0.016	176.919
(ν=0.859)		(27.0404)	(0.0020)	(62.4201)
	24-35	14.953	0.006	
		(39.3962)	(0.0028)	
Normal contaminada	1-9	29.553	0.004	22.164
(ν=0.611,γ=0.012)		(19.9741)	(0.0017)	(5.9694)
	10-35	-37.138	0.012	
		(13.3001)	(0.0010)	

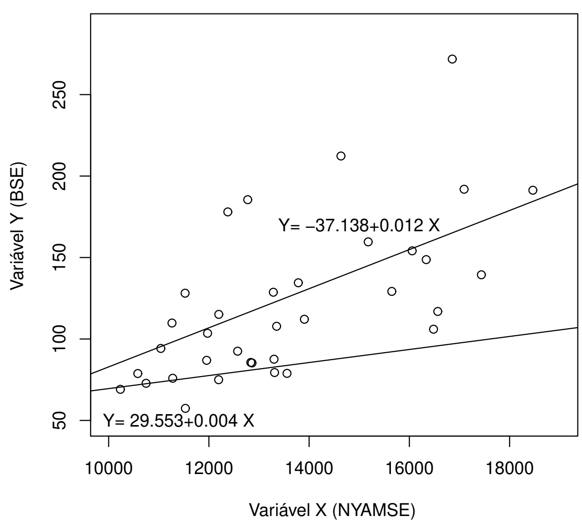


Figura 5.2: Dados Holbert: Ajuste do modelo de regressão com ponto de mudança nos coeficientes de regressão sob distribuição Normal Contaminada.

O método de influência local proposto no Capítulo 4 é aplicado sobre o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância. Primeramente, a identificação das observações atípicas será feita considerando a distância de Mahalanobis. A Figura 5.4 mostra estas distâncias para os modelos estudados. Nota-se que a observação 22 aparece como um "*outlier*"para todos os modelos considerados.

Tabela 5.3: Dados Holbert: Resultados para critério SIC para modelo de regressão linear com mudança nos coeficientes de regressão e variância.

Modelo	$minSIC(k)$	$SIC(n)$	$\widehat{k}$
Normal	337.876	361.496	9
t de Student	341.442	361.462	9
	(ν=50.000)	(ν=2.939)	
Slash	341.432	361.591	9
	(ν=20.389)	(ν=0.980)	
Normal contaminada	336.503	362.812	9
	(ν=0.751,γ=0.006)	(ν=0.263,γ=0.097)	

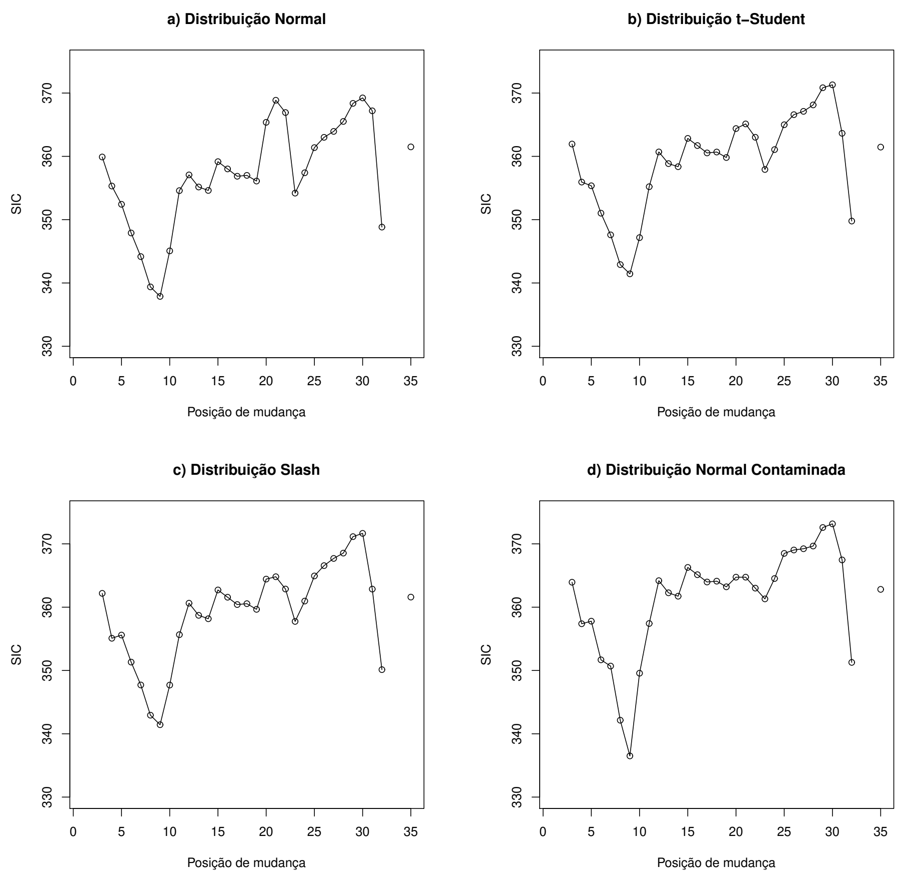


Figura 5.3: Dados Holbert: SIC para ajustes com diversas distribuições. Modelo de regressão com mudança nos coeficientes de regressão e variância.

A flexibilidade das estimativas de máxima verossimilhança dos modelos sob distribuições com

Tabela 5.4: Dados Holbert: Estimativa dos parâmetros para modelos de regressão com ponto de mudança nos coeficientes de regressão e variância sob distintas distribuições.

Modelo	Casos	β_0	β_1	σ^2
Normal	1-9	31.341	0.004	19.042
		(15.4557)	(0.0013)	(8.9766)
	10-35	-30.697	0.012	1532.308
		(49.7261)	(0.0035)	(424.9858)
t de Student	1-9	30.981	0.004	18.774
(ν=50.000)		(15.5486)	(0.0013)	(9.0015)
	10-35	-28.651	0.012	1460.001
		(49.2495)	(0.0035)	(418.7303)
Slash	1-9	31.321	0.004	18.127
(ν=20.389)		(15.4559)	(0.0013)	(8.5483)
	10-35	-30.548	0.012	1455.489
		(49.6858)	(0.0035)	(404.3160)
Normal contaminada	1-9	7.644	0.006	0.180
(ν=0.751,γ=0.006)		(2.2461)	(0.0002)	(0.0859)
	10-35	-42.170	0.013	11.586
		(11.9427)	(0.0008)	(3.7929)

caudas mais pesadas contra observações atípicas é evidenciada na Figura 5.5 que apresenta os pesos estimados (q_i), $i = 1, \ldots, n$, provenientes da etapa E do algoritmo, versus as distâncias de Mahalanobis, onde se observa pesos menores para a observação 22. Note que as estimativas dos modelos sob distribuição t de Student e Slash apresentam graus de liberdade altos que indicam uma aproximação à distribuição Normal, ainda assim os pesos estimados conservam uma relação inversa à distância de Mahalanobis e desta forma, o processo de estimação de parâmetros atribui pesos menores as observações anômalas. Para o modelo sob distribuição Normal, temos que $q_i = 1$ para $i = 1, \ldots, n$, mostrados como uma linha tracejada nos gráficos apresentados.

Finalmente, as observações influentes no conjunto de dados Holbert serão identificadas utilizando as quantidades $M(0)_l$ para $l = 1, .., g$ calculadas através da curva conformal $B_{f_Q,\mathbf{d}_l}$ segundo a metodologia detalhada na Seção 4.2 e considerando os esquemas ponderação de casos, perturbação na variável explicativa e perturbação na variável resposta. Em todos os casos, considera-se a marca de referência proposta por Lee & Xu (2004) com $c* = 3$. Os esquemas especiais de perturbação para modelos com ponto de mudança são utilizados em todas as situações.

As Figuras 5.6-5.8 apresentam os gráficos de $M(0)$ para o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância sob as quatro distribuições estudadas para os três esquemas mencionados e considerando o segundo esquema especial (para as observações 10-35). Os resultados correspondentes ao primeiro esquema especial não identificam observações influentes por isso não são apresentados, neste trabalho, por brevidade.

Nas figuras observamos que não existe nenhuma observação influente para as estimativas de máxima verossimilhança sob ponderação de casos para os modelos sob distribuição Normal, t de

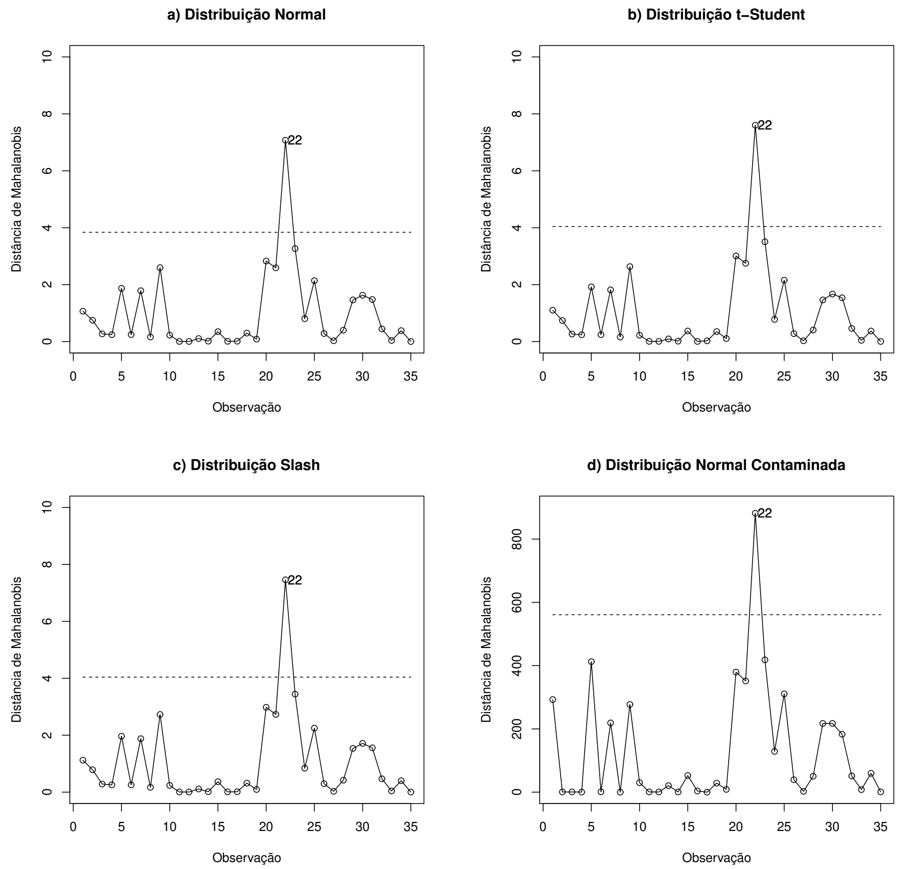


Figura 5.4: Dados Holbert: Distância de Mahalanobis para o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância sob distintas distribuições, considerando $\xi = 0.95$.

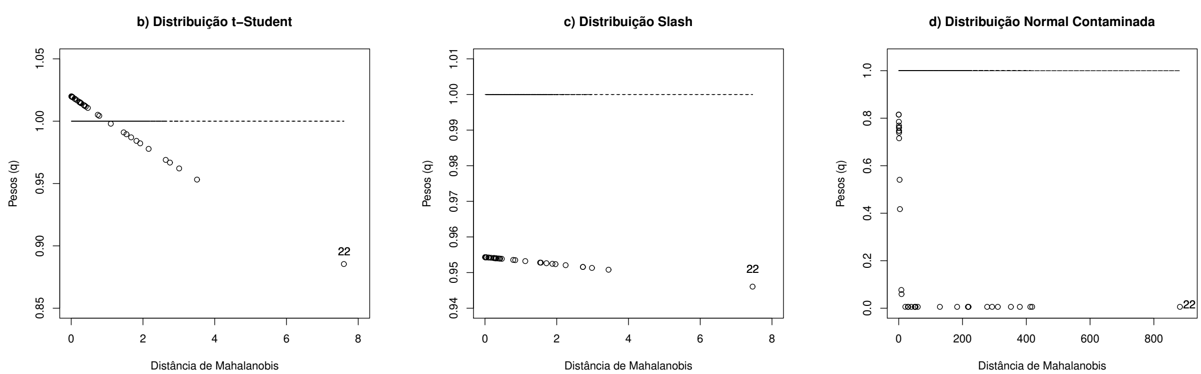


Figura 5.5: Dados Holbert: Distância de Mahalanobis versus q_i para o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância sob distintas distribuições.

Student e Slash. Por outro lado, o modelo sob distribuição Normal Contaminada mostra que a

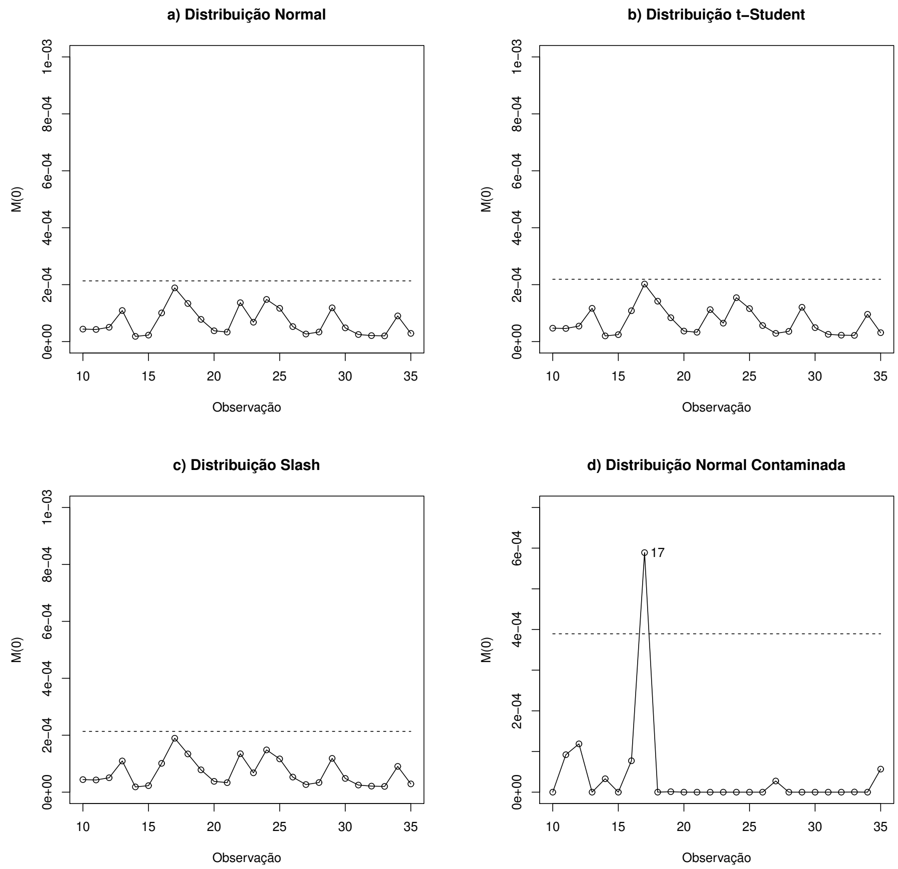


Figura 5.6: Dados Holbert: Gráficos $M(0)$ considerando esquema ponderação de casos na segunda situação especial para modelos de regressão com ponto de mudança nos coeficientes de regressão e variância sob distintas distribuições consideradas.

observação 17 é influente para as estimativas de máxima verossimilhança nos três esquemas de perturbação avaliados. Para os modelos sob distribuição Normal, t de Student e Slash a observação 22 aparece como influente para o esquema de perturbação na variável resposta enquanto que nenhuma observação aparece para o esquema de perturbação na variável explicativa.

As observações 17 e 22 apresentam os valores extremos no conjunto de dados, a observação 17 apresenta o maior valor observado na variável explicativa (NYAMSE) e a observação 22 apresenta o maior valor observado na variável resposta (BSE). No gráfico de dispersão (Figura 5.2), estas observações correspondem aos pontos extremos da parte superior direita e que certamente apresentam comportamentos diferenciados do resto de observações do conjunto de dados estudado.

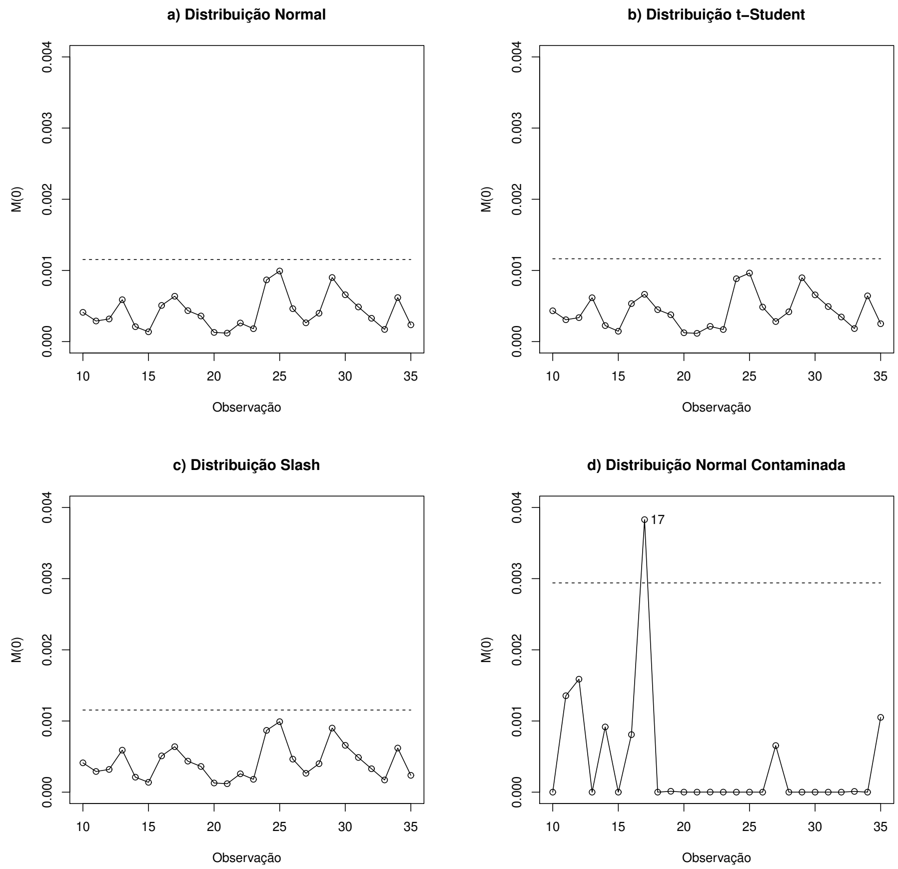


Figura 5.7: Dados Holbert: Gráficos $M(0)$ considerando esquema perturbação da variável explicativa na segunda situação especial para modelos de regressão com ponto de mudança nos coeficientes de regressão e variância sob distintas distribuições consideradas.

5.2 Conjunto de Dados Ibope

Dentro de um processo contínuo para a obtenção de audiências de televisão realizada pelo Instituto Brasileiro de Opinião Pública e Estatística (IBOPE) a atualização nos procedimentos de medição são executadas com frequência. Uma atualização específica ocorre quando os instrumentos de medição são atualizados, quando isto ocorre torna-se necessário conhecer a relação numérica existente entre as duas formas de medição.

Um estudo interno foi conduzido por vários dias para a obtenção destes resultados. Para um área geográfica especificada foram avalidados dois instrumentos de medição (denominados Caderno e Meter) sobre duas amostras representativas de domicílios. Em cada amostra foi utilizado um instrumento diferente. Os mesmos procedimentos de controle e processamento de informação definidos foram aplicados sobre as duas amostras e foram obtidas as audiências respectivas.

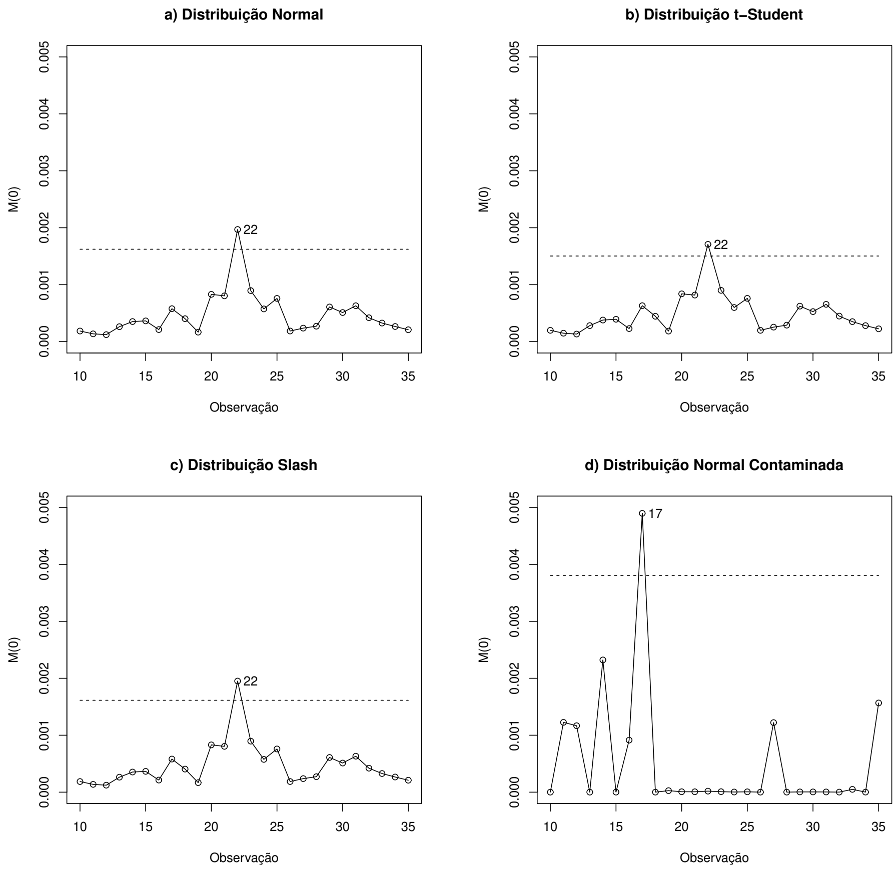


Figura 5.8: Dados Holbert: Gráficos $M(0)$ considerando esquema perturbação da variável resposta na segunda situação especial para modelos de regressão com ponto de mudança nos coeficientes de regressão e variância sob distintas distribuições consideradas.

Uma característica importante é a forma de medição de cada instrumento: as audiências medidas via Meter (que é uma medição eletrônica e mais recente) podem ser obtidas minuto a minuto enquanto que as audiências medidas via Caderno são obtidas por preenchimento manual das informações e realizado em faixas de 15 minutos. Espera-se que a medição via Meter seja mais precisa.

Dois conjuntos de informações de audiências de televisão domiciliares consolidadas correspondentes a dias diferentes são considerados. As audiências são apresentadas em faixas de 15 minutos, estas se iniciam às 06:00 hs. e terminam às 24:00 hs. do dia de medição e que corresponde a 72 dados ordenados. Duas variáveis são consideradas, a primeira corresponde às audiências obtidas a partir da aplicação do instrumento denominado Caderno (CAD) e a segunda corresponde às audiências obtidas a partir do instrumento eletrônico denominado Meter (MET). No anexo desta dissertação encontram-se os dados em questão.

O modelo proposto pretende obter uma relação entre os valores de audiência obtidos pelo uso do Meter e pelo uso do Caderno. Especificamente é importante manter um histórico de dados. Por ser o Caderno, a forma mais antiga de medição, precisamos estimar valores de Audiência via Meter a partir de informações de audiência obtidas a partir de Cadernos de medição, isto é, devemos considerar o modelo

$$MET_i = \beta_0 + \beta_1 CAD_i + \epsilon_i$$

Adicionalmente, é sabido que os comportamentos de audiência mudam a partir de certo horário. Desta forma a inclusão de um ponto de mudança no modelo proposto é necessária e a avaliação de observações atípicas torna-se importante. A metodologias para determinação do ponto de mudança e o análise de influência local propostas nesta dissertação serão aplicadas nestes conjuntos de dados.

1) **Conjunto de dados IBOPE 01**

Para este conjunto de dados, consideramos o modelo de regressão com mudança nos coeficientes de regressão. O procedimento SIC proposto é aplicado para a determinação do ponto de mudança, sob as quatro distribuições pertencentes à classe de Mistura de Escala Normal. O resumo dos resultados se apresentam na Tabela 5.5.

Tabela 5.5: Dados Ibope 01: Resultados para critério SIC para modelo de regressão linear com mudança nos coeficientes de regressão.

Modelo	$minSIC(k)$	$SIC(n)$	$\widehat{k}$
Normal	411.599	470.134	62
t de Student	408.026	466.810	62
	(ν=3.114)	(ν=1.871)	
Slash	408.750	469.055	62
	(ν=1.068)	(ν=0.679)	
Normal contaminada	409.575	462.596	62
	(ν=0.319,γ=0.130)	(ν=0.442,γ=0.054)	

Observa-se que ao considerar o modelo sob qualquer uma das distribuições estudadas o ponto de mudança encontra-se na posição $k = 62$. Utilizando o critério SIC, observamos que os modelos sob distribuições com caudas pesadas são os mais adequados e destes, o modelo sob distribuição t de Student com ponto de mudança na posição $k = 62$ é o que melhor se ajusta aos dados. Por sua vez, este ponto corresponde ao intervalo horário de $21:15-21:30$ horas. A Figura 5.9 apresenta as audiências medidas por cada um dos instrumentos ao longo do dia. Note que a posição $k = 62$ identifica claramente o ponto a partir da qual se produz uma inversão nos comportamentos de audiência medidos pelos dois instrumentos, Até o ponto em questão a medição via Meter apresenta valores menores e logo após o ponto, estes valores são superiores as medições realizadas via caderno. A Figura 5.10 mostra os valores do critério de informação de Schwarz (SIC) contra a posição de mudança para todas as distribuições consideradas.

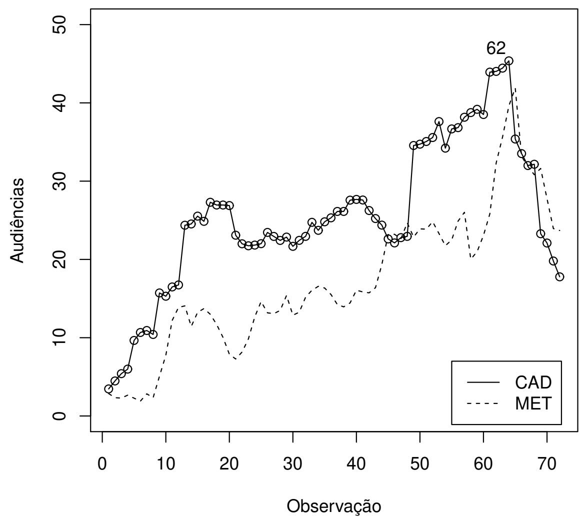


Figura 5.9: Dados Ibope 01: Medições de audiência via instrumentos Meter (MET) e Caderno (CAD) por observações ordenadas segundo horário de medição.

A Tabela 5.6 contém as estimativas de máxima verossimilhança para os parâmetros dos quatro modelos propostos acompanhadas de seus respectivos desvios padrões (entre parênteses) obtidos via matriz de informação observada. A Figura 5.11 apresenta o gráfico de dispersão com o modelo de regressão com ponto de mudança nos coeficientes de regressão ajustado sob a distribuição t de Student. Por outro lado, o modelo de regressão com ponto de mudança nos coeficientes de regressão e variância também foi avaliado e não se mostrou melhor.

O método de influência local proposto no Capítulo 4 é utilizado sobre o modelo de regressão com ponto de mudança nos coeficientes de regressão. Primeramente, a identificação das observações atípicas será feita considerando a distância de Mahalanobis. A Figura 5.12 mostra estas distâncias para os modelos estudados. Nota-se que as observações 20, 45, 46, 47, 48 e 65 aparecem como "*outliers*"para todos os modelos considerados.

A flexibilidade das estimativas de máxima verossimilhança dos modelos sob distribuições com caudas mais pesadas contra observações atípicas é evidenciada na Figura 5.13 que apresenta os pesos estimados (q_i) para ($i = 1, \dots, n$) provenientes da aplicação do algoritmo EM versus as distâncias de Mahalanobis, onde se observa pesos menores para as observações 20, 45, 46, 47, 48 e 65 consideradas "*outliers*". Desta forma, o processo de estimação de parâmetros atribui pesos menores as observações anômalas. Para o modelo sob distribuição Normal,

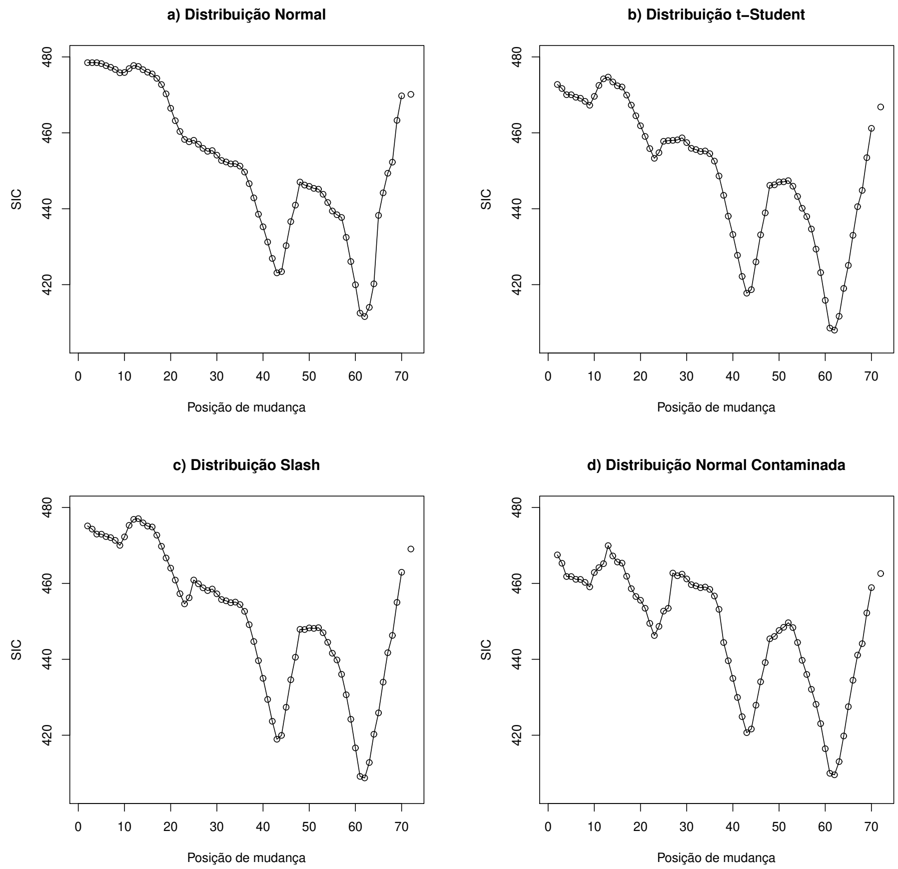


Figura 5.10: Dados Ibope 01: SIC para ajustes com diversas distribuições. Modelo de regressão com mudança nos coeficientes de regressão.

temos que $q_i = 1$ para $i = 1, \ldots, n$, mostrados como uma linha tracejada nos gráficos apresentados.

Finalmente, as observações influentes no conjunto de dados Ibope 01 serão identificadas utilizando a quantidades $M(0)_l$ para $l = 1, 2, .., g$ calculadas através da curva conformal $B_{f_Q, \mathbf{d}_l}$ segundo a metodologia detalhada na Seção 4.2 e considerando os esquemas ponderação de casos, perturbação na variável explicativa e perturbação na variável resposta. Em todos os casos considera-se a marca de referência proposta por Lee & Xu (2004) com $c* = 3$. Os esquemas especiais de perturbação para modelos com ponto de mudança são utilizados em todas as situações.

As figuras 5.14-5.16 apresentam os gráficos de $M(0)$ para o modelo de regressão com ponto de mudança nos coeficientes de regressão sob as quatro distribuições estudadas e os três esquemas mencionados, considerando o primeiro esquema especial (para as observações 1-62). Os resultados correspondentes ao segundo esquema especial não serão apresentados neste trabalho por brevidade, pois nenhuma observação influente foi identificada.

Tabela 5.6: Dados Ibope 01: Estimativa dos parâmetros para modelos de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições.

Modelo	Casos	β_0	β_1	σ^2
Normal	1-62	-1.650	0.670	13.221
		(1.3126)	(0.0498)	(2.2035)
	63-72	16.081	0.521	
		(3.9785)	(0.1245)	
t de Student	1-62	-2.305	0.685	5.996
(ν=3.114)		(1.0840)	(0.0412)	(1.4155)
	63-72	15.741	0.509	
		(2.9129)	(0.0921)	
Slash	1-62	-2.395	0.686	3.389
(ν=1.068)		(1.0271)	(0.0389)	(0.7987)
	63-72	16.126	0.498	
		(2.8944)	(0.0912)	
Normal contaminada	1-62	-2.387	0.688	4.278
(ν=0.319,γ=0.130)		(1.0025)	(0.0385)	(0.9094)
	63-72	16.107	0.494	
		(2.7636)	(0.0856)	

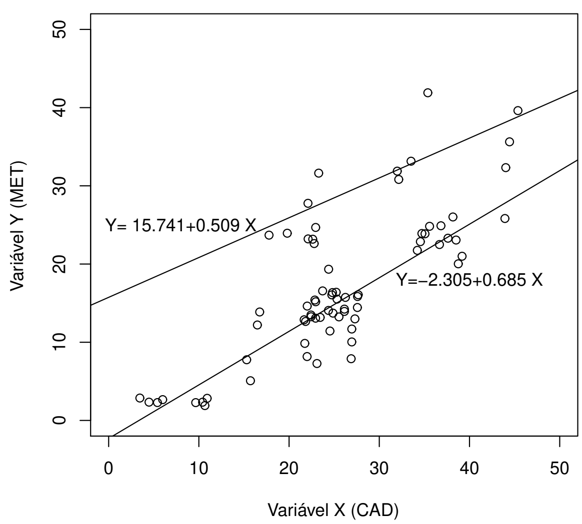


Figura 5.11: Dados Ibope 01: Ajuste do modelo de regressão com ponto de mudança nos coeficientes de regressão sob distribuição t de Student.

Observando as figuras, temos que a observação 61 é influente para as estimativas de máxima verossimilhança sob ponderação de casos e perturbação na variável resposta para os modelos sob distribuição t de Student, Slash e Normal Contaminada. Enquanto que sob distribuição Normal, as observações 61 e 62 são consideradas influentes somente para ponderação de casos. Por outro lado, quando avaliamos o esquema de perturbação na variável explicativa, somente o modelo sob

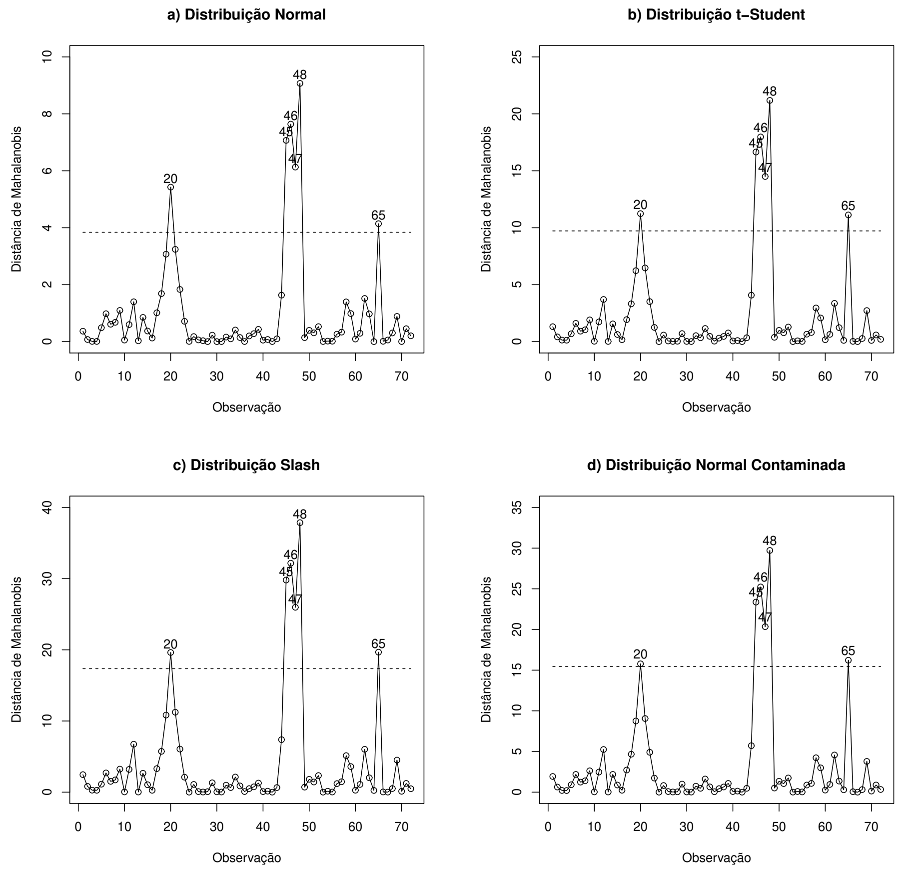


Figura 5.12: Dados Ibope 01: Distância de Mahalanobis para o modelo de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições, considerando $\xi = 0.95$.

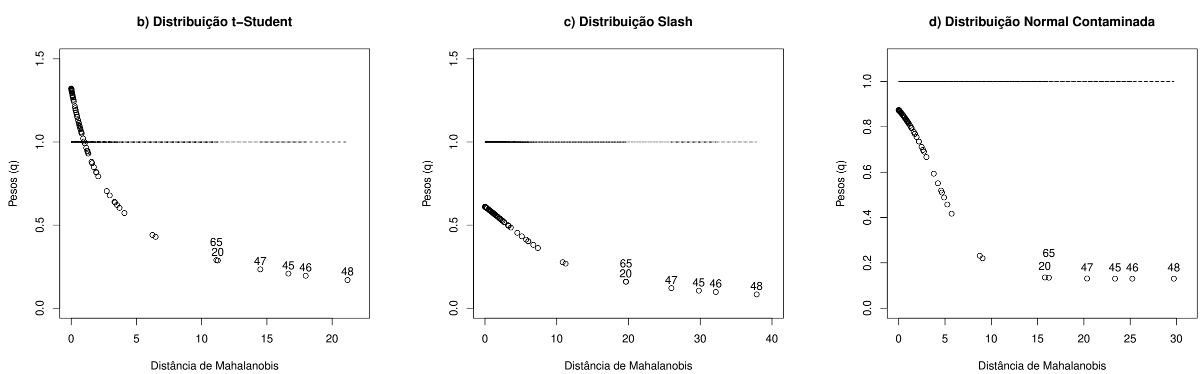


Figura 5.13: Dados Ibope 01: Distância de Mahalanobis versus q_i para o modelo de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições.

distribuição Normal mostra observações influentes (observações 46 e 48).

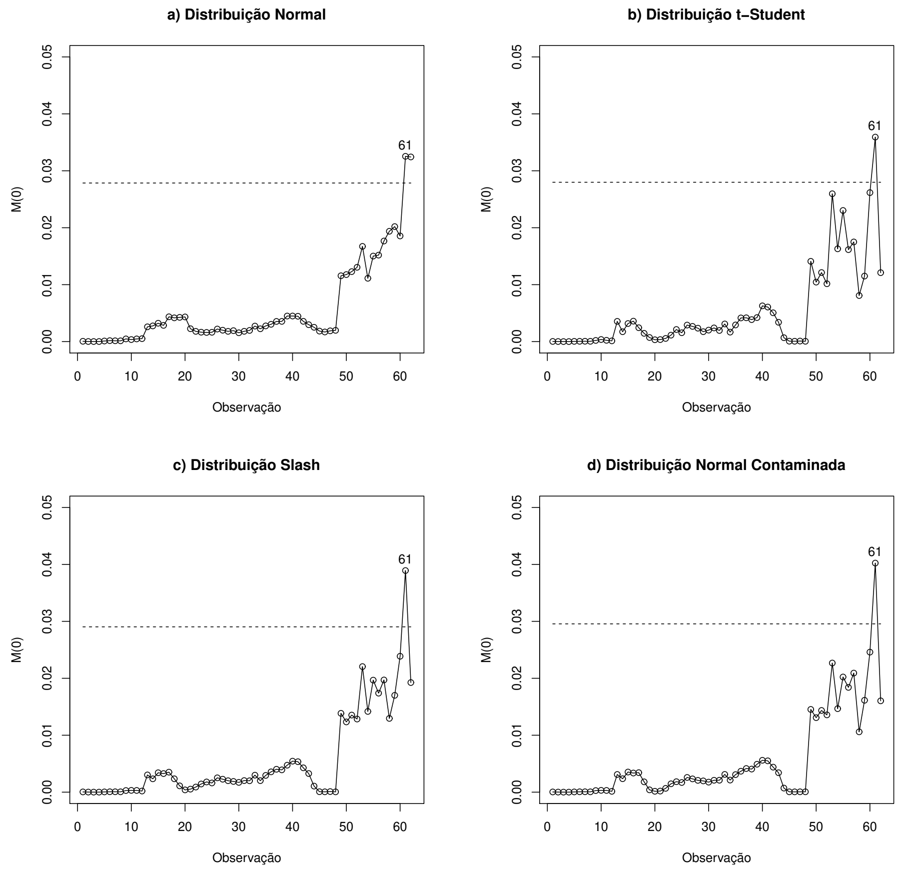


Figura 5.14: Dados Ibope 01: Gráficos $M(0)$ considerando esquema ponderação de casos na primeira situação especial para modelos de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições consideradas.

As observações 61 e 62 apresentam os valores mais altos de audiência até esse momento do dia segundo a medição via Caderno (CAD) e próximos as valores de audiência das observações 63 e 64. No gráfico de dispersão (Figura 5.11), estas observações correspondem aos pontos extremos à direita.

2) Conjunto de dados IBOPE 02

Para este conjunto de dados, consideramos o modelo de regressão com mudança na variância. O procedimento SIC proposto é aplicado para a determinação do ponto de mudança, sob as quatro distribuições pertencentes à classe de Mistura de Escala Normal estudadas nesta dissertação. O resumo dos resultados se apresentam na Tabela 5.7.

Observa-se que ao considerar o modelo sob qualquer uma das distribuições consideradas o ponto

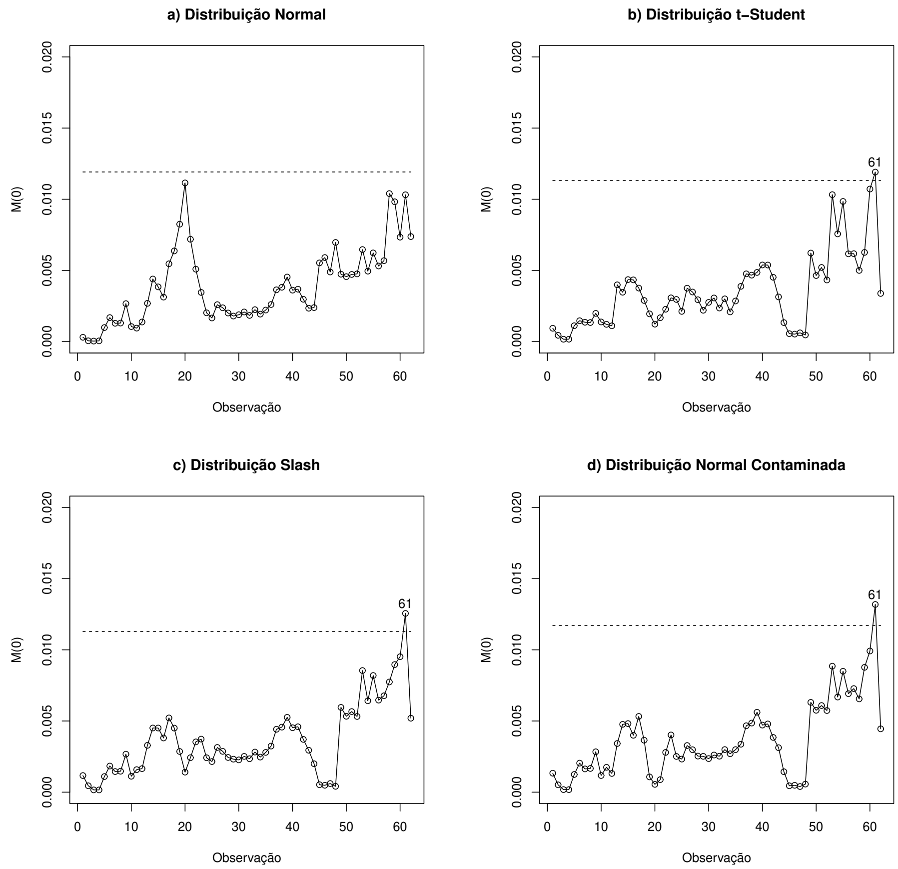


Figura 5.15: Dados Ibope 01: Gráficos $M(0)$ considerando esquema perturbação da variável explicativa na segunda situação especial para modelos de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições consideradas.

Tabela 5.7: Dados Ibope 02: Resultados para critério SIC para modelo de regressão linear com mudança na variância.

Modelo	$minSIC(k)$	$SIC(n)$	$\widehat{k}$
Normal	395.859	403.161	8
t de Student	398.751	404.235	8
	(ν=12.104)	(ν=7.113)	
Slash	398.193	403.983	8
	(ν=2.476)	(ν=2.021)	
Normal contaminada	401.627	407.655	8
	(ν=0.046,γ=0.154)	(ν=0.057,γ=0.155)	

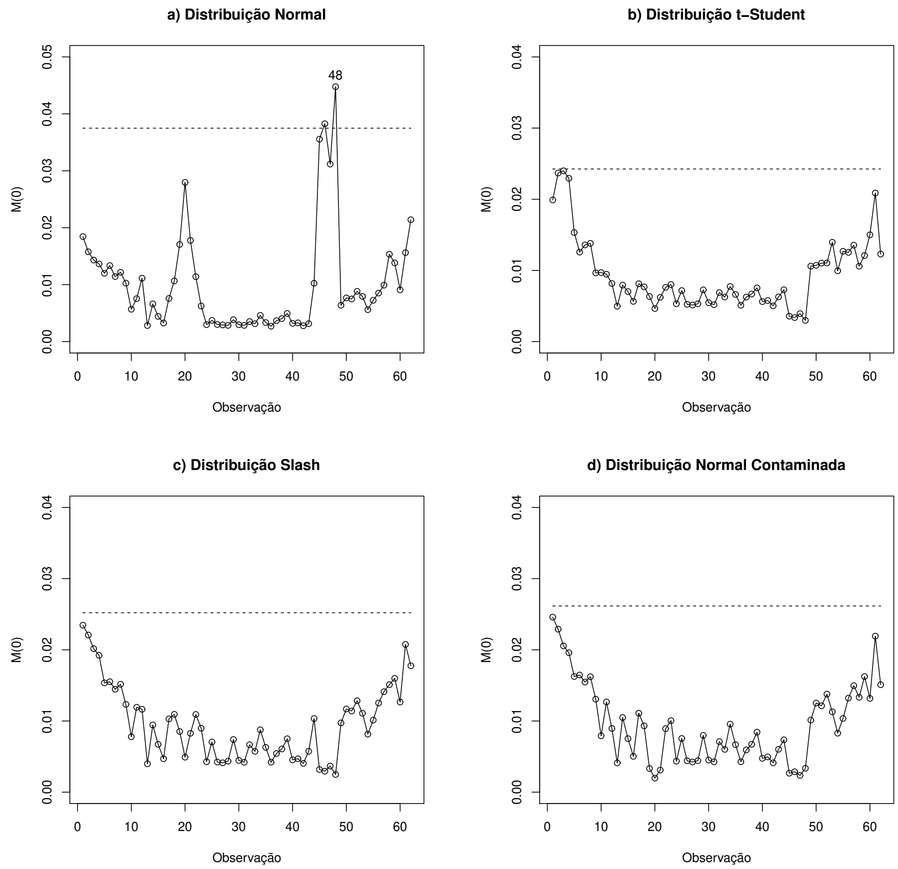


Figura 5.16: Dados Ibope 01: Gráficos $M(0)$ considerando esquema perturbação da variável resposta na segunda situação especial para modelos de regressão com ponto de mudança nos coeficientes de regressão sob distintas distribuições consideradas.

de mudança encontra-se na posição $k = 8$. Utilizando o critério SIC, observamos que qualquer dos modelos considerados são adequados e destes, o modelo sob distribuição Normal com ponto de mudança na posição $k = 8$ é o que melhor se ajusta aos dados. Por sua vez, este ponto corresponde ao intervalo horário de $07:45 - 08:00$ horas. A Figura 5.17 apresenta as audiências medidas por cada um dos instrumentos ao longo do dia, a posição $k = 8$ identifica claramente o ponto a partir da qual se produz uma diferença nos comportamentos de audiência medidos pelos dois instrumentos. Até o ponto em questão a medição via Meter apresenta os valores mais baixos do dia e mais próximos dos valores medidos via Caderno. A Figura 5.18 mostra os valores do critério de informação de Schwarz (SIC) contra a posição de mudança para todas as distribuições consideradas.

A Tabela 5.8 contém as estimativas de máxima verossimilhança para os parâmetros dos quatro modelos propostos acompanhados de seus respectivos desvios padrões (entre parênteses) obtidos via matriz de informação observada. A Figura 5.19 apresenta o gráfico de dispersão com o modelo de regressão com ponto de mudança na variância ajustado. Observe que as oito

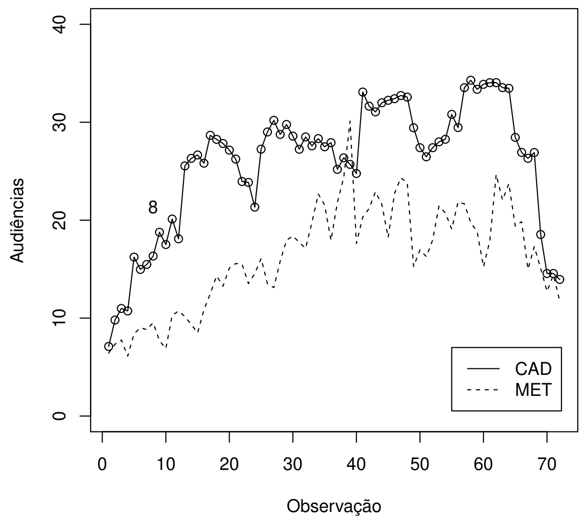


Figura 5.17: Dados Ibope 02: Medições de audiência via instrumentos Meter (MET) e Caderno (CAD) por observações ordenadas segundo horário de medição.

primeiras observações (identificadas no gráfico com os números 1-8) mostram uma variabilidade muito menor que o resto de observações do conjunto de dados.

Tabela 5.8: Dados Ibope 02: Estimativa dos parâmetros para modelos de regressão com ponto de mudança na variância sob distintas distribuições.

Modelo	Casos	β_0	β_1	σ^2
Normal	1-8	0.602	0.596	1.247
		(0.7559)	(0.0378)	(0.6556)
	9-72			14.847
				(2.6421)
t de Student	1-8	0.589	0.596	1.170
(ν=12.104)		(0.7784)	(0.0376)	(0.6519)
	9-72			12.112
				(2.4051)
Slash	1-8	0.614	0.594	0.852
(ν=2.476)		(0.7583)	(0.0368)	(0.4646)
	9-72			8.688
				(1.7295)
Normal contaminada	1-8	0.644	0.592	1.165
(ν=0.046,γ=0.154)		(0.7363)	(0.0358)	(0.6175)
	9-72			11.677
				(2.2196)

O método de influência local proposto no Capítulo 4 é utilizado sobre o modelo de regressão

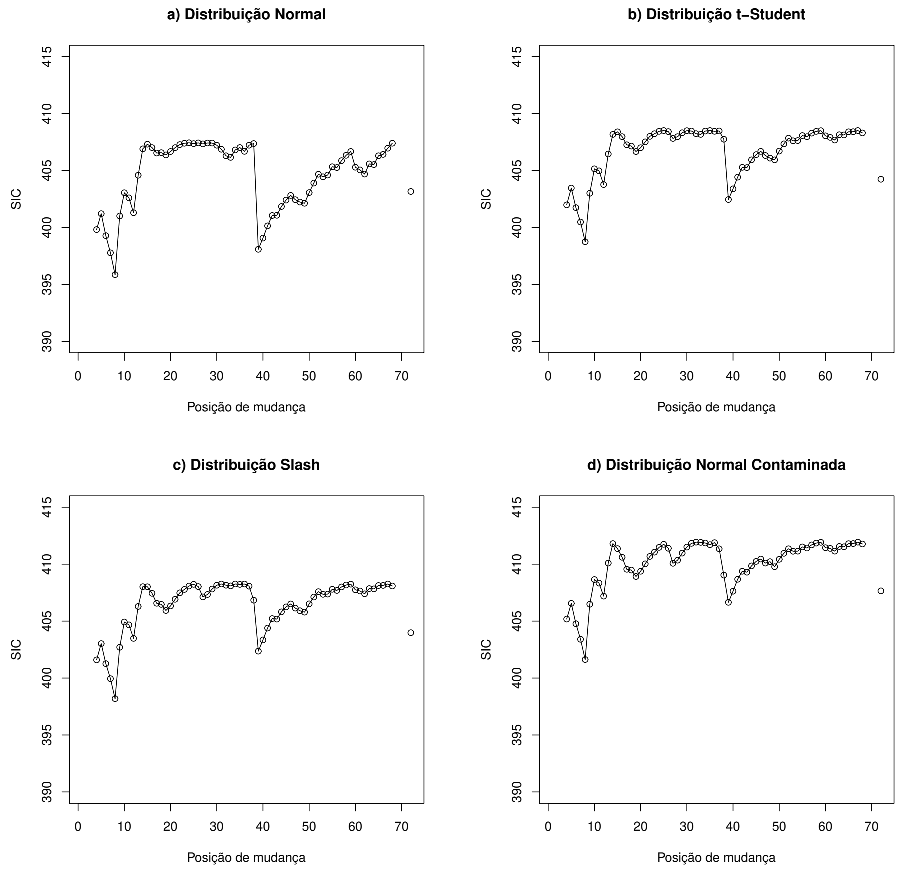


Figura 5.18: Dados Ibope 02: SIC para ajustes com diversas distribuições. Modelo de regressão com mudança na variância.

com ponto de mudança na variância. Primeramente, a identificação das observações atípicas será feita considerando a distância de Mahalanobis. A Figura 5.20 mostra estas distâncias para os modelos estudados sob as quatro distribuições consideradas. Nota-se que as observações 15 e 39 aparecem como "*outliers*"para todos os modelos considerados.

A flexibilidade das estimativas de máxima verossimilhança dos modelos sob distribuições com caudas mais pesadas contra observações atípicas é evidenciada na Figura 5.21 que apresenta os pesos estimados (q_i) para ($i = 1, \ldots, n$) provenientes da aplicação do algoritmo EM, versus as distâncias de Mahalanobis, onde se observa pesos menores para as observações 15 e 39 identificadas como "*outliers*". Desta forma, o processo de estimação de parâmetros atribui pesos menores as observações anômalas. Para o modelo sob distribuição Normal, temos que $q_i = 1$ para $i = 1, \ldots, n$, mostrados como uma linha tracejada nos gráficos apresentados.

Finalmente, as observações influentes no conjunto de dados Ibope 02 serão identificadas utilizando a quantidades $M(0)_l$ para $l = 1, 2, .., g$ calculadas através da curva conformal $B_{f_Q,\mathbf{d}_l}$

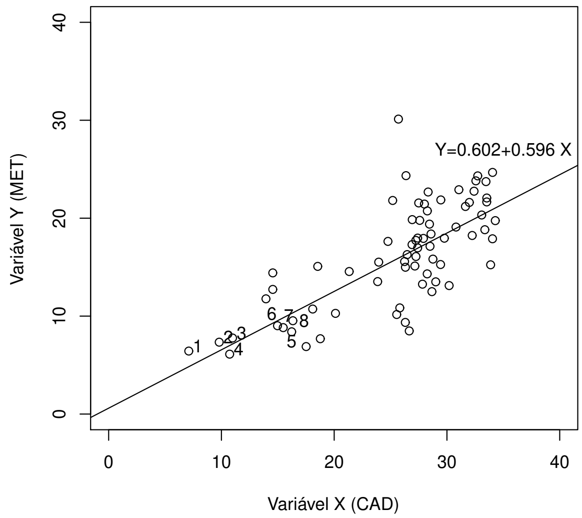


Figura 5.19: Dados Ibope 02: Ajuste do modelo de regressão com ponto de mudança na variância sob distribuição t de Student.

segundo metodologia detalhada na Seção 4.2 e considerando os esquemas ponderação de casos, perturbação na variável explicativa e perturbação na variável resposta. Em todos os casos considera-se a marca de referência proposta por Lee & Xu (2004) com $c* = 3$. Os esquemas especiais de perturbação para modelos com ponto de mudança são utilizados em todas as situações.

As Figuras 5.22-5.24 apresentam os gráficos de $M(0)$ para o modelo de regressão com ponto de mudança na variância sob as quatro distribuições estudadas e três esquemas mencionados, considerando o segundo esquema especial (para as observações 9-72). Os resultados correspondentes ao primeiro esquema especial não identificam observações influentes por isso não são apresentados, neste trabalho, por brevidade.

Observando as figuras vemos que não existe nenhuma observação influente para as estimativas de máxima verossimilhança sob ponderação de casos para os modelos sob distribuição Normal, t de Student e Slash e Normal Contaminada. Por outro lado, quando avaliamos o esquema de perturbação na variável explicativa e perturbação na variável resposta somente o modelo sob distribuição Normal mostra uma observação influênte (observação 39), também identificada ao observar as distâncias de Mahalanobis). Nenhum dos modelos sob caudas pesadas apresenta observações influentes.

A observação 39 corresponde ao valor mais alto de audiência do dia segundo a medição via Meter (MET) o qual é o único que supera às medições obtidas via Caderno (CAD). Note que os modelos sob caudas pesadas lidam melhor com a presença desta observação. No gráfico de dispersão (Figura 5.19), esta observação corresponde ao ponto extremo na parte superior.

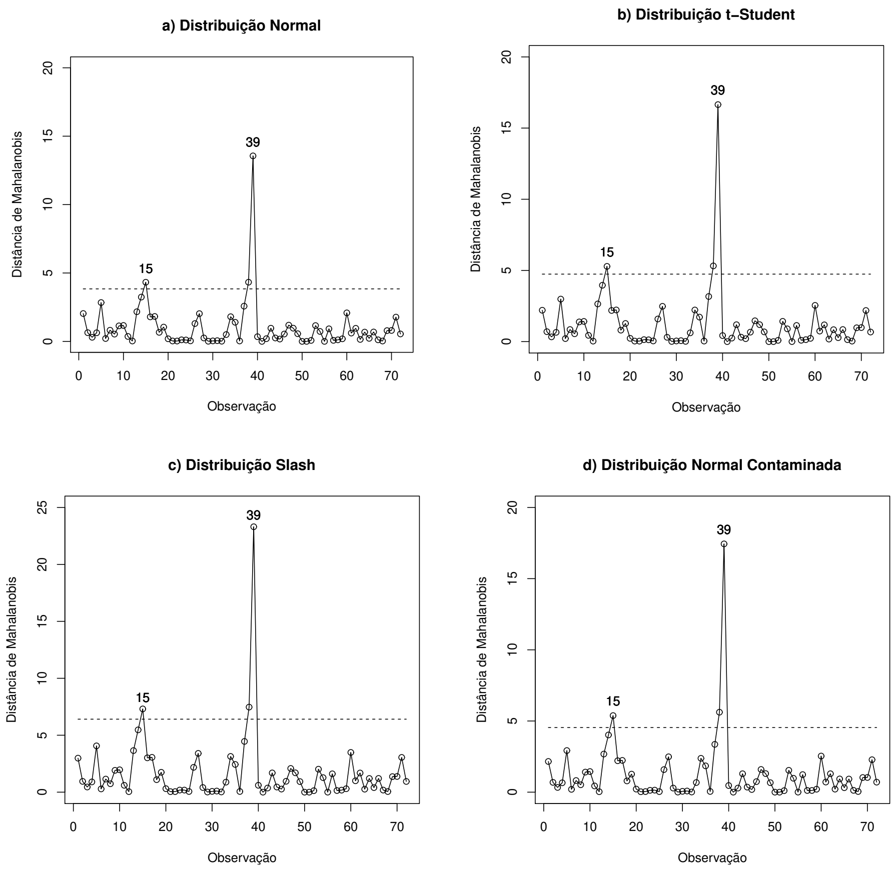


Figura 5.20: Dados Ibope 02: Distância de Mahalanobis para o modelo de regressão com ponto de mudança na variância sob distintas distribuições, considerando $\xi = 0.95$.

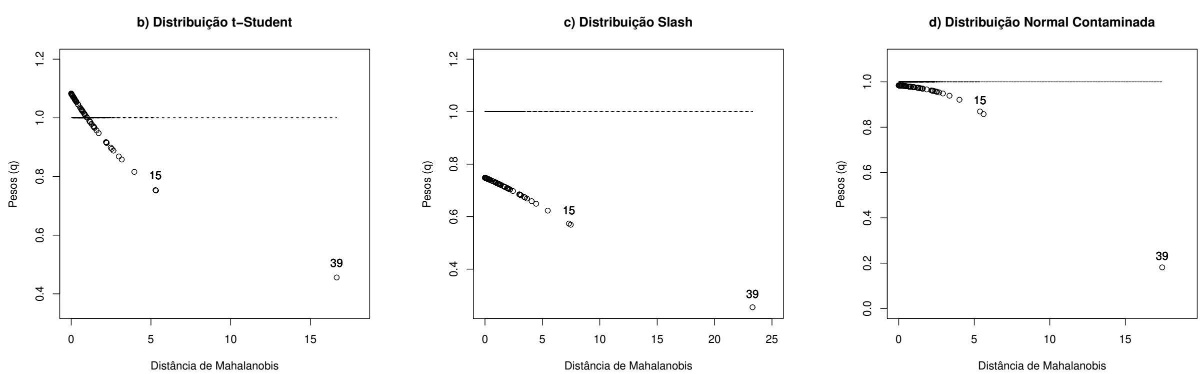


Figura 5.21: Dados Ibope 02: Distância de Mahalanobis versus q_i para o modelo de regressão com ponto de mudança na variância sob distintas distribuições.

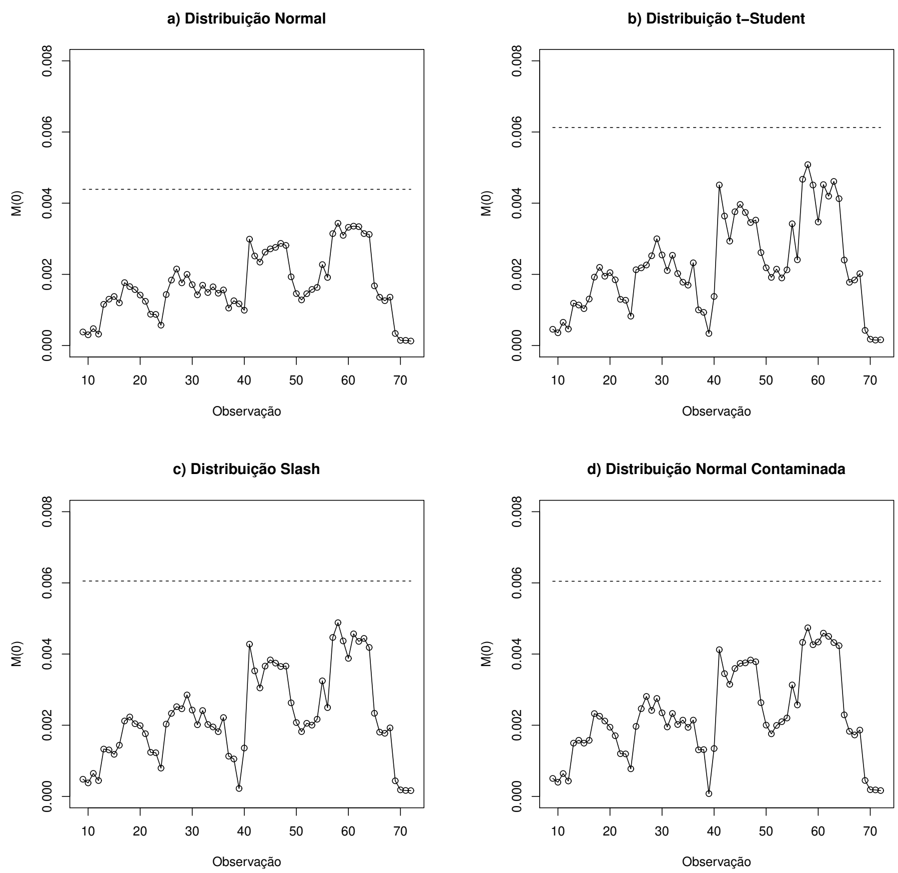


Figura 5.22: Dados Ibope 02: Gráficos $M(0)$ considerando esquema ponderação de casos na primeira situação especial para modelos de regressão com ponto de mudança na variância sob distintas distribuições consideradas.

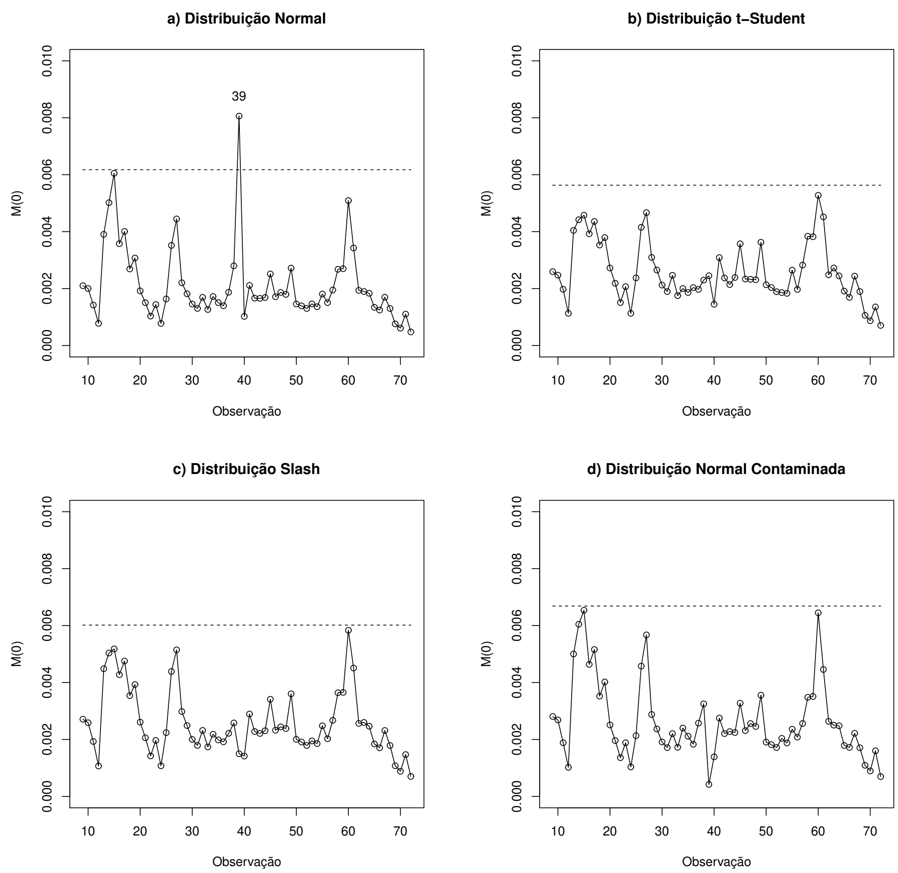


Figura 5.23: Dados Ibope 02: Gráficos $M(0)$ considerando esquema perturbação da variável explicativa na segunda situação especial para modelos de regressão com ponto de mudança na variância sob distintas distribuições consideradas.

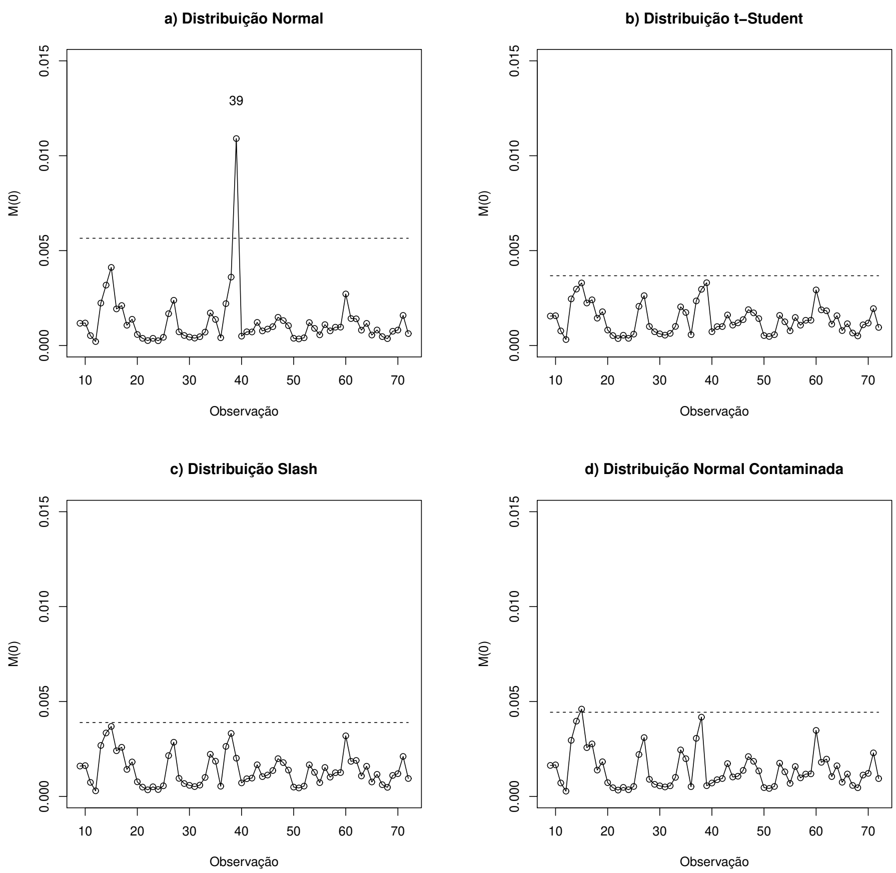


Figura 5.24: Dados Ibope 02: Gráficos $M(0)$ considerando esquema perturbação da variável resposta na segunda situação especial para modelos de regressão com ponto de mudança na variancia sob distintas distribuições consideradas.

Capítulo 6

Considerações Finais

6.1 Conclusões

Neste trabalho considera-se o estudo estimação e diagnóstico em modelos de regressão Mistura de Escala Normal com um ponto de mudança. Os modelos baseados nesta classe de distribuições apresentam alternativas robustas para modelos desenvolvidos sob normalidade. Três casos com relação ao ponto de mudança são estudados: modelos de regressão com mudança nos coeficientes de regressão, modelos de regressão com mudança na variância e modelos de regressão com mudança simultânea nos coeficientes de regressão e variância.

No Capítulo 3, apresenta-se a estimação de máxima verossimilhança no modelo de regressão linear com ponto de mudança sob Mistura de Escala Normal. Resultados específicos para cada um dos modelos estudados assim como a implementação de algoritmo EM utilizando a formulação hierárquica são desenvolvidos para estes modelos. Também é apresentada a metodologia para seleção do modelo e determinação do ponto de mudança através do critério de informação de Schwarz (SIC) que se mostra válida para os três tipos de modelos estudados.

No Capítulo 4, um enfoque de influência (Cook, 1986) é considerado no modelo de regressão Mistura de Escala Normal com ponto de mudança. A complexidade na função de verossimilhança dos dados observados leva ao uso da função Q (Zhu & Lee, 2001), proveniente do algoritmo EM. Para cada um dos modelos estudados foram obtidos os elementos para obter a curvatura normal considerando três esquemas de perturbação: ponderação de casos, perturbação sobre a variável explicativa e perturbação sobre a variável resposta. Uma avaliação sobre valores de referência ("*benchmark*") utilizados para avaliar se uma observação do conjunto de dados é potencialmente influente foi realizada, o valor proposto por Lee & Xu (2004) foi comparado com uma alternativa computacional via Simulação Monte Carlo. Pelo tempo computacional o uso da proposta de Lee & Xu (2004) parece ser uma alternativa computacional apropriada.

Finalmente, a metodologia desenvolvida é aplicada a exemplos reais que incluem um conjunto de dados conhecido na literatura estatística sobre volume de vendas dos mercados de ações (Holbert, 1982) e dois conjuntos de dados sobre medição de audiência de televisão no Brasil. Estas aplicações, apresentadas no Capítulo 5, mostram que os resultados desenvolvidos neste trabalho são úteis para a análise de dados e espera-se que desperte o interesse de estudantes, pesquisadores e profissionais pelo tema, que consideramos ser de grande aplicabilidade.

6.2 Perspectivas Para Trabalhos Futuros

O modelo de regressão linear simples com ponto de mudança considerado neste trabalho assume que a variável dependente é fixa. No entanto, no caso específico dos dados de medição de audiência de televisão é possível supor que ambas variáveis (independente e indepentente) possuem um erro por proceder de medições de dois instrumentos diferentes. Nesse caso, uma opção alternativa para este problema é considerar um modelo de regressão linear com erro de medida. A aplicação nos dados sobre volume de vendas dos mercados de ações (Holbert, 1982) já foi realizada por Chang & Huang (1997) considerando um modelo de regressão com erro de medida normal. Assim, uma extensão para modelos de regressão linear Mistura de escala normal com erro de medida e ponto de mudança deverá ser realizada.

Da mesma forma, o trabalho apresentado desenvolve a metodologia para a presença de um único ponto de mudança, assim este pode ser estendido para modelos com mais de um ponto de mudança.

Por outro lado, o método de estimação proposto para modelos de regressão linear com ponto de mudança Mistura de Escala Normal inclui uma optimização dos graus de liberdade das distribuições considerando o conjunto de dados total. Um estudo considerando a optimização dos graus de liberdade separadamente em cada partição do conjunto de dados definida pela posição do ponto de mudança pode ser realizado para avaliar os efeitos sobre a estimação. Também, podem ser realizados estudos similares para modelos de regressão multivariados.

Finalmente, a metodologia de estimação e diagnóstico apresentada para modelos de regressão com um ponto de mudança pode ser estendida considerando outra classe de distribuições que incluem classes de distribuições assimétricas.

Bibliografia

[1] Andrews, D. F. & Mallows S. L. (1974). Scale mixtures of normal distributions. *Jornal of the Royal Statistical Society*, **ser. B**, 36, 99-102.

[2] Arellano-Valle, R. B. (1994). *Distribuições Elipticas: Propriedades, Inferência e Aplicações a Modelos de Regressão*. Tese de doutorado - IME - Universidade de São Paulo.

[3] Bhatti, M. I. & Wang, J. (2000). On Testing for a Change-Point in Variance of Normal Distribution. *Biometrical Journal*, **Vol. 42**, 8, 1021-1032.

[4] Box, G. E. P. & Tiao, G. C. (1973)*Bayesian Inference in Statical Analysis*. Reading, Mass: Addison-Wisley.

[5] Brodsky, B. E. & Darkhovsky, B. S. (1993)*Nonparametric Methods in Change Point Problems* . Kluwer Academic Publishers, Netherlands.

[6] Brown, R. L., Durbin, J. & Evans J. M. (1975). Techniques for testing tue constancy of regression relationships over time (with discussion). *Jornal of the Royal Statistical Society*, **ser. B**, 37, 149-192.

[7] Chan Y. P. & Huang, W. T. (1997). Inferences for the linear errors-in-variables with change-point models. *Journal of the American Statistical Association*, **92**, 10, 171-178.

[8] Chen, J. (1998). Testing for a change point in linear regression models. *Comunication in Statistics-Theory and Methods*, **vol. 27**, 10, 2481-2493.

[9] Chen, J. & Gupta, A. K. (1995). Likelihood procedure for testing change points hypothesis for multivariate gaussian model. *Random Operators and Stochastic Equations*, **3**, 235-244.

[10] Chen, J. & Gupta, A. K. (1997). Testing and locating variance change point with application to stock prices. *Journal of the American Statistical Association*, **92**, 739-747.

[11] Chen, J. & Gupta, A. K. (1999). Change point analysis of a Gaussian model. *Statistical Papers*, **40**, 323-333.

[12] Chen, J. & Gupta, A. K. (2001). On change point detection and estimation. *Communications in Statistics - Simulation and Computation*, **vol. 30**, 3, 665-697.

[13] Chernoff, H. & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *The Annals of Mathematical Statistics*, **35**, 999-1018.

[14] Chin Choy, J. H., & Broemeling, L. D. (1980). Some Bayesian inferences for a changing linear model. *Technometrics*, **22**, 71-78.

[15] Cohen, Y. & Cohen, J. Y. (2008)*Statistics and Data with R: An applied approach through examples*. Wiley, United Kingdom.

[16] Cook, R. D. (1986). Assessment of local influence. *Jornal of the Royal Statistical Society*, **ser. B**, 48, 133-169.

[17] Cook, R. D. & Weisberg, S. (1982)*Residual and Influence in Regression*. Wiley, New York.

[18] Csörgő, M. & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, New York.

[19] Dempster, A. P., Lair, N. M. & Rubin, D. B. (1977). Maximum Likelihood from incomplete data via the EM algorithm. *Jornal of the Royal Statistical Society*, **ser. B**, 1, 1-22.

[20] Dempster, A. P., Lair, N. M. & Rubin, D. B. (1980). Iteratively reweighted least squares for linear regression when errors are normal/independent distributed. In: V. Krishnaiah (Ed.). *In Multivariate Analysis*, **Vol. V**, 35-57, North-Holland, Amsterdam.

[21] Draper, N. R. & Smith, H. (1998)*Applied Regression Analysis*. Wiley, New York.

[22] Everitt, B. S. & Hothorn, T. (2010)*A handbook of statistical analyses using R*. Taylor and Francis, New York.

[23] Fang, K. T., Kotz, S. & Ng, K. W. (1990)*Symmetric multivariate and related distribution*. Chapman and Hall, London.

[24] Fernandez, C. & Steel, M. (1999). Multivariate student t regression models: pitfalls and inference. *Biometrika*, **86**, 153-167.

[25] Fernandez, C. & Steel, M. (2000). Bayesian regression analysis whith Scale Mixture of Normals. *Econometric Theory*, **16**, 80-101.

[26] Ferreira, P. E. (1975). A Bayesian analysis of a switching regression model: Known number of regimes. *Journal of the American Statistical Association*, **70**, 370-374.

[27] Gardner, L. A. (1969). On detecting change in the mean of normal variates . *Annals of Mathematical Statistics*, **40**, 116-126.

[28] Gomes de Souza Tu, D. (2004). *Regressão com Erros de Medida e Pontos de Mudança Utilizando Metodologia Bayesiana*. Tese de doutorado. IME - Universidade de São Paulo.

[29] Hawkins, D. L. (1989). A U-I aproach to retrospective testing for shifting parameters in a linear model. *Comunication in Statistics*, **18**, 3117-3134.

[30] Hofrichter, J. (2007). *Change Point Detection in Generalized Linear Models.* Dissertation zur Erlangung des akademischen Grades Doktor der technischen Wissenschaften. Graz University of Technology.

[31] Holbert, D. (1982). A bayesian analysis of a switching linear model. *Journal of Econometrics*, **19**, 77-87.

[32] Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics*, **21**, 671-680.

[33] Kim, H. J. & Siegmund, D. (1989). Likelihood Ratio Tests for a Change-Point in Simple Linear Regression. *Biometrika*, **76**, 409-423.

[34] Kim, H. J. (1994). The Likelihood Ratio and Cusum Tests for a Change-Point in Linear Regression. *Journal of Multivariate Analysis*, **51**, 54-70.

[35] Lachos, V. H. (2002). *Influência e diagnóstico em modelos de Grubbs.* Tese de mestrado. IMECC - Universidade Estadual de Campinas.

[36] Lachos, V. H., Angolini, T., & Abanto-Valle, C. A. (2011). On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers*, **vol. 52**, 3, 567-590.

[37] Lange, K. & Sinsheimer, J. S. (1993). Normal/independent distribution and their applications in robust regression. *Journal of the American Statistics Association*, **2**, 175-198.

[38] Lange, K. L., Little, J. A. & Taylor, M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881-896.

[39] Lee, S. Y. & Xu, L. (2004). Influence analysis of nolinear mixed-effects models. *Computational Statistics and Data Analysis*, **45**, 321-341.

[40] McGee, V. E. & Carleton, W. T. (1970). Piecewise Regression. *Journal of the American Statistical Association*, **65**, 1109-1124.

[41] Meng, X. & D. B. Rubin, (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80**, 267-278.

[42] Ortega, E. M., Bolfarine, H. & Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis*, **42**, 165-186.

[43] Osorio, F. (2006). *diagnóstico de influência em modelos elípticos com efeitos mistos.* Tese de doutorado. IME - Universidade de São Paulo.

[44] Osorio, F. & Galea, M. (2006). Detection of a change-point in student-t linear regression models. *Statistical Papers*, **47**, 31-48.

[45] Pinheiro, J. C., Liu, C. H. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixes-effects using multivariate t-distribution. *Jornal of Computational and Graphical Statistics*, **10**, 249-276.

[46] Poon, W & Poon Y. S. (1999). Conformal normal curvature and assessment of local influence. *Jornal of the Royal Statistical Society*, **ser. B**, 61, 51-61.

[47] Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**, 873-880.

[48] Quandt, R. E. (1960). Tests of the hypothesis that a linear regression obeys two separate regimes. *Journal of the American Statistical Association*, **55**, 324-330.

[49] Rosa, G., Padovani, C. & Gianola, D. (2003). Robust Linear Mixed Models whith Normal/Independent Distribution and Bayesian MCMC Implementation. *Biometrical Journal*, **45**, 573-590.

[50] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

[51] Srivastava, M. S. (1975). On Tests for Detecting Change in Mean. *Annals of Statistics*, **3**, 98-108.

[52] Srivastava, M. S. & Worsley, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *Journal of the American Statistical Association*, **81**, 199-204.

[53] Verveke, G. & Molenberghs, G. (1997)*Linear Mixed Models in Practice*. Springer-Verlag. New York.

[54] Worsley, K. J. (1979). On the likelihood ratio test for a shift in locations of normal population. *Journal of the American Statistical Association*, **74**, 365-367.

[55] Zeller, C. B. (2009). *Distribuições de Misturas de Escala Skew-Normal: Estimação e diagnóstico em Modelos Lineares*. Tese de doutorado. IMECC - Universidade Estadual de Campinas.

[56] Zevallos, M., Santos, B. & Hotta, L. K. (2012). A note on influence diagnostics in AR(1) time series models. *Jornal of Statistical Planning and Inference*, **142**, 2999-3007.

[57] Zhang, X. & King, M. L. (2005). Influence diagnostics in generalize autoregressive condicional heteroscedasticity processes. *Jornal of Business and Economics Statistics*, **ser. 23**, 118-129.

[58] Zhu, H. T. & Lee, S. Y. (2001). Local Influence for incomplete data models. *Jornal of the Royal Statistical Society*, **ser. B**, 63, 111-126.

Apêndice A

Dados Utilizados em Aplicações

São apresentados os três conjuntos de dados utilizados nas aplicações do Capítulo 5. Cada um deles e descrito detalhadamente.

A Tabela A.1 apresenta o conjunto de dados referido ao volume de vendas (em milhões de dólares) dos mercados de ações de New York (NYAMSE) e de Boston (BSE), denominado também Dados Holbert. As informações correspondem a registros mensais entre Janeiro de 1967 e Novembro de 1969.

A Tabelas A.2-A.3 apresentam os dois conjuntos de informações de audiências de televisão (em número de domicílios $\times$ 10000) medidos via Caderno (CAD) e Meter (MET) referentes a dois dias de medição em uma área geográfica definida para um estudo interno específico. As informações correspondem a registros de intervalos de 15 minutos contínuos iniciados às 06:00 horas até às 24:00 horas do dia de medição. Estes conjuntos de dados são denominados Dados Ibope 01 e 02, respetivamente.

Tabela A.1: Dados de volume de vendas de mercado de ações.

Ordem	Data	NYAMSE	BSE
1	jan 67	10581.6	78.8
2	fev 67	10234.3	69.1
3	mar 67	13299.5	87.6
4	abr 67	10746.5	72.8
5	mai 67	13310.7	79.4
6	jun 67	12835.5	85.6
7	jul 67	12194.2	75.0
8	ago 67	12860.4	85.3
9	set 67	11955.6	86.9
10	out 67	13351.5	107.8
11	nov 67	13285.9	128.7
12	dez 67	13784.4	134.5
13	jan 68	16336.7	148.7
14	fev 68	11040.5	94.2
15	mar 68	11525.3	128.1
16	abr 68	16056.4	154.1
17	mai 68	18464.3	191.3
18	jun 68	17092.2	191.9
19	jul 68	15178.8	159.6
20	ago 68	12774.8	185.5
21	set 68	12377.8	178.0
22	out 68	16856.3	271.8
23	nov 68	14635.3	212.3
24	dez 68	17436.9	139.4
25	jan 69	16482.2	106.0
26	fev 69	13905.4	112.1
27	mar 69	11973.7	103.5
28	abr 69	12573.6	92.5
29	mai 69	16566.8	116.9
30	jun 69	13558.7	78.9
31	jul 69	11530.9	57.4
32	ago 69	11278.0	75.9
33	set 69	11263.7	109.8
34	out 69	15649.5	129.2
35	nov 69	12197.1	115.1

Tabela A.2: Dados de audiências de televisão (Parte 1).

		Conjunto de Dados 01		Conjunto de Dados 02	
Ordem	Horário	CAD	MET	CAD	MET
1	06:00-06:15	3.4578	2.8644	7.1051	6.4323
2	06:15-06:30	4.4766	2.3379	9.8041	7.3395
3	06:30-06:45	5.4072	2.2868	10.9862	7.7661
4	06:45-07:00	5.9894	2.6599	10.7304	6.1158
5	07:00-07:15	9.6501	2.2786	16.2302	8.3997
6	07:15-07:30	10.6689	1.9000	14.9732	9.0191
7	07:30-07:45	10.9204	2.8289	15.4805	8.8261
8	07:45-08:00	10.4131	2.3353	16.3362	9.5295
9	08:00-08:15	15.7186	5.0769	18.7703	7.6908
10	08:15-08:30	15.2996	7.7590	17.5134	6.8916
11	08:30-08:45	16.4859	12.2066	20.1155	10.2806
12	08:45-09:00	16.7417	13.8694	18.0955	10.7227
13	09:00-09:15	24.3670	14.0671	25.5447	10.1636
14	09:15-09:30	24.5302	11.4360	26.3078	9.3565
15	09:30-09:45	25.5313	13.2406	26.6562	8.4876
16	09:45-10:00	24.8651	13.7227	25.8182	10.8406
17	10:00-10:15	27.2998	12.9871	28.6676	12.4985
18	10:15-10:30	26.9514	11.6891	28.2486	14.3143
19	10:30-10:45	26.9514	10.0387	27.8296	13.2613
20	10:45-11:00	26.8808	7.8930	27.1548	15.1137
21	11:00-11:15	23.0924	7.2787	26.2420	15.5744
22	11:15-11:30	21.9986	8.1699	23.9485	15.5128
23	11:30-11:45	21.7428	9.8512	23.8559	13.5258
24	11:45-12:00	21.8311	12.6583	21.3330	14.5556
25	12:00-12:15	22.0120	14.6258	27.2512	16.0858
26	12:15-12:30	23.4498	13.1750	29.0068	13.5021
27	12:30-12:45	22.9383	13.0762	30.1931	13.1194
28	12:45-13:00	22.4267	13.4820	28.7553	15.8215
29	13:00-13:15	22.8543	15.3976	29.7699	17.9554
30	13:15-13:30	21.6722	12.8734	28.5878	18.3871
31	13:30-13:45	22.4353	13.2583	27.2426	17.7334
32	13:45-14:00	22.9469	15.1758	28.4995	17.1295
33	14:00-14:15	24.7374	16.0589	27.5867	19.7747
34	14:15-14:30	23.7185	16.5639	28.3321	22.6735
35	14:30-14:45	24.8080	16.3483	27.4941	21.5543
36	14:45-15:00	25.3152	15.5167	27.9131	17.9323
37	15:00-15:15	26.1312	14.2347	25.1872	21.8074
38	15:15-15:30	26.1312	13.9400	26.3693	24.3425
39	15:30-15:45	27.5733	14.4414	25.6988	30.1176
40	15:45-16:00	27.6659	16.0691	24.7682	17.6281
	Continua				

Tabela A.3: Dados de audiências de televisão (Parte 2).

		Conjunto de Dados 01		Conjunto de Dados 02	
Ordem	Horário	CAD	MET	CAD	MET
41	16:00-16:15	27.5910	15.8475	33.0817	20.3249
42	16:15-16:30	26.2458	15.7324	31.6396	21.1955
43	16:30-16:45	25.2226	16.3906	31.0574	22.9043
44	16:45-17:00	24.3847	19.3355	31.9879	21.6066
45	17:00-17:15	22.6205	23.1748	32.2394	18.2322
46	17:15-17:30	22.1089	23.2148	32.4026	22.7537
47	17:30-17:45	22.7837	22.6191	32.7290	24.3196
48	17:45-18:00	22.9469	24.6766	32.5658	23.8119
49	18:00-18:15	34.5549	22.8624	29.4344	15.2675
50	18:15-18:30	34.7224	23.9044	27.4015	16.9675
51	18:30-18:45	35.0665	23.8662	26.4710	16.2800
52	18:45-19:00	35.5737	24.8231	27.4015	17.9690
53	19:00-19:15	37.6114	23.3022	28.0014	21.4426
54	19:15-19:30	34.2156	21.7515	28.2615	20.7416
55	19:30-19:45	36.6723	22.4929	30.8021	19.0941
56	19:45-20:00	36.8355	24.8940	29.4569	21.8637
57	20:00-20:15	38.1630	26.0185	33.5280	21.6590
58	20:15-20:30	38.7586	20.0274	34.2868	19.7492
59	20:30-20:45	39.1776	20.9952	33.3605	18.8218
60	20:45-21:00	38.5114	23.0675	33.8721	15.2401
61	21:00-21:15	43.9272	25.8214	34.0396	17.8974
62	21:15-21:30	44.0198	32.3248	34.0439	24.6721
63	21:30-21:45	44.4387	35.6288	33.5323	22.0599
64	21:45-22:00	45.3693	39.6195	33.4617	23.7331
65	22:00-22:15	35.3800	41.9002	28.4515	19.3920
66	22:15-22:30	33.5189	33.1494	26.9211	19.8503
67	22:30-22:45	31.9928	31.8799	26.3169	14.9889
68	22:45-23:00	32.1559	30.8172	26.9034	17.3008
69	23:00-23:15	23.2824	31.6222	18.5413	15.0780
70	23:15-23:30	22.0960	27.7664	14.5543	12.7270
71	23:30-23:45	19.7983	23.9403	14.5543	14.4135
72	23:45-24:00	17.7783	23.6954	13.9501	11.7654