



Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica - IMECC
Departamento de Estatística



**MODELOS NÃO LINEARES SOB A CLASSE DE
DISTRIBUIÇÕES MISTURAS DA ESCALA
SKEW-NORMAL**

Dissertação de Mestrado

Aldo William Medina Garay

Orientador: Prof. Dr. Víctor Hugo Lachos Dávila

Co-Orientador: Prof. Dr. Filidor E. Vilca Labra

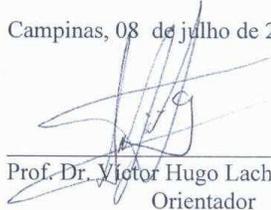
CAMPINAS

2010

MODELOS NÃO LINEARES SOB A CLASSE DE DISTRIBUIÇÕES
MISTURAS DA ESCALA SKEW-NORMAL

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Aldo William Medina Garay e aprovada pela comissão julgadora.

Campinas, 08 de julho de 2010


Prof. Dr. Victor Hugo Lachos Dávila.
Orientador


Prof. Dr. Filidor Edilfonso Vilca Labra
Co-orientador

Banca Examinadora:

- 1 Prof. Dra. Mariana Rodrigues Motta (IMECC-UNICAMP)
- 2 Prof. Dra. Márcia D'Elia Branco (IME-USP)
- 3 Prof. Dr. Filidor Edilfonso Vilca Labra (IMECC-UNICAMP)

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de **MESTRE** em **ESTATÍSTICA**.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**
Bibliotecária: Maria Fabiana Bezerra Müller – CRB8 / 6162

Medina Garay, Aldo William

M468m Modelos não lineares sob a classe de distribuições misturas da escala skew-normal/Aldo William Medina Garay-- Campinas, [S.P. : s.n.], 2010.

Orientadores : Victor Hugo Lachos Dávila ; Filidor Edilson Vilca Labra

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1.Distribuição normal assimétrica. 2.Misturas de escala. 3.Modelos não-lineares (Estatística). 4.Heteroscedasticidade. 5.Algoritmos de expectativa de maximização. 6.Modelo de eliminação de casos. 7.Influência local. I. Lachos Dávila, Victor Hugo. II.Labra, Filidor Edilson Vilca. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Título em inglês: Nonlinear models based on scale mixtures of skew-normal distributions

Palavras-chave em inglês (Keywords): 1. Skew normal distributions. 2. Scale mixtures 3. Nonlinear models (Statistics). 4. Heteroscedasticity. 5. EM algorithms. 6. Case-deletion model. 7. Local influence.

Área de concentração: Métodos Estatísticos

Titulação: Mestre em Estatística

Banca examinadora: Prof. Dr. Filidor Edilson Vilca Labra (IMECC – UNICAMP)
Profa. Dra. Mariana Rodrigues Motta (IMECC – UNICAMP)
Profa. Dra. Márcia D'Elia Branco (IME - USP)

Data da defesa: 08/07/2010

Programa de Pós-Graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 08 de julho de 2010 e aprovada

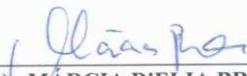
Pela Banca Examinadora composta pelos Profs. Drs.



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



Prof(a). Dr(a). MARIANA RODRIGUES MOTTA



Prof(a). Dr(a). MÁRCIA D'ELIA BRANCO

DEDICATÓRIA

*À minha mãe Alejita, ao
espírito eterno de meu
pai Carlos (in memoriam),
a meu tio Fabián Garay,
a meu inseparável irmão Carlos
e as minhas irmãs Nora e Lili.*

Agradecimentos

- Em primeiro lugar agradeço a minha mãe por seu sacrifício, entrega, esforço e dedicação para me educar, guiar e ser meu melhor exemplo de vida.
- Ao meu orientador o professor Victor Hugo Lachos, por sua dedicação, paciência, incentivo e em especial a amizade nesta experiência do mestrado. Encontrei nele não só um excelente pesquisador, se não uma grande pessoa, um excelente amigo e melhor profissional.
- Ao professor Filidor Vilca pela co-orientação, sua experiência foi muito importante na obtenção dos resultados finais da pesquisa. Agradeço sua disposição para me ajudar nos momentos difíceis da elaboração desta dissertação.
- Ao professor Edwin Ortega, por sua dedicação e contribuição importante para que este trabalho pudesse ser realizado.
- Aos professores do Departamento de Estatística do IMECC, pelos ensinamentos concedidos.
- À Rafael (Rafiti), Werner, Rafael (Café) e André (Dezão), a famosa família G6, por sua amizade, apoio, disposição para me ajudar com meu português e em especial por ser minha família brasileira nestes dois últimos anos.
- Aos meus colegas de turma, Pedro Moraes, Rodrigo (Goiano), Diego, Marley, Karina, Martha, Marcio Diniz, Bia, Guilherme, Jonas, Walter, Rafael, Sheila,

Nívea e Mariana, uma amostra representativa mas não aleatória do Brasil, por sua amizade e pelos momentos compartilhados.

- À família Possani: Romildo, Silvana e Fernando pela amizade, disposição e ajuda nos momentos mais difíceis nestes últimos dois anos.
- Aos meus amigos Iván e Michael Gonzales e a minhas vizinhas da casa G12: Stella, Cris, Milena e Andrea pelos momentos compartilhados.
- À Gina Tasso por seu carinho, apoio e disposição para me ajudar e acompanhar sempre.
- Aos professores da Universidad Nacional Mayor de San Marcos, em especial à Prof. Violeta Nolberto e Fátima Medina por seus ensinamentos e amizade.
- Aos professores da Universidad Agraria La Molina em especial aos Prof. Víctor Maehara, Luz Bullón, Fernando Rosas, Jesús Salinas e Iván Soto, pela amizade e oportunidade de ser parte da família Molinera.
- À Fapesp pelo apoio financeiro.
- Acima de tudo agradeço a Deus, pela oportunidade de estar aqui e pela dádiva da vida.

A todos, o meu muito obrigado.

Resumo

Neste trabalho estudamos alguns aspectos de estimação e diagnóstico de influência global e local de modelos não lineares sob a classe de distribuição misturas da escala skew-normal, baseado na metodologia proposta por [Cook \(1986\)](#) e [Poon & Poon \(1999\)](#). Os modelos não lineares heteroscedásticos também são discutidos.

Esta nova classe de modelos constitui uma generalização robusta dos modelos de regressão não linear simétricos, que têm como membros particulares distribuições com caudas pesadas, tais como skew-t, skew-slash, skew-normal contaminada, entre outras. A estimação dos parâmetros será obtida via o algoritmo EM proposto por [Dempster et al. \(1977\)](#). Estudos de testes de hipóteses são considerados utilizando as estatísticas de escore e da razão de verossimilhança, para testar a homogeneidade do parâmetro de escala. Propriedades das estatísticas do teste são investigadas através de simulações de Monte Carlo.

Exemplos numéricos considerando dados reais e simulados são apresentados para ilustrar a metodologia desenvolvida.

Palavras-chave: Modelo de eliminação de casos; Algoritmo EM; Homogeneidade; Influência Local; Modelo de regressão não linear; Distribuição misturas da escala skew normal.

Abstract

In this work, we studied some aspects of estimation and diagnostics on the global and local influence in nonlinear models under the class of scale mixtures of the skew-normal (SMSN) distribution, based on the methodology proposed by [Cook \(1986\)](#) e [Poon & Poon \(1999\)](#). Heteroscedastic nonlinear models are also discussed.

This new class of models are a robust generalization of non-linear regression symmetrical models, which have as members individual distributions with heavy tails, such as skew-t, skew-slash, and skew-contaminated normal, among others. The parameter estimation will be obtained with the EM algorithm proposed by [Dempster *et al.* \(1977\)](#). Studies testing hypotheses are considered using the score statistics and the likelihood ratio test to test the homogeneity of scale parameter. Properties of test statistics are investigated through Monte Carlo simulations.

Numerical examples considering real and simulated data are presented to illustrate the methodology.

Key-words: Case-deletion model; EM algorithm; Homogeneity; Local influence; Nonlinear regression models; Scale mixtures of skew-normal distributions.

Sumário

Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	1
1.1 Motivação	1
1.2 O algoritmo EM	4
1.3 Diagnóstico	5
1.4 Definição dos Objetivos	6
1.5 Apresentação dos Capítulos	7
2 Distribuições misturas da escala skew-normal (MESN)	9
2.1 Distribuição skew-normal padrão	11
2.2 Distribuição skew-normal com três parâmetros	12
2.3 Distribuições misturas da escala skew-normal	13
2.4 Estimação dos parâmetros nos modelos MESN	17
2.4.1 Estimação dos parâmetros via algoritmo EM	17
2.4.2 Estudo de simulação	20
2.5 Considerações finais	22
3 Modelos não lineares sob a classe de distribuições MESN	25
3.1 Descrição do modelo	28
3.2 Função de log-verossimilhança	29

3.3	Estimação dos parâmetros via algoritmo tipo EM	31
3.3.1	Considerações para a implementação	33
3.3.2	Seleção de modelos	34
3.3.3	Análise de resíduos	34
3.3.4	Estudo de simulação	35
3.4	Análise de diagnóstico	37
3.4.1	Modelo de eliminação de casos (CDM)	38
3.4.2	Influência local	39
3.5	Aplicação: Produtividade de óleo de palma	42
3.5.1	Ajuste do modelo	43
3.5.2	Análise de diagnóstico	44
3.5.3	Análise de resíduos	46
3.5.4	Influência de um valor outlier	48
3.6	Considerações finais	50
4	Modelos não lineares heteroscedásticos sob a classe de distribuições	
	MESN	51
4.1	Descrição do modelo	53
4.2	Função de log-verossimilhança	54
4.3	Estimação dos parâmetros via algoritmo tipo EM	55
4.3.1	Análise de resíduos	57
4.3.2	Estudo de simulação	58
4.4	Testes de Hipóteses	60
4.4.1	Teste de escore	62
4.4.2	Teste da razão de verossimilhança	64
4.4.3	Estudo de simulação	65
4.5	Aplicação 1: Produtividade de óleo de palma	69
4.5.1	Teste de heteroscedasticidade	69
4.6	Aplicação 2: Calibração ultrasônica	70

4.6.1	Ajuste do modelo	71
4.6.2	Análise de diagnóstico	73
4.6.3	Análise de resíduos	75
4.6.4	Teste de heteroscedasticidade	75
4.7	Considerações finais	78
5	Conclusões Finais e Perspectivas Futuras	79
5.1	Conclusões finais	79
5.2	Perspectivas Futuras	80
A	Resultados adicionais do Capítulo 3 e Capítulo 4	83
A.1	Elementos da matriz de perturbação Δ^*	83
A.2	Primeira e segunda derivada no modelo hete-roscedástico	85
	Referências Bibliográficas	87

Lista de Figuras

2.1	Função de densidade da SN, ST, SSL e SCN.	17
2.2	Viés dos EMV de $\beta_1, \beta_2, \sigma^2$ e λ para o modelo $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n.$	21
2.3	EQM dos EMV $\beta_1, \beta_2, \sigma^2$, e λ para o modelo $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n.$	22
3.1	Mudança relativa média nas estimações de $\beta_1, \beta_2, \beta_3$ e σ^2 para o modelo de crescimento não linear com distribuições SN, ST, SSL, SCN.	36
3.2	Mudança relativa média nas estimações de $\beta_1, \beta_2, \beta_3$ e σ^2 para o modelo de crescimento não linear com distribuições SN, ST($\nu = 3, 6, 10$)	37
3.3	Log-verossimilhança perfilada do parâmetro ν considerando os modelos ST-NLM e SSL-NLM.	43
3.4	Gráfico de GD_i^l (1ra linha) e LD_i^l (2da linha) sob os modelos SN-NLM, ST-NLM e SSL-NLM.	45
3.5	Pesos estimados (\hat{u}_i) para os modelos ST-NLM e SSL-NLM.	45
3.6	Gráfica de $M(0)$ sob os modelos SN-NLM, ST-NLM e SSL-NLM. Perturbação de ponderação de casos (1ra linha) e perturbação do parâmetro de escala (2da linha).	47
3.7	(a) Gráfico dos resíduos \mathbf{r}_i em contraposição à variável explanatória x_i e (b) Valores preditos.	48
3.8	Envelopes simulados para os resíduos de Pearson sob os modelos SN-NLM, ST-NLM e SSL-NLM.	48

3.9	Mudanças relativas nas estimações de β_2 e β_3 , sob os modelos SN-NLM, ST-NLM e SSL-NLM considerando diferentes perturbações ∇ na observação 5.	49
4.1	Mudança relativa média nas estimações de β_2 , β_3 , σ^2 e ρ para o modelo de crescimento não linear com distribuição SN, ST com $\nu = 3$ e SSL com $\nu = 2$	60
4.2	Mudança relativa média nas estimações de β_2 , β_3 , σ^2 e ρ para o modelo de crescimento não linear com distribuição SN e ST com $\nu = 3, 6, 10, 12$	61
4.3	Viés na estimação de β_3 , σ^2 , ρ e λ no modelo NLMH-MESN definido em (4.13).	63
4.4	EQM nas estimações de β_3 , σ^2 , ρ e λ no modelo NLMH-MESN definido em (4.13).	64
4.5	Poder do teste de escore para diferentes valores de n e ρ , considerando as distribuições SN, ST e SSL no estudo de simulação.	66
4.6	Comparação entre a função de distribuição empírica da SR e a distribuição χ_1^2 , usando as distribuições SN, ST e a SSL (1ra, 2da e 3ra linha respectivamente).	68
4.7	Gráfico de dispersão para os dados de calibração ultrasônica.	71
4.8	Log-verossimilhança perfilada do parâmetro ν considerando os modelos ST-NLMH e SSL-NLMH.	72
4.9	Gráfico de a) GD_i^l e b) LD_i^l sob os modelos SN-NLMH, ST-NLMH e SSL-NLMH.	74
4.10	Pesos estimados (\hat{u}_i) para os modelos ST-NLMH e SSL-NLMH.	74
4.11	Gráfica de $M(0)$ com os modelos SN-NLMH, ST-NLMH e SSL-NLMH, sob perturbação de ponderação de casos.	75
4.12	Envelopes simulados para os resíduos de Pearson sob os modelos N-NLMH, SN-NLMH, T-NLMH e ST-NLMH.	76

Lista de Tabelas

2.1	Viés dos EMV dos parâmetros com diferentes tamanhos de amostras. . .	23
2.2	EQM dos EMV dos parâmetros com diferentes tamanhos de amostras. . .	24
3.1	EMV e EP obtidos por ajustar vários modelos de misturas nos dados de produtividade de óleo de palma.	44
4.1	Viés na estimação dos parâmetros para diferentes tamanhos de amostras no modelo NLMH-MESN definido em (4.13).	62
4.2	Taxas de rejeição da $H_0 : \rho = 0$, sob as distribuições SN, ST e a SSL . . .	67
4.3	Estatística de escore (SR) e <i>p-valor</i> para os modelos MESN.	70
4.4	EMV e EP obtidos ao ajustar vários modelos de misturas nos dados de calibração ultrasônica.	72
4.5	ERV e <i>p-valor</i> nos modelos NLMH-MESN.	77
4.6	Estatística de Escore(SR) e <i>p-valor</i> nos modelos NLMH-MESN.	77

Capítulo 1

Introdução

1.1 Motivação

O modelo de regressão não linear com distribuição normal (N-NLM) é frequentemente utilizado em diferentes áreas de pesquisa com o intuito de modelar dados com distribuição simétrica; as funções não lineares dos parâmetros são utilizadas com o objetivo de explicar ou descrever os fenômenos sob estudo. Entretanto, é bastante conhecido que vários fenômenos nem sempre podem ser modelados com a distribuição normal, seja pela falta de simetria ou a presença de valores atípicos.

Em décadas passadas, mesmo quando o fenômeno sobre o estudo não apresentava resposta para a qual a suposição de normalidade da variável resposta fosse razoável, procurava-se algum tipo de transformação no sentido de obter pelo menos, o comportamento simétrico dos dados. A mais conhecida e usada destas técnicas foi a proposta por [Box & Cox \(1964\)](#). Recentemente, muitos autores têm discutido o problema da transformação de variáveis, veja por exemplo, [Azzalini & Capitanio \(1999\)](#).

Diante desta problemática, nos últimos anos, distribuições alternativas são consideradas para resolver este tipo de problema. As distribuições elípticas (Fang *et al.*, 1990) são, talvez, as mais conhecidas dentre as propostas por preservar a estrutura simétrica da distribuição Gaussiana e apresentarem caudas mais ou menos pesadas que a distribuição normal. Esta ótima propriedade têm sido recentemente considerada em muitos estudos. Por exemplo, Cysneiros & Vanegas (2008) estudaram o modelo de regressão não linear com erros simétricos apresentando um estudo analítico e empírico para descrever o comportamento dos resíduos padronizados. Vanegas & Cysneiros (2009) propõem procedimentos de diagnóstico baseados no método de eliminação de casos (CDM) para modelos de regressão não linear simétricos. Recentemente, Cysneiros *et al.* (2009) introduzem a classe de modelos de regressão não linear heteroscedásticos, no qual apresentam um processo iterativo para estimar a média e o parâmetro de dispersão do modelo.

Apesar desta nova classe de distribuições representar uma boa alternativa à distribuição normal, a mesma parece não ser apropriada quando a distribuição dos dados observados ou erros do modelo é assimétrica (Hill & Dixon, 1982). Neste contexto, as distribuições skew-elípticas vêm sendo usadas com sucesso. Esta nova família de distribuições têm a vantagem de ser analiticamente tratável e é uma alternativa viável para modelar conjunto de dados que apresentam comportamento assimétrico. A distribuição skew-normal, introduzida por Azzalini (1985) é talvez a pioneira dentro dessa nova estratégia de modelagem que contém como caso particular a distribuição normal. Usando a idéia de Azzalini (1985), outras distribuições assimétricas foram desenvolvidas, como por exemplo, a skew-t, skew-cauchy, skew-slash, skew-normal contaminada, skew-exponencial potência, entre outras. No contexto de modelos não lineares, podemos mencionar alguns trabalhos recentes onde foram desenvolvidos estudos com estes tipos de distribuições. Por exemplo, Cancho *et al.* (2009) apresentam o modelo de regressão não linear skew-normal (SN-NLM), na qual uma análise completa baseada na verossi-

milhança é desenvolvida, incluindo o algoritmo EM para encontrar as estimativas de máxima verossimilhança (EMV). Recentemente, [Xie et al. \(2009a\)](#) e [Xie et al. \(2009b\)](#) usaram o teste de escore para testar a homogeneidade das variâncias no modelo SN-NLM. Entretanto, sabe-se que as estimativas dos parâmetros de um modelo baseado na distribuição skew-normal também são sensíveis às observações atípicas como é discutido em [Montenegro et al. \(2009\)](#). Uma solução para o problema dos dados atípicos em um contexto assimétrico foi apresentado por [Branco & Dey \(2001\)](#) que propuseram a utilização da classe de distribuições misturas da escala skew normal (MESN), a fim de lidar simultaneamente com a assimetria e diferentes graus de curtose. Por outro lado, é interessante notar que esta classe contém como caso especial a família de distribuições mistura da escala normal, como é estudada e discutida em [Lange & Sinsheimer \(1993\)](#).

No entanto, considerando que o modelo assumido é apropriado para um conjunto de dados, a avaliação dos aspectos de robustez nas estimativas dos parâmetros têm sido uma preocupação importante de vários pesquisadores nos últimos anos. O método de eliminação de casos (CDM), que consiste em estudar o impacto de cada uma das observações nas estimativas dos parâmetros depois de ser excluída do conjunto de dados, é, provavelmente, a técnica mais utilizada para detectar observações atípicas ([Cook & Weisberg, 1983](#)). Apesar disso, a investigação sobre a influência de pequenas perturbações, no modelo ou nos dados, sobre as estimativas dos parâmetros têm recebido crescente atenção nos últimos anos. Esta pode ser alcançada através da análise de influência local ([Cook, 1986](#)). Esta técnica é bastante geral, sendo utilizada para avaliar a estabilidade das estimativas com relação ao modelo teórico e tem sido aplicada recentemente com sucesso em modelos lineares e não lineares sob a classe de distribuições simétricas misturas da escala normal, veja por exemplo, [Osorio \(2006\)](#) e as referências citadas no trabalho dele. Além disso, alguns resultados também podem ser encontrados, considerando a distribuição skew-normal, veja por exemplo os trabalhos de [Lachos et al. \(2005\)](#) e [Lachos et al. \(2008\)](#) que aplicam a técnica de influência local no modelo de calibração comparativa e no modelo de regressão com erros nas variáveis com inter-

cepto nulo, respectivamente.

Tendo como base, o que foi discutido até aqui, apresentaremos uma breve descrição do algoritmo EM, que será utilizado para encontrar as estimativas de máxima verossimilhança (EMV) dos parâmetros nos modelos não lineares sob a classe de distribuições MESN. Também apresentaremos um resumo de análise de influência aplicados nos modelos não lineares e finalmente descreveremos os objetivos e a organização do presente trabalho.

1.2 O algoritmo EM

O Algoritmo EM (Dempster *et al.*, 1977) é um enfoque aplicado no cálculo iterativo de estimativas dos parâmetros do modelo pelo método de máxima verossimilhança. Este algoritmo é aplicado ao problema de estimação a partir de dados incompletos, aumentando o vetor de dados observados (\mathbf{y}_{obs}) com a inclusão de variáveis latentes (\mathbf{y}_{nobs}), as quais não são observadas diretamente. Obtendo-se, deste modo, o vetor de dados completos $\mathbf{y}_{comp} = (\mathbf{y}_{obs}, \mathbf{y}_{nobs})$, de tal forma que a função de log-verossimilhança é representada por $\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}_{comp}) = \log(f(\mathbf{y}_{comp}|\boldsymbol{\theta}))$. Cada iteração do algoritmo EM consiste em dois passos:

- **Passo E (Esperança):**

Este passo consiste em calcular a esperança da log-verossimilhança completa, denotada por $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(i-1)})$, condicionada no vetor de dados observados. Isto é, para a i -ésima iteração temos que, dado $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(i-1)}$,

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(i-1)}) = E\{\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}_{comp})|\mathbf{y}_{obs}, \hat{\boldsymbol{\theta}}^{(i-1)}\};$$

- **Passo M (Maximização):**

Consiste em maximizar a log-verossimilhança completa em relação aos parâmetros do modelo, substituindo os dados latentes por seus valores esperados condicionais obtidos no passo E. Para a i -ésima iteração obtemos $\hat{\boldsymbol{\theta}}^{(i)}$ que maximize $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(i-1)})$, tal que,

$$Q(\hat{\boldsymbol{\theta}}^{(i)}|\hat{\boldsymbol{\theta}}^{(i-1)}) > Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(i-1)}), \quad \forall \boldsymbol{\theta} \in \Theta.$$

Estes passos alternam-se até convergir e geralmente é considerado como critério de convergência $\|\hat{\boldsymbol{\theta}}^{(i)} - \hat{\boldsymbol{\theta}}^{(i-1)}\| < \epsilon$, com $\epsilon > 0$.

1.3 Diagnóstico

Os modelos estatísticos são ferramentas importantes para extrair e compreender características essenciais de um conjunto de dados. Uma etapa importante na análise é a verificação de possíveis afastamentos das suposições feitas no modelo, como por exemplo, a existência de observações extremas com alguma interferência nos resultados do ajuste. Os elementos do conjunto de dados que efetivamente controlam aspectos da análise são ditos *influentes* se elas produzem alterações no resultado de análise quando excluídas ou submetidas a algum tipo de perturbação.

Existem basicamente duas abordagens para a detecção de observações que influenciam seriamente os resultados de uma análise estatística: **i)** a abordagem **método de eliminação de casos** (tradução livre de “case-deletion model”), em que o impacto de se excluir uma observação na previsão é diretamente avaliada por medidas, tais como, o afastamento pela verossimilhança e a distância de Cook (ver, [Cook, 1977](#)) e **ii)** a abordagem em que a estabilidade dos resultados com relação ao modelo teórico é estudado através de diferentes níveis de perturbação no modelo. Uma técnica desenvolvida por [Cook \(1986\)](#) conhecida como método de **influência local** teve uma grande recep-

tividade entre os pesquisadores, gerando inúmeras publicações em que a metodologia é aplicada, seja em classes particulares ou propondo extensões desta técnica. Por exemplo, [Thomas & Cook \(1990\)](#) aplicam o método de influência local para predições em modelos lineares generalizados; [Laurent & Cook \(1993\)](#) consideram o método de influência local em um modelo de regressão não linear, enfatizando a relação entre o jacobiano do *leverage* e o método de influência local sobre os valores do modelo ajustado, entre outros. Nos últimos anos, vários autores também utilizam este método para modelos de regressão linear com erros não normais; veja por exemplo, [Galea et al. \(2005\)](#), [Cysneiros & Vanegas \(2008\)](#), [Lin et al. \(2009\)](#). Recentemente [Zeller et al. \(2010\)](#) aplicaram a técnica de influência nos modelos lineares mistos skew-normal/independente. No entanto, estudos de inferência e análise de diagnóstico para os modelos não lineares sob a classe de distribuições MESN não têm sido desenvolvidos.

1.4 Definição dos Objetivos

O objetivo principal deste trabalho é apresentar um estudo de estimação e diagnóstico de influência nos modelos não lineares sob a classe de distribuições assimétricas MESN. Neste estudo, a estimação dos parâmetros é feita via o algoritmo EM e a análise de diagnóstico é baseada na metodologia proposta por [Cook \(1986\)](#) e [Poon & Poon \(1999\)](#). Definimos os seguintes pontos como os objetivos específicos do nosso trabalho:

- 1 Descrever a classe de distribuições assimétricas de misturas da escala skew-normal, como discutido em [Branco & Dey \(2001\)](#), [Lachos & Vilca \(2007\)](#), [Kim \(2008\)](#), [Basso et al. \(2009\)](#), entre outros, e desenvolver algoritmos para estimação dos parâmetros do modelo.
- 2 Estender o modelo de regressão não linear SN-NLM apresentado por [Cancho et al. \(2009\)](#) assumindo que a distribuição dos erros pertence à classe MESN e desenvolver o algoritmo EM para estimar os parâmetros de interesse deste modelo.

- 3 Fazer uma análise de influência local e global, baseada na metodologia proposta por [Cook \(1986\)](#) e [Poon & Poon \(1999\)](#) nos modelos não lineares sob a classe MESN.
- 4 Propor o teste de escore e o teste da razão de verossimilhança para verificar o pressuposto de homocedasticidade do parâmetro de escala nos modelos não lineares sob MESN, que representam uma extensão dos trabalhos apresentados por [Cysneiros *et al.* \(2007\)](#) e [Lin *et al.* \(2009\)](#).
- 5 Apresentar e descrever o modelo de regressão não linear heteroscedástico sob a classe MESN. Assim como descrever o algoritmo EM para estimação dos parâmetros deste modelo.

1.5 Apresentação dos Capítulos

No Capítulo 2, apresentaremos a classe de distribuições misturas da escala skew-normal (MESN), originalmente introduzida por [Branco & Dey \(2001\)](#) e, posteriormente, estudada por [Lachos & Vilca \(2007\)](#), [Kim \(2008\)](#), [Basso *et al.* \(2009\)](#), entre outros. Tal apresentação tem a finalidade de descrever algumas propriedades, tais como representação estocástica e os momentos condicionais, que serão úteis para a implementação do algoritmo EM na estimação dos parâmetros. Além disso, apresentaremos um estudo de simulação para avaliar a performance dos EMV dos parâmetros.

No Capítulo 3, estenderemos o modelo não linear SN-NLM, apresentado por [Canchito *et al.* \(2009\)](#), assumindo que a a distribuição dos erros pertence à classe MESN. Em seguida, desenvolveremos o algoritmo EM para estimar os parâmetros do modelo apresentado, a função escore e a matriz de informação observada. Para avaliar a performance dos EMV dos parâmetros realizaremos um estudo de simulação. Desenvolveremos, também, uma análise de diagnóstico de influência local e global, baseado na metodologia proposta por [Cook \(1986\)](#) e [Poon & Poon \(1999\)](#) nos modelos não lineares

sob MESN. Finalizaremos este capítulo com uma aplicação dos dados trabalhados por [Cancho *et al.* \(2009\)](#), para ilustrar a metodologia proposta.

No Capítulo 4, faremos uma apresentação e descrição do modelo de regressão não linear heteroscedástico sob a classe MESN, propondo o algoritmo EM para a estimação dos parâmetros neste modelo. Um estudo de simulação será realizado para avaliar a robustez e as propriedades assintóticas das estimativas. Além disso, estenderemos os trabalhos realizados por [Cysneiros *et al.* \(2007\)](#) e [Lin *et al.* \(2009\)](#) apresentando o teste de score e o teste da razão de verossimilhança para verificar o pressuposto de homocedasticidade do parâmetro escalar (σ^2) nos modelos não lineares sob MESN. Posteriormente, discutiremos um estudo de simulação com o intuito de verificar as propriedades assintóticas da estatística de score e o poder do teste. Finalmente, apresentaremos uma aplicação dos dados trabalhados por [Lin *et al.* \(2009\)](#) para ilustrar a metodologia proposta.

No Capítulo 5, finalizaremos esta dissertação com conclusões finais e perspectivas para trabalhos futuros.

Capítulo 2

Distribuições misturas da escala skew-normal (MESN)

Em muitas aplicações de diferentes áreas de pesquisa como, por exemplo, finanças, economia, sociologia, climatologia e ciências biomédicas, notamos que a estrutura dos dados e/ou modelos utilizados não satisfazem a suposição de normalidade devido, entre outras causas, à presença de alta assimetria ou de observações atípicas. Por esta razão, a construção e estudo de famílias paramétricas flexíveis, que são capazes de acomodar valores práticos de assimetria e curtose, recentemente tem recebido considerável atenção na literatura.

Nesse contexto, a família elíptica, que agrupa um conjunto de distribuições simétricas com caudas mais/menos pesadas do que a distribuição normal, possui um grande destaque. A principal vantagem desta classe é que permite estender os modelos desenvolvidos sob normalidade para um conjunto mais amplo de distribuições. Esta classe de distribuições inclui a distribuição t-student, slash, normal contaminada, entre outras. Uma descrição detalhada da classe de distribuições elípticas multivariadas é apresen-

tada, por exemplo, em [Fang *et al.* \(1990\)](#) e [Arellano-Valle \(2004\)](#).

No entanto, apesar desta classe representar uma interessante alternativa à distribuição normal, ela pode não ser apropriada quando a distribuição dos dados é assimétrica. Para tratar este problema, [Branco & Dey \(2001\)](#) propõem a família de distribuições skew elípticas, que inclui como caso especial a distribuição skew normal ([Azzalini, 1985](#)). Ao longo deste capítulo discutimos sobre uma sub-classe da família de distribuições skew elípticas, especificamente, a família de distribuições misturas da escala skew-normal (MESN). Como veremos, esta classe possui interessantes propriedades que permitem fácil implementação de algoritmos de otimização numérica. Além disso, a mesma contém como um caso especial a família de distribuições misturas da escala normal (MEN) ([Andrews & Mallows, 1974](#))

Iniciamos este capítulo com a distribuição skew normal univariada e uniparamétrica, mostrando algumas de suas propriedades. Em seguida, apontamos uma relação entre a forma uniparamétrica com o modelo skew normal de três parâmetros (locação-escala), usada para construir a classe de distribuições misturas da escala skew normal (MESN). Para esta classe, também é apresentada algumas de suas propriedades, assim como, sua representação estocástica, que é útil na implementação do algoritmo EM. Em seguida, são apresentados alguns exemplos de distribuições pertencentes a esta classe e o algoritmo EM para estimação de máxima verossimilhança dos parâmetros em modelos de locação-escala MESN. Um estudo de simulação é apresentado para avaliar as propriedades assintóticas dos EMV dos parâmetros, mostrando que nosso algoritmo proposto é adequado. Finalmente, algumas considerações gerais são discutidas.

2.1 Distribuição skew-normal padrão

Uma variável aleatória Z tem distribuição skew-normal padrão se sua função de densidade de probabilidade (fdp) é definida por

$$f_Z(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R}, \quad (2.1)$$

onde $\phi(\cdot)$ e $\Phi(\cdot)$ indicam a função de densidade de probabilidade e a função de distribuição acumulada de uma normal padrão. Aqui $\lambda \in \mathbb{R}$ é o parâmetro de assimetria, o qual caracteriza a forma da distribuição, tal que valores positivos de λ indicam assimetria positiva e valores negativos indicam assimetria negativa; note que quando $\lambda = 0$ em (2.1), $f_Z(z)$ é a densidade de uma normal padrão.

Para considerações deste trabalho denotaremos $Z \sim SN(\lambda)$ para representar uma variável aleatória com densidade skew-normal padrão com parâmetro de assimetria λ , e as principais propriedades que tem esta variável Z são:

- Se $Z \sim SN(\lambda)$, então $|Z| \sim HN(0, 1)$, onde HN é a distribuição half normal padrão,
- Quando $\lambda \rightarrow \infty$ a densidade (2.1) converge para uma $HN(0, 1)$,
- Se $Z \sim SN(\lambda)$, então $-Z \sim SN(-\lambda)$,
- A densidade (2.1) é log-côncava,
- $1 - F_Z(z; \lambda) = F_Z(z; -\lambda)$,
- $F_Z(z; \lambda) = [\Phi(z)]^2$,
- $Z \sim SN(\lambda)$ então $Z^2 \sim \chi_1^2$.

Uma discussão mais detalhada das propriedades e proposições podem ser vistas em [Bayes \(2005\)](#) e [Basso \(2009\)](#).

Proposição 2.1.1 *Seja uma variável aleatória $Z \sim SN(\lambda)$. Então,*

$$i) E[Z] = \sqrt{\frac{2}{\pi}}\delta, \quad Var[Z] = 1 - \frac{2}{\pi}\delta^2,$$

$$ii) \text{ O coeficiente de assimetria é } \gamma = \sqrt{\frac{2}{\pi}} \left(\frac{4}{\pi} - 1\right) \delta^3 \left(1 - \frac{2}{\pi}\delta^2\right)^{-\frac{3}{2}},$$

$$iii) \text{ O coeficiente de curtose é } \kappa = \frac{8}{\pi^2}(\pi - 3)\delta^4 \left(1 - \frac{2}{\pi}\delta\right)^{-2}.$$

Proposição 2.1.2 *Seja $Y_0, Y_1 \stackrel{iid}{\sim} N(0, 1)$. Então,*

$$Z = \delta|Y_0| + (1 - \delta^2)^{1/2}Y_1 \sim SN(\lambda), \quad (2.2)$$

em que $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$

2.2 Distribuição skew-normal com três parâmetros

Uma variável aleatória Y tem distribuição skew-normal com parâmetro de posição $\mu \in \mathbb{R}$ e escala $\sigma^2 > 0$, se sua função densidade de probabilidade (fdp) é definida por:

$$f_Y(y) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right), \quad y \in \mathbb{R}. \quad (2.3)$$

e escrevemos $Y \sim SN(\mu, \sigma^2, \lambda)$. Considerando as proposições desenvolvidas para a distribuição skew-normal padrão, podemos generalizá-las para distribuição skew-normal com três parâmetros.

Proposição 2.2.1 *Se $Z \sim SN(\lambda)$ e $Y = \mu + \sigma Z$. Então $Y \sim SN(\mu, \sigma^2, \lambda)$.*

Proposição 2.2.2 *Seja uma variável aleatória $Y \sim SN(\mu, \sigma^2, \lambda)$. Então,*

$$i) E[Y] = \mu + \sigma\delta\sqrt{\frac{2}{\pi}}, \quad Var[Y] = \sigma^2 \left(1 - \frac{2}{\pi}\delta\right),$$

ii) *Os coeficientes de assimetria e de curtose são os mesmos do caso uni paramétrico¹.*

¹Detalhes desta seção podem ser encontradas em [Basso et al. \(2009\)](#)

2.3 Distribuições misturas da escala skew-normal

A fim de introduzir as notações necessárias que serão utilizadas ao longo do trabalho, esta seção é iniciada com a definição das distribuições misturas da escala skew-normal (MESN). Uma variável aleatória Y é dito ter distribuição MESN, se esta pode ser escrita como:

$$Y = \mu + \kappa^{1/2}(U)Z, \quad (2.4)$$

em que μ é o parâmetro de posição, Z é uma variável aleatória skew-normal com parâmetro de posição 0, escala σ^2 e de assimetria λ ; além disso, $\kappa(u)$ é uma função de pesos positiva; U é uma variável aleatória com função de distribuição $H(\cdot; \boldsymbol{\nu})$ e densidade $h(\cdot; \boldsymbol{\nu})$ e $\boldsymbol{\nu}$ é um escalar ou vetor de parâmetros da distribuição de U , o qual pode ser conhecido ou desconhecido (Lange *et al.*, 1989). Este parâmetro está relacionado com a curtose, podendo controlar as caudas da distribuição sob investigação. Embora posamos lidar com qualquer função κ , neste trabalho restringiremos a nossa atenção para o caso em que $\kappa(U) = 1/U$, já que sob esta restrição interessantes propriedades matemáticas são obtidas. Dado $U = u$, temos que $Y|U = u \sim SN(\mu, u^{-1}\sigma^2, \lambda)$. Assim, a densidade de Y é dada por:

$$f(y) = 2 \int_0^\infty \phi(y; \mu, u^{-1}\sigma^2) \Phi\left(\frac{u^{1/2}\lambda(y - \mu)}{\sigma}\right) dH(u; \boldsymbol{\nu}), \quad (2.5)$$

em que $\phi(\cdot; \mu, \sigma^2)$ denota a densidade de uma distribuição normal univariada com média μ e variância $\sigma^2 > 0$ e $\Phi(\cdot)$ é a função de distribuição da normal padronizada (univariada). A notação $Y \sim MESN(\mu, \sigma^2, \lambda; H)$ será usada quando Y tem função de densidade de probabilidade como em (2.5).

Quando H é degenerada com $u = 1$, obtemos de (2.4) a distribuição $SN(\mu, \sigma^2, \lambda)$; quando $\lambda = 0$, a distribuição MESN é reduzida à classe de distribuições misturas da escala normal(MES) representada pela função de densidade de probabilidade:

$$f_0(y) = \int_0^\infty \phi_p(y; \mu, u^{-1}\sigma^2) dH(u; \boldsymbol{\nu}).$$

Apresentaremos em seguida alguns resultados importantes para as distribuições MESN, cujas demonstrações foram desenvolvidas em [Lachos & Vilca \(2007\)](#), [Basso et al. \(2009\)](#) e [Lachos et al. \(2010\)](#).

Proposição 2.3.1 *A representação estocástica da variável $Y \sim MESN(\mu, \sigma^2, \lambda; H)$ é dada por:*

$$Y = \mu + \Delta U^{-1/2}|T_0| + U^{-1/2}T_1\Gamma^{1/2}, \quad (2.6)$$

em que $\Delta = \sigma\delta$, $\Gamma = \sigma^2 - \Delta^2$, $|T_0|$ denota o valor absoluto de T_0 , $U \sim H(\cdot; \nu)$, $T_0 \sim N(0, 1)$ e $T_1 \sim N(0, 1)$ são mutuamente independentes.

Proposição 2.3.2 *Seja $Y \sim MESN(\mu, \sigma^2, \lambda, H)$, então a função geradora de momentos (fgm) é dada por:*

$$M_Y(t) = 2 \int_0^\infty e^{t\mu + \frac{1}{2}u^{-1}t^2\sigma^2} \Phi(u^{-1/2}\delta\sigma t) dH(u), \quad t \in \mathbb{R}. \quad (2.7)$$

Proposição 2.3.3 *Seja $Y \sim MESN(\mu, \sigma^2, \lambda, H)$. Então,*

$$1) \text{ Se } E[U^{-1/2}] < \infty, \text{ então } E[Y] = \mu + \sqrt{\frac{2}{\pi}}\kappa_1\Delta,$$

$$2) \text{ Se } E[U^{-1}] < \infty, \text{ então } Var[Y] = \kappa_2\sigma^2 - \frac{2}{\pi}\kappa_1^2\Delta^2, \text{ com } \kappa_m = E[U^{-m/2}] \text{ e } \Delta = \sigma\delta.$$

Um resultado importante para a implementação do algoritmo EM é o cálculo dos seguintes momentos condicionais $\kappa_r = E[U^r|\mathbf{y}]$ e $\tau_r = E[U^{r/2}W_\Phi(U^{1/2}A)|\mathbf{y}]$, com $W_\Phi(x) = \phi(x)/\Phi(x)$, $x \in \mathbb{R}$. As demonstrações destes resultados podem ser encontradas em [Zeller \(2009\)](#).

Proposição 2.3.4 *Seja $Y \sim MESN(\mu, \sigma^2, \lambda, H)$ e o fator de mistura de escala $U \sim H$. Então,*

$$\kappa_r = \frac{2\psi_0(y)}{\psi(y)} E[U_y^r \Phi(U_y^{1/2}A)] \quad \text{e} \quad \tau_r = \frac{2\psi_0(y)}{\psi(y)} E[U_y^{r/2} \phi(U_y^{1/2}A)],$$

em que ψ_0 é a fdp de $Y_0 \sim MEN(\mu, \sigma^2, H)$, $U_y \stackrel{d}{=} U|Y_0$ e $A = \lambda\sigma^{-1}(y - \mu)$.

A seguir apresentaremos alguns exemplos de distribuições que pertencem à classe das distribuições MESN:

- **Distribuição skew-t com ν graus de liberdade.** Neste caso consideramos $U \sim \text{Gamma}(\nu/2, \nu/2)$, $\nu > 0$, em (2.5), onde $\text{Gamma}(a, b)$ denota a distribuição gamma com média a/b . A densidade de Y tem a seguinte forma:

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left(1 + \frac{d}{\nu}\right)^{-\frac{\nu+1}{2}} T\left(\sqrt{\frac{\nu+1}{d+\nu}}A; \nu+1\right), \quad y \in \mathbb{R}, \quad (2.8)$$

em que $d = (y - \mu)^2/\sigma^2$ e $T(\cdot; \nu)$ denota a função de distribuição da t-student, com parâmetro de posição zero, escala um e ν graus de liberdade, denotada por $t(0, 1, \nu)$. Usaremos a notação $Y \sim \text{ST}(\mu, \sigma^2, \lambda; \nu)$ para representar a distribuição skew-t. Um caso particular da distribuição skew-t é a distribuição skew-cauchy, quando $\nu = 1$. Quando $\nu \rightarrow \infty$, temos a distribuição skew-normal como caso limite². Considerando a Proposição 2.3.3, a média e a variância da variável Y são dados por:

$$\begin{aligned} E[Y] &= \mu + \sqrt{\frac{\nu}{\pi}} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right) \Delta, \quad \nu > 1 \text{ e} \\ \text{Var}[Y] &= \left(\frac{\nu}{\nu-2}\right) \sigma^2 - \frac{\nu}{\pi} \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\right)^2 \Delta^2, \quad \nu > 2, \end{aligned}$$

em que $\Delta = \sigma\delta$ e $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

- **Distribuição skew-slash.** Neste caso temos $U \sim \text{Beta}(\nu, 1)$ - onde $\text{Beta}(a, b)$ denota a distribuição beta com parâmetros a e b - com parâmetro de assimetria ν . Usaremos a notação $Y \sim \text{SSL}(\mu, \sigma^2, \lambda; \nu)$ para representar a distribuição skew-slash. A densidade de Y é dado por:

$$f(y) = 2\nu \int_0^1 u^{\nu-1} \phi(y; \mu, u^{-1}\sigma^2) \Phi(u^{1/2}A) du, \quad y \in \mathbb{R}. \quad (2.9)$$

A skew-slash é uma distribuição de cauda pesada, tendo como limite a distribuição skew-normal (quando $\nu \rightarrow \infty$)³.

²Aplicações podem ser obtidas em Lin *et al.* (2007) e Azzalini & Genton (2008)

³Aplicações podem ser encontradas em Wang & Genton (2006)

Da Proposição 2.3.3 também é possível verificar que:

$$\begin{aligned} E[Y] &= \mu + \sqrt{\frac{2}{\nu}} \left(\frac{2\nu}{2\nu-1} \right) \Delta, \quad \nu > 1/2 \text{ e} \\ \text{Var}[Y] &= \left(\frac{\nu}{\nu-1} \right) \sigma^2 - \left(\frac{2}{\pi} \right) \left(\frac{2\nu}{2\nu-1} \right)^2 \Delta^2, \quad \nu > 1. \end{aligned}$$

- **Distribuição skew-normal contaminada.** Aqui U é uma variável aleatória discreta que toma um de dois estados considerando que a função de probabilidade de U é dada por

$$h(u|\boldsymbol{\nu}) = \nu \mathbb{I}_{(u=\gamma)} + (1-\nu) \mathbb{I}_{(u=1)}, \quad 0 < \nu < 1, \quad 0 < \gamma \leq 1,$$

tal que $\boldsymbol{\nu} = (\nu, \gamma)^\top$.

Denotamos por $Y \sim SCN(\mu, \sigma^2, \lambda; \nu, \gamma)$ para representar a distribuição skew-normal contaminada. Da definição (2.5), temos imediatamente:

$$f(y) = 2\{\nu\phi(y; \mu, \gamma^{-1}\sigma^2)\Phi(\gamma^{1/2}A) + (1-\nu)\phi(y; \mu, \sigma^2)\Phi(A)\}. \quad (2.10)$$

Os parâmetros ν and γ podem ser interpretados como a proporção de outliers e o fator escala, respectivamente. A distribuição skew-normal contaminada reduz a skew-normal quando $\gamma = 1$. Assim, da Proposição 2.3.3, temos que a média e a variância são dadas por,

$$\begin{aligned} E[Y] &= \mu + \sqrt{\frac{2}{\nu}} \left(\frac{\nu}{\gamma^{1/2}} + 1 - \nu \right) \Delta \text{ e} \\ \text{Var}[Y] &= \left(\frac{\nu}{\gamma} + 1 - \nu \right) \sigma^2 - \frac{2}{\pi} \left(\frac{\nu}{\gamma^{1/2}} + 1 - \nu \right)^2 \Delta^2. \end{aligned}$$

Com o objetivo de comparar estas quatro distribuições, na Figura 2.1 mostramos as densidades da $SN(\mu, \sigma^2, \lambda)$, $ST(\mu, \sigma^2, \lambda; 2)$, $SSL(\mu, \sigma^2, \lambda; 1)$ e $SCN(\nu, \sigma^2, \lambda; 0.5, 0.5)$, com $\mu = 0$, $\sigma^2 = 1$ e $\lambda = -2.5, 0, 2.5$. Note que alteramos a escala das densidades para que as mesmas possuam o mesmo valor na origem. Podemos observar na Figura 2.1 que as distribuições skew-slash, skew-t e skew-normal contaminada têm caudas mais pesadas que a distribuição skew-normal.

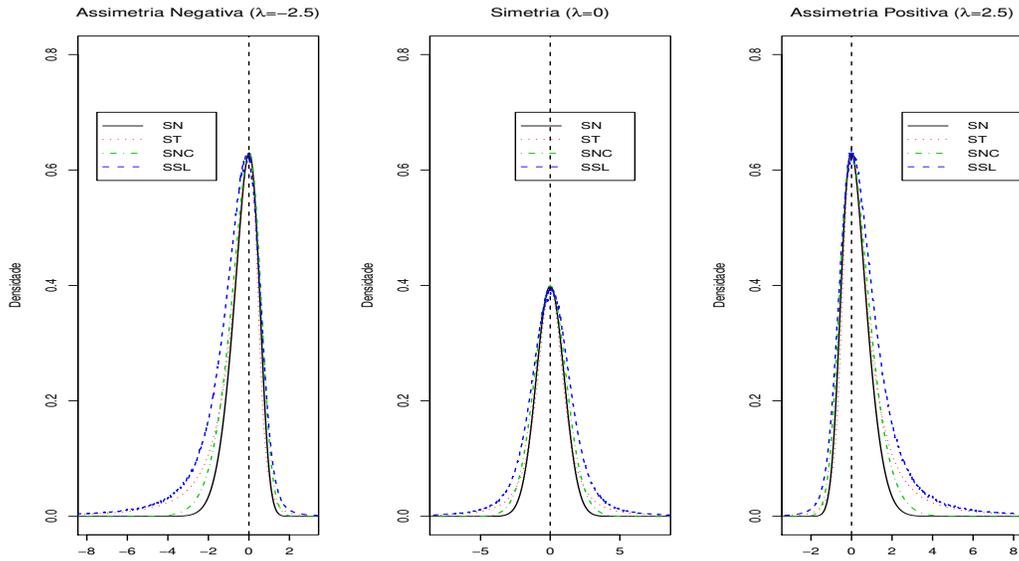


Figura 2.1: Função de densidade da SN, ST, SSL e SCN.

2.4 Estimação dos parâmetros nos modelos MESN

Com o intuito de estimar os parâmetros pelo método de máxima verossimilhança nos modelos sob a classe de distribuições MESN, nesta seção adaptamos o algoritmo EM, proposto por [Dempster *et al.* \(1977\)](#), para o problema em estudo. Para fazer isso, apresentaremos a representação hierárquica dos modelos MESN sob um ponto de vista de dados incompletos.

2.4.1 Estimação dos parâmetros via algoritmo EM

O modelo hierárquico para uma variável aleatória com distribuição MESN é dado por:

$$Y_i|U_i = u_i, T_i = t_i \sim N(\mu + \Delta t_i, u_i^{-1}\Gamma), \quad (2.11)$$

$$T_i|U_i = u_i \sim NT(0, u_i^{-1}; (0, \infty)), \quad (2.12)$$

$$U_i \sim H(u_i; \nu). \quad (2.13)$$

em que $\Delta = \sigma\delta$, $\Gamma = (1 - \delta^2)\sigma^2$ e $NT(a, b, (p, q))$ denota a distribuição normal truncada no intervalo (p, q) . Um resultado importante e de muita utilidade é que a distribuição de $T_i|Y_i = y_i, U_i = u_i$ é $NT(\mu_{T_i}, u_i^{-1}M_T^2; (0, \infty))$, com

$$\mu_{T_i} = \frac{\Delta}{\Delta^2 + \Gamma}(y_i - \mu), \quad M_T^2 = \frac{\Gamma}{\Delta^2 + \Gamma}.$$

Considerando o passo E do algoritmo EM, representamos a estimativa dos parâmetros $\xi = g(\boldsymbol{\theta})$ por $\hat{\xi} = g(\hat{\boldsymbol{\theta}})$, onde $g(\cdot)$ é uma função genérica de $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda)^\top$. Dado $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{t} = (t_1, \dots, t_n)^\top$ e $\mathbf{u} = (u_1, \dots, u_n)^\top$, a função de log-verossimilhança completa com $(\mathbf{y}, \mathbf{t}, \mathbf{u})$ é expressa por:

$$\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{t}, \mathbf{u}) = c - \frac{n}{2} \log \Gamma - \frac{1}{2\Gamma} \sum_{i=1}^n u_i (y_i - \mu - \Delta t_i)^2 + \sum_{i=1}^n \log(h(u_i; \nu)), \quad (2.14)$$

em que c é uma constante que não depende de $\boldsymbol{\theta}$.

Seja $\hat{u}_i = E[U_i|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, y_i]$, $\hat{ut}_i = E[U_i t_i|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, y_i]$, $\hat{ut}_i^2 = E[U_i t_i^2|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, y_i]$ e usando a propriedade recursiva do valor esperado, obtemos

$$\hat{ut}_i = \hat{u}_i(\hat{\mu}_{T_i} + b) + \hat{M}_T \hat{\tau}_{1_i}, \quad \hat{ut}_i^2 = \hat{u}_i(\hat{\mu}_{T_i} + b)^2 + \hat{M}_T^2 + \hat{M}_T(\hat{\mu}_{T_i} + 2b)\hat{\tau}_{1_i},$$

em que

$$\hat{\tau}_{1_i} = E \left[U_i^{1/2} W_\Phi \left(\frac{U_i^{1/2} \hat{\mu}_{T_i}}{\hat{M}_T} \right) | \hat{\boldsymbol{\theta}}, y_i \right].$$

As esperanças condicionais: $\hat{u}_i = \hat{u}_{1_i}$ e $\hat{\tau}_{1_i}$, podem ser derivadas dos resultados da Subseção 2.1 dado por [Basso et al. \(2009\)](#).

Assim, o valor esperado condicional da log-verossimilhança completa é:

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) &= E[\ell_{comp}(\boldsymbol{\theta})|y, \hat{\boldsymbol{\theta}}^{(k)}] = c - \frac{n}{2} \log(\Gamma) - \frac{1}{2\Gamma} \sum_{i=1}^n \left[\hat{u}_i^{(k)} (y_i - \mu)^2 \right. \\ &\quad \left. - 2\Delta(y_i - \mu)\hat{ut}_i^{(k)} + \Delta^2 \hat{ut}_i^2^{(k)} + E \left[\log(h(u_i; \nu)) | y, \hat{\boldsymbol{\theta}}^{(k)} \right] \right], \end{aligned}$$

onde $\hat{\boldsymbol{\theta}}^{(k)}$ é o valor atualizado de $\hat{\boldsymbol{\theta}}$.

O passo M contém expressões matemáticas analiticamente intratáveis, que podem ser substituídas por uma seqüência de passos de maximização condicional (CM). Este procedimento é conhecido como **Algoritmo ECM** ([Meng & Rubin \(1993\)](#)). A seguir

descreveremos este tipo de algoritmo EM para a estimativa de máxima verossimilhança dos parâmetros da distribuição MESN.

Passo E: Dado $\widehat{\boldsymbol{\theta}}^{(k)}$, calculamos $\widehat{u}_i^{(k)}$, $\widehat{ut}_i^{(k)}$, $\widehat{ut}_i^2^{(k)}$, para $i = 1, \dots, n$.

Passo CM: Atualizamos $\widehat{\boldsymbol{\theta}}^{(k)}$ maximizando $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, o qual resulta nas seguintes expressões:

$$\begin{aligned}\widehat{\mu}^{(k+1)} &= \frac{\sum_{i=1}^n (\widehat{u}_i^{(k)} y_i - \widehat{ut}_i^{(k)} \widehat{\Delta}^{(k)})}{\sum_{i=1}^n \widehat{u}_i^{(k)}}, \\ \widehat{\Gamma}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \left[\widehat{u}_i^{(k)} (y_i - \widehat{\mu}^{(k+1)})^2 - 2(y_i - \widehat{\mu}^{(k+1)}) \widehat{ut}_i^{(k)} \widehat{\Delta}^{(k+1)} + \widehat{ut}_i^2^{(k)} (\widehat{\Delta}^2)^{(k+1)} \right], \\ \widehat{\Delta}^{(k+1)} &= \frac{\sum_{i=1}^n \widehat{ut}_i^{(k)} (y_i - \widehat{\mu}^{(k+1)})}{\sum_{i=1}^n \widehat{ut}_i^2^{(k)}},\end{aligned}$$

em que $\psi(y; \boldsymbol{\theta})$ é a densidade MESN.

Os valores de $\widehat{\sigma}^2^{(k+1)}$ e $\widehat{\lambda}^{(k+1)}$ podem ser recuperados pelas relações $\widehat{\lambda} = \widehat{\Delta}/\sqrt{\widehat{\Gamma}}$ e $\widehat{\sigma}^2 = \widehat{\Delta}^2 + \widehat{\Gamma}$, respectivamente, tal que o processo de estimação é iterado até cumprir alguma regra de convergência, por exemplo se $\|\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}\| < \epsilon$, com ϵ suficientemente pequeno. Neste trabalho, denominaremos essa variante do algoritmo EM por algoritmo tipo EM

2.4.2 Estudo de simulação

Com o intuito de avaliar a performance do método proposto para estimar os parâmetros no modelo com distribuição MESN, apresentamos o seguinte estudo de simulação, onde mostramos que a estimação dos parâmetros através do algoritmo tipo EM proporciona boas propriedades assintóticas, como esperado.

Realizamos um estudo de simulação Monte Carlo com o seguinte modelo:

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.15)$$

onde $\varepsilon_i \sim MESN(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma_i^2, \lambda; H)$ são variáveis aleatórias iid com média zero. A variável x_i foi gerada considerando uma distribuição uniforme no intervalo $(0.1, 20)$ e os valores são os mesmos em todas as simulações. Os valores dos parâmetros foram estabelecidos como $\beta_1 = 37, \beta_2 = 0.6, \sigma^2 = 0.5, \lambda = -3$.

Assim, com o objetivo de avaliar as estimativas dos parâmetros para o modelo (2.15) comparamos o valor do desvio relativo médio dos parâmetros para diferentes tamanhos de amostra, sendo o principal foco de nosso interesse a avaliação do viés e do erro quadrático médio (EQM) das estimações obtidas pelo algoritmo tipo EM. Os tamanhos de amostras fixados são $n = 30, 50, 100, 150, 200, 250, 300, 400, 500$ e para cada combinação dos parâmetros e tamanhos de amostra, foram geradas 1000 amostras dos modelos MESN, isto é: skew normal(SN-LM), a skew-t (ST-LM) com $\nu = 3$, e a skew slash (SSL-LM) com $\nu = 2$. Então calculamos o viés e o erro quadrático médio (EQM) sobre todas as réplicas.

Para $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) = (\beta_1, \beta_2, \sigma^2, \lambda)$, temos que

$$Vies(\theta_j) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_j^{(i)} - \theta_j) \quad \text{e} \quad EQM(\theta_j) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_j^{(i)} - \theta_j)^2, \quad j = 1, \dots, 4. \quad (2.16)$$

Nas Tabelas 2.1 e 2.2 e nas Figuras 2.2 e 2.3, podemos observar que os valores médios do viés e do EQM são pequenos em todos os parâmetros, a exceção do λ . Assim, temos que a medida que os tamanhos de amostras aumentam os valores do viés e EQM se aproximam a zero.

Podemos concluir através dos resultados do estudo de simulação que a performance dos EMV obtidos pelo algoritmo tipo EM é aceitável, já que tanto o viés como o EQM tendem a zero à medida que o tamanho da amostra n aumenta. No entanto ressaltamos que para tamanhos não muito grandes de amostra, o parâmetro λ tem estimativas viesadas.

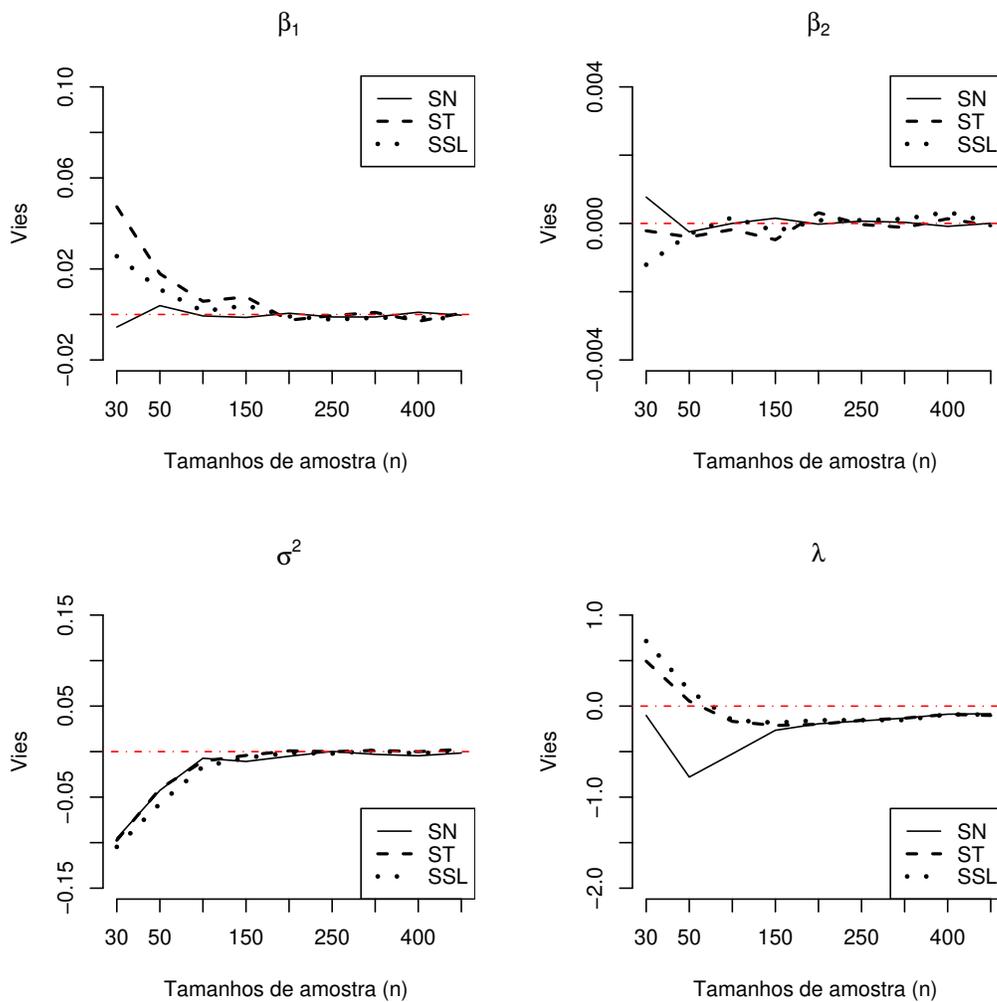


Figura 2.2: Viés dos EMV de β_1 , β_2 , σ^2 e λ para o modelo $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \dots, n$.

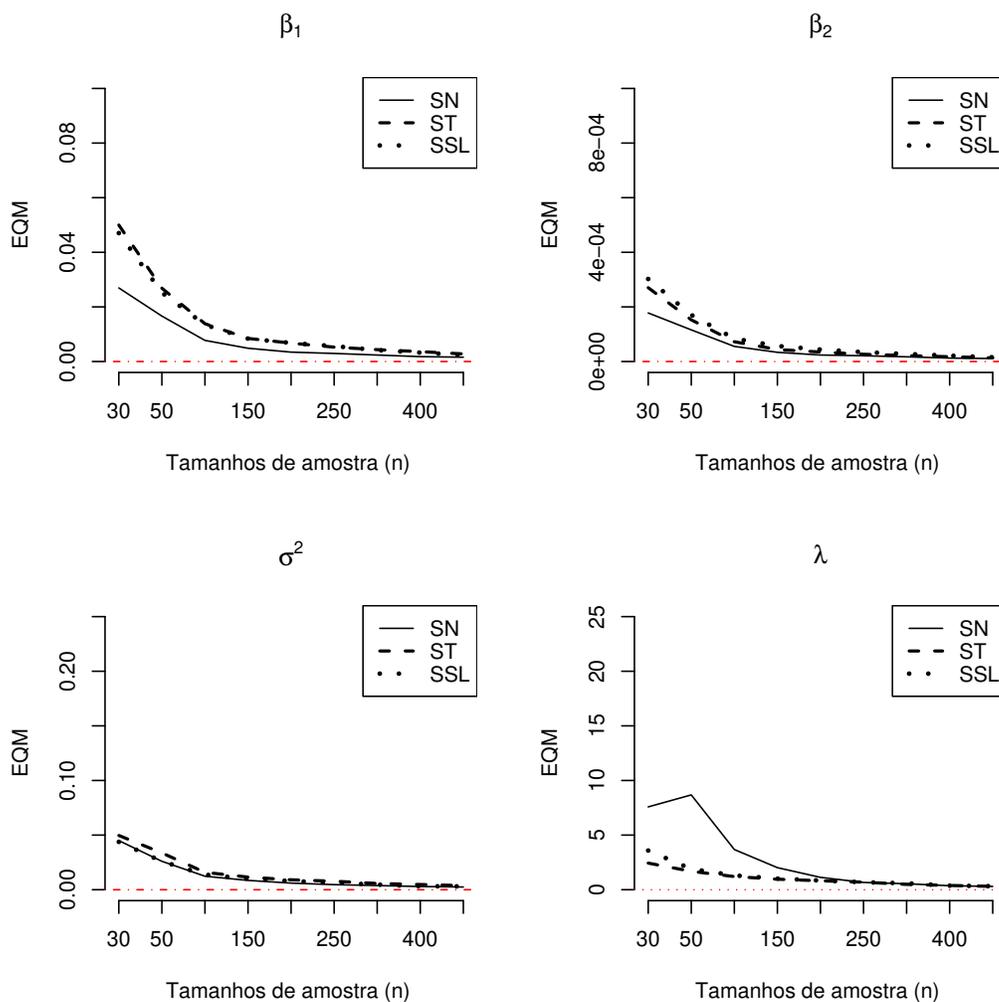


Figura 2.3: EQM dos EMV β_1 , β_2 , σ^2 , e λ para o modelo $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \dots, n$.

2.5 Considerações finais

Neste capítulo apresentamos a classe de distribuições MESN, desenvolvendo suas propriedades mais relevantes, assim como, o algoritmo tipo EM para estimação de máxima verossimilhança (EMV) dos parâmetros. Mostramos por meio de um estudo de simulação que as estimações do modelo locação-escala na classe de distribuições MESN têm boas propriedades assintóticas, no sentido que o viés e o erro quadrático médio tendem a zero. Esta classe de distribuições é bastante abrangente e tem sido

Tabela 2.1: Viés dos EMV dos parâmetros com diferentes tamanhos de amostras.

SN				
n	β_1	β_2	σ^2	λ
30	-0.005538	7.704056e-04	-0.095174	-0.102838
100	-0.000642	1.242774e-06	-0.007267	-0.524677
200	0.000543	-2.224958e-05	-0.005089	-0.194227
300	-0.001105	3.248876e-05	-0.003005	-0.132154
500	-0.000292	3.221047e-06	-0.001549	-0.083597
ST				
n	β_1	β_2	σ^2	λ
30	0.047329	-2.157411e-04	-0.097386	0.493508
100	0.005783	-1.790982e-04	-0.010565	-0.167699
200	-0.002517	3.138173e-04	0.000871	-0.199745
300	0.000892	-1.189951e-04	0.001325	-0.132872
500	0.000531	-6.684754e-05	0.002441	-0.102653
SSL				
n	β_1	β_2	σ^2	λ
30	0.025595	-1.20951e-03	-0.104648	0.714446
100	0.001434	1.78818e-04	-0.016474	-0.154642
200	-0.000813	9.20431e-05	-0.001448	-0.155439
300	-0.001458	1.30775e-04	0.001162	-0.153093
500	-0.001877	2.91470e-05	0.001742	-0.093730

a base de alguns estudos recentes como, por exemplo em [Basso et al. \(2009\)](#) usam esta classe de distribuições em misturas finitas de distribuições; [Zeller et al. \(2010\)](#) que estudam esta classe de distribuições em modelos lineares mistos. No entanto, surge a inquietude de estender estes métodos em modelos não lineares com erros sob a classe de distribuições MESN.

Tabela 2.2: EQM dos EMV dos parâmetros com diferentes tamanhos de amostras.

SN				
n	β_1	β_2	σ^2	λ
30	0.026953	1.779123e-04	0.044952	7.580480
100	0.007735	5.535795e-05	0.012346	3.677420
200	0.003441	2.386031e-05	0.006051	1.130500
300	0.002374	1.724803e-05	0.003811	0.545057
500	0.001584	1.092020e-05	0.002467	0.300480
ST				
n	β_1	β_2	σ^2	λ
30	0.049981	2.703588e-03	0.049676	2.437909
100	0.013850	7.246418e-05	0.015904	1.203927
200	0.006733	3.479262e-05	0.009069	0.817658
300	0.004349	2.155480e-05	0.005716	0.486148
500	0.002789	1.327071e-05	0.003720	0.302482
SSL				
n	β_1	β_2	σ^2	λ
30	0.047029	3.023485e-04	0.043794	3.586345
100	0.013468	8.405593e-05	0.013784	1.307183
200	0.006815	4.386702e-05	0.007931	0.842009
300	0.004190	2.759513e-05	0.004641	0.588331
500	0.002387	1.5668120e-05	0.003073	0.323788

Capítulo 3

Modelos não lineares sob a classe de distribuições MESN

Os modelos normais de regressão linear e não-linear podem ser considerados como uma das ferramentas mais populares para a análise estatística de dados. No entanto, os fenômenos sob investigação nem sempre estão de acordo com o modelo normal, seja devido a falta de simetria ou a presença de observações atípicas. Para lidar com este problema, algumas propostas foram elaboradas na literatura substituindo a suposição de normalidade por classes paramétricas (e não paramétricas) mais flexíveis de distribuições.

Dentro deste contexto, e sendo o foco de nosso trabalho os modelos não lineares podemos citar os seguintes trabalhos: [Cordeiro *et al.* \(2000\)](#) que estudaram as estimativas de máxima verossimilhança corrigidos na classe simétrica de modelos de regressão não linear. [Cysneiros & Vanegas \(2008\)](#) que estudaram os modelos simétricos de regressão não linear, realizando um estudo analítico e empírico para descrever o comportamento dos resíduos padronizados. [Lemonte & Cordeiro \(2009\)](#) que estudaram

os modelos não lineares considerando que os erros seguem uma distribuição Birnbaum-Saunders. Por outro lado, [Cancho *et al.* \(2009\)](#) introduzem os modelos de regressão não linear skew normal (SN-NLM) e apresentam uma análise completa baseada na verossimilhança, incluindo o algoritmo EM para encontrar os EMV. No entanto, não têm sido estudado os modelos não lineares sob a classe MESN.

A análise estatística se assenta em três pilares fundamentais: a modelagem que estamos considerando, a estabilidade dos resultados e inferências e a avaliação das observações influentes, que são os elementos dos dados que controlam os aspectos da análise. Sobre este último, existem duas principais abordagens para detecção de observações influentes: **i)** a abordagem **modelo de eliminação de casos (CDM)**, em que o impacto de excluir uma observação sobre a previsão está diretamente avaliada por medidas tais como o afastamento pela log-verossimilhança e a distância de Cook (ver, [Cook, 1977](#)) e **ii)** a abordagem conhecida como **influência local**, que foi desenvolvida inicialmente por [Cook \(1986\)](#). Nesse contexto, nos últimos anos, muitos autores estudaram estas duas abordagens para os modelos de regressão linear e não linear com distribuições não normais, como por exemplo [Galea *et al.* \(1997\)](#) que estudaram a análise de influência local em modelos de regressão linear elípticos. [Ortega *et al.* \(2003\)](#) que estudaram a análise de diagnóstico em modelos de regressão log-gamma generalizada. [Lachos *et al.* \(2005\)](#) e [Lachos *et al.* \(2008\)](#) que aplicaram a técnica de influência local no modelo de calibração comparativa e no modelo de regressão com erros nas variáveis com intercepto nulo, respectivamente. [Ferreira \(2008\)](#) que fez uma análise de inferência e diagnóstico em modelos assimétricos. [Xie *et al.* \(2009a\)](#) que estudaram a análise de diagnóstico para os modelos de regressão não linear skew normal com erros AR(1). [Lin *et al.* \(2009\)](#) que estudaram a análise de diagnóstico para os modelos não lineares com distribuição skewt-t-normal e recentemente [Zeller *et al.* \(2010\)](#) que fazem o análise de influência nos modelos lineares mistos com skew normal. No entanto, até agora não existe a análise de diagnóstico nos modelos não lineares sob a classe de distribuição MESN.

Neste capítulo apresentaremos os modelos não lineares sob a classe de distribuições misturas da escala skew normal, desenvolvendo as duas abordagens da análise de diagnóstico para esta classe de modelos. Os resultados deste trabalho, serão o complemento da teoria apresentada no capítulo anterior e servirá como uma alternativa para avaliar e estudar os modelos não lineares, sobre um enfoque diferente das técnicas existentes.

Iniciaremos este capítulo com a descrição dos modelos não lineares sob a classe de distribuições mistura da escala skew normal (NLM-MESN). Definiremos a função de log-verossimilhança, a função score e a matriz de informação observada (MIO). Descreveremos o algoritmo tipo EM para obtenção das estimativas de máxima verossimilhança (EMV) dos parâmetros do modelo. Apresentaremos também, alguns critérios de seleção de modelos e a construção dos envelopes simulados ([Atkinson, 1981](#)) com o intuito de analisar os resíduos de Pearson, além disso, avaliaremos por meio de um estudo de simulação o desempenho dos EMV dos parâmetros dos modelos não lineares sob a classe de distribuições MESN.

Desenvolveremos de uma maneira detalhada o análise de diagnóstico para os modelos não lineares sob a classe de distribuições mistura da escala skew normal (NLM-MESN), considerando o modelo de eliminação de casos (CDM) e a influência local. Estudaremos o conjunto de dados, *Produtividade de óleo de palma*, que foi apresentado em [Foong \(1999\)](#) e trabalhado por [Cancho et al. \(2009\)](#), com a finalidade de estimar os parâmetros considerando os modelos não lineares sob a classe MESN, fazendo também uma análise de diagnóstico. Por fim, apresentaremos as considerações gerais do capítulo.

3.1 Descrição do modelo

O modelo de regressão não linear sob a classe de distribuições MESN, denotado por NLM-MESN, é definido por:

$$Y_i = \eta(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.1)$$

em que Y_i é a variável resposta, $\eta(\cdot)$ é uma função injetiva e duas vezes diferenciável com respeito ao vetor de parâmetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, \mathbf{x}_i é um vetor de valores das variáveis explicativas. Neste trabalho assumimos que os erros aleatórios $\varepsilon_i \sim MESN(-\sqrt{\frac{2}{\pi}}\kappa_1\Delta, \sigma^2, \lambda; H)$, com $\kappa_1 = E\{U^{-1/2}\}$, $\Delta = \sigma\delta$ e $\delta = \lambda/(1 + \lambda^2)^{1/2}$, que correspondem ao modelo em que os erros têm média zero. Pela propriedade de linearidade temos que $Y_i \sim MESN(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + b\Delta, \sigma^2, \lambda; H)$, para $i = 1, \dots, n$. Então:

$$E[Y_i] = \eta(\boldsymbol{\beta}, \mathbf{x}_i), \quad Var[Y_i] = \kappa_2\sigma^2 - b^2\Delta^2, \quad \text{em que } b = -\sqrt{\frac{2}{\pi}}\kappa_1.$$

Estes resultados podem ser provados usando o Lema 2 dado em [Basso et al. \(2009\)](#) e a Proposição 2.3.3, isto é,

$$\begin{aligned} E[Y_i] &= E[\eta(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i] = \eta(\boldsymbol{\beta}, \mathbf{x}_i) + E[\varepsilon_i] \\ &= \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \sqrt{\frac{2}{\pi}}\kappa_1\Delta + \sqrt{\frac{2}{\pi}}\kappa_1\Delta \\ &= \eta(\boldsymbol{\beta}, \mathbf{x}_i) \\ Var[Y_i] &= Var[\eta(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i] = Var[\varepsilon_i] \\ &= \kappa_2\sigma^2 - b^2\sigma^2\delta^2 = \kappa_2\sigma^2 - b^2\Delta^2 \end{aligned}$$

Seguindo as sugestões dadas por [Lange et al. \(1989\)](#) e [Berkane et al. \(1994\)](#), assumimos que o valor de ν é conhecido ou fixado de antemão. Estes autores apontaram dificuldades na estimação de ν , devido a problemas na função de verossimilhança.

3.2 Função de log-verossimilhança

A função de log-verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$, dada uma amostra $\mathbf{y} = (y_1, \dots, y_n)^\top$, é definida por $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$, em que

$$\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \log K_i,$$

com $K_i = \int_0^\infty u_i^{1/2} \exp\{-\frac{1}{2}u_i d_i\} \Phi(u_i^{1/2} A_i) dH(u_i)$, em que $d_i = (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta)^2 / \sigma^2$ é a distância de Mahalanobish e $A_i = (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta)\lambda / \sigma$.

- *Função escore*

A função escore é dada por $\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n U_i(\boldsymbol{\theta})$, em que $U_i(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (U_i(\boldsymbol{\beta})^\top, U_i(\sigma^2), U_i(\lambda))^\top$ e para $\boldsymbol{\gamma} = \boldsymbol{\beta}, \sigma^2$ ou λ , $U_i(\boldsymbol{\gamma})$ tem a forma:

$$U_i(\boldsymbol{\gamma}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2} \frac{\partial \log \sigma^2}{\partial \boldsymbol{\gamma}} + \frac{1}{K_i} \frac{\partial K_i}{\partial \boldsymbol{\gamma}}, \quad i = 1, \dots, n, \quad (3.2)$$

em que

$$\frac{\partial K_i}{\partial \boldsymbol{\gamma}} = I_i^\phi(1) \frac{\partial A_i}{\partial \boldsymbol{\gamma}} - \frac{1}{2} I_i^\Phi \left(\frac{3}{2} \right) \frac{\partial d_i}{\partial \boldsymbol{\gamma}}.$$

- *Matriz de informação observada*

A matriz de informação observada $\mathbf{J}(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$, tem elementos dados por $J_{\boldsymbol{\gamma}\boldsymbol{\tau}} = -\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}$, para $\boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\beta}, \sigma^2$ ou λ , onde

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = -\frac{1}{2} \frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} - \frac{1}{K_i^2} \frac{\partial K_i}{\partial \boldsymbol{\gamma}} \frac{\partial K_i}{\partial \boldsymbol{\tau}^\top} + \frac{1}{K_i} \frac{\partial^2 K_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}.$$

e

$$\begin{aligned} \frac{\partial^2 K_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} &= \frac{1}{4} I_i^\Phi \left(\frac{5}{2} \right) \frac{\partial d_i}{\partial \boldsymbol{\gamma}} \frac{\partial d_i}{\partial \boldsymbol{\tau}^\top} - \frac{1}{2} I_i^\Phi \left(\frac{3}{2} \right) \frac{\partial^2 d_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} - \frac{1}{2} I_i^\phi(2) \left(\frac{\partial A_i}{\partial \boldsymbol{\gamma}} \frac{\partial d_i}{\partial \boldsymbol{\tau}^\top} + \frac{\partial d_i}{\partial \boldsymbol{\gamma}} \frac{\partial A_i}{\partial \boldsymbol{\tau}^\top} \right) \\ &\quad - I_i^\phi(2) A_i \frac{\partial A_i}{\partial \boldsymbol{\gamma}} \frac{\partial A_i}{\partial \boldsymbol{\tau}^\top} + I_i^\phi(1) \frac{\partial^2 A_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}, \end{aligned}$$

com,

$$I_i^\Phi(w) = \int_0^\infty u_i^w \exp\left(-\frac{1}{2}u_i d_i\right) \Phi_1(u_i^{1/2} A_i) dH(u_i)$$

e

$$I_i^\phi(w) = \frac{1}{\sqrt{2\pi}} \int_0^\infty u_i^w \exp\left(-\frac{1}{2}u_i(d_i + A_i^2)\right) dH(u_i).$$

Note que podemos escrever $K_i = I_i^\Phi(\frac{1}{2})$, para $i = 1, \dots, n$.

De [Basso et al. \(2009\)](#), temos os seguintes resultados para as principais distribuições MESN consideradas neste trabalho:

- *Skew-t*:

$$\begin{aligned} I_i^\Phi(w) &= \frac{2^w \nu^{\nu/2} \Gamma(w + \nu/2)}{\Gamma(\nu/2)(\nu + d_i)^{\nu/2+w}} T\left(\sqrt{\frac{\nu + 2w}{d_i + \nu}} A_i; \nu + 2w\right) \text{ e} \\ I_i^\phi(w) &= \frac{2^w \nu^{\nu/2} \Gamma(\frac{\nu+2w}{2})}{\sqrt{2\pi} \Gamma(\nu/2)(d_i + A_i^2 + \nu) \frac{\nu + 2w}{2}}. \end{aligned}$$

- *Skew-slash*:

$$\begin{aligned} I_i^\Phi(w) &= \frac{\nu 2^{\nu+w} \Gamma(\nu + w)}{d_i^{\nu+w}} P_1\left(\nu + w, \frac{d_i}{2}\right) E\{\Phi(S_i^{1/2} A_i)\} \text{ e} \\ I_i^\phi(w) &= \frac{\nu 2^{\nu+w} \Gamma(\nu + w)}{\sqrt{2\pi} (d_i + A_i^2)^{\nu+w}} P_1\left(\nu + w, \frac{d_i + A_i^2}{2}\right), \end{aligned}$$

onde $S_i \sim \text{Gamma}(\nu + w, \frac{d_i}{2}) \mathbb{I}_{(0,1)}$.

As derivadas de d_i e A_i envolvem manipulações algébricas padronizadas e podem ser encontradas em [Cancho et al. \(2009\)](#).

Note que desde que temos expressões para a matriz de informação observada e função escore para $\boldsymbol{\theta}$, o método de Newton-Raphson também pode ser facilmente implementado para obter os EMV. Contudo, neste trabalho discutiremos os passos necessários para a implementação do algoritmo tipo EM.

3.3 Estimação dos parâmetros via algoritmo tipo EM

Com o objetivo de estimar os parâmetros nos modelos NLM-MESN via algoritmo tipo EM, seguimos a seqüência apresentada na Subseção 2.4.1, para a qual utilizamos a seguinte representação hierárquica para Y_i

$$Y_i|U_i = u_i, T_i = t_i \sim N(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + \Delta t_i, u_i^{-1}\Gamma), \quad (3.3)$$

$$T_i|U_i = u_i \sim NT(b, u_i^{-1}; (b, \infty)), \quad (3.4)$$

$$U_i \sim H(., \boldsymbol{\nu}), \quad (3.5)$$

em que,

$$\Gamma = (1 - \delta^2)\sigma^2, \quad \Delta = \sigma\delta, \quad b = -\sqrt{\frac{2}{\pi}}\kappa_1, \quad e \quad \kappa_1 = E\{U^{-1/2}\}. \quad (3.6)$$

Segue que a distribuição de $T_i|Y_i = y_i, U_i = u_i$ é $NT((\mu_{T_i} + b, u_i^{-1}M_T^2); (b, \infty))$, com

$$M_T^2 = \frac{\Gamma}{\Delta^2 + \Gamma}, \quad e \quad \mu_{T_i} = \frac{\Delta}{\Delta^2 + \Gamma}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \Delta b).$$

Passo E:

Sejam $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{t} = (t_1, \dots, t_n)^\top$ e $\mathbf{u} = (u_1, \dots, u_n)^\top$. A função de log-verossimilhança completa para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$ associada com $(\mathbf{y}, \mathbf{t}, \mathbf{u})$ é dada por:

$$\ell_{comp}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{t}, \mathbf{u}) = c - \frac{n}{2} \log \Gamma - \frac{1}{2\Gamma} \sum_{i=1}^n u_i (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \Delta t_i)^2, \quad (3.7)$$

em que c é uma constante que não depende de $\boldsymbol{\theta}$. Assim, dado $\widehat{\boldsymbol{\theta}}^{(k)}$, encontramos $\widehat{u}_i^{(k)}$, $\widehat{ut}_i^{(k)}$ e $\widehat{ut}_i^2^{(k)}$ para $i = 1, \dots, n$, através das seguintes relações:

$$\begin{aligned} \widehat{u}_i^{(k)} &= E[U_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}, y_i], \\ \widehat{ut}_i^{(k)} &= E[U_i t_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}, y_i] = \widehat{u}_i^{(k)}(\widehat{\mu}_{T_i}^{(k)} + b) + \widehat{M}_T^{(k)} \widehat{\tau}_{1_i}^{(k)}, \\ \widehat{ut}_i^2^{(k)} &= E[U_i t_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}, y_i] = \widehat{u}_i^{(k)}(\widehat{\mu}_{T_i}^{(k)} + b)^2 + \widehat{M}_T^2{}^{(k)} + \widehat{M}_T^{(k)}(\widehat{\mu}_{T_i}^{(k)} + 2b)\widehat{\tau}_{1_i}^{(k)}, \end{aligned}$$

em que

$$\widehat{\tau}_{1_i}^{(k)} = E \left[U_i^{1/2} W_{\Phi} \left(\frac{U_i^{1/2} \widehat{\mu}_{T_i}^{(k)}}{\widehat{M}_T^{(k)}} \right) | \widehat{\boldsymbol{\theta}}^{(k)}, y_i \right].$$

Para a distribuição skew-t e skew-slash temos expressões computacionais atrativas que podem ser facilmente implementadas (vide [Basso, 2009](#)).

Passo CM:

A Q-função é dada por:

$$\begin{aligned} Q(\boldsymbol{\theta} | \widehat{\boldsymbol{\theta}}^{(k)}) &= E[\ell_{comp}(\boldsymbol{\theta}) | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}] \\ &= c - \frac{n}{2} \log(\Gamma) - \frac{1}{2\Gamma} \sum_{i=1}^n \left[\widehat{u}_i^{(k)} (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i))^2 - 2\Delta (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i)) \widehat{ut}_i^{(k)} + \Delta^2 \widehat{ut}_i^2^{(k)} \right], \end{aligned}$$

em que $\widehat{\boldsymbol{\theta}}^{(k)}$ é o valor atualizado de $\widehat{\boldsymbol{\theta}}$. O passo CM consiste em atualizar $\widehat{\boldsymbol{\theta}}^{(k)}$ pela maximização $Q(\boldsymbol{\theta} | \widehat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, que leva as seguintes expressões:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \operatorname{argmin}_{\boldsymbol{\beta}} (\mathbf{z}^{(k)} - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}))^\top \widehat{\mathbf{U}}^{(k)} (\mathbf{z}^{(k)} - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x})), \quad (3.8)$$

$$\widehat{\Delta}^{(k+1)} = \frac{\sum_{i=1}^n \widehat{ut}_i^{(k)} (y_i - \eta(\boldsymbol{\beta}^{(k+1)}, \mathbf{x}_i))}{\sum_{i=1}^n \widehat{ut}_i^2^{(k)}}, \quad (3.9)$$

$$\begin{aligned} \widehat{\Gamma}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \left((y_i - \eta(\boldsymbol{\beta}^{(k+1)}, \mathbf{x}_i))^2 \widehat{u}_i^{(k)} - 2\Delta^{(k+1)} (y_i - \eta(\boldsymbol{\beta}^{(k+1)}, \mathbf{x}_i)) \widehat{ut}_i^{(k)} \right. \\ &\quad \left. + (\Delta^2)^{(k+1)} \widehat{ut}_i^2^{(k)} \right), \end{aligned} \quad (3.10)$$

em que $\widehat{\mathbf{U}}^{(k)} = \operatorname{diag}(\widehat{u}_1^{(k)}, \dots, \widehat{u}_n^{(k)})$ e $\mathbf{z}^{(k)}$ é a resposta observada corrigida, dada por $\mathbf{z}^{(k)} = \mathbf{y} - \widehat{\Delta}^{(k)} \widehat{\boldsymbol{\tau}}^{(k)}$, com $\widehat{\boldsymbol{\tau}}^{(k)} = (\widehat{\tau}_1^{(k)}, \dots, \widehat{\tau}_n^{(k)})^\top$ e $\widehat{\tau}_i^{(k)} = \widehat{ut}_i^{(k)} / \widehat{u}_i^{(k)}$. Este processo é iterado até que uma regra de convergência definida seja satisfeita, isto é, se $\|\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}\|$ é suficientemente pequeno ou até que uma certa diferença envolvendo duas avaliações sucessivas do log-verossimilhança $\|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})\|$ ou $\|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) / \ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1\|$, seja suficientemente pequena. Uma observação interessante é que no passo CM as estimações de $\boldsymbol{\beta}$ são equivalentes às estimativas obtidas pelo método de mínimos quadrados ponderados em um modelo não linear da forma $\mathbf{z} = \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}) + \boldsymbol{\epsilon}$, permitindo uma eficiente implementação do algoritmo em softwares estatísticos como SAS, R, Ox e Matlab. Note

tambm que as estimativas para $\sigma^{2(k+1)}$ e $\lambda^{(k+1)}$ podem ser obtidas usando (3.6), isto , $\hat{\lambda} = \hat{\Delta}/\sqrt{\hat{\Gamma}}$ e $\hat{\sigma}^2 = \hat{\Delta}^2 + \hat{\Gamma}$.

3.3.1 Consideraces para a implementaco

Sabe-se que os mtodos computacionais para estimativas de mxima verossimilhana de funces no-lineares podem enfrentar alguns obstculos, no sentido de que o mtodo no necessariamente pode fornecer solues globais, se os valores iniciais esto longe dos valores verdadeiros dos parmetros. Assim, a escolha dos valores iniciais para o algoritmo tipo EM no contexto no-linear tem um grande papel na estimativa dos parmetros. Em nossos exemplos consideraremos o seguinte procedimento para o caso SN-NLM:

- Obter as estimativas $\hat{\beta}^{(0)}$ usando o mtodo de mnimos quadrados ponderados
- Calcular os valores iniciais $(\hat{\sigma}^2)^{(0)}$ e $\hat{\lambda}^{(0)}$ usando os residuais e o mtodo dos estimadores de momentos, apresentados em Basso *et al.* (2009) e Lin *et al.* (2007), com $M_1 = 0$. O intervalo para o coeficiente de assimetria γ_1 da distribuo SN  aproximadamente $(-0.9953, 0.9953)$, como apresentado em Azzalini (2005). No entanto, os mtodos de momentos podem produzir valores iniciais de $\hat{\gamma}_1^{(0)}$ que no esto neste intervalo. Neste caso, usamos como pontos iniciais os valores -0.99 (se $\gamma_1^{(0)} \leq -0.9953$) ou 0.99 (se $\gamma_1^{(0)} \geq 0.9953$).

No entanto, quando modelamos dados usando a ST-NLM, SCN-NLM ou a SSL-NLM, adotamos a seguinte estratgia:

- Obter os valores iniciais atravs do mtodo dos momentos para o SN-NLM, como foi descrito no item anterior;
- Obter as estimaces dos parmetros da SN-NLM, via o algoritmo tipo EM;
- Usar as estimativas obtidas no item anterior como valores iniciais do algoritmo tipo EM para os correspondentes parmetros dos modelos ST-NLM e SSL-NLM;

- Para estimar ν no modelo ST-NLM e SSL-NLM, fixamos valores inteiros para ν desde 3 até 100 e desde 2 até 100, respectivamente, escolhendo o valor de ν que maximiza a função de verossimilhança.

3.3.2 Seleção de modelos

Alguns critérios de escolha de modelos geralmente utilizados são: o AIC ([Akaike, 1974](#)) e o Critério de Determinação de Eficiência (EDC) ([Bai et al., 1989](#)). Estes são definidos por

$$-2\ell(\hat{\boldsymbol{\theta}}) + \gamma c_n,$$

em que $\ell(\cdot)$ é a log-verossimilhança, γ é o número de parâmetros livres que são estimados no modelo e o termo de penalização c_n ; que é uma seqüência (conveniente) de números positivos. Temos que $c_n = 2$ para o AIC. Para o critério EDC, c_n é escolhido de modo que satisfaz as condições $c_n/n \rightarrow 0$ e $c_n/(\log \log n) \rightarrow 0$, quando $n \rightarrow \infty$. Aqui usamos $c_n = 0.2\sqrt{n}$, que foi considerado em [Bai et al. \(1989\)](#).

3.3.3 Análise de resíduos

A análise de resíduos tem como finalidade identificar observações atípicas, assim como verificar a correta especificação do modelo ajustado. Para isto, neste trabalho definimos o seguinte resíduo padronizado, comumente chamado de resíduo de Pearson,

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(y_i)}}, \quad (3.11)$$

em que $\widehat{Var}(y_i) = \kappa_2 \hat{\sigma}^2 - \frac{2}{\pi} \kappa_1^2 \hat{\sigma}^2 \hat{\delta}^2$, $\hat{\mu}_i = \eta(\hat{\boldsymbol{\beta}}, \mathbf{x}_i)$, $i = 1, \dots, n$, com $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ e $\hat{\delta}$ sendo as EMV de $\boldsymbol{\beta}$, σ^2 e δ , respectivamente.

Com estes resíduos padronizados construímos envelopes simulados, como foi sugerido por [Atkinson \(1981\)](#), para detectar uma incorreta especificação da distribuição do erro, da componente sistemática $\eta(\boldsymbol{\beta}, \mathbf{x}_i)$ assim como a presença de observações atípicas.

3.3.4 Estudo de simulaco

Na Subsecco 2.4.2 desenvolvemos um estudo de simulaco para avaliar a performance dos EMV obtidos pelo algoritmo tipo EM nos modelos MESN. Nesta Subsecco, nosso objetivo  comparar o desempenho das estimativas dos parmetros nos modelos no lineares na classe MESN com presenca de *outliers* na varivel resposta. Para isto, realizamos um estudo de simulaco Monte Carlo usando o seguinte modelo:

$$Y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \varepsilon_i, \quad i = 1, \dots, n = 50, \quad (3.12)$$

onde $\varepsilon_i \stackrel{\text{iid}}{\sim} MESN(-\sqrt{\frac{2}{\pi}}\kappa_1\Delta, \sigma^2, \lambda; H)$ so variveis aleatrias independentes e idnticamente distribudas com mdia zero. Cada valor de x_i foi gerado considerando uma distribuo uniforme no intervalo (4, 53) e so os mesmos em todas as simulaces.

Os valores verdadeiros dos parmetros considerados para nosso estudo foram: $\beta_1 = 37.3, \beta_2 = 44.5, \beta_3 = 0.73, \sigma^2 = 2.95, \lambda = -2$ (*i.e.*, $\Delta = 1.537$). Geramos 2000 amostras da $SN(-\sqrt{2/\pi}\Delta, \sigma^2, \lambda)$ e para garantir a presenca de um *outlier*, fizemos $Y_i^* = Y_i - \vartheta$, onde Y_i  o valor central da amostra e $\vartheta = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. Em cada rplica obtemos os parmetros estimados com e sem a perturbao considerada, denotadas por $\hat{\theta}$ e $\hat{\theta}_{(i)}$, respectivamente. Em todos os casos consideramos o modelo skew normal (SN-NLM), skew-t (ST-NLM) com valores de $\nu = (3, 6, 10)$, skew-slash (SSL-NLM) com $\nu = 2$ e skew normal contaminada (SCN-NLM) com $\nu = (0.2, 0.2)$. Por fim, calculamos as mudanas relativas das estimativas de $\beta_1, \beta_2, \beta_3$ e σ^2 usando $\left| \frac{\hat{\theta}_{(i)}}{\hat{\theta}} - 1 \right|$.

Nas Figuras 3.1 e 3.2, podem ser encontrados os valores mdios das mudanas relativas nas estimativas para as 2000 rplicas, onde podemos observar que para todos os modelos as mudanas relativas aumentam conforme aumentam os valores de ϑ . No modelo skew-t a influncia nas estimativas aumentam quando ν tambm aumenta. No entanto, para os modelos MESN com caudas pesadas, estas medidas variam pouco, o que indica que esses modelos so mais robustos do que o modelo SN-NLM na presenca de observaes atpicas. Note tambm que, ao contrrio dos outros modelos considerados, nos modelos ST-NLM e SSL-NLM a influncia dos *outliers* em $\hat{\beta}$ e $\hat{\sigma}^2$ no  estritamente crescente quando ϑ aumenta. Estes grficos mostram tambm que, embora exista uma observao

atípica em todos os modelos considerados, esta observação parece ser mais influente nas estimativas do modelo skew-normal e do modelo skew-normal contaminada, o que reflete a capacidade dos modelos com caudas mais pesadas do que o modelo skew-normal para reduzir e controlar a influência de *outliers* nas estimativas dos parâmetros. Estes resultados são semelhantes aos apresentados em Vanegas & Cysneiros (2009), em um contexto simétrico.

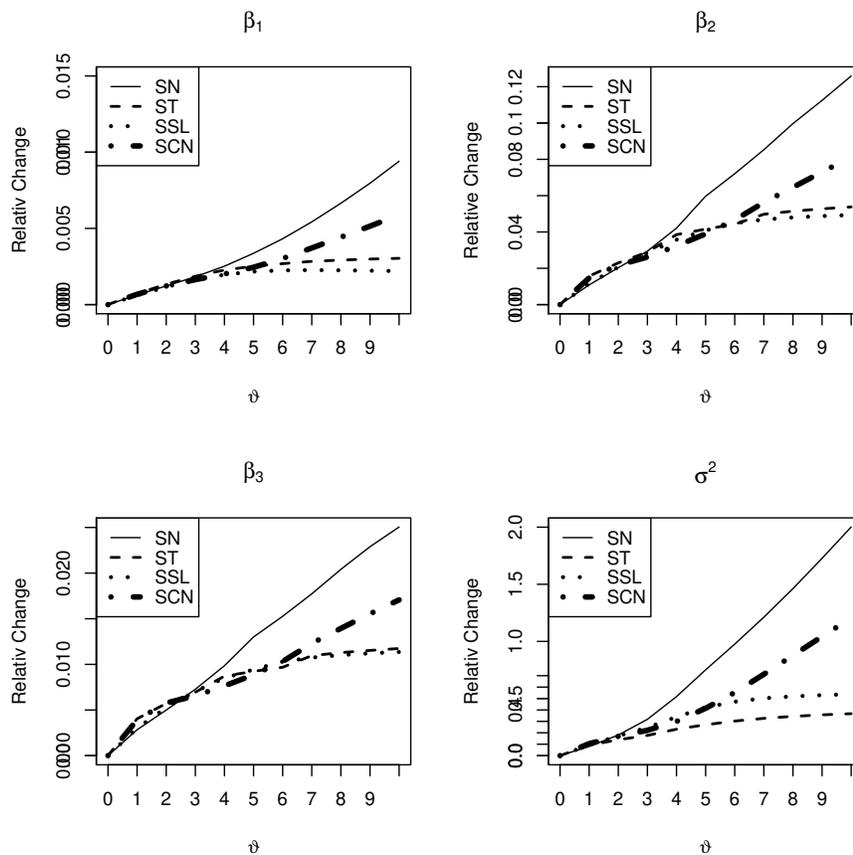


Figura 3.1: Mudança relativa média nas estimações de β_1 , β_2 , β_3 e σ^2 para o modelo de crescimento não linear com distribuições SN, ST, SSL, SCN.

3.4 Análise de diagnóstico

Nesta seção desenvolveremos os dois principais tipos de abordagem da análise de diagnóstico para os modelos não lineares sob a classe de distribuições mistura da escala skew normal (NLM-MESN), o modelo de eliminação de casos (CDM) e a análise de influência local.

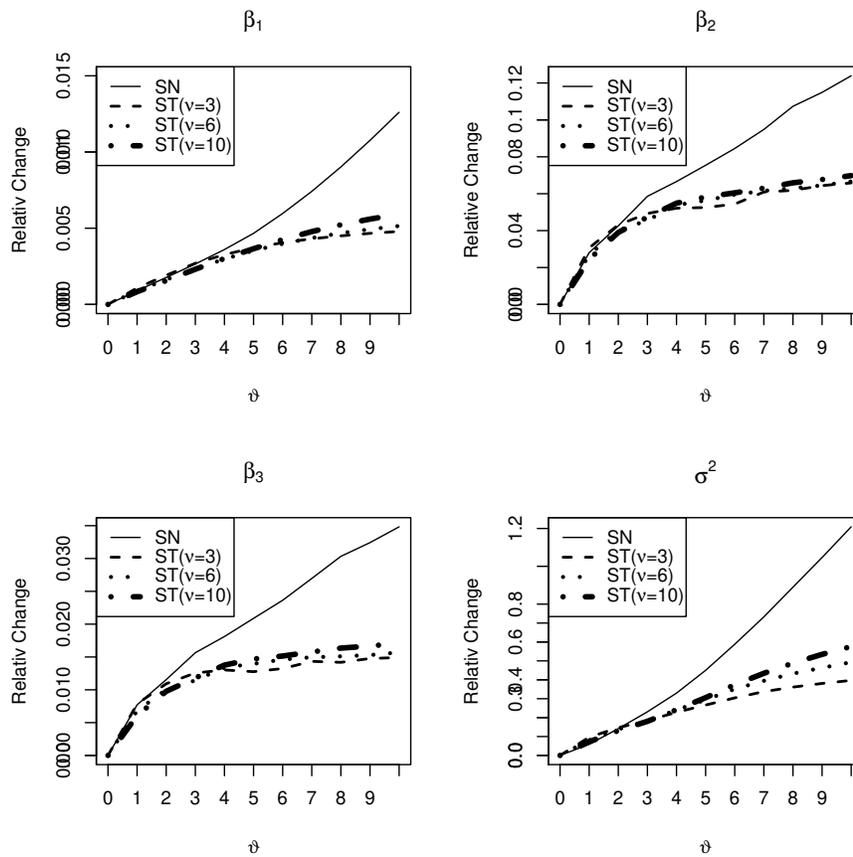


Figura 3.2: Mudança relativa média nas estimções de β_1 , β_2 , β_3 e σ^2 para o modelo de crescimento não linear com distribuições SN, ST($\nu = 3, 6, 10$)

3.4.1 Modelo de eliminação de casos (CDM)

A identificação de observações que influenciam desproporcionalmente nas estimativas dos parâmetros é uma etapa fundamental no processo de validação do modelo, já que a presença deste tipo de observações, pode gerar inferências inadequadas. Uma abordagem importante na identificação de observações atípicas é baseada na metodologia conhecida como Modelo de Eliminação de Casos⁴, para os modelos de regressão linear com distribuição normal.

Para estudar o impacto da i -ésima observação na estimativa de máxima verossimilhança de $\boldsymbol{\theta}$, é usual comparar as estimativas dos parâmetros obtidos com todas as observações, denotada por $\hat{\boldsymbol{\theta}}$, e as estimativas dos parâmetros obtidos quando a i -ésima observação foi excluída, denotada por $\hat{\boldsymbol{\theta}}_{(i)}$. Esta metodologia corresponde ao modelo de eliminação de casos (CDM), o qual pode ser expresso como:

$$Y_j = \eta(\boldsymbol{\beta}, \mathbf{x}_j) + \varepsilon_j, \quad j \neq i,$$

em que a função de log-verossimilhança para $\boldsymbol{\theta}$ sem a i -ésima observação é denotada por $\ell_{(i)}(\boldsymbol{\theta}) = \sum_{j \neq i} \ell_j(\boldsymbol{\theta})$. Calcular $\hat{\boldsymbol{\theta}}_{(i)} = (\boldsymbol{\beta}_{(i)}^\top, \sigma_{(i)}^2, \lambda_{(i)})^\top$ para todo i e comparar $\hat{\boldsymbol{\theta}}_{(i)}$ com $\hat{\boldsymbol{\theta}}$, pode envolver um extenso trabalho computacional, especialmente quando o tamanho da amostra n é muito grande. Afortunadamente, o seguinte resultado obtido por [Cook & Weisberg \(1982\)](#) proporciona uma fórmula aproximada sobre o CDM para estimar os parâmetros, em cada um dos n casos, isto é,

$$\hat{\boldsymbol{\theta}}_{(i)} = \hat{\boldsymbol{\theta}} + \{\mathbf{J}(\hat{\boldsymbol{\theta}})\}^{-1} \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}), \quad (3.13)$$

em que $\dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}) = \partial \ell_{(i)}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ com,

$$\frac{\partial \ell_{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = \sum_{j \neq i} \frac{\partial \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}, \quad e \quad (3.14)$$

$$\frac{\partial \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2} \frac{\partial \log \sigma^2}{\partial \boldsymbol{\gamma}} + \frac{1}{K_j} \frac{\partial K_j}{\partial \boldsymbol{\gamma}}, \quad (3.15)$$

⁴ Tradução livre de “case-deletion model”, proposto por [Cook \(1977\)](#)

para $j = 1, \dots, n$, em que $\boldsymbol{\gamma} = \boldsymbol{\beta}, \sigma^2$ ou λ e K_i é definido e calculado na Seção 3.2.

Deste resultado, podemos ver as diferenças entre as estimativas dos parâmetros, com todas as observações e sem a i -ésima observação. Conseqüentemente podemos obter medidas para avaliar o impacto das observações influentes nos modelos NLM-MESN.

- **Distância generalizada de Cook**

A distância generalizada de Cook é definida como a distância padronizada entre $\hat{\boldsymbol{\theta}}_{(i)}$ e $\hat{\boldsymbol{\theta}}$, isto é,

$$GD_i = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top \mathbf{M}(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}), \quad (3.16)$$

que avalia a combinação ponderada dos elementos para a diferença $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$. Cook & Weisberg (1982) consideram várias opções para \mathbf{M} . Uma escolha muito usada é a matriz de informação observada, isto é $\mathbf{M} = \mathbf{J}$. Agora, substituindo (3.13) em (3.16), obtemos a seguinte aproximação:

$$GD_i^l = \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}})^\top \{\mathbf{J}(\hat{\boldsymbol{\theta}})\}^{-1} \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}), \quad i = 1, \dots, n.$$

- **Afastamento pela log-verossimilhança**

O afastamento pela log-verossimilhança (Cook & Weisberg, 1982) é definido como:

$$LD_i = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{(i)})\}. \quad (3.17)$$

Substituindo (3.13) em (3.17), obtemos a seguinte aproximação para LD_i .

$$LD_i^l = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}} + \{\mathbf{J}(\hat{\boldsymbol{\theta}})\}^{-1} \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}))\}, \quad i = 1, \dots, n.$$

3.4.2 Influência local

Nesta subseção apresentaremos em forma resumida o método de influência local proposto por Cook (1986) e outro método alternativo que tem sido utilizado para avaliar a influência de determinada observação. O método de influência local permite avaliar o impacto sob as EMV, de pequenas perturbações no modelo e/ou nos dados, baseado no afastamento pela log-verossimilhança.

A perturbação no modelo é introduzida através de um vetor $\boldsymbol{\omega} \in \Omega \subseteq \mathbb{R}^q$, em que Ω é um conjunto aberto. Seja $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ a função de log-verossimilhança do modelo perturbado. Assumiremos que existe $\boldsymbol{\omega}_0 \in \Omega$ tal que $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$, para todo $\boldsymbol{\theta} \in \Theta$. Considerando $\boldsymbol{\theta}$ e $\boldsymbol{\theta}_{\boldsymbol{\omega}}$ as EMV sob o modelo postulado e perturbado, respectivamente, [Cook \(1986\)](#) sugere que a comparação entre $\boldsymbol{\theta}$ e $\boldsymbol{\theta}_{\boldsymbol{\omega}}$ seja feita através do afastamento pela log-verossimilhança definido por,

$$LD(\boldsymbol{\omega}) = 2 \left\{ \ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})) \right\}, \quad \boldsymbol{\omega} \in \Omega. \quad (3.18)$$

A idéia de [Cook \(1986\)](#) é investigar o comportamento da função $LD(\boldsymbol{\omega})$ numa vizinhança de $\boldsymbol{\omega}_0$. Para isso, é analisada a curvatura normal na direção \mathbf{d} , da superfície $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, LD(\boldsymbol{\omega}))^\top$, chamada de gráfico de influência. Em seu artigo, [Cook \(1986\)](#) mostra que a curvatura normal para $\boldsymbol{\theta} \in \mathbb{R}^{p+2}$ na direção de $\mathbf{d} \in \mathbb{R}^q$, $\|\mathbf{d}\| = 1$ é dada por $C_d(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^{*\top} \mathbf{J}^{-1} \boldsymbol{\Delta}^* \mathbf{d}|$, onde \mathbf{J} é a matriz de informação observada e $\boldsymbol{\Delta}^*$ é a $(p+2) \times q$ matriz com elementos $\Delta_{rs}^* = \partial^2 \ell(\boldsymbol{\theta}) / \partial \theta_r \partial \omega_s$, para $r = 1, \dots, (p+2)$ e $s = 1, \dots, q$; ambos avaliados em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_o$ (modelo postulado).

A sugestão aqui é examinar os elementos do autovetor associado com o maior autovalor da matriz $\ddot{\mathbf{T}} = \boldsymbol{\Delta}^{*\top} \{\mathbf{J}\}^{-1} \boldsymbol{\Delta}^*$. Alternativamente, podemos avaliar o impacto da i -ésima observação mediante a medida de influência local total definida por $C_i = C_{d_i}(\boldsymbol{\theta})$, onde \mathbf{d}_i é o vetor $q \times 1$ de zeros com o valor de 1 na i -ésima posição. Apesar de que o método de influência local proposto por [Cook \(1986\)](#) tem sido aplicado com sucesso em muitos problemas estatísticos, alguns aspectos são questionados, como por exemplo, a falta de um critério objetivo para avaliar a grandeza da curvatura normal e o tamanho relativo das componentes da direção correspondente à maior curvatura normal. Ante isso, [Poon & Poon \(1999\)](#) propõem um método de influência local baseado na curvatura normal conformalizada $B_d(\boldsymbol{\theta}) = C_d(\boldsymbol{\theta}) / \text{tr}(2\ddot{\mathbf{T}})$, que é uma medida padronizada da curvatura normal utilizada por [Cook \(1986\)](#).

Uma propriedade interessante da curvatura normal conformalizada é que para alguma direção unitária \mathbf{d} temos, $0 \leq B_d(\boldsymbol{\theta}) \leq 1$, o que permitirá uma comparação das curvaturas obtidas entre os diferentes modelos misturas da escala skew normais. [Poon & Poon \(1999\)](#) propuseram que a i -ésima observação seria considerada como influente se $M(0)_i = B_{d_i}$, é maior que o valor:

$$M\bar{(0)} + c^* SM(0),$$

em que $M\bar{(0)} = 1/q$ e $SM(0)$ é o erro padrão da amostra de $\{M(0)_k, k = 1 \dots, q\}$ e c^* é uma constante selecionada. Dependendo do caso, c^* pode ser qualquer valor, veja [Montenegro et al. \(2009a\)](#) para mais detalhes. A seguir, avaliamos a matriz Δ^* sobre dois esquemas de perturbação considerando um modelo NLM-MESN.

• Perturbação de ponderação de casos

Considerando uma atribuição arbitrária de ponderações (pesos) para as unidades experimentais, a função log-verossimilhança pode ser definida por,

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \left[\log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \log K_i \right],$$

em que K_i é definida em [\(3.2\)](#) e $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ é um vetor de pesos das contribuições de cada observação para a log-verossimilhança; $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$ é o ponto de não perturbação, isto é, $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$. Este esquema de perturbação procura determinar se a contribuição das observações com diferentes ponderações (pesos) afetam as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$. Depois de algumas manipulações algébricas temos que a matriz Δ^* , é dada por:

$$\Delta^* = (\Delta_1^*, \dots, \Delta_n^*),$$

em que $\Delta_i^* = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, com os elementos especificados em [\(3.2\)](#).

• **Perturbação do parâmetro de escala**

Para estudar o efeito do afastamento da suposição de homogeneidade, considerando o parâmetro de escala σ^2 , apresentamos a seguinte perturbação $\sigma_{\omega_i}^2 = \sigma^2/\omega_i$, com $\omega_i > 0$. Esta perturbação considera que a distribuição de Y_i é heteroscedástica, isto é,

$$\text{Var}(Y_i) = \sigma_{\omega_i}^2 (k_2 - b\delta^2),$$

em que k_2 , δ e b estão definidas na Seção 3.2. Sobre este esquema de perturbação, o modelo não perturbado é obtido quando $\boldsymbol{\omega}_o = (1, \dots, 1)^\top$. Além disso, a função log-verossimilhança perturbada tem a seguinte forma:

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \left[\log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log \omega_i + \log K_i^w \right],$$

em que $d_i^w = \omega_i^{1/2}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta\omega_i^{-1/2})^2/\sigma^2$, $A_i^w = \lambda\omega_i^{1/2}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta\omega_i^{-1/2})/\sigma$ e K_i^w é definido como na Seção 3.2, com σ^2/ω_i no lugar de σ^2 . A matriz de perturbação é dada por $\boldsymbol{\Delta}^* = (\boldsymbol{\Delta}_\beta^{*\top}, \boldsymbol{\Delta}_\sigma^{*\top}, \boldsymbol{\Delta}_\lambda^{*\top})^\top$, cujos elementos são apresentados no Apêndice A.1. Um estudo mais detalhado dos modelos não lineares heteroscedásticos sob a classe MESN será abordado no Capítulo 4.

3.5 Aplicação: Produtividade de óleo de palma

Nesta seção, usamos os dados referentes a produtividade de óleo de palma, apresentados por Foong (1999) e também estudados por Cancho *et al.* (2009) usando SN-NLM e assumindo o modelo de crescimento não linear (ver Figura 3.7b). Com o intuito de ilustrarmos nossos métodos, substituímos o pressuposto skew normal pela classe de distribuições MESN, como segue:

$$Y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} \text{MESN} \left(-\sqrt{\frac{2}{\pi}} k_1 \Delta, \sigma^2, \lambda; H \right), \quad (3.19)$$

para $i = 1, \dots, 19$, em que H denota a função de distribuição da variável de mistura U .

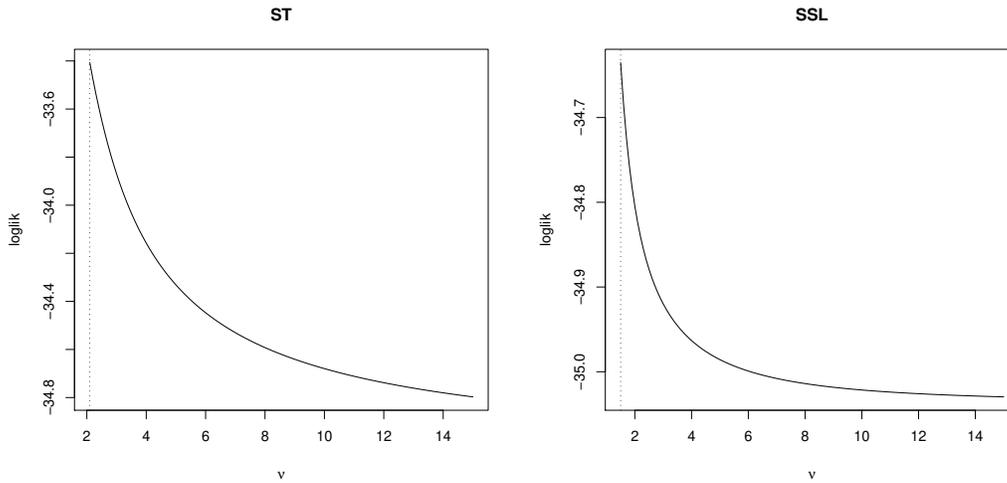


Figura 3.3: Log-verossimilhança perfilada do parâmetro ν considerando os modelos ST-NLM e SSL-NLM.

3.5.1 Ajuste do modelo

Para a estimação do modelo consideramos as distribuições SN-NLM, ST-NLM e SSL-NLM; escolhendo para as distribuições ST e SSL o valor inteiro de ν que maximiza a função de log-verossimilhança, como é mostrado na Figura 3.3. Para o modelo ST encontramos que o valor de $\nu = 3$ e para a SSL encontramos o valor de $\nu = 2$. Com os valores de $\nu = 3$ e $\nu = 2$ as variâncias da distribuição skew-t e a skew slash são finita. A Tabela 3.1 contém as EMV obtidas pelo algoritmo tipo EM para os parâmetros dos três modelos assim como seus correspondentes erros padrão assintóticos (EP), obtidos através da matriz de informação observada. O critério de seleção do modelo AIC indica que a distribuição ST apresenta o melhor ajuste para este conjunto de dados.

Embora as estimativas dos parâmetros de regressão são semelhantes nos três modelos, os erros padrão dos modelos ST-NLM e SSL-NLM são menores que os obtidos sob o modelo SN-NLM. Isto sugere que os dois modelos com caudas mais pesadas parecem produzir EMV mais precisas do que o modelo SN-NLM. As estimações para os componentes da variância (σ^2 e λ) não são comparáveis, pois eles estão em escala diferente.

Tabela 3.1: EMV e EP obtidos por ajustar vários modelos de misturas nos dados de produtividade de óleo de palma.

Parâmetros	SN-NLM		ST-NLM		SSL-NLM	
	EMV	EP	EMV	EP	EMV	EP
β_1	37.351	0.462	37.529	0.441	37.463	0.486
β_2	44.576	17.039	43.483	10.364	43.373	14.982
β_3	0.731	0.070	0.732	0.045	0.728	0.063
σ^2	6.919	2.655	1.644	1.152	3.105	1.708
λ	-4.453	3.125	-1.871	1.332	-3.489	2.481
ν	-	-	3	-	2	-
log-verossimilhança	-35.037		-33.829		-34.781	
AIC	80.074		79.659		81.562	

3.5.2 Análise de diagnóstico

- *Modelo de eliminação de casos (CDM)*

Na Figura 3.4, são apresentadas a distância generalizada de Cook (GD_i^l) e o afastamento pela log-verossimilhança (LD_i^l), definidas na Subseção 3.4.1. Os resultados mostrados indicam que as observações #10, #13, #15, #16 e #18 são identificadas como influentes nas estimações dos parâmetros sob todos os modelos. No entanto, baseados na (GD_i^l), vemos que a observação #13 destaca-se entre elas. Nota-se da Figura 3.5 que quando usamos distribuições com caudas mais pesadas do que a SN-NLM, o algoritmo tipo EM permite acomodar as observações influentes (#13, #15, #16 e #18), atribuindo-lhes pequenos pesos no processo de estimação. Os pesos estimados para a distribuição skew-normal ($\hat{u}_i, i = 1, \dots, 19$) são indicados como uma linha contínua. Portanto, esta classe de distribuições (MESN) permite, naturalmente, atribuir pesos diferentes a cada observação e, conseqüentemente, controlar a influência de uma observação sobre as estimativas dos parâmetros. Estes resultados concordam com os apresentados em Osorio *et al.* (2007), no contexto simétrico.

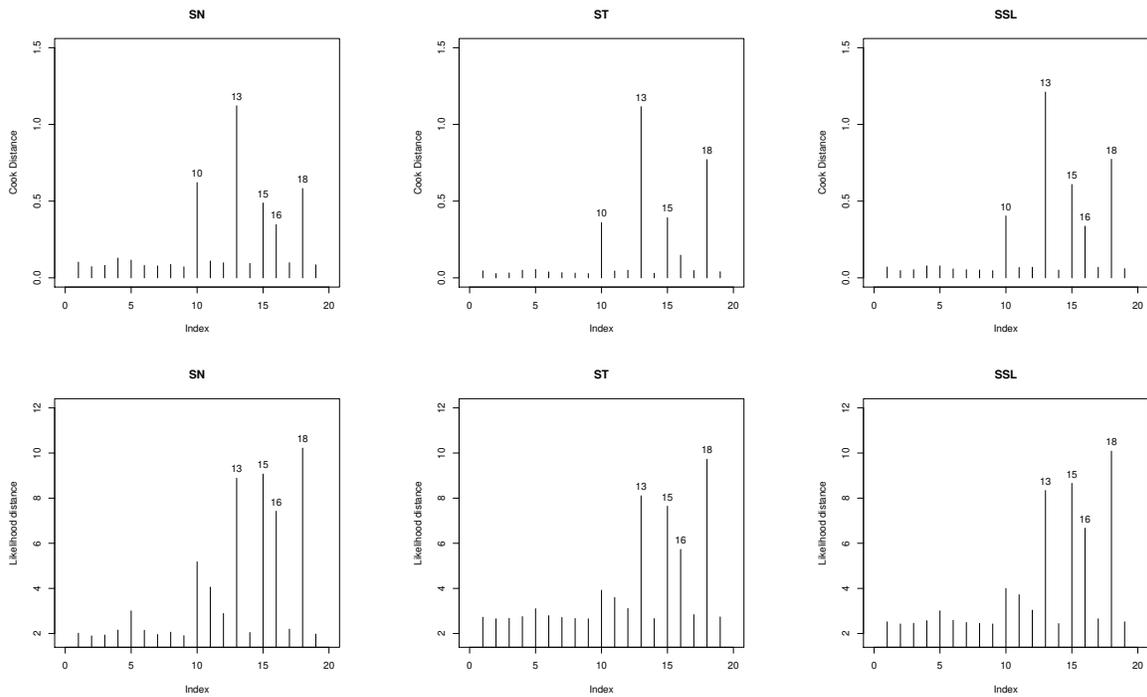


Figura 3.4: Gráfico de GD_i^l (1ra linha) e LD_i^l (2da linha) sob os modelos SN-NLM, ST-NLM e SSL-NLM.

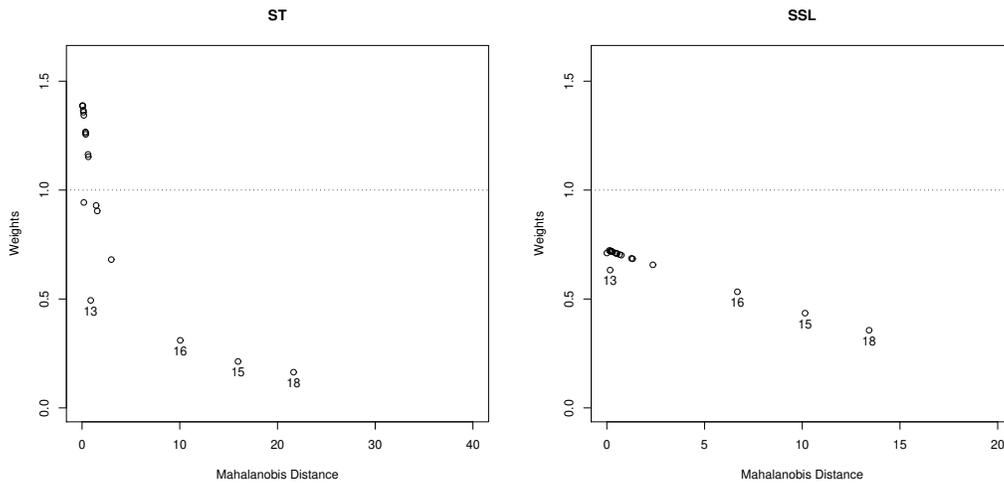


Figura 3.5: Pesos estimados (\hat{u}_i) para os modelos ST-NLM e SSL-NLM.

A seguir, focamos nossa análise na aplicação do método de influência local, baseado nos esquemas de perturbação propostos. Em todos os casos, consideraremos o valor de $c^* = 1.96$.

- *Perturbação de ponderação de casos*

Sob o esquema de perturbação de ponderação de casos, encontramos os valores de curvatura máxima $C_{\mathbf{d}_{maxSN}} = 3.55$, $C_{\mathbf{d}_{maxST}} = 2.28$ e $C_{\mathbf{d}_{maxSSL}} = 2.13$.

Na primeira linha da Figura 3.6 notamos que sob os modelos SN-NLM, ST-NLM e SSL-NLM, a observação #13 é identificada como influente. Além disso, a observação #18 também é identificada como influente sob o modelo SN-NLM.

- *Perturbação do parâmetro de escala*

Neste caso, os valores da curvatura máxima são dados por $C_{\mathbf{d}_{maxSN}} = 6.40$, $C_{\mathbf{d}_{maxST}} = 2.10$ e $C_{\mathbf{d}_{maxSSL}} = 4.82$. Sob este esquema de perturbação (na Figura 3.6- segunda linha) mostra-se influente quando assumimos os modelos SN-NLM e SSL-NLM.

Sob este esquema de perturbação, que considera o modelo heteroscedástico, é importante notar que, como esperado, a influência das observações é reduzida quando consideramos distribuições com caudas mais pesadas do que a SN-NLM. Para este conjunto de dados, o modelo ST-NLM, controla melhor as observações influentes.

3.5.3 Análise de resíduos

Considerando que com o modelo ST-NLM se obtém os melhores ajustes para os dados de produtividade de óleo de palma (Ver Subseção 3.5.1), avaliamos o gráfico dos resíduos de Pearson r_i (3.11) contrapondo-o à variável explanatória x_i , como se observa na Figura 3.7a. Este gráfico mostra que os resíduos das observações #13, #15, #16 e #18 são possíveis outliers, já que se encontram fora das linha de referência (-1,1). Note que

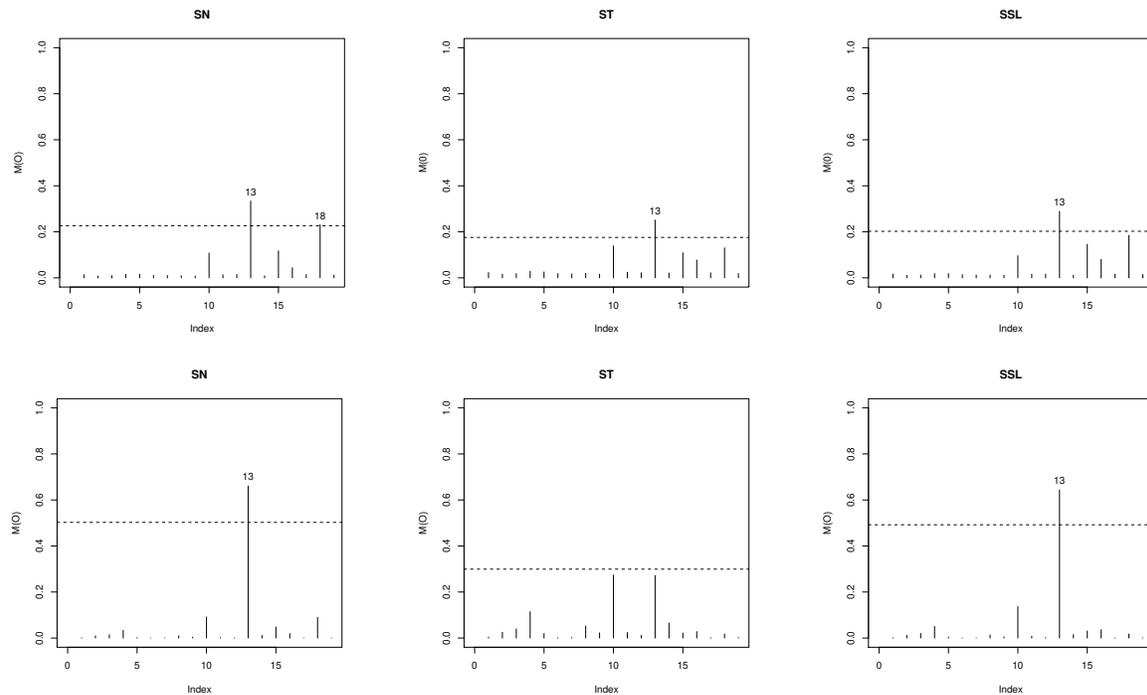


Figura 3.6: Gráfica de $M(i)$ sob os modelos SN-NLM, ST-NLM e SSL-NLM. Perturbação de ponderação de casos (1ra linha) e perturbação do parâmetro de escala (2da linha).

estas observações são as mesmas que foram encontradas na análise CDM.

Finalmente, na Figura 3.7b mostramos os valores preditos, com todas as observações e sem as observações influentes ($\#15$, $\#16$, $\#18$) para o modelo ST-NLM; notamos que quando as observações influentes não são consideradas, pode ser visto uma ligeira mudança na curva ao compará-la com a curva obtida com todas as observações.

Com o intuito de verificar se existe uma especificação incorreta da distribuição dos dados e a componente sistemática (3.19), na Figura 3.8 apresentamos os gráficos dos envelopes simulados para os resíduos de Pearson. Esta figura claramente indica que a ST-NLM é mais adequada para modelar os dados de produtividade de óleo de palma do que os modelos SN-NLM e SSL-NLM, já que não há observações que estão fora dos envelopes.

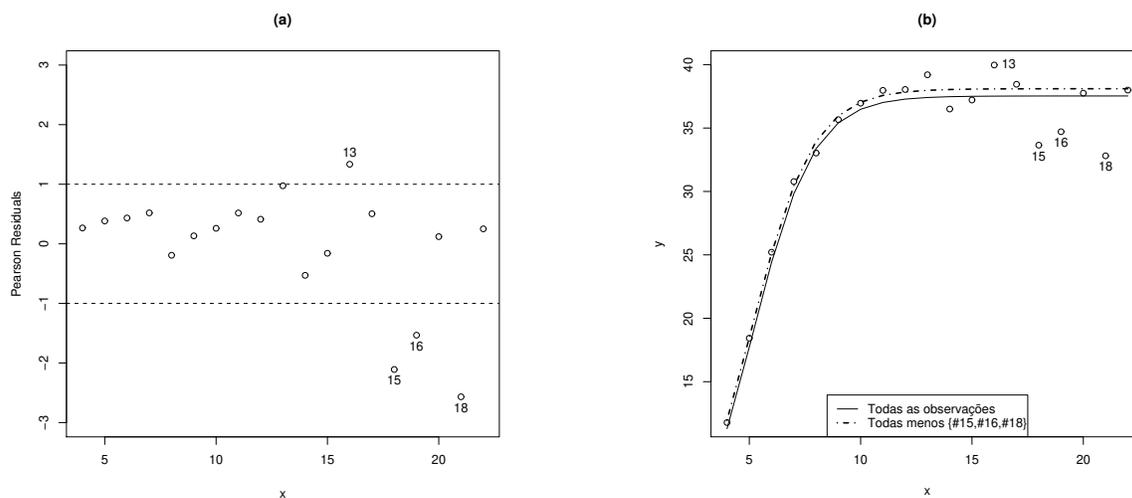


Figura 3.7: (a) Gráfico dos resíduos r_i em contraposição à variável explanatória x_i e (b) Valores preditos.

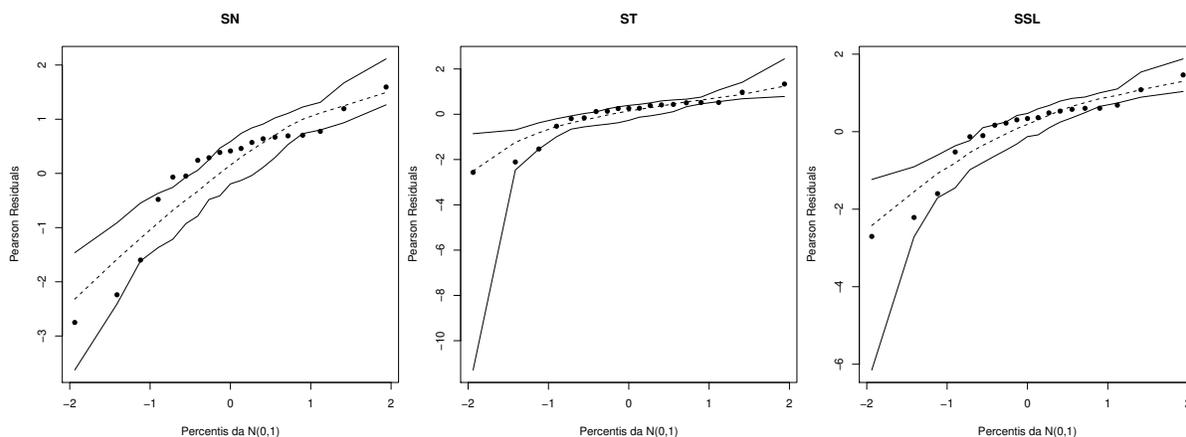


Figura 3.8: Envelopes simulados para os resíduos de Pearson sob os modelos SN-NLM, ST-NLM e SSL-NLM.

3.5.4 Influência de um valor outlier

A estimação robusta dos modelos ST-NLM e SSL-NLM pode ser também estudada através da influência de uma única observação outlier nas EMV de θ . Em particular, podemos avaliar quanto as estimativas de θ estão sendo influenciadas por uma mudança de ∇ unidades em uma observação Y_k .

Como é considerado em [Zeller et al. \(2010\)](#), substituímos a observação Y_k por $Y_k(\nabla) = Y_k - \nabla$, e observamos as mudanças relativas nas estimativas $\left(\left| \frac{\widehat{\theta}(\nabla) - \widehat{\theta}}{\widehat{\theta}} \right| \right)$, onde $\widehat{\theta}$ denota as estimativas baseadas nos dados originais e $\widehat{\theta}(\nabla)$ denota a estimação dos parâmetros, para os dados contaminados. Neste exemplo, selecionamos aleatoriamente a observação 5 e a contaminamos considerando a variação $\nabla = 0, 0.4, 0.8, \dots, 5$.

Na Figura 3.9, temos os resultados das mudanças relativas das estimações de β_2 e β_3 para os valores de contaminação ∇ , sob os modelos SN-NLM, ST-NLM e SSL-NLM. Como esperávamos, os modelos ST-NLM e SSL-NLM, estão sendo menos afetados pelas variações ∇ do que o modelo SN-NLM.

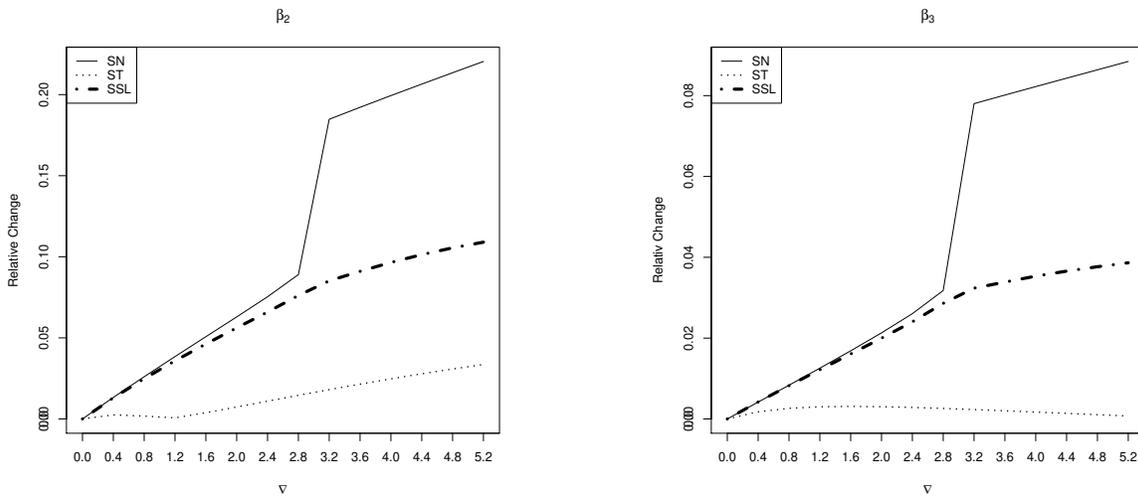


Figura 3.9: Mudanças relativas nas estimações de β_2 e β_3 , sob os modelos SN-NLM, ST-NLM e SSL-NLM considerando diferentes perturbações ∇ na observação 5.

3.6 Considerações finais

Neste capítulo, foi discutido um modelo de regressão não linear no qual assumimos que as observações seguem uma distribuição sob a classe de distribuições misturas da escala skew-normal (NLM-MESN). Apresentando também o algoritmo tipo EM para obter as EMV dos parâmetros do modelo.

No estudo de simulação apresentado na Subseção 3.3.4 encontramos que as estimativas dos parâmetros sob os modelos ST-NLM e SSL-NLM são menos sensíveis à presença de observações atípicas do que a distribuição SN-NLM. Além disso, baseado no estudo de diagnóstico de influência e nos tipos de perturbações - perturbação de caso e do parâmetro escalar (σ^2) - aplicados no conjunto de dados produtividade de óleo de palma, notamos que sob os modelos ST-NLM e SSL-NLM os resultados foram mais apropriados.

Dos resultados obtidos neste capítulo, vemos que os modelos NLM-MESN surgem como uma boa alternativa para estudar os modelos não lineares sob um enfoque diferente das técnicas existentes.

Contudo, baseado na perturbação de escala, que representa o modelo heteroscedástico, acreditamos que um modelo mais adequado para os dados analisados é o modelo heteroscedástico sob as distribuições MESN, o que será o foco principal no próximo capítulo.

Capítulo 4

Modelos não lineares heteroscedásticos sob a classe de distribuições MESN

Uma consideração importante na análise estatística nos modelos não lineares com erros normais é o pressuposto que todas as observações têm variâncias iguais. O não cumprimento deste pressuposto afeta a eficiência das estimativas dos parâmetros, motivo pelo qual, é importante o desenvolvimento de testes que nos permitam determinar a presença ou ausência de tal homogeneidade. Nos últimos anos, vários autores propuseram testes para verificar a homogeneidade da variância em modelos diferentes, entre eles, [Cook & Weisberg \(1983\)](#) que desenvolveram um teste de escore para heteroscedasticidade em modelos de regressão. [Chih-Ling \(1986\)](#) desenvolveu o teste de escore para o modelo autoregressivo de primeira ordem, AR(1), com heteroscedasticidade. [Lin & Wei \(2003\)](#) que consideraram testes de heteroscedasticidade em modelos não lineares, [Cysneiros *et al.* \(2007\)](#) que desenvolveram testes de diagnóstico para detectar heteroscedasticidade em modelos simétricos de regressão linear e, mais recentemente,

Lin *et al.* (2009) que desenvolveram o teste de escore para testar a homogeneidade do parâmetro escalar no modelo ST-Normal-NLM, que foi introduzido por Gómez *et al.* (2007).

Desse modo, este capítulo tem como intuito preencher essa lacuna, desenvolvendo tais testes, os quais complementarão o que foi apresentado nos capítulos anteriores.

Contudo, surge a inquietude natural para desenvolver as técnicas que permitam encontrar as estimativas dos parâmetros, uma vez que a presença de heteroscedasticidade tem sido comprovada para o modelo sob investigação. Este capítulo tenta responder esta inquietude, apresentando para isto, uma extensão do modelo NLM-MESN. Propomos assim, os modelos não lineares heteroscedásticos sob a classe de distribuições MESN (NLMH-MESN), desenvolvendo o algoritmo ECM para encontrar as estimativas de máxima verossimilhança dos parâmetros de interesse, a função escore e a matriz de informação observada (MIO). Assim também, apresentaremos um estudo de simulação para avaliar a robustez e as propriedades assintóticas das estimativas obtidas.

Desenvolveremos o teste de escore e o teste da razão de verossimilhança para avaliar a heteroscedasticidade do parâmetro escalar nos modelos NLM-MESN, para o qual, também apresentaremos um estudo de simulação que avaliará a distribuição assintótica e o poder do teste da estatística de escore. Em seguida, apresentaremos uma aplicação a um conjunto de dados reais previamente analisados por Lin *et al.* (2009).

Acreditamos que este capítulo será um aporte importante para a análise de dados assimétricos com presença de heteroscedasticidade.

4.1 Descrição do modelo

O modelo não linear heteroscedástico sob a classe de distribuições MESN (NLMH-MESN), é definido por:

$$Y_i = \eta(\boldsymbol{\beta}, \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.1)$$

em que Y_i é a variável resposta, $\eta(\cdot)$ é uma função injetiva e duas vezes continuamente diferenciável com respeito ao vetor de parâmetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, \mathbf{x} é um vetor ou matriz que contém os valores das variáveis explicativas e os erros aleatórios $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ são independentes.

Neste capítulo, assumimos que $\varepsilon_i \sim MESN\left(-\sqrt{\frac{2}{\pi}}\kappa_1\Delta m_i^{1/2}, \sigma_i^2, \lambda; H\right)$, em que $\sigma_i^2 = \sigma^2 m_i$ e $m_i = m_i(\boldsymbol{\rho}, \mathbf{z}_i)$ é uma função conhecida e duas vezes continuamente diferenciável e $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$ contém valores das variáveis explanatórias $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^\top$. Como apontado por [Lin et al. \(2009\)](#), as variáveis \mathbf{z}_i são em geral, embora não seja necessário, um subconjunto de $\{x_1, x_2, \dots, x_n\}$. Da Proposição [2.3.3](#), segue que,

$$E[\varepsilon_i] = 0 \quad \text{e} \quad Var[\varepsilon_i] = \sigma^2 m_i (\kappa_2 - b^2 \delta^2).$$

Considerando as propriedades das distribuições MESN desenvolvidas na Capítulo [2](#) temos que $Y_i \sim MESN(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + b\Delta m_i^{1/2}, \sigma^2 m_i, \lambda; H)$, para $i = 1, \dots, n$; onde $b = -\sqrt{\frac{2}{\pi}}\kappa_1$. Para efeitos de estimação consideraremos que $\boldsymbol{\nu}$ é conhecido. Da Proposição [2.3.3](#), segue que,

$$E[Y_i] = \eta(\boldsymbol{\beta}, \mathbf{x}_i) \quad \text{e} \quad Var[Y_i] = \sigma^2 m_i (\kappa_2 - b^2 \delta^2).$$

Como apontado por [Lin et al. \(2009\)](#), quando a variância depende de um número de variáveis explanatórias \mathbf{z}_i , duas formas de m_i são usualmente consideradas. A saber,

- (i) o modelo log-linear, isto é, $m_i = m_i(\boldsymbol{\rho}, \mathbf{z}_i) = \exp\left(\sum_{j=1}^q \rho_j z_{ij}\right)$
- (ii) o modelo produto potência, isto é, $m_i = m_i(\boldsymbol{\rho}, \mathbf{z}_i) = \prod_{j=1}^q z_{ij}^{\rho_j}$

Certamente, (ii) requer que z_{ij} seja estritamente positivo; no entanto, esta restrição não é necessária em (i).

Cook & Weisberg (1983) e Simonoff & Tsai (1994) usaram o modelo log-linear e/ou modelo produto potência como a função m da variância nos modelos normais. Barroso & Cordeiro (2005) aplicaram o modelo log-linear para modelar a heteroscedasticidade no modelo de regressão t. Os resultados de Chen (1983) mostram que o teste de heteroscedasticidade não é muito sensível com a forma funcional m_i considerada.

4.2 Função de log-verossimilhança

Seja $\theta_1 = \boldsymbol{\rho}$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$ definido em (3.2). Então de (4.1) temos que a função de log-verossimilhança para $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}^\top)^\top$ dada uma amostra observada $\mathbf{y} = (y_1, \dots, y_n)^\top$ pode ser escrita como:

$$\ell(\boldsymbol{\theta}_2) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}_2) = \sum_{i=1}^n \left[\log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log m_i + \log K_i^m \right]. \quad (4.2)$$

- *Função escore*

A função escore é dada por:

$$\mathbf{U}(\boldsymbol{\theta}_2) = \frac{\partial \ell(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} = \sum_{i=1}^n \left[-\frac{1}{2} \frac{\partial \log \sigma^2}{\partial \boldsymbol{\theta}_2} - \frac{1}{2} \frac{\partial \log m_i}{\partial \boldsymbol{\theta}_2} + \frac{1}{K_i^m} \frac{\partial K_i^m}{\partial \boldsymbol{\theta}_2} \right], \quad (4.3)$$

em que

$$\frac{\partial K_i^m}{\partial \boldsymbol{\theta}_2} = I_i^\phi(1) \frac{\partial A_i}{\partial \boldsymbol{\theta}_2} - \frac{1}{2} I_i^\Phi \left(\frac{3}{2} \right) \frac{\partial d_i}{\partial \boldsymbol{\theta}_2},$$

com $K_i^m = \int_0^\infty u_i^{1/2} \exp\{-\frac{1}{2}u_i d_i\} \Phi(u_i^{1/2} A_i) dH(u_i)$, $A_i = \lambda(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta m_i^{1/2}) / (m_i \sigma^2)^{1/2}$ e $d_i = (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - b\Delta m_i^{1/2})^2 / (m_i \sigma^2)$ que é a distância de Mahalanobish. As quantidades $I_i^\phi(\cdot)$ e $I_i^\Phi(\cdot)$ são definidas como na Seção 3.2 com $\sigma^2 m_i$ no lugar de σ^2 .

- *Matriz de informao observada*

A matriz de informao observada para $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}^\top)$   dada por:

$$\mathbf{J}(\boldsymbol{\theta}_2) = \begin{bmatrix} \mathbf{J}_{\rho\rho} & \mathbf{J}_{\rho\boldsymbol{\theta}} \\ \mathbf{J}_{\rho\boldsymbol{\theta}} & \mathbf{J}(\boldsymbol{\theta}) \end{bmatrix} = - \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} \quad (4.4)$$

$$= - \sum_{i=1}^n \left[-\frac{1}{2} \frac{\partial^2 \log \sigma^2}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} - \frac{1}{2} \frac{\partial^2 \log m_i}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} - \frac{1}{(K_i^m)^2} \frac{\partial K_i^m}{\partial \boldsymbol{\theta}_2} \frac{\partial K_i^m}{\partial \boldsymbol{\theta}_2^\top} + \frac{1}{K_i^m} \frac{\partial^2 K_i^m}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} \right], \quad (4.5)$$

em que

$$\begin{aligned} \frac{\partial^2 K_i^m}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} &= \frac{1}{4} I_i^\Phi \left(\frac{5}{2} \right) \frac{\partial d_i}{\partial \boldsymbol{\theta}_2} \frac{\partial d_i}{\partial \boldsymbol{\theta}_2^\top} \\ &\quad - \frac{1}{2} I_i^\Phi \left(\frac{3}{2} \right) \frac{\partial^2 d_i}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} - \frac{1}{2} I_i^\phi (2) \left(\frac{\partial A_i}{\partial \boldsymbol{\theta}_2} \frac{\partial d_i}{\partial \boldsymbol{\theta}_2^\top} + \frac{\partial d_i}{\partial \boldsymbol{\theta}_2} \frac{\partial A_i}{\partial \boldsymbol{\theta}_2^\top} \right) \\ &\quad - I_i^\phi (2) A_i \frac{\partial A_i}{\partial \boldsymbol{\theta}_2} \frac{\partial A_i}{\partial \boldsymbol{\theta}_2^\top} + I_i^\Phi (1) \frac{\partial^2 A_i}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top}. \end{aligned}$$

As derivadas de d_i e A_i envolvem rotineiras manipulaes alg bricas e so dadas no Ap ndice A.2.

4.3 Estimaco dos parmetros via algoritmo tipo EM

Com o objetivo de estimar os parmetros do modelo NLMH-MESN via algoritmo tipo EM, seguimos a seq ncia descrita na Subseo 2.4.1, pela qual escrevemos a seguinte representao hier rquica para Y_i

$$Y_i | U_i = u_i, T_i = t_i \sim N(\eta(\boldsymbol{\beta}, \mathbf{x}_i) + \Delta m_i^{1/2} t_i, u_i^{-1} m_i \Gamma), \quad (4.6)$$

$$T_i | U_i = u_i \sim NT((b, u_i^{-1}); (b, \infty)), \quad (4.7)$$

$$U_i \sim H(.; \boldsymbol{\nu}), \quad (4.8)$$

em que a distribuo de $T_i | Y_i = y_i, U_i = u_i$   $NT((\mu_{T_i} + b, u_i^{-1} M_T^2); (b, \infty))$, com $M_T^2 = \frac{\Gamma}{\Delta^2 + \Gamma}$ e $\mu_{T_i} = \frac{\Delta}{m_i^{1/2} (\Delta^2 + \Gamma)} (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \Delta b m_i^{1/2})$, $i = 1, \dots, n$.

Assim, os passos necessários para a implementação do algoritmo tipo EM são:

Passo E:

Sejam $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{t} = (t_1, \dots, t_n)^\top$ e $\mathbf{u} = (u_1, \dots, u_n)^\top$. Então, a função de log-verossimilhança completa para $\boldsymbol{\theta}_2 = (\boldsymbol{\rho}^\top, \boldsymbol{\beta}^\top, \sigma^2, \lambda)^\top$ associada com $(\mathbf{y}, \mathbf{t}, \mathbf{u})$ é dada por:

$$\ell_{comp}(\boldsymbol{\theta}_2 | \mathbf{y}, \mathbf{t}, \mathbf{u}) = c - \frac{n}{2} \log \Gamma - \frac{1}{2} \sum_{i=1}^n \log m_i - \frac{1}{2\Gamma} \sum_{i=1}^n \frac{u_i}{m_i} (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i) - \Delta m_i^{1/2} t_i)^2,$$

tal que c é uma constante que não depende do vetor de parâmetros $\boldsymbol{\theta}_2$. Assim, dado $\widehat{\boldsymbol{\theta}}_2^{(k)}$ obtemos $\widehat{u}_i^{(k)}$, $\widehat{ut}_i^{(k)}$, $\widehat{ut}_i^2^{(k)}$, para $i = 1, \dots, n$, através das seguintes expressões:

$$\begin{aligned} \widehat{u}_i^{(k)} &= E[U_i | \boldsymbol{\theta}_2 = \widehat{\boldsymbol{\theta}}_2^{(k)}, y_i], \\ \widehat{ut}_i^{(k)} &= E[U_i t_i | \boldsymbol{\theta}_2 = \widehat{\boldsymbol{\theta}}_2^{(k)}, y_i] = \widehat{u}_i^{(k)} (\widehat{\mu}_{T_i}^{(k)} + b) + \widehat{M}_T^{(k)} \widehat{\tau}_{1_i}^{(k)}, \\ \widehat{ut}_i^2^{(k)} &= E[U_i t_i^2 | \boldsymbol{\theta}_2 = \widehat{\boldsymbol{\theta}}_2^{(k)}, y_i] = \widehat{u}_i^{(k)} (\widehat{\mu}_{T_i}^{(k)} + b)^2 + \widehat{M}_T^{2(k)} + \widehat{M}_T^{(k)} (\widehat{\mu}_{T_i}^{(k)} + 2b) \widehat{\tau}_{1_i}^{(k)}, \end{aligned}$$

em que

$$\widehat{\tau}_{1_i}^{(k)} = E \left[U_i^{1/2} W_\Phi \left(\frac{U_i^{1/2} \widehat{\mu}_{T_i}^{(k)}}{\widehat{M}_T^{(k)}} \right) | \widehat{\boldsymbol{\theta}}_2^{(k)}, y_i \right].$$

Como no modelo homocedástico, para a distribuição skew-t e skew-slash temos expressões computacionais atrativas que podem ser facilmente implementadas.

Passo CM:

Neste caso a função Q é dada por:

$$\begin{aligned} Q(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}}_2^{(k)}) &= E[\ell_{comp}(\boldsymbol{\theta}_2) | \mathbf{y}, \widehat{\boldsymbol{\theta}}_2^{(k)}] = c - \frac{n}{2} \log(\Gamma) - \frac{1}{2} \sum_{i=1}^n \log m_i \\ &\quad - \frac{1}{2\Gamma} \sum_{i=1}^n \left[\frac{\widehat{u}_i^{(k)}}{\widehat{m}_i} (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i))^2 - 2\Delta (y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i)) \frac{\widehat{ut}_i^{(k)}}{\widehat{m}_i^{1/2}} + \Delta^2 \widehat{ut}_i^2^{(k)} \right], \end{aligned}$$

em que $\widehat{\boldsymbol{\theta}}_2^{(k)}$ é o valor atualizado de $\widehat{\boldsymbol{\theta}}_2$. Assim, o passo CM, consiste em atualizar $\widehat{\boldsymbol{\theta}}_2^{(k)}$

maximizando $Q(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}}_2^{(k)})$ sob $\boldsymbol{\theta}_2$, o que leva s seguintes expresses:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \operatorname{argmin}_{\boldsymbol{\beta}} (\mathbf{z}^{(k)} - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}))^\top \widehat{\mathbf{U}}^{(k)} (\mathbf{z}^{(k)} - \boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x})), \quad (4.9)$$

$$\widehat{\Delta}^{(k+1)} = \frac{\sum_{i=1}^n \widehat{ut}_i^{*(k)} (y_i - \eta(\boldsymbol{\beta}^{(k+1)}, \mathbf{x}_i))}{\sum_{i=1}^n \widehat{ut}_i^2^{(k)}}, \quad (4.10)$$

$$\begin{aligned} \widehat{\Gamma}^{(k+1)} = & \frac{1}{n} \sum_{i=1}^n \left((y_i - \eta(\boldsymbol{\beta}^{(k+1)}, \mathbf{x}_i))^2 \widehat{u}_i^{*(k)} - 2\Delta^{(k+1)} (y_i - \eta(\boldsymbol{\beta}^{(k+1)}, \mathbf{x}_i)) \widehat{ut}_i^{*(k)} \right. \\ & \left. + (\Delta^2)^{(k+1)} \widehat{ut}_i^2^{(k)} \right), \end{aligned} \quad (4.11)$$

$$\widehat{\boldsymbol{\rho}}^{(k+1)} = \operatorname{argmax}_{\boldsymbol{\rho}} \sum_{i=1}^n \log(f(y_i | \boldsymbol{\beta}^{(k+1)}, \sigma^2(k+1), \lambda^{(k+1)}, \boldsymbol{\nu})), \quad (4.12)$$

em que $\widehat{u}_i^{*(k)} = \widehat{u}_i^{(k)} / \widehat{m}_i^{(k)}$, $\widehat{ut}_i^{*(k)} = \widehat{ut}_i^{(k)} / \widehat{m}_i^{1/2(k)}$, $\widehat{\mathbf{U}}^{(k)} = \operatorname{diag}(\widehat{u}_1^{*(k)}, \dots, \widehat{u}_n^{*(k)})$, $\mathbf{z}^{(k)}$  a resposta observada corrigida, dada por $\mathbf{z}^{(k)} = \mathbf{y} - \widehat{\Delta}^{(k)} \widehat{\boldsymbol{\tau}}^{(k)}$, com $\widehat{\boldsymbol{\tau}}^{(k)} = (\widehat{\tau}_1^{(k)}, \dots, \widehat{\tau}_n^{(k)})^\top$, $\widehat{\tau}_i^{(k)} = \widehat{ut}_i^{*(k)} / \widehat{u}_i^{*(k)}$ e $\boldsymbol{\eta}(\boldsymbol{\beta}, \mathbf{x}) = (\eta(\boldsymbol{\beta}, \mathbf{x}_1), \dots, \eta(\boldsymbol{\beta}, \mathbf{x}_n))^\top$. Este processo  iterado at que uma regra de convergncia definida seja satisfeita, por exemplo, se $\|\widehat{\boldsymbol{\theta}}_2^{(k+1)} - \widehat{\boldsymbol{\theta}}_2^{(k)}\|$  suficientemente pequeno ou at que uma certa diferena envolvendo duas avaliaes sucessivas do log-verossimilhana $\|\ell(\widehat{\boldsymbol{\theta}}_2^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}_2^{(k)})\|$ ou $\|\ell(\widehat{\boldsymbol{\theta}}_2^{(k+1)}) / \ell(\widehat{\boldsymbol{\theta}}_2^{(k)}) - 1\|$, seja suficientemente pequena. Note tambm que estimativas para $\sigma^2(k+1)$ e $\lambda^{(k+1)}$ podem ser obtidas usando (3.6), isto , $\widehat{\lambda} = \widehat{\Delta} / \sqrt{\widehat{\Gamma}}$ e $\widehat{\sigma}^2 = \widehat{\Delta}^2 + \widehat{\Gamma}$.

4.3.1 Anlise de resduos

Para a anlise dos resduos nos modelos NLMH-MESN consideraremos o resduo padronizado de Pearson, definido em (3.11), onde $\widehat{\mu}_i = \eta(\widehat{\boldsymbol{\beta}}, \mathbf{x}_i)$ e $\widehat{Var}(y_i) = \widehat{\sigma}^2 \widehat{m}_i (k_2 - \frac{2}{\pi} k_1^2 \delta^2)$, com $\widehat{m}_i = m(\widehat{\boldsymbol{\rho}}, \mathbf{z}_i)$ e $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_i^2$, $\widehat{\delta}$ e $\widehat{\boldsymbol{\rho}}$ que so as EMV de $\boldsymbol{\beta}$, σ_i^2 , δ e $\boldsymbol{\rho}$ respectivamente. Aqui, tambm geraremos os envelopes seguindo o mesmo procedimento desenvolvido na Subseo 3.3.3.

4.3.2 Estudo de simulação

Nesta seção desenvolveremos um estudo de simulação para avaliar a performance e as propriedades assintóticas das estimativas dos parâmetros obtidas pelo algoritmo tipo EM nos modelos NLMH-MESN. Para isto, realizaremos um estudo de simulação Monte Carlo com o seguinte modelo:

$$Y_i = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 x_i)} + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.13)$$

em que $\varepsilon_i \sim MESN(-\sqrt{\frac{2}{\pi}}\kappa_1 \Delta m_i^{1/2}, \sigma_i^2, \lambda; H)$, com $\sigma_i^2 = \sigma^2 m(z_i, \rho)$, são variáveis aleatórias independentes e identicamente distribuídas com média zero. Consideramos o modelo log linear $\exp(\rho z_i)$ para representar a função m_i . Cada valor de x_i foi gerado considerando uma distribuição uniforme no intervalo $(0.1, 20)$ e se definiu $z_i = x_i/10$; estes valores são os mesmos em todas as simulações. Os valores verdadeiros dos parâmetros considerados para nosso estudo foram: $\beta_1 = 37, \beta_2 = 43, \beta_3 = 0.6, \sigma^2 = 0.5, \lambda = -3$ e $\rho = -0.2$.

- *Robustez das estimativas*

O objetivo deste estudo de simulação é comparar o desempenho das estimativas dos parâmetros nos modelos NLMH-MESN com presença de *outliers* na variável resposta. Para isto, geramos 1000 amostras de tamanho $n = 50$, considerando $\varepsilon_i \sim MESN(-\sqrt{2/\pi}\Delta m_i, \sigma_i^2, \lambda)$ em (4.13). Para garantir a presença de um *outlier* na data gerada, construímos $Y_i^* = Y_i - \vartheta$, onde Y_i é o valor central da amostra e $\vartheta = 1, 2, 3, 4, 5, 6, 7, 8, 9$ e 10 . Em cada réplica obtemos as estimações dos parâmetros com e sem a perturbação considerada, denotados por $\hat{\theta}$ e $\hat{\theta}_{(i)}$, respectivamente. Os modelos considerados são a skew-normal (SN-NLMH), a skew-t (ST-NLMH) com valores de $\nu = (3, 6, 10, 12)$ e a skew-slash (SSL-NLMH) com $\nu = 2$. Por fim, calculamos as mudanças relativas $\left| \frac{\hat{\theta}_{(i)}}{\hat{\theta}} - 1 \right|$, onde $\theta = \beta_1, \beta_2, \beta_3, \sigma^2$ ou ρ .

As Figuras 4.1 e 4.2 mostram as médias das mudanças relativas na estimação para as 1000 replicações. Podemos observar que no modelo skew normal a mudança

relativa aumenta conforme aumentam os valores de ϑ . No entanto, para os modelos MESN com caudas pesadas essas mudanças são muito pequenas, o que indica que, da mesma forma que no caso simétrico, modelos de caudas pesadas são mais robustos que o modelo SN-NLMH na presença de observações atípicas. Nota-se também, que ao contrário dos outros modelos considerados, na ST-NLMH e SSL-NLMH a influência do *outlier* em $\hat{\beta}$ e $\hat{\sigma}^2$ não é estritamente crescente quando ϑ aumenta. Concluimos, então, que embora exista uma observação extrema em todos os modelos considerados, esta observação parece ser mais influente nos erros do modelo skew-normal, o que reflete a capacidade dos modelos com caudas mais pesadas do que skew-normal para reduzir e controlar a influência de *outlier* nas estimativas dos parâmetros.

- *Propriedades assintóticas*

O intuito deste estudo de simulação é a avaliação do viés e do erro quadrático médio (EQM) das estimações dos parâmetros obtidas pelo algoritmo tipo EM nos modelos NLMH-MESN. Os tamanhos de amostra considerados são $n = 30, 50, 100, 150, 200, 250, 300, 400$ e 500 . Para cada combinação dos parâmetros e tamanhos de amostra, foram geradas 500 réplicas dos modelos: skew normal (SN-NLMH), skew-t (ST-NLMH) com $\nu = 3$ e skew slash (SSL-NLMH) com $\nu = 2$. Logo, calculamos o viés e o EQM sobre todas as réplicas.

Sendo $\theta = (\rho, \beta_1, \beta_2, \beta_3, \sigma^2, \lambda)$, como em (2.16) definimos:

$$Vies(\theta_j) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_j^{(i)} - \theta_j) \quad \text{e} \quad EQM(\theta_j) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_j^{(i)} - \theta_j)^2, \quad j = 1, \dots, 6.$$

Dos resultados mostrados na Tabela 4.1 e nas Figuras 4.3 e 4.4, observamos que o viés e o EQM tendem a se aproximar a zero quando os tamanhos de amostras aumentam. Portanto, podemos concluir que as estimativas obtidas pelo algoritmo tipo EM para os modelos NLMH-MESN são consistentes.

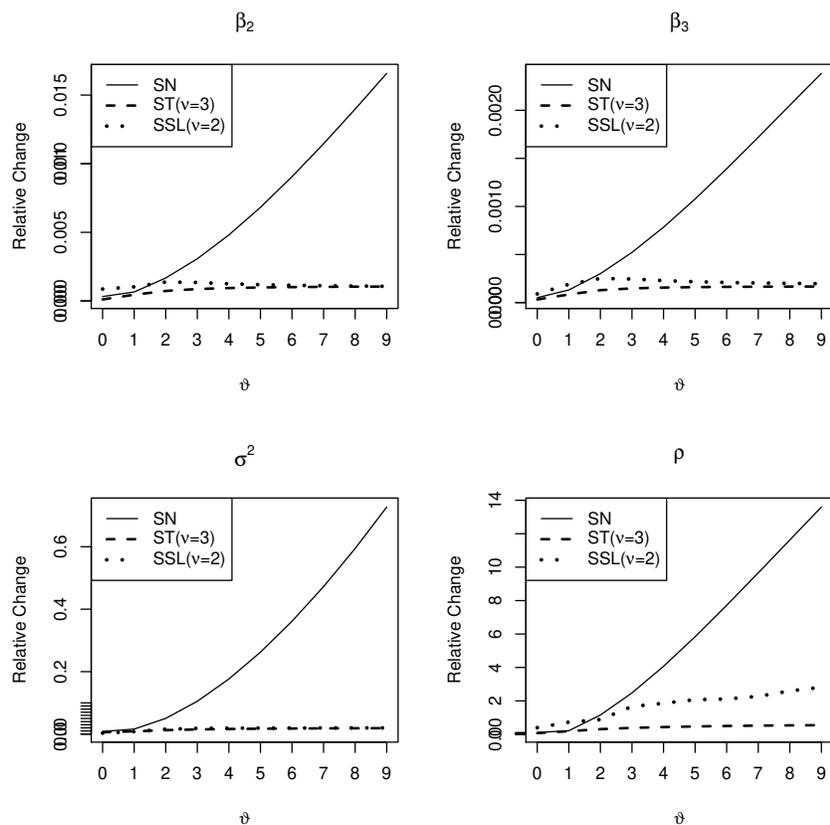


Figura 4.1: Mudança relativa média nas estimações de β_2 , β_3 , σ^2 e ρ para o modelo de crescimento não linear com distribuição SN, ST com $\nu = 3$ e SSL com $\nu = 2$.

4.4 Testes de Hipóteses

O modelo definido em (3.1) assume que a $Var(Y_i) = \sigma^2 [k_2 - b\delta^2]$ é constante, isto é, o parâmetro de escala é constante. No entanto, de forma semelhante ao parâmetro de dispersão descrito por Lin *et al.* (2009), o parâmetro de escala atual pode estar relacionado com a i -ésima observação. Assim, não podemos obter inferências para o modelo sem verificar os pressupostos, visto que existem muitos parâmetros desconhecidos envolvidos. Assim, é necessário testar a homogeneidade do parâmetro escalar. Nesta segunda parte do capítulo nos concentraremos em desenvolver o teste de score e o teste da razão de verossimilhança para detectar a presença de homogeneidade do

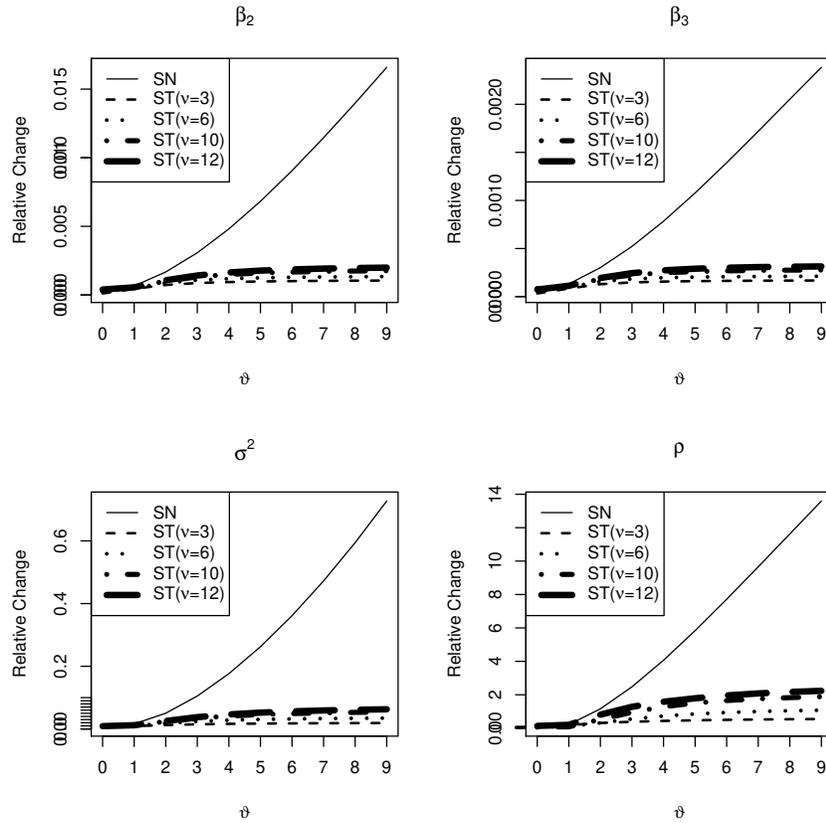


Figura 4.2: Mudança relativa média nas estimações de β_2 , β_3 , σ^2 e ρ para o modelo de crescimento não linear com distribuição SN e ST com $\nu = 3, 6, 10, 12$.

parâmetro de escala nos modelos NLM-MESN.

Seguindo o estudo feito por [Lin et al. \(2009\)](#), generalizamos primeiro o parâmetro de escala σ^2 por σ_i^2 , onde σ_i^2 é modelado por

$$\sigma_i^2 = \sigma^2 m(\mathbf{z}_i, \boldsymbol{\rho}). \quad (4.14)$$

Assumimos que existe um único valor $\boldsymbol{\rho}_0$ de $\boldsymbol{\rho}$, tal que $m(\mathbf{z}_i, \boldsymbol{\rho}_0) = 1$, para todo i . Obviamente, se $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ então $\sigma_i^2 = \sigma^2$ e Y_i tem variância constante. Conseqüentemente, temos o modelo de regressão não linear homocedástico sob as distribuições MESN, como descrito no Capítulo 3. Testar a homocedasticidade dos erros é equivalente testar a

Tabela 4.1: Viés na estimação dos parâmetros para diferentes tamanhos de amostras no modelo NLMH-MESN definido em (4.13).

SN-NLMH						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	0.018142	0.170583	2.13575e-05	-0.084223	-0.118008	0.796860
100	0.006792	0.100482	3.91535e-05	-0.043359	0.029257	-0.811886
200	0.002079	0.064363	1.92594e-04	-0.021745	0.013798	-0.358806
300	0.002328	0.050321	1.20312e-04	-0.007175	0.003215	-0.192273
500	0.001778	-0.033103	-9.26027e-05	0.007755	-0.006792	-0.092239
ST-NLMH						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	0.067886	0.069632	-7.45632e-04	-0.147138	0.008791	-0.074835
100	0.012163	0.200567	2.69659e-04	-0.056125	0.035765	-0.748369
200	0.007974	0.017946	-1.40273e-04	-0.014836	0.011951	-0.359338
300	0.000959	-0.045411	-2.45914e-04	-0.000850	0.005841	-0.231270
500	-0.001173	0.020128	9.32806e-05	0.003383	0.000531	-0.087790
SSL-NLMH						
n	β_1	β_2	β_3	ρ	σ^2	λ
30	0.024367	0.373273	2.46772e-04	-0.128733	0.006428	-0.824971
100	0.002395	0.255741	6.62715e-04	-0.045246	0.025884	-0.716985
200	0.004480	0.026543	-1.24525e-04	-0.023700	0.017441	-0.348624
300	-0.000788	0.057916	1.28792e-04	-0.010133	0.011916	-0.208165
500	-0.000653	0.030382	3.68057e-05	-0.005529	0.002849	-0.062118

hipótese $\boldsymbol{\rho} = \boldsymbol{\rho}_0$, isto é,

$$H_0 : \boldsymbol{\rho} = \boldsymbol{\rho}_0 \text{ vs } H_1 : \boldsymbol{\rho} \neq \boldsymbol{\rho}_0. \quad (4.15)$$

Para testar as hipóteses definidas acima (4.15), podemos usar a estatística de escore ou a estatística da razão de verossimilhança, as quais definiremos a seguir.

4.4.1 Teste de escore

Como indicamos na apresentação deste capítulo, o teste de escore foi utilizado nos últimos anos por muitos autores para avaliar a presença de heteroscedasticidade nos modelos de regressão. Como por exemplo, veja [Cook & Weisberg \(1983\)](#), [Chih-Ling \(1986\)](#), [Lin & Wei \(2003\)](#), [Cysneiros et al. \(2007\)](#), [Lin et al. \(2009\)](#), entre outros. É

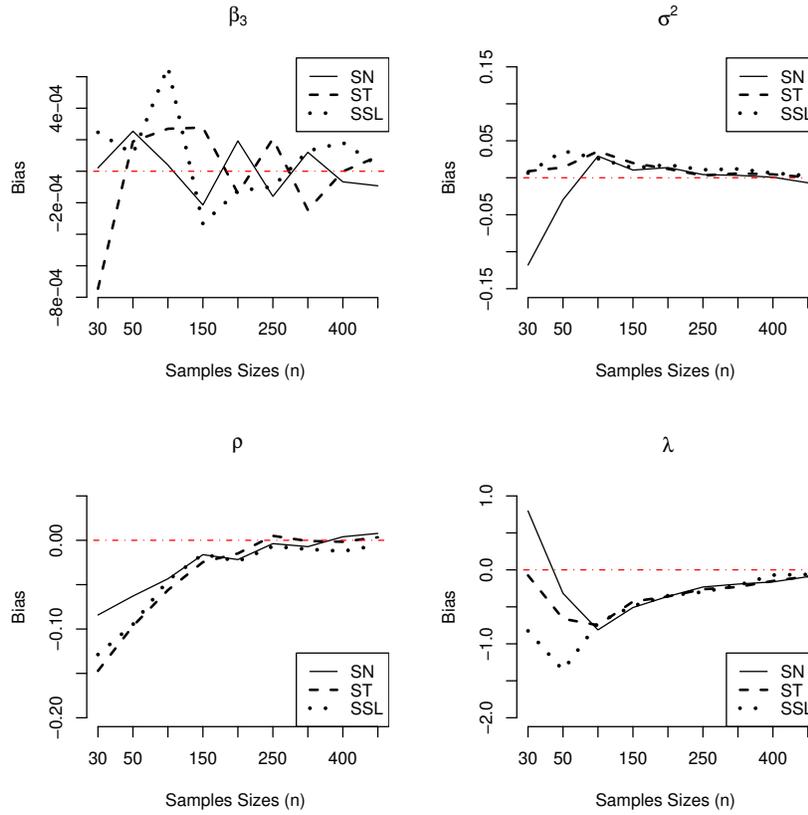


Figura 4.3: Viés na estimação de β_3 , σ^2 , ρ e λ no modelo NLMH-MESN definido em (4.13).

mais atraente usar o teste de escore, já que para obter a estatística do teste somente se requer calcular as EMV sob a hipótese nula (H_0).

Sejam $\theta_1 = \rho$ e $\theta = (\beta^\top, \sigma^2, \lambda)^\top$, sendo θ_1 o parâmetro de interesse. Então, a estatística do teste de escore para testar H_0 (Cox & Hinkley, 1974) é da seguinte forma:

$$SR = U_\rho(\hat{\theta}_2^0)^\top \mathbf{J} \rho \rho(\hat{\theta}_2^0) U_\rho(\hat{\theta}_2^0), \quad (4.16)$$

em que $U_\rho(\hat{\theta}_2^0) = \frac{\partial \ell(\theta_2)}{\partial \rho} \Big|_{\theta_2 = \hat{\theta}_2^0}$, $\mathbf{J} \rho \rho(\hat{\theta}_2^0)$ é a parte superior esquerda de $\mathbf{J}^{-1}(\theta_2) \Big|_{\theta_2 = \hat{\theta}_2^0}$ correspondente a ρ e $\hat{\theta}_2^0 = (\rho_0, \hat{\beta}^\top, \hat{\sigma}^2, \hat{\lambda})^\top$ são as EMV de $\theta_2 = (\rho, \beta^\top, \sigma^2, \lambda)^\top$ sob H_0 . Sob H_0 , a distribuição assintótica da estatística SR segue uma χ_q^2 .

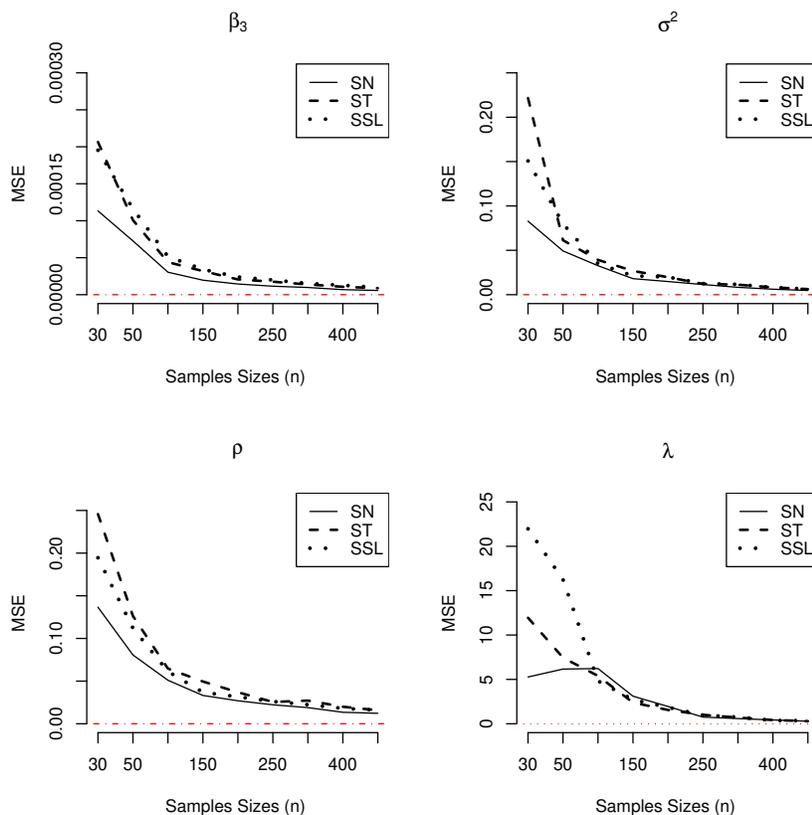


Figura 4.4: EQM nas estimações de β_3 , σ^2 , ρ e λ no modelo NLMH-MESN definido em (4.13).

4.4.2 Teste da razão de verossimilhança

A estatística da razão de verossimilhança (ERV) para testar H_0 , é definida por:

$$ERV = 2 \left(\ell(\hat{\theta}_2) - \ell(\tilde{\theta}_2^0) \right), \quad (4.17)$$

em que $\tilde{\theta}_2^0 = (\rho_0, \tilde{\beta}^\top, \tilde{\sigma}^2, \tilde{\lambda})^\top$ denota as EMV de θ_2 sobre H_0 e $\hat{\theta}_2 = (\hat{\rho}, \hat{\beta}^\top, \hat{\sigma}^2, \hat{\lambda})^\top$ denota as EMV sobre o modelo irrestrito. Neste caso, a distribuição assintótica da estatística da razão de verossimilhança segue uma χ_q^2 .

Nota-se que as EMV de θ_2 sob H_0 , que correspondem ao modelo homocedástico, são obtidas via algoritmo tipo EM⁵. Apesar da implementação da estatística de escore (SR) ser mais simples do que a ERV , visto que a primeira requer apenas a obtenção das estimativas sob H_0 , enquanto que a segunda necessita a obtenção das EMV sob H_0 e sob o modelo irrestrito. Como desenvolvemos o algoritmo tipo EM para obter as EMV sob H_0 e sob o modelo irrestrito, concluímos que a ERV não tem sido muito difícil de implementar.

4.4.3 Estudo de simulação

Nesta seção, desenvolveremos um estudo de simulação com o intuito de estudar a performance da distribuição assintótica e o poder do teste da estatística de escore, considerando os modelos NLM-MESN. Escolhemos a estatística do teste de escore para este estudo de simulação, já que o mesmo teste é comumente usado para avaliar a heteroscedasticidade do parâmetro de escala nos modelos de regressão não linear.

Como em Xie *et al.* (2009b), o modelo usado para este estudo de simulação é

$$Y_i = \exp(\beta x_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (4.18)$$

onde $\epsilon_i \sim MESN(-\sqrt{\frac{2}{\pi}}k_1\Delta, \sigma_i^2, \lambda; H)$, $\sigma_i^2 = \sigma^2 m_i(x_i, \rho)$, com $m_i(x_i, \rho) = \exp(\rho x_i)$. Vamos assumir que $\beta = 2$, $\sigma^2 = 0.5$, $\lambda = 1$.

Para este estudo de simulação, trabalharemos com a distribuição skew-normal, skew-t com $\nu = 3$ e a skew slash com $\nu = 3$; considerando diferentes tamanhos de amostra ($n = 30, 50, 70, 90$, e 150). Cada x_i foi gerado considerando uma distribuição uniforme no intervalo $(0.2, 2)$, e são os mesmos em cada simulação.

- *Poder do teste de escore*

Neste caso, os valores da variável aleatória Y_i são gerados seguindo o modelo (4.18), considerando os verdadeiros valores dos parâmetros, os valores de x_i fixos, os diferentes tamanhos de amostra (n) e os valores de $\rho = 0, 0.2, 0.4, 0.6, 0.8, 1$;

⁵Como foi discutido no Capítulo 3

repetindo este procedimento 2000 vezes.

Para cada valor de n e ρ , encontramos a proporção de vezes que foi rejeitada a hipótese nula (H_0), tal que esta proporção representa o valor simulado do poder do teste. Em cada caso, as estatísticas de escore foram comparadas com o valor crítico χ_1^2 ao nível $\alpha = 0.05$. A Tabela 4.2 apresenta o poder do teste estimado da estatística SR , para os diferentes tamanhos de amostra n e ρ . Note que para uma amostra de tamanho $n = 30$ o poder do teste estimado da estatística SR no modelo SN é 0.082, 0.0875 no modelo ST e 0.0945 no modelo SSL, ao nível de 0.05. Portanto, este poder do teste é muito conservador em todos os casos. Contudo, quando $n \geq 50$, os valores do poder do teste estão mais perto do 0.05. Quando ρ e n aumenta, o poder do teste da estatística SR aumenta rapidamente como se mostra na Figura 4.5. Nesta figura é possível observar que se o tamanho da amostra é moderada ou grande, a SR proposto detecta muito bem a heteroscedasticidade do parâmetro escalar.

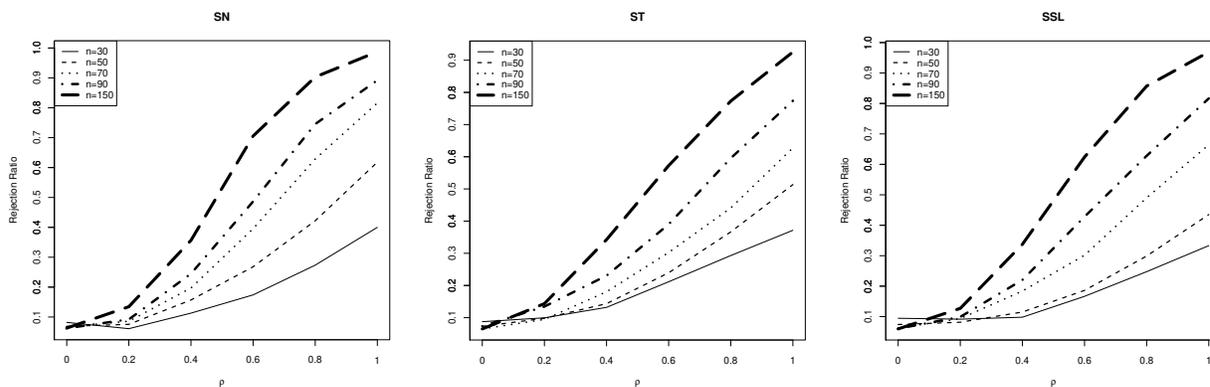


Figura 4.5: Poder do teste de escore para diferentes valores de n e ρ , considerando as distribuições SN, ST e SSL no estudo de simulação.

- *Distribuição empírica da estatística de escore*

A performance da distribuição assintótica da estatística de escore SR é avaliada comparando a distribuição empírica com a distribuição teórica via simulação Monte Carlo. Os valores da variável aleatória Y_i , são gerados seguindo o modelo (4.18), considerando os verdadeiros valores dos parâmetros, os valores de x_i , os diferentes tamanhos de amostra (n) e $\rho = 0$ (sob H_0), repetindo esse procedimento 2000 vezes. Em cada repetição é obtida a função de distribuição empírica do teste de escore. A comparação entre a função de distribuição empírica da estatística do escore e a função de distribuição χ_1^2 estão representadas na Figura 4.6 e mostram que a função de distribuição empírica da estatística do escore é aproximada à distribuição teórica χ_1^2 .

Tabela 4.2: Taxas de rejeição da $H_0 : \rho = 0$, sob as distribuições SN, ST e a SSL

SN-NLM						
n	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
30	0.0820	0.0610	0.1130	0.1740	0.2735	0.4000
50	0.0690	0.0755	0.1570	0.2680	0.4220	0.6180
70	0.0670	0.0860	0.1970	0.3955	0.6285	0.8160
90	0.0625	0.0920	0.2450	0.4865	0.7455	0.8920
150	0.0630	0.1345	0.3565	0.7055	0.9030	0.9875

ST-NLM						
n	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
30	0.0875	0.0995	0.1320	0.2115	0.2930	0.3715
50	0.0730	0.0980	0.1435	0.2395	0.3665	0.5135
70	0.0650	0.0945	0.1805	0.3025	0.4410	0.6275
90	0.0735	0.1345	0.2305	0.3900	0.5955	0.7740
150	0.0650	0.1435	0.3425	0.5725	0.7730	0.9255

SSL-NLM						
n	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$
30	0.0945	0.0915	0.0980	0.1665	0.2475	0.3330
50	0.0740	0.0815	0.1150	0.1860	0.2990	0.4350
70	0.0620	0.0935	0.1840	0.3015	0.4900	0.6645
90	0.0605	0.0990	0.2205	0.4280	0.6280	0.8165
150	0.0595	0.1270	0.3375	0.6240	0.8570	0.9690

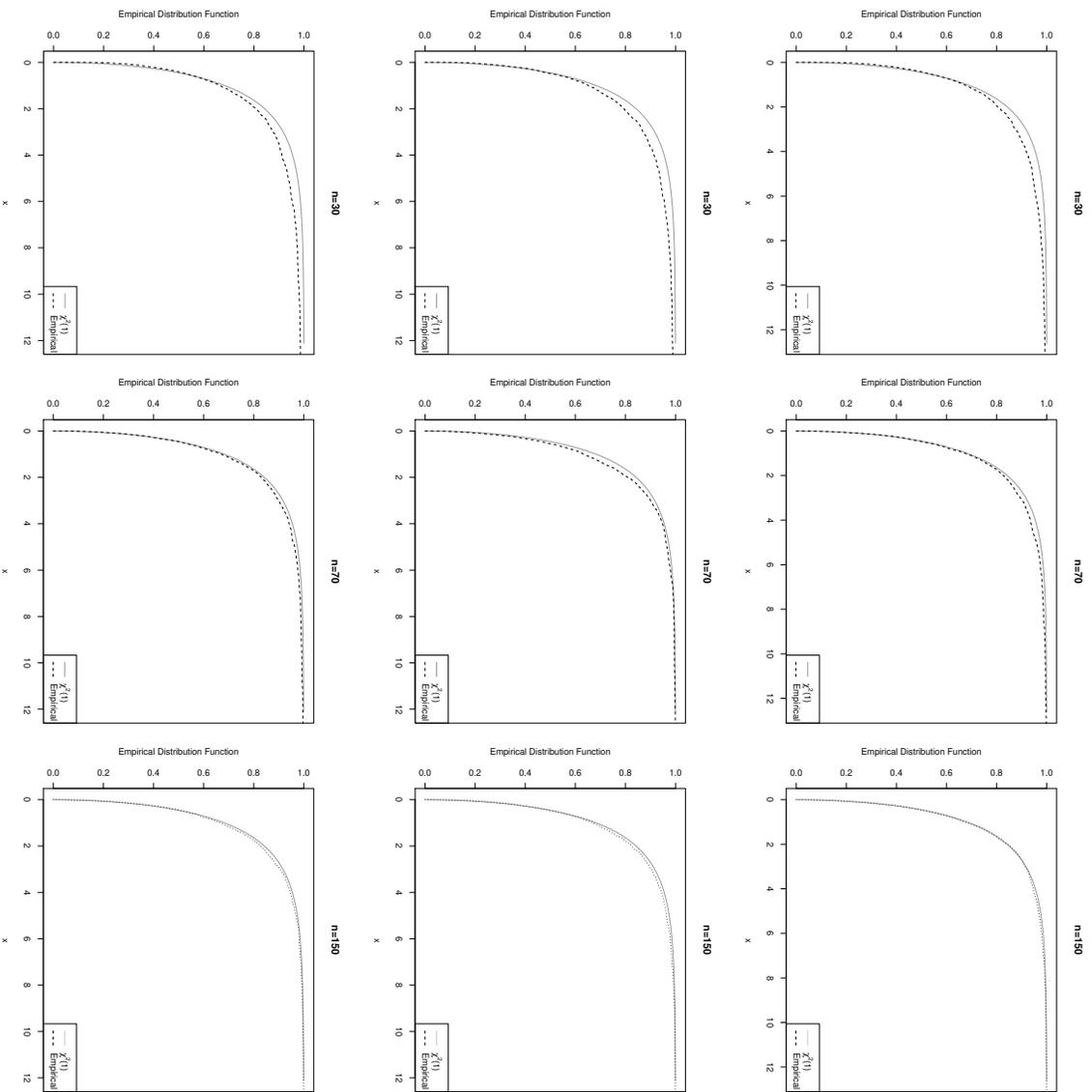


Figura 4.6: Comparação entre a função de distribuição empírica da SR e a distribuição X_1^2 , usando as distribuições SN , ST e a SSL (1ra, 2da e 3ra linha respectivamente).

4.5 Aplicação 1: Produtividade de óleo de palma

Nesta seção consideraremos os dados de produtividade de óleo de palma, que foram avaliados na Seção 3.5 sob o modelo homocedástico, onde concluímos que o modelo ST-MESN é mais adequado para ajustar os dados. Encontrou-se também que as observações #15, #16 e #18 são influentes.

4.5.1 Teste de heteroscedasticidade

Para fazer o teste de heteroscedasticidade do parâmetro de escala no modelo ST-NLM, com os dados de produtividade de óleo de palma, aplicamos o teste de escore descrito na Subseção 4.4.1 considerando $x_i = z_i$ e o modelo log-linear para a função m_i , isto é, $m_i = \exp(\rho x_i)$. Note que, quando $\rho = 0$, temos que $\sigma_i^2 = \sigma^2$ para todo i .

Assim, aplicando o teste de escore para testar as hipóteses $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$, encontramos que o valor da estatística de escore é $SR_{ST} = 19.79$ e seu correspondente p -valor = 8.642e-06, que nos permite rejeitar a hipótese de homocedasticidade. A seguir, avaliamos o impacto das observações #15, #16 e #18 na decisão do teste de hipótese de homocedasticidade do parâmetro de escala. Refazendo o teste sem estas observações encontramos que $SR_{ST} = 3.02$ (p -valor = 0.0818) o qual indica que as observações excluídas mudam a decisão para não rejeitar H_0 , isto é, a homocedasticidade do parâmetro de escala é preservada. Esse resultado concorda com a análise gráfica representada na Figura 3.7a, onde vemos um padrão constante dos resíduos, no intervalo de referência $(-1, 1)$.

A Tabela 4.3 mostra os valores das SR para os modelos SN-NLM, SSL-NLM e a ST-NLM, com seus correspondentes p -valor. Notamos que para os modelos SN-NLM e SSL-NLM são rejeitadas a hipótese H_0 . A presença de observações influentes afetam significativamente nossa decisão sobre heteroscedasticidade do parâmetro de escala, e esta decisão pode ser modificada dependendo do modelo usado. Similares conclusões são obtidas quando usamos o modelo produto potência para a função m , isto é, $m_i(x_i; \rho) = x_i^\rho$.

Tabela 4.3: Estatística de escore (SR) e p -valor para os modelos MESN.

Modelo	Todas as observações		Todas - #15, #16 e #18	
	SR	p -valor	SR	p -valor
SN-NLM	38.87257	4.5239e-10	5.324405	0.02102
ST-NLM	19.79010	8.6429e-06	3.028285	0.08182
SSL-NLM	32.12216	1.4477e-08	3.910459	0.04798

4.6 Aplicação 2: Calibração ultrasônica

Nesta seção, usaremos os dados referentes a calibração ultrasônica, analisados por Chwirut (1979). Estes dados são resultados de um estudo de NIST envolvendo calibração ultrasônica. Recentemente [Lin et al. \(2009\)](#) avaliaram o mesmo conjunto de dados sob o modelo homocedástico, considerando a distribuição skew-t-normal. A variável dependente é a resposta ultrasônica (y) e a variável explicativa é a distância metálica (x); a relação não linear entre as variáveis pode ser observada na Figura 4.7. Ilustraremos nossos métodos, substituindo o pressuposto skew-t-normal pela classe geral de distribuições MESN, como segue:

$$Y_i = \frac{\exp(-\beta_1 x_i)}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} \text{MESN}\left(-\sqrt{\frac{2}{\pi}}\kappa_1\Delta, \sigma_i^2, \lambda; H\right), \quad (4.19)$$

onde $\sigma_i^2 = \sigma^2 m_i$, para $i = 1, \dots, 214$.

Seguindo [Lin et al. \(2009\)](#), consideraremos para os dados o modelo não linear definido em (4.19) e para a função m_i o modelo produto potência, isto é, $m_i = x_i^\rho$

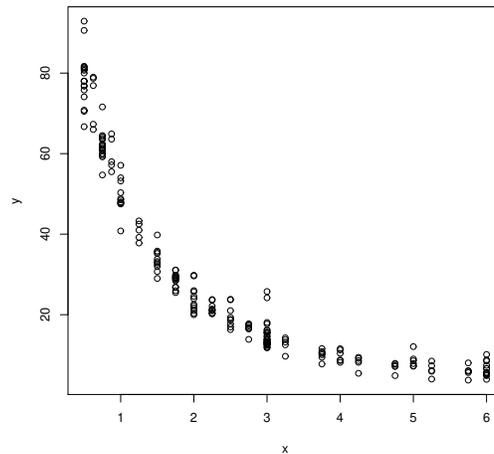


Figura 4.7: Gráfico de dispersão para os dados de calibração ultrasônica.

4.6.1 Ajuste do modelo

Para ajustar o modelo consideraremos as distribuições assimétricas SN-NLMH, ST-NLMH e SSL-NLMH e, além disso, consideraremos também as distribuições simétricas normal (N-HNLM) e t-student(T-HNLM). Sob a distribuição ST e SSL é escolhido o valor inteiro de ν que maximiza a função de log-verossimilhança⁶, como é mostrado na Figura 4.8. Para o modelo ST encontramos que o valor de $\nu = 4$ e para o modelo SSL encontramos o valor de $\nu = 2$.

A Tabela 4.4 mostra as EMV dos parâmetros para os cinco modelos, com seus correspondentes erros padrões assintóticos (EP) obtidos através da matriz de informação observada. O critério de seleção do modelo AIC indica que as distribuições assimétricas com caudas pesadas apresentam melhores ajustes, sendo a distribuição ST-NLMH significativamente melhor. Se compararmos o critério AIC dos modelos assimétricos com o obtido por Lin *et al.* (2009) usando a distribuição skew-t normal (AIC=1069.3), temos que os modelos assimétricos produzem melhores ajustes.

⁶Procedimento discutido em Lange *et al.* (1989)

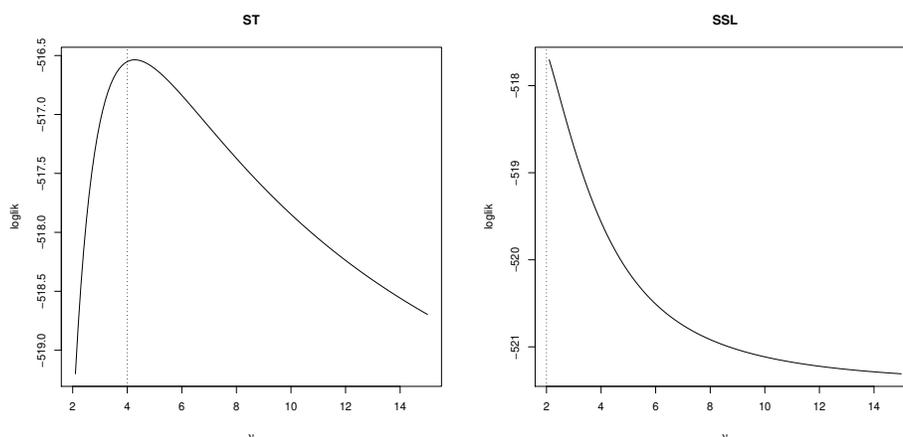


Figura 4.8: Log-verossimilhança perfilada do parâmetro ν considerando os modelos ST-NLMH e SSL-NLMH.

Embora as estimativas dos parâmetros de regressão são semelhantes para as distribuições consideradas, os erros padrões obtidos, sob distribuições com caudas pesadas são menores que os obtidos sob o modelo SN-NLMH (na maioria dos casos). Isto sugere que a ST-NLMH e a SSL-NLMH produzem estimativas de máxima verossimilhança mais eficientes do que o modelo SN-NLMH.

Tabela 4.4: EMV e EP obtidos ao ajustar vários modelos de misturas nos dados de calibração ultrasônica.

Parâmetros	N-NLMH		T-NLMH		SN-NLMH		ST-NLMH		SSL-NLMH	
	EMV	EP	EMV	EP	EMV	EP	EMV	EP	EMV	EP
β_1	0.148	0.01595	0.157	0.01354	0.154	0.01512	0.156	0.01415	0.155	0.01449
β_2	0.005	0.00041	0.005	0.00033	0.006	0.00038	0.005	0.00036	0.005	0.00037
β_3	0.012	0.00076	0.012	0.00064	0.012	0.00071	0.012	0.00066	0.012	0.00068
ρ	-0.959	0.12627	-1.035	0.08299	-0.987	0.12717	-0.959	0.18024	-1.025	0.15795
σ^2	16.20	2.10449	8.837	0.61650	33.785	5.82211	11.323	3.04540	13.86	3.29217
λ					2.248	0.43980	0.885	0.37991	1.477	0.42380
ν	-	-	4			-	4	-	2	
log-verossimilhança	-531.076		-519.328		-521.454		-516.5663		-517.507	
AIC	1072.152		1050.656		1054.908		1047.133		1049.014	

4.6.2 Análise de diagnóstico

- *Modelo de eliminação de casos (CDM)*

Baseado na distância generalizada de Cook (GD_i^l), os resultados mostrados na Figura 4.9, indicam que as observações (#146, #147 e #176) são identificados como as mais influentes na estimação dos parâmetros sob o modelo SN-NLMH; no entanto, nenhuma observação é considerada influente sob o modelo ST-NLMH e SSL-NMLH. O mesmo padrão pode ser observado quando usamos o afastamento pela log-verossimilhança (LD_i^1), onde apenas as observações #146, #147 e #176 são identificados como as mais influentes na estimação dos parâmetros nos modelos SN-NLMH, ST-NLMH e SSL-NLMH. Nota-se na Figura 4.10, que as distribuições com caudas mais pesadas do que a SN-NLMH, permitem acomodar as observações influentes (#146, #147 e #176), atribuindo-lhes pequenos pesos no processo de estimação; os pesos estimados para a distribuição skew-normal ($\hat{u}_i, i = 1, \dots, 214$) são indicados como uma linha contínua. Portanto, esta classe de distribuições (MESN) permite, naturalmente, atribuir pesos diferentes a cada observação e, conseqüentemente, controlar a influência de uma observação sobre as estimativas dos parâmetros.

A seguir, apresentamos um estudo de influência local baseado no esquema de perturbação de ponderação de casos, sob os modelos SN-NLMH, ST-NLMH e SSL-NLMH, considerando o valor de $c^* = 2$ para $M(0)$.

- *Perturbação de ponderação de casos*

Sob o esquema de perturbação de ponderação de casos, encontramos os valores de curvatura máxima $C_{d_{maxSN}} = 1.93$, $C_{d_{maxST}} = 1.64$ e $C_{d_{maxSSL}} = 1.78$.

Da Figura 4.11 notamos que nas três distribuições a observação #176 é identificada como influente. Além disso, a observação #146 também é identificada como influente sob o modelo SN-NLMH.

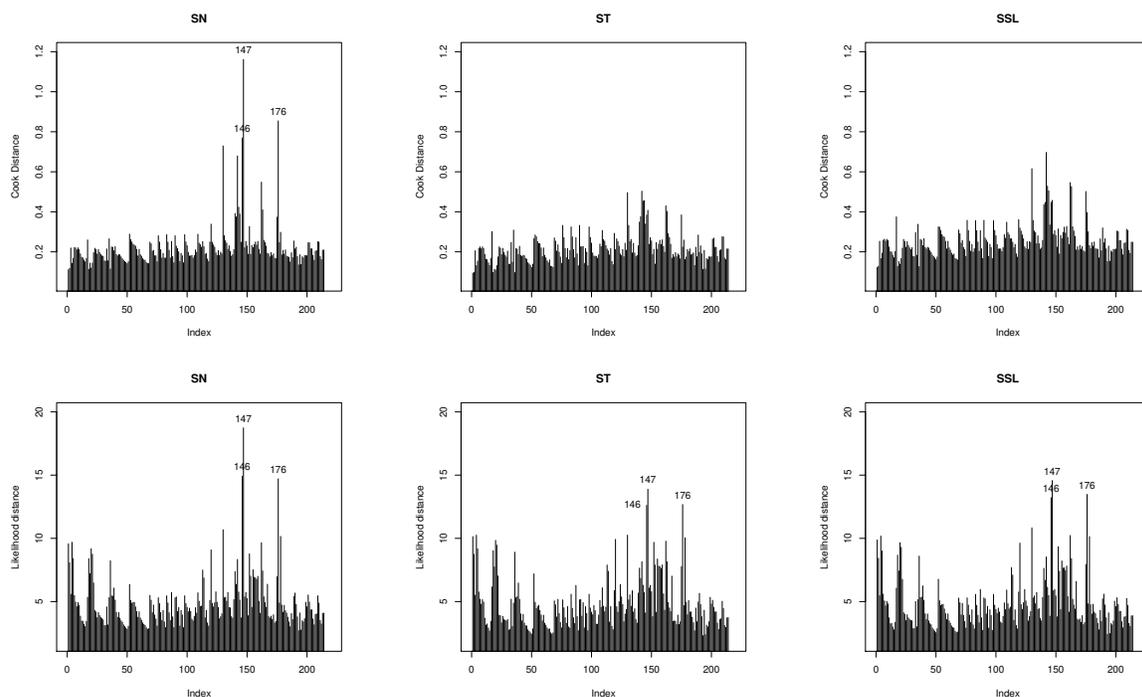


Figura 4.9: Gráfico de a) GD_i^l e b) LD_i^l sob os modelos SN-NLMH, ST-NLMH e SSL-NLMH.

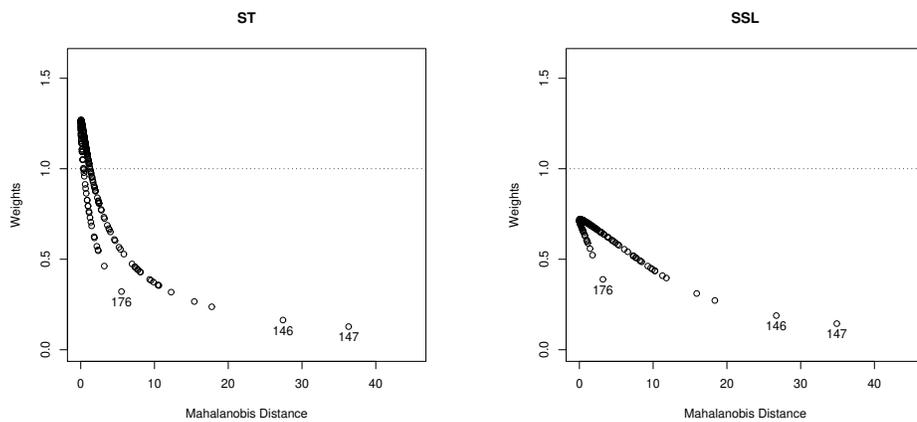


Figura 4.10: Pesos estimados (\hat{u}_i) para os modelos ST-NLMH e SSL-NLMH.

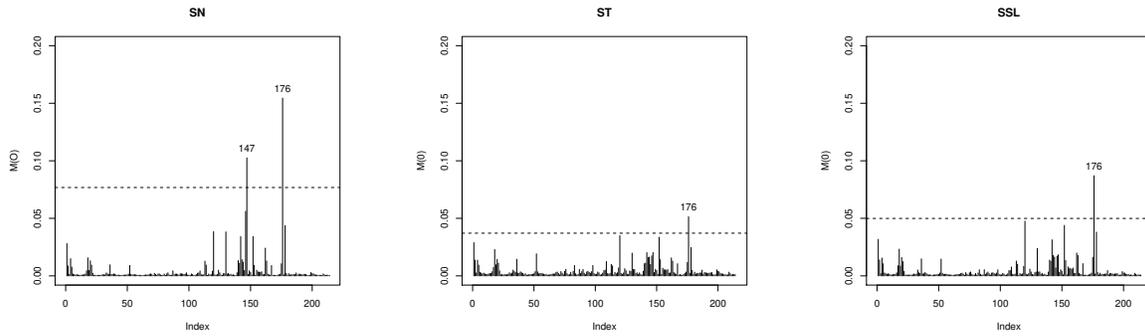


Figura 4.11: Gráfica de $M(0)$ com os modelos SN-NLMH, ST-NLMH e SSL-NLMH, sob perturbação de ponderação de casos.

4.6.3 Análise de resíduos

A fim de verificar se existe uma especificação incorreta da distribuição dos erros e a componente sistemática (4.19), na Figura 4.12 é mostrado o gráfico dos envelopes simulados para os resíduos de Pearson. Esta figura indica claramente que o modelo ST-NLMH é mais apropriado para ajustar os dados de calibração ultrasônica do que os modelos N-NLMH, SN-NLMH e T-NLMH, dado que não há observações que estão fora dos limites dos envelopes simulados.

4.6.4 Teste de heteroscedasticidade

Com a finalidade de verificar a presença de heteroscedasticidade do parâmetro escalar no conjunto de dados de calibração ultrasônica, aplicaremos o teste da razão de verossimilhança no modelo ST-NLMH, sendo este modelo o mais apropriado para a modelagem deste conjunto de dados⁷, com $\sigma_i^2 = \sigma^2 x_i^\rho$. Nota-se que, se $\rho = 0$, então $\sigma_i^2 = \sigma^2$ para todo i , isto é, a estatística da razão de verossimilhança para testar as hipóteses $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$, é definida por:

$$ERV = 2 \left(\ell(\hat{\theta}_2) - \ell(\tilde{\theta}_2^0) \right)$$

⁷Ver Subseção 4.6.1 e 4.6.3

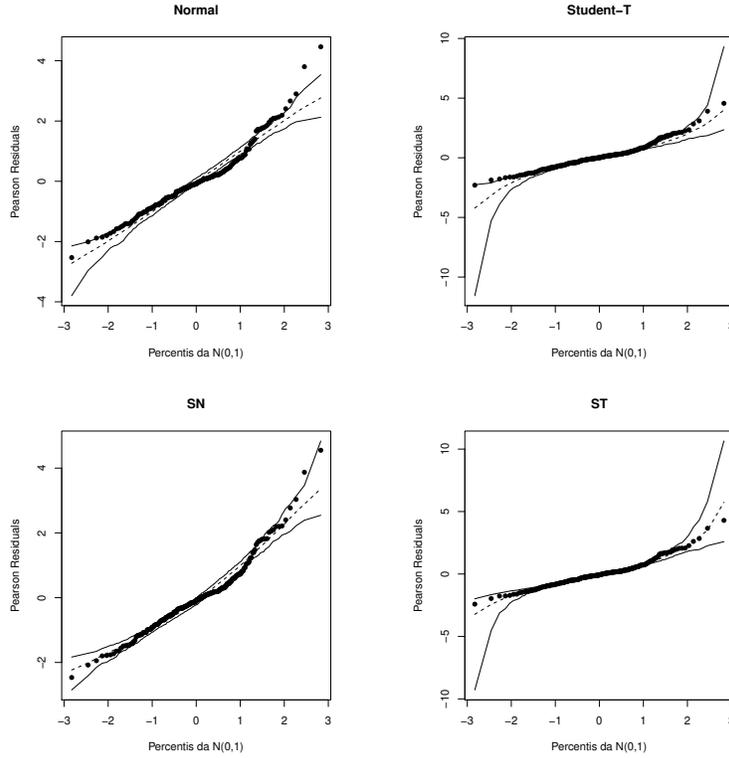


Figura 4.12: Envelopes simulados para os resíduos de Pearson sob os modelos N-NLMH, SN-NLMH, T-NLMH e ST-NLMH.

onde $\tilde{\boldsymbol{\theta}}_2^0 = (0, \tilde{\boldsymbol{\beta}}^\top, \tilde{\sigma}^2, \tilde{\lambda})^\top$ denota as EMV sob H_0 e $\hat{\boldsymbol{\theta}}_2 = (\hat{\rho}, \hat{\boldsymbol{\beta}}^\top, \hat{\sigma}^2, \hat{\lambda})^\top$ denota as EMV sob o modelo irrestrito.

Neste caso, temos que $ERV_{ST} = 33.04$ e seu correspondente $p\text{-valor} = 8.994e-09$, com o qual rejeitamos a hipótese de homocedasticidade, conclusão que coincide com a obtida por [Lin et al. \(2009\)](#) usando a estatística de score. Portanto, concluímos que a suposição de homogeneidade do parâmetro de escala não é adequado para os dados de calibração ultrasônica.

Não obstante, com o intuito de verificar se a presença das observações influentes obtidas no análise de diagnóstico (#146, #147 e #176) afetam a heteroscedasticidade do parâmetro de escala, aplicamos novamente o teste da razão de verossimilhança no conjunto de dados sem considerar estas observações influentes; encontrando que $ERV_{ST} = 38.05$ com

seu correspondente p -valor ≈ 0 , com o qual concluímos que a heteroscedasticidade do parâmetro de escala é preservada. Neste conjunto de dados a presença de possíveis observações influentes não afetam significativamente a heteroscedasticidade do parâmetro de escala do modelo.

A Tabela 4.5 apresenta os valores da ERV e p -valor para os modelos SN-NLMH, SSL-NLMH e ST-NLMH, mostrando que em todas as distribuições avaliadas, a hipótese de homocedasticidade do parâmetro de escala é rejeitada, considerando todas as observações ou sem considerar aquelas que são influentes. Similares resultados foram obtidos quando consideramos o teste de escore para avaliar a heteroscedasticidade do parâmetro de escala (vide Tabela 4.6).

Estes resultados reforçam o fato que existe uma forte presença de heteroscedasticidade do parâmetro de escala, a qual deve ser considerada para fazer as inferências no conjunto de dados de calibração ultrasônica. Similares resultados foram obtidos quando consideramos o modelo log-linear para a função m , isto é, $m_i = \exp(z_i\rho)$, por tanto, $\sigma_i^2 = \sigma^2 m_i = \sigma^2 \exp(z_i\rho)$

Tabela 4.5: ERV e p -valor nos modelos NLMH-MESN.

Modelo	Todas as observações		Todas - #146, #147 e #176	
	ERV	p -valor	ERV	p -valor
SN-NLMH	68.40134	1.110223e-16	81.19974	≈ 0
ST-NLMH	33.04722	8.994735e-09	38.05352	≈ 0
SSL-NLMH	44.34826	2.748557e-11	49.71340	≈ 0

Tabela 4.6: Estatística de Escore(SR) e p -valor nos modelos NLMH-MESN.

Modelo	Todas as observações		Todas - #146, #147 e #176	
	SR	p -valor	SR	p -valor
SN-NLMH	74.69615	≈ 0	75.34588	≈ 0
ST-NLMH	33.67623	≈ 0	31.56924	≈ 0
SSL-NLMH	49.86896	≈ 0	44.72882	≈ 0

4.7 Considerações finais

Apresentamos neste capítulo o modelo não linear heteroscedástico sob a classe de distribuições MESN, como uma extensão do modelo NLM-MESN apresentado no Capítulo 3. Foi discutido o algoritmo tipo EM para encontrar as estimativas de máxima verossimilhança dos parâmetros. Estudos de simulação desenvolvidos mostraram que as estimativas obtidas via algoritmo tipo EM nos modelos com caudas mais pesadas do que a skew-normal reduzem e controlam a influência de observações influentes na estimativa dos parâmetros. Além disso, foi discutido o teste de escore e o teste da razão de verossimilhança para avaliar a heteroscedasticidade do parâmetro de escala. Estes testes, foram complementados com um estudo de simulação para estudar a distribuição assintótica da estatística de escore, assim como o poder do teste.

Na aplicação 1, trabalhamos com os dados de produtividade de óleo de palma, apresentados no Capítulo 3, mostrando que a presença de observações influentes podem afetar significativamente nossa decisão sobre heteroscedasticidade do parâmetro escalar. Na aplicação 2, avaliamos os dados de calibração ultrasônica, encontrando que os modelos com caudas mais pesadas do que o modelo SN-NLMH, produzem melhores ajustes. O teste da razão de verossimilhança e o teste de escore evidenciaram que o modelo heteroscedástico é mais apropriado para ajustar os dados.

Capítulo 5

Conclusões Finais e Perspectivas Futuras

5.1 Conclusões finais

Nesta dissertação discutimos os modelos de regressão não linear sob a classe de distribuições misturas da escala skew-normal. Desenvolvemos o algoritmo tipo EM para encontrar as EMV dos parâmetros para as distribuições skew-normal (SN-NLM), skew-t (ST-NLM) e skew slash (SSL-NLM). Os desvios padrões assintóticos foram obtidos baseado na matriz de informação observada tanto no modelo homocedástico quanto no modelo heteroscedástico. Apresentamos os resíduos de Pearson para avaliar o ajuste para os dados via gráfico de envelopes. Estudos de simulação foram feitos com o intuito de investigar a robustez dos resultados.

Conhecendo a importância de avaliar a presença das observações influentes, utilizamos a análise de diagnóstico nos modelos NLM-MESN através do *modelo de elimi-*

nação de casos(CDM)- considerando, para isso, o afastamento pela log-verossimilhança e a distância de Cook - e um método mais geral, conhecido na literatura como *método de influência local* (Cook, 1986), que estuda a estabilidade dos resultados sob o esquema de perturbação de ponderação de casos e a perturbação ao parâmetro de escala.

Outro ponto importante, que foi abordado é um estudo de teste de hipótese baseado nas estatísticas de escore e da razão de verossimilhança, para testar a homocedasticidade do parâmetro de escala nos modelos. Além disso, estudos de simulação foram feitos para estudar a performance do teste de escore.

Os resultados obtidos foram aplicados em dois conjuntos de dados reais: um conjunto de dados que trata sobre a produtividade de óleo de palma - que foi considerado por Cancho *et al.* (2009) baseado no modelo SN-NLM - e outro conjunto de dados que trata sobre calibração ultrasônica - que foi analisado recentemente por Lin *et al.* (2009). Foram realizados estudos de simulação utilizando os programas estatísticos como o R e Matlab.

Finalmente, os resultados deste trabalho representam uma extensão dos resultados apresentados por Cancho *et al.* (2009), Lin *et al.* (2009), Xie *et al.* (2009a) e Xie *et al.* (2009b), num contexto assimétrico, assim como, dos resultados apresentados por Osorio (2006), Vanegas & Cysneiros (2009), num contexto simétrico.

5.2 Perspectivas Futuras

Vários trabalhos de pesquisa podem ser derivadas e/ou direcionadas a partir dos resultados deste trabalho, entre elas podemos mencionar as seguintes:

- Estudar os resultados de estimação e diagnóstico encontrados nos modelos não-lineares sob a classe MESN desde uma perspectiva bayesiana.

-
- Estender os resultados de estimação e diagnóstico encontrados nos modelos NLM-MESN para o caso multivariado.
 - Estudar os modelos mistos não lineares sob a classe MESN, com a finalidade de introduzir mais propriedades e discutir alguns aspectos inferenciais, incluindo o algoritmo tipo EM para a obtenção das EMV dos parâmetros de interesse, bem como a questão de identificabilidade.
 - Realizar um estudo de inferência e diagnóstico nos modelos lineares generalizados mistos sob a classe MESN, desde uma perspectiva clássica e bayesiana.
 - Realizar um estudo de inferência e diagnóstico nos modelos para dados de contagem com excesso de zeros e efeitos aleatórios assimétricos, desde uma perspectiva clássica e bayesiana.

Apêndice A

Resultados adicionais do Capítulo 3 e Capítulo 4

A.1 Elementos da matriz de perturbação Δ^*

A fim de avaliar o esquema de perturbação $\sigma^2(\omega_i) = \sigma^2/\omega_i$, vamos apresentar os elementos de Δ para este esquema. Em este caso a função de log-verossimilhança é dado por $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\omega_i)$, onde que

$$\ell_i(\boldsymbol{\theta}|\omega_i) = \log 2 - \frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log \omega_i + \log K_{\omega_i},$$

e K_{ω_i} é definida na seção com $\sigma^2(\omega_i)$ em vez de σ^2 . Neste caso $d_{\omega_i} = B^2(\omega_i)$ e $A_{\omega_i} = \lambda B(\omega_i)$, em que $B(\omega_i) = \omega_i^{1/2}(y_i - \eta(\boldsymbol{\beta}, \mathbf{x}_i))/\sigma - b\delta\sigma$.

Sobre este esquema de perturbação temos $\Delta_{\boldsymbol{\omega}_0}^* = (\Delta_{\boldsymbol{\beta}}^{*\top}, \Delta_{\sigma^2}^{*\top}, \Delta_{\lambda}^{*\top})^\top$, em que

$$\Delta_{i\gamma}^* = -\frac{1}{K_{\omega_i}^2} \frac{\partial}{\partial \omega_i} K_{\omega_i} \frac{\partial}{\partial \gamma} K_{\omega_i} + \frac{1}{K_{\omega_i}} \frac{\partial^2}{\partial \omega_i \partial \gamma} K_{\omega_i}, \quad \gamma = \boldsymbol{\beta}, \sigma^2, \lambda,$$

com

$$\begin{aligned}\frac{\partial}{\partial \omega_i} K_{\omega_i} &= -\frac{1}{2} I_i^\Phi(3/2) \frac{\partial}{\partial \omega_i} d_{\omega_i} + I_i^\phi(1) \frac{\partial}{\partial \omega_i} A_{\omega_i}, \\ \frac{\partial}{\partial \gamma} K_{\omega_i} &= -\frac{1}{2} I_i^\Phi(3/2) \frac{\partial}{\partial \gamma} d_{\omega_i} + I_i^\phi(1) \frac{\partial}{\partial \gamma} A_{\omega_i}, \\ \frac{\partial^2}{\partial \omega_i \partial \gamma} K_{\omega_i} &= -\frac{1}{2} \left[-\frac{1}{2} I_i^\Phi(5/2) \frac{\partial}{\partial \gamma} d_{\omega_i} + I_i^\phi(2) \frac{\partial}{\partial \omega_i} A_{\omega_i} \right] \frac{\partial}{\partial \gamma} d_{\omega_i} - \frac{1}{2} I_i^\Phi(3/2) \frac{\partial^2}{\partial \omega_i \partial \gamma} d_{\omega_i} \\ &\quad - \frac{1}{2} I_i^\phi(2) \left[\frac{\partial}{\partial \gamma} d_{\omega_i} + 2A_{\omega_i} \frac{\partial}{\partial \omega_i} A_{\omega_i} \right] \frac{\partial}{\partial \gamma} A_{\omega_i} + I_i^\phi(1) \frac{\partial^2}{\partial \omega_i \partial \gamma} A_{\omega_i}.\end{aligned}$$

A continuação, mostramos as derivadas de d_{ω_i} e A_{ω_i} em relação a θ e ω e para simplificar a notação consideramos $\eta_i = \eta(\beta, \mathbf{x}_i)$. Se consideramos $\omega_i = \omega_{i0} = 1$, obtemos:

- d_i :

$$\begin{aligned}\frac{\partial d_{\omega_i}}{\partial \omega} &= C_i B_i, \\ \frac{\partial d_{\omega_i}}{\partial \beta} &= -\frac{2}{\sigma} B_i \frac{\partial \eta_i}{\partial \beta}, \quad \frac{\partial d_{\omega_i}}{\partial \sigma^2} = -\frac{1}{\sigma^2} B_i C_i, \quad \frac{\partial d_{\omega_i}}{\partial \lambda} = -2b\delta' B_i, \\ \frac{\partial^2 d_{\omega_i}}{\partial \beta \partial \omega_i} &= -\frac{1}{\sigma} (2B_i + b\delta) \frac{\partial \eta_i}{\partial \beta}, \quad \frac{\partial^2 d_{\omega_i}}{\partial \omega \partial \sigma^2} = -\frac{1}{2\sigma^2} (2B_i + b\delta), \quad \frac{\partial^2 d_{\omega_i}}{\partial \lambda \partial \omega_i} = -b\delta' C_i,\end{aligned}$$

- A_i :

$$\begin{aligned}\frac{\partial A_{\omega_i}}{\partial \omega_i} &= \frac{\lambda}{2} C_i, \\ \frac{\partial A_{\omega_i}}{\partial \beta} &= -\frac{\lambda}{\sigma} \frac{\partial \eta_i}{\partial \beta}, \quad \frac{\partial A_{\omega_i}}{\partial \sigma^2} = -\frac{\lambda}{2\sigma^2} C_i, \quad \frac{\partial A_{\omega_i}}{\partial \lambda} = B_i - b\lambda\delta', \\ \frac{\partial^2 A_{\omega_i}}{\partial \beta \partial \omega_i} &= -\frac{\lambda}{2\sigma} \frac{\partial \eta_i}{\partial \beta}, \quad \frac{\partial^2 A_{\omega_i}}{\partial \sigma^2 \partial \omega_i} = -\frac{\lambda}{4\sigma^2} C_i, \quad \frac{\partial^2 A_{\omega_i}}{\partial \lambda \partial \omega_i} = \frac{1}{2} C_i.\end{aligned}$$

A.2 Primeira e segunda derivada no modelo hete-roscedástico

Em este apêndice a primeira e segunda derivada de $d_i = B_i^2$ e $A_i = \lambda B_i$ são obtidas, em que $B_i = (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}) - b\sigma_i\delta)/\sigma_i = C_i - b\delta$, com $C_i = (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}))/\sigma_i$ e $\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}$. Aqui $\sigma_i^2 = \sigma^2 m(\mathbf{x}_i, \boldsymbol{\rho}) = \sigma^2 m_i$ e assim $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda, \boldsymbol{\rho}^\top)^\top$.

- d_i :

$$\begin{aligned} \frac{\partial d_i}{\partial \boldsymbol{\beta}} &= -2 \frac{B_i}{\sigma_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial d_i}{\partial \sigma^2} = -\frac{B_i}{\sigma^2} C_i, \quad \frac{\partial d_i}{\partial \lambda} = -2bB_i\delta', \quad \frac{\partial d_i}{\partial \boldsymbol{\rho}} = -\frac{B_i}{m_i} C_i \frac{\partial m_i}{\partial \boldsymbol{\rho}}, \\ \frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= 2 \left[\frac{1}{\sigma_i^2} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}^\top} + \frac{B_i}{\sigma_i} \frac{\partial^2 \eta_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right], \\ \frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \sigma^2} &= \frac{1}{\sigma^2 \sigma_i} [2B_i + b\delta] \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \\ \frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \lambda} &= \frac{2b}{\sigma_i} \delta' \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \\ \frac{\partial^2 d_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\rho}^\top} &= \frac{1}{\sigma_i m_i} [2B_i + b\delta] \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top}, \\ \frac{\partial^2 d_i}{\partial \sigma^2 \partial \sigma^2} &= \frac{1}{2\sigma^4} (2B_i + \frac{1}{2}b\delta) C_i, \quad \frac{\partial^2 d_i}{\partial \sigma^2 \partial \lambda} = \frac{b\delta'}{\sigma^2} C_i, \quad \frac{\partial^2 d_i}{\partial \sigma^2 \partial \boldsymbol{\rho}^\top} = \frac{1}{2\sigma^2 m_i} (2B_i + b\delta) C_i \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top}, \\ \frac{\partial^2 d_i}{\partial \lambda \partial \lambda} &= -2b[\delta'' B_i - b(\delta')^2], \quad \frac{\partial^2 d_i}{\partial \lambda \partial \boldsymbol{\rho}} = \frac{b\delta'}{m_i} (2B_i + b\delta) \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top}, \\ \frac{\partial^2 d_i}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^\top} &= \left[\frac{1}{2m_i^2} (4B_i + b\delta) \frac{\partial m_i}{\partial \boldsymbol{\rho}} \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top} - \frac{B_i}{m_i} \frac{\partial^2 m_i}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^\top} \right] C_i, \end{aligned}$$

onde δ' e δ'' são a primeira e segunda derivada de δ .

- A_i :

$$\begin{aligned} \frac{\partial A_i}{\partial \boldsymbol{\beta}} &= -\frac{\lambda}{\sigma_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial A_i}{\partial \sigma^2} = -\frac{\lambda}{2\sigma^2} C_i, \quad \frac{\partial A_i}{\partial \lambda} = B_i - b\lambda\delta', \quad \frac{\partial A_i}{\partial \boldsymbol{\rho}} = -\frac{\lambda}{2m_i} C_i \frac{\partial m_i}{\partial \boldsymbol{\rho}}, \\ \frac{\partial^2 A_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\frac{\lambda}{\sigma_i} \frac{\partial^2 \eta_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}, \quad \frac{\partial^2 A_i}{\partial \boldsymbol{\beta} \partial \sigma^2} = \frac{\lambda}{\sigma^2 \sigma_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \quad \frac{\partial^2 A_i}{\partial \boldsymbol{\beta} \partial \lambda} = -\frac{1}{\sigma_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}, \\ \frac{\partial^2 A_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\rho}^\top} &= \frac{\lambda}{2m_i \sigma_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top}, \\ \frac{\partial^2 A_i}{\partial \sigma^2 \partial \sigma^2} &= \frac{3\lambda}{4\sigma^4} C_i, \quad \frac{\partial^2 A_i}{\partial \sigma^2 \partial \lambda} = -\frac{1}{2\sigma^2} C_i, \quad \frac{\partial^2 A_i}{\partial \sigma^2 \partial \boldsymbol{\rho}^\top} = \frac{\lambda}{4m_i \sigma^2} C_i \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top}, \\ \frac{\partial^2 A_i}{\partial \lambda \partial \lambda} &= -b[2\delta' + \lambda\delta''], \quad \frac{\partial^2 A_i}{\partial \lambda \partial \boldsymbol{\rho}^\top} = -\frac{1}{2m_i} C_i \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top}, \\ \frac{\partial^2 A_i}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^\top} &= \frac{\lambda}{2} \left[\frac{3}{2m_i^2} \frac{\partial m_i}{\partial \boldsymbol{\rho}} \frac{\partial m_i}{\partial \boldsymbol{\rho}^\top} - \frac{1}{m_i} \frac{\partial^2 m_i}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^\top} \right] C_i. \end{aligned}$$

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 716–723.
- Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.
- Arellano-Valle, R.B., G. H. W. Q. F. A. (2004). A new of class of skew-normal distributions. *Communication in Statistics - Theory and Methods*, **33**, 1465–1480.
- Atkinson, A. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.
- Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, **61**, 579–602.
- Azzalini, A. & Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, **76**, 1490–1507.
- Bai, Z. D., Krishnaiah, P. R. & Zhao, L. C. (1989). On rates of convergence of efficient detection criteria in signal processing with white noise. *IEEE Trans. Info. Theory*, **35**, 380–388.
- Barroso, L. P. & Cordeiro, G. M. (2005). Bartlett corrections in heteroskedastic t regression models. *Statistics & Probability Letters*, **75**(2), 86–96.

- Basso, M. R. (2009). *Misturas Finitas de Misturas de Escala Skew-Normal*. Dissertação do mestrado, Instituto de Matemática, Estatística e Computação Científica, IMECC - UNICAMP.
- Basso, R., Lachos, V., Cabral, C. & Ghosh, P. (2009). Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics and Data Analysis*, **doi:10.1016/j.csda.2009.09.031**.
- Bayes, C. L. (2005). *Inferência Bayesiana no modelo normal assimétrico*. Dissertação do mestrado, Instituto de Matemática e Estatística, IME-USP.
- Berkane, M., Kano, Y. & Bentler, P. M. (1994). Pseudo maximum likelihood estimation in elliptical theory: effects of misspecification. *Computational Statistics & Data Analysis*, **18**, 255–267.
- Box, G. & Cox, D. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society*, **26**, 211–252.
- Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113.
- Cancho, V. C., Lachos, V. H. & Ortega, E. M. M. (2009). A nonlinear regression model with skew-normal errors. *Statistical Papers*, **doi:10.1007/s00362-008-0139-y**.
- Chen, C.-F. (1983). Score tests for regression models. *Journal of the American Statistical Association*, **78**(381), 158–161.
- Chih-Ling, T. (1986). Score test for the first-order autoregressive model with heteroscedasticity. *Biometrika*, **73**, 455–460.
- Cook, R. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**(1), 15–18.
- Cook, R. & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, pages 1–10.

- Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.
- Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.
- Cordeiro, G. M., Ferrari, S., Uribe-Opazo, M. & Vasconcellos, K. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters*, **46**, 317–328.
- Cox, D. & Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall/CRC, New York, NY.
- Cysneiros, F., Paula, G. & Galea, M. (2007). Heteroscedastic symmetrical linear models. *Statistics and Probability Letters*, **77**(11), 1084–1090.
- Cysneiros, F., Cordeiro, G. & Cysneiros, A. (2009). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, -(999), 1–11.
- Cysneiros, F. J. A. & Vanegas, L. H. (2008). Residuals and their statistical properties in symmetrical nonlinear models. *Statistics & Probability Letters*, **78**, 3269–3273.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–22.
- Fang, K. T., Kotz, S. & Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall, London.
- Ferreira, C. S. (2008). *Inferência e diagnóstico em modelos assimétricos*. Tese de doutorado, Departamento de Estatística, IME-USP.
- Foong, F. S. (1999). Impact of mixture on potential evapotranspiration, growth and yield of palm oil. *PORIM Interl. Palm Oil Cong. (Agric.)*, pages 265–287.
- Galea, M., Paula, G. A. & Bolfarine, H. (1997). Local influence in elliptical linear regression models. *Journal of the Royal Statistical Society*, **46**, 71–79.

- Galea, M., Paula, G. A. & Cysneiros, F. J. A. (2005). On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters*, pages 459–467.
- Gómez, H., Venegas, O. & Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, **18**(4), 395–408.
- Hill, M. A. & Dixon, W. J. (1982). Robustness in real life: a study of clinical laboratory data. *Biometrics*, **38**. 377-396.
- Kim, H. M. (2008). A note on scale mixtures of skew normal distribution. *Statistics and Probability Letters*, **78**. 1694-1701.
- Lachos, V. H. & Vilca, L. F. (2007). Skew-normal/ independent distributions, with applications. RT-IMECC 02, IMECC-UNICAMP.
- Lachos, V. H., Bolfarine, H., Vilca, L. F. & Galea-Rojas, M. (2005). Estimation and influence diagnostic for structural comparative calibration models under the skew-normal distribution. RT-MAE 12, IME-USP.
- Lachos, V. H., Montenegro, L. C. & Bolfarine, H. (2008). Inference and influence diagnostics for skew-normal null intercept measurement errors models. *Journal of Statistical Computation and Simulation*, **78**. 395-419.
- Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**. SS-08-045.
- Lange, K. L. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**, 175–198.
- Lange, K. L., Little, R. & Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, **84**, 881–896.
- Laurent, R. T. & Cook, R. D. (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika*, **80**, 99–106.
- Lemonte, A. J. & Cordeiro, G. M. (2009). Birnbaum-saunders nonlinear regression models. *Computational Statistics & Data Analysis*, **53**, 4441–4452.

- Lin, J. & Wei, B. (2003). Testing for heteroscedasticity in nonlinear regression models. *Communications in statistics. Theory and methods*, **32**(1), 171–192.
- Lin, J., Xie, F. & Wei, B. (2009). Statistical Diagnostics for Skew-t-Normal Nonlinear Models. *Communications in Statistics-Simulation and Computation*, **38**(10), 2096–2110.
- Lin, T. I., Lee, J. C. & Hsieh, W. J. (2007). Robust mixture modelling using the skew t distribution. *Statistics and Computing*, **17**, 81–92.
- Meng, X. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **81**, 633–648.
- Montenegro, L. C., Bolfarine, H. & Lachos, V. (2009). Local influence analysis of skew-normal linear mixed models. *Communication in Statistics–Theory and Methods*, **38**, 484–496.
- Montenegro, L. C., Bolfarine, H. & Lachos, V. (2009a). Local influence analysis of skew-normal linear mixed models. *Communication in Statistics–Theory and Methods*, **38**, 484–496.
- 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.
- Osorio, F. (2006). *Diagnóstico de influência em modelos elípticos com efeitos mistos*. Tese de doutorado, Departamento de Estatística, IME-USP.
- Osorio, F., Paula, G. A. & Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**, 4354–4368.
- Poon, W. & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **61**, 51–61.
- Simonoff, J. S. & Tsai, C.-L. (1994). Use of modified profile likelihood for improved tests of constancy of variance in regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **43**(2), 357–370.
- Thomas, W. & Cook, R. D. (1990). Assessing influence on prediction from generalized linear models. *Technometrics*, **32**, 59–65. Preprint.

- Vanegas, L. H. & Cysneiros, J. A. (2009). Assesment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*, **54**, 1002–1016.
- Wang, J. & Genton, M. G. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209–220.
- Xie, F. C., Lin, J. G. & Wei, B. C. (2009a). Diagnostics for skew-normal nonlinear regression models with ar(1) errors. *Computational Statistics & Data Analysis*, (**In press**).
- Xie, F. C., Wei, B. C. & Lin, J. G. (2009b). Homogeneity diagnostics for skew-normal nonlinear regression models. *Statistics & Probability Letters*, **79**, 821–827.
- Zeller, C. B. (2009). *Distribuição de Misturas Escala Skew Normal: Estimação e Diagnóstico em Modelos Lineares*. Tese de doutorado, Departamento de Estatística, IMECC-UNICAMP.
- Zeller, C. B., Labra, F. V. & Lachos, V. H. ; BALAKRISHNAN, N. (2010). Influence analyses of skew-normal/independent linear mixed models. *Computational Statistics & Data Analysis*, **54**, 1266–1280.