RIGNALDO RODRIGUES CARVALHO

DIAGNÓSTICO DE INFLUÊNCIA EM MODELOS COM ERROS NA VARIÁVEL SKEW-NORMAL/INDEPENDENTE

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, da Univeridade Estadual de Campinas para a obtenção do Título de Mestre em Estatística.

Orientador: Prof. Dr. Victor Hugo Lachos Dávila
Co-orientador: Prof. Dr. Filidor E. Vilca Labra

CAMPINAS
2010

DIAGNÓSTICO DE INFLUÊNCIA EM MODELOS COM ERROS NA VARIÁVEL SKEW-NORMAL/INDEPENDENTE

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Rignaldo Rodrigues Carvalho e aprovada pela comissão julgadora.

Campinas, 16 de dezembro de 2010

Prof. Dr: Victor Hugo Lachos Dávila
Orientador

Prof. Dr: Filidor Edilfonso Vilca Labra
Co-orientador

Banca Examinadora:

1 Prof. Dr. Filidor Edilfonso Vilca Labra (IMECC-UNICAMP)
2 Prof. Dr. Caio Lucidius Naberezny Azevedo (IMECC-UNICAMP)
3 Prof. Dr. Heleno Bolfarine (IME-USP)

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

Carvalho, Rignaldo Rodrigues
C253d Diagnóstico de influência em modelos com erros na variável skew-normal/independente/Rignaldo Rodrigues Carvalho-- Campinas, [S.P. : s.n.], 2010.

Orientadores : Victor Hugo Lachos Dávila; Filidor Edilfonso Vilca Labra

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

1.Algoritmos de expectativa de maximização. 2.Distribuição normal assimétrica. 3.Influência local. 4.Misturas de escala. 5.Distância de Mahalanobis. I. Lachos Dávila, Victor Hugo. II.Labra, Filidor Edilfonso Vilca. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Título em inglês: Influence of diagnostic in models with errors in variable skew-normal/independent

Palavras-chave em inglês (Keywords): 1. EM algorithms. 2. Skew normal distributions. 3. Local influence. 4. Scale mixtures. 5. Mahalanobis distance.

Área de concentração: Métodos Estatísticos

Titulação: Mestre em Estatística

Banca examinadora: Prof. Dr. Filidor Edilfonso Vilca Labra (IMECC – UNICAMP)
Prof. Dr. Caio Lucidius Naberezny Azevedo (IMECC – UNICAMP)
Prof. Dr. Heleno Bolfarine (IME - USP)

Data da defesa: 16/12/2010

Programa de Pós-Graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 16 de dezembro de 2010 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

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

Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO

Prof(a). Dr(a). HELENO BOLFARINE

Agradecimentos

Agradeço à minha família: minha mãe Zilda, meu pai Reginaldo por todo apoio que sempre me deram. Agradeço também à minha namorada, Erica Mantovanello, pela compreensão e toda a paciência durante minhas eternas revisões desse trabalho. A conquista desse título só foi possível graças a eles.

Ao meu orientador Prof. Dr. Victor Hugo Lachos Dávila e ao meu co-orientador Prof. Dr. Filidor E. Vilca Labra, por todo o conhecimento e experiência transmitidos durante a relização da dissertação e pela confiança depositada em mim. A Prof. Dra. Camila Borelli Zeller que foi uma pessoa fundamental na conclusão desse trabalho. Ao Prof. Dr.José Marcos Pinto da Cunha do NEPO-Unicamp pelo apoio que me foi dado e pela amizade até hoje. E a todos os professores do Departamento de Estatística pelos ensinamentos concedidos.

A toda banca examinadora, por terem aceito o convite e por todas sugestões e correções. Aos meus amigos Julio César e Daniel Nascimento pela convivência harmônica de 3 anos, ao meu amigo André pelo apoio quando eu nem pensava ser capaz de chegar tão longe, ao meu amigo Rodrigo Manfredini pelas discussões estatísticas intermináveis sobre nossos estudos e a todos os meus amigos que aqui não foram citados mas que de alguma forma fizeram ou fazem parte da minha vida, agradeço a todos por todo apoio dado durante a realização do curso.

“ *No futuro, o pensamento estatístico*
será tão necessário para a cidadania eficiente
como saber ler e escrever.”
Herbert George Wells

“ *Uma resposta aproximada da questão certa*
é mais valiosa do que uma resposta certa
de um problema aproximado.”
John Tukey

“ *Para entender as ideias de Deus,*
precisamos estudar estatística,
porque essa é a medida de Seu propósito.”
Florence Nightingale

Resumo

O modelo de medição de Barnett é frequentemente usado para comparar vários instrumentos de medição. É comum assumir que os termos aleatórios têm uma distribuição normal. Entretanto, tal suposição faz a inferência vulnerável a observações atípicas por outro lado distribuições de misturas de escala skew-normal tem sido uma interessante alternativa para produzir estimativas robustas tendo a elegância e simplicidade da teoria da máxima verossimilhança. Nós usamos resultados de Lachos et al. (2008) para obter a estimação dos parâmetros via máxima verossimilhança, baseada no algoritmo EM, o qual rende expressões de forma fechada para as equações no passo M. Em seguida desenvolvemos o método de influência local de Zhu e Lee (2001) para avaliar os aspectos de estimação dos parâmetros sob alguns esquemas de perturbação. Os resultados obtidos são aplicados a conjuntos de dados bastante estudados na literatura, ilustrando a utilidade da metodologia proposta.

Palavras-Chave: *Distribuições Misturas de Escala Skew-Normal; Algoritmo EM; Modelo de Barnett; Distância de Mahalanobis; Influência Local.*

Abstract

The Barnett measurement model is frequently used to comparing several measuring devices. It is common to assume that the random terms have a normal distribution. However, such assumption makes the inference vulnerable to outlying observations whereas scale mixtures of skew-normal distributions have been an interesting alternative to produce robust estimates keeping the elegancy and simplicity of the maximum likelihood theory. We used results in Lachos et al. (2008) for obtaining parameter estimation via maximum likelihood, based on the EM-algorithm, which yields closed form expressions for the equations in the M-step. Then we developed the local influence method to assessing the robustness aspects of these parameter estimates under some usual perturbation schemes. Results obtained for one real data set are reported, illustrating the usefulness of the proposed methodology.

Key-Words: *Scale Mixtures of Skew-Normal Distribution; EM Algorithm; Barnett Model; Mahalanobis Distance; Local Influence.*

Sumário

Lista de Figuras

Lista de Tabelas

1 *Introdução*

1.1 Modelos com Erros de Medição na Variável

A regressão linear é uma das ferramentas estatísticas mais amplamente utilizadas e tem sido assunto para extensivas pesquisas na literatura por mais de um século. Em várias aplicações a variável exploratória (x) não é observada diretamente, ou seja, observa-se $X = x + u$, onde u é a medida de erro. Essa medição com erros pode gerar um ruído na estimação dos parâmetros se o modelo não levar em consideração esse erro. O modelo de erro de medição (MEM) é útil em várias áreas, incluindo modelos de regressão com erros lineares e erros não-lineares, modelos de análise fatorial, modelo estrutural latente, modelos de equações simultâneas e problemas de comparação de instrumentos (veja, Barnett, 1969, Theobald e Mallison, 1978, Shyr e Gleser, 1986, Bolfarine e Galea-Rojas, 1995 e Chipkevitch et al., 1996). No Capítulo 5 apresentamos 2 casos de problemas de comparação de intrumentos.

Solano (2000) e Lachos (2002) estudam o problema de comparação de instrumentos sob o modelo de comparação comparativa e o modelo de Grubbs respectivamente. A comparação de instrumentos é necessária devido a variação de preço, tempo gasto para medição e outras características, tal como eficiência. No modelo de Barnett (1969), frequentemente usado para comparar instrumentos ou métodos de medição é comumente assumido que as observações seguem uma distribuição normal. Porém é bem conhecido que tal suposição faz a inferência vulnerável na presença de observações atípicas. Uma alternativa é considerar distribuições com caudas mais pesadas que a distribuição normal, como por exemplo as distribuições t-student, normal contaminada, slash entre outras que são membros da família misturas de escala normal (MEN). Para mais detalhes desta classe de distribuição veja Lange e Sinsheimer (1993).

Embora os modelos com erros nas variáveis sob as distribuições MEN representem um interessante extensão robusta, nem sempre é a melhor alternativa quando a distribuição da variável explicativa x tem distribuição assimétrica.

Recentemente, Lachos et al. (2008) prôpos uma solução para acomodar assimetria e caudas pesadas simultaneamente usando a misturas de escala skew-normal (MESN). Essa extensão resulta em uma classe flexível para modelos multivariados robustos com erros de medição que contém como elementos próprios, a distribuição skew-normal (SN), a skew-t (ST), a skew-slash (SSL) e a distribuição skew-normal contaminada (SCN), todas essas distribuições têm caudas mais pesadas que a normal, e então podem ser usadas para se obter inferências robustas em vários tipos de modelos.

Esse trabalho é motivado pelo fato que alguns conjuntos de dados considerados na literatura apresentam comportamento de não normalidade, tal como assimetria e caudas pesadas.

Segundo Branco e Dey (2001), um vetor aleatório p-dimensional $\mathbf{Y}$ tem distribuição MESN se

$$\mathbf{Y} = \boldsymbol{\mu} + k^{1/2}(U)\mathbf{Z}, \tag{1.1}$$

em que $\boldsymbol{\mu}$ é o vetor de locação, $\mathbf{Z}$ é um vetor aleatório normal assimétrico (Azzalini e Dalla-Valle, 1996), com vetor locação $\mathbf{0}$, matriz escala $\boldsymbol{\Sigma}$ e vetor de parâmetros de assimetria $\boldsymbol{\lambda}$. Considerando a notação usual, nós escrevemos $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. $k(\cdot)$ é uma função de peso e U é uma variável aleatória mista positiva com função distribuição acumulada (fda) $H(u|\boldsymbol{\nu})$ e função densidade de probabilidade (fdp) $h(u|\boldsymbol{\nu})$, independente de $\mathbf{Z}$, onde $\boldsymbol{\nu}$ é um escalar ou vetor de parâmetros indexando a distribuição de U. Dado U, $\mathbf{Y}$ segue a distribuição skew-normal multivariada com vetor $\boldsymbol{\mu}$ de locação, matriz escala $k(u)\boldsymbol{\Sigma}$ e vetor de parâmetros de assimetria $\boldsymbol{\lambda}$, i.e., $\mathbf{Y}|U = u \sim \mathrm{SN}_p(\boldsymbol{\mu}, k(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Então a fdp de $\mathbf{Y}$ é dada por

$$f(\mathbf{y}) = 2\int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, k(u)\boldsymbol{\Sigma})\Phi(k^{-1/2}(u)A)dH(u), \tag{1.2}$$

em que $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $\phi_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ padrão para fdp da distribuição normal p-variada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, $\Phi(\cdot)$ representa a fda da distribuição normal univariada padrão. Um caso particular dessa distribuição é a distribuição skew-normal, para qual H é degenerada, com $k(u) = 1$, $u > 0$. Nós usamos a notação $\text{MESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ para dizer que $\mathbf{Y}$ tem distribuição com fdp (1.2). Nós notamos que quando $k(u) = u^{-1}$ em (1.1), a distribuição de $\mathbf{Y}$ se reduz a skew-normal/independente (SNI) discutida, por exemplo, em Ghosh et al.(2008). Nesse trabalho nós usaremos a classe de distribuições MESN para desenvolver um modelo de medição de Barnett mais flexível. Quando $\boldsymbol{\lambda} = \mathbf{0}$, a distribuição MESN se reduz para distribuição MEN a qual as propriedades são bem estudadas na literatura. Nós usaremos a notação $\mathbf{Y} \sim \text{MEN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$ quando $\mathbf{Y}$ pertencer a uma distribuição da classe misturas de escala normal MEN. Note que quando $\boldsymbol{\lambda} = \mathbf{0}$ e $k(u) = u^{-1}$ em (1.1), a distribuição de $\mathbf{Y}$ se reduz para a família de distribuições normal/independente (NI) discutida, por exemplo, em Lange e Sinsheimer (1993).

1.2 Diagnóstico

Uma análise de influência é um importante passo em análises de dados. Essa análise pode ser realizada através de uma análise de influência local, uma técnica estatística geral usada para avaliar a estabilidade da estimação das “saídas” com relação as “entradas” do modelo. “Entrada” de modelos pode incluir dados, parâmetros ou outras informações em relação ao modelo (exemplo, heterocedásticidade ou uma covariância medida com erro). “Saídas” de modelos pode incluir os parâmetros estimados, funções, valores, objetivo final, estimação de resíduos, etc. Seguindo o trabalho pioneiro de Cook (1986) essa área de pesquisa tem recebido considerável atenção na recente literatura estatística: veja Lachos et al. (2007b), De Castro et al. (2008), entre outros. Lachos et al. (2008) apresenta um estudo de estimação no modelo com erros nas variáveis sob a misturas de escala da skew-normal. A aplicação da metodologia de Cook (1986) neste modelo envolve extensas manipulações algébricas, pois esta depende das primeiras e segundas derivadas da log-verossimilhança sob a distribuição MESN que em geral é muito complexa. Recentemente Zhu e Lee (2001) desenvolveram um método para a análise de influência local para modelos de estatística com dados perdidos (*missings*, do inglês) através da utilização da função Q-afastamento fechada relacionada a esperança condicional da log-verossimilhança para dados completos no passo E do algoritmo EM. Inpirados por trabalhos anteriores de Osório et al. (2009) e Montenegro et al. (2009b), que estudaram o método de influência local no contexto de misturas de escala do modelo de Grubbs normal e modelo de Grubbs skew-normal, respectivamente, estudaremos o método de influência local de Zhu e Lee (2001) para a classe de modelos MESN-MEM.

1.3 Objetivo do Trabalho

O objetivo deste trabalho é realizar um estudo de estimação e diagnóstico de influência em modelos de regressão com erros na variável considerando que as observações seguem uma distribuição dentro da classe de distribuições de misturas de escala skew-normal. Para o processo de estimação de máxima verossimilhança nós usamos o algoritmo-EM. Valendo-se da esperança condicional da função de verossimilhança completa derivamos medidas de diagnóstico baseadas no método de influência local proposto por Zhu e Lee (2001). Vale a pena ressaltar que os resultados deste projeto visam contribuir positivamente para o desenvolvimento nesta área de pesquisa, apontando novos resultados em modelos de interesse prático e estendendo resultados recentemente encontrados, por exemplo, em Arellano-Valle et al. (2005b), Osorio (2006) e Lachos et al. (2007a). Assim os objetivos específicos deste trabalho podem ser resumidos como segue:

1. Apresentar um estudo de estimação e diagnóstico, no modelo com erros nas variáveis multivariado (Barnett, 1969).

2. Aplicar o método de influência local proposto por Zhu e Lee (2001) no modelo com erros de medição, considerando alguns esquemas de perturbação.

1.4 Organização do Trabalho

Este trabalho é composto por 6 capítulos e 5 apêndices. No Capítulo 2 apresentamos uma breve revisão da classe de distribuições de misturas de escala skew-normal proposta por Ghosh et al. (2008). Um destaque especial para as propriedades aqui apresentadas, que serão utéis em todo o estudo, além dos exemplos de distribuições pertencentes a essa classe e da estimação de máxima verossimilhança via algoritmo EM.

No Capítulo 3 apresentamos o modelo com erros nas variáveis multivariado. Uma revisão sobre variável latente, estimação dos parâmetros via algoritmo EM e obtenção da matriz observada são os assuntos tratados em detalhes nesse capítulo.

No Capítulo 4 tratamos da aplicação do método de influência local proposto por Zhu e Lee (2001). Consideramos 3 tipós de perturbação: ponderação de casos, perturbação aditiva na variável resposta e perturbação multiplicativa na variável resposta. Uma revisão com outros métodos de estudo de influência local, como o de Cook (1986), também são apresentados ao leitor.

O Capítulo 5 apresenta duas aplicações dos resultados obtidos com conjuntos de dados reais encontrados na literatura.

O Capítulo 6 é dedicado à conclusões e alguns direcionamentos para estudos subsequentes.

Finalmente, nos apêndices apresentamos alguns lemas utilizados nesse trabalho, detalhes sobre a matriz de informação observada, um programa de implementação no software Matlab R2007a utilizado para obter os estimadores de máxima verossimilhança para os parâmetros do modelo, os critérios de seleção de modelos, além dos conjuntos de dados utilizados no capítulo 6.

Quando mencinarmos manipulações algébricas no decorrer da dissertação, estamos referindo à manipulações algébricas desenvolvidas “manualmente” e conferidas através do software Matlab R2007a.

2 Distribuições Misturas de Escala Skew-Normal

2.1 Introdução

Existe uma tendência geral na literatura estatística de encontrar modelos flexíveis que representem características dos dados tão adequadas quanto possível, reduzindo suposições não realistas. A motivação deve-se ao crescente poder computacional e, também, a inúmeros conjuntos de dados em distintos campos da ciência, que frequentemente não satisfazem algumas suposições, tais como independência e normalidade. Para modelar desvios destas suposições e particularmente da suposição de normalidade, que desempenha um papel importante na análise estatística, diferentes enfoques podem ser encontrados na literatura. Do ponto de vista prático, o método mais comum adotado para alcançar a normalidade (ou simetria) é a transformação de variáveis. Embora tal método possa dar resultados empíricos razoáveis, este deve ser evitado se um modelo mais conveveniente possa ser encontrado. Azzalini e Capitanio (1999) apresentam algumas razões para evitar este procedimento:

a) A transformação não fornece informação útil para entender o mecanismo de gerações dos dados;

b) A transformação de variáveis dificulta a interpretação, especialmente quando cada variável é transformada usando diferentes funções;

c) A transformação para um conjunto de dados pode frequentemente não ser aplicável para outros conjuntos de dados;

d) Quando a suposição de homocedasticidade é necessária, algumas vezes a transformação requerida difere da transformação para alcançar a normalidade.

Um enfoque alternativo à transformação de variáveis para modelagem de dados de um ponto de vista paramétrico e apropriado para o tratamento de observações contínuas multivariadas consiste em construir classes paramétricas flexíveis de distribuições multivariadas que exibam assimetria e curtose diferentes da distribuição normal. A classe de distribuições elípticas é provavelmente o exemplo mais conhecido deste enfoque, incluindo um vasto conjunto de distribuições, como por exemplo, **normal, normal contaminada, slash, t de Student, exponencial potência**, entre outros, com a principal vantagem de representar uma extensão natural do conceito de simetria no contexto multivariado.

A classe de distribuições de misturas de escala skew-normal (MESN), proposta por Ghosh et al. (2008), pode ser utilizada como uma alternativa para inferência robusta em vários tipos de modelos. Esta classe é construída a partir da mistura de uma v.a. skew-normal e uma v.a. positiva, esta última indexada por um parâmetro (ou vetor de parâmetros) adicional $\boldsymbol{\nu}$ que regula o comportamento das caudas da distribuição. A principal virtude das distribuições MESN é que as mesmas possuem uma representação estocástica que permite uma fácil implementação do algoritmo EM quando aplicados a modelos de interesse prático (Lachos et al., 2007b).

Neste capítulo será proposta uma revisão sobre algumas das propriedades e resultados referêntes à distribuição skew-normal multivariada. Essa distribuição será utilizada nos demais capítulos para compor uma nova classe de distribuições de misturas de escala skew-normal (MESN), classe essa menos sensível a valores extremos. a Figura 1 apresenta o comportamento da distribuição skew-normal univariada de acordo com o paraâmetro λ.

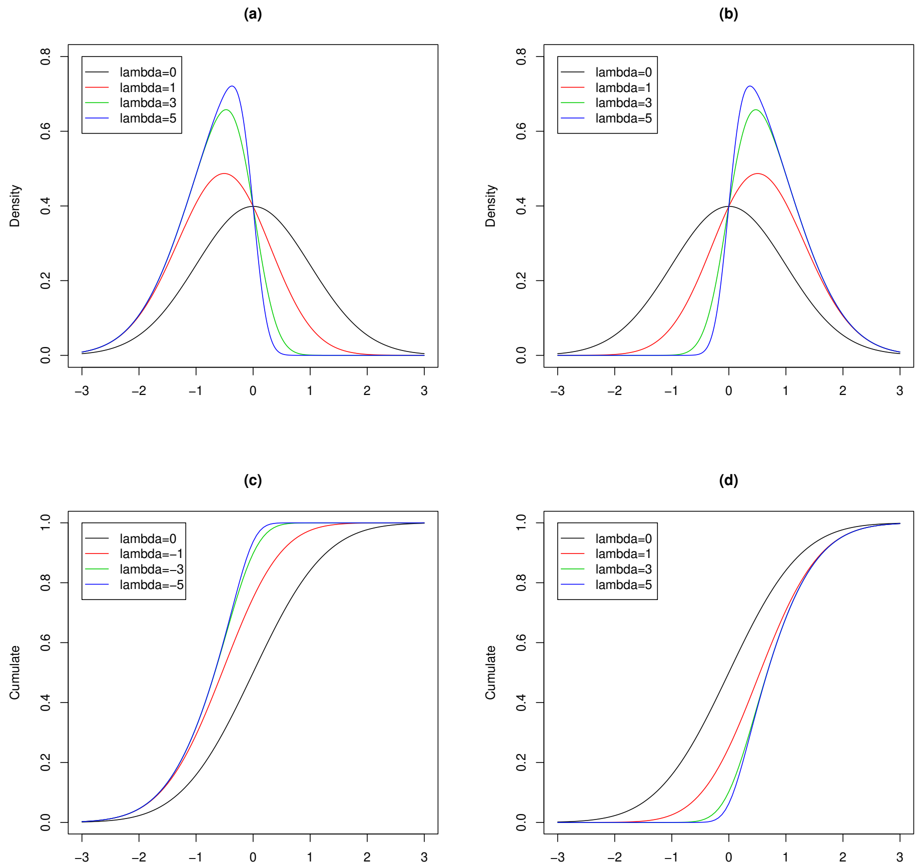

Figura 1: Função densidade de probabilidade e função distribuição acumulada da skew-normal padrão. (a) Função distribuição de probabilidade da SN com λ negativo (b) Função distribuição de probabilidade da SN com λ positivo (c) Função distribuição acumuladada da SN com λ negativo (d) Função distribuição acumuladada da SN com λ positivo.

2.2 A Distribuição Skew-Normal Multivariada

Muitas definições da distribuição skew-normal multivariada são encontradas na literatura, com diferentes parametrizações e interpretações. Nesta seção, será considerada uma definição unificada das definições descritas em Arellano-Valle et al. (2005a). A escolha dessa definição para esse trabalho se deve ao fato de que muitas propriedades, caracterizações e manipulações algébricas são obtidas facilmente a partir desta e, além disso, com essa definição o caso univariado pode ser visto como um caso particular derivado da representação multivariada, o que pode não ocorrer com outras definições dessa dis-

tribuição, fazendo-se necessária algumas reparametrizações. A seguir, será introduzida uma notação que será utilizada ao decorrer desse trabalho.
Sejam $\phi_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\Phi_p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a função distribuição de probabilidade (fdp) e a função distribuição acumulada (fda), respectivamente, da distribuição $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ avaliada em $\mathbf{x}$. Quando $\boldsymbol{\mu} = 0$ e $\boldsymbol{\Sigma} = \mathbb{I}_p$ (a matriz identidade $p \times p$), denotaremos essas funções como $\phi_p(\mathbf{x})$ e $\Phi_p(\mathbf{x})$.

2.2.1 Função Densidade de Probabilidade e Função de Distribuição Acumulada

Definição 1. *Um vetor aleatório p-dimensional $\mathbf{Y}$ segue uma distribuição skew-normal com vetor de locação $\boldsymbol{\mu} \in \mathbb{R}^p$, matriz escala $\boldsymbol{\Sigma}$ ($p \times p$ e definida positiva) e vetor de assimetria $\boldsymbol{\lambda} \in \mathbb{R}^p$, se sua fdp é dada por*

$$f_Y(\mathbf{y}) = 2\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^p, \tag{2.1}$$

a qual nós podemos denotar por $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Quando $\boldsymbol{\mu} = 0$ e $\boldsymbol{\Sigma} = \mathbb{I}_p$, nós denotamos por $\mathbf{Y} \sim SN_P(\boldsymbol{\lambda})$. A representação estocástica para a distribuição em (2.1) é dada por

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\delta}|X_0| + (\mathbb{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{X}_1), \quad com \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}}, \tag{2.2}$$

em que $X_0 \sim N_1(0, 1)$ independente de $\mathbf{X}_1 \sim N_p(\mathbf{0}, \mathbb{I}_p)$ e "$\stackrel{d}{=}$" significa "distribuído como".

Essa representação é uma generalização da representação univariada de Henze (1986). Como no caso univariado, várias propriedades da distribuição multivariada em (2.2) podem ser derivadas de sua representação estocástica. Para uma leitura complementar sobre o assunto, sugiro ao leitor que veja Arellano-Valle e Genton (2005) e Arellano-Valle et al. (2005a).

2.2.2 Propriedades

A seguir é apresentado ao leitor algumas propriedades da distribuição skew-normal multivariada que serão utéis no decorrer desse trabalho. A prova dos resultados pode ser encontrada em Lachos (2004).

Proposição 1. *Seja* $\mathbf{Z} \sim SN_p(\boldsymbol{\lambda})$ *e considere a transformação linear* $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$*, em que* $\boldsymbol{\mu} \in \mathbb{R}^p$ *e* $\boldsymbol{\Sigma}$ *é uma matriz* $p \times p$ *definida positiva. Então*

$$\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}).$$

Demonstração. A prova segue do fato que $f_Y(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} f_Z(\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))$. □

No próximo resultado será apresentada a função geratriz de momentos (fgm) da distribuição skew-normal multivariada padronizada. Propriedades adicionais para essa distribuição serão obtidas a partir dessa.

Proposição 2. *Seja* $\mathbf{Z} \sim SN_p(\boldsymbol{\lambda})$*. Então sua fgm é dada por*

$$M_Z(s) = 2e^{\frac{1}{2}s^\top s}\Phi_1(\delta s), \quad s \in \mathbb{R}^p.$$

Demonstração. Note que $e^{s^\top \mathbf{z}}\phi_p(z) = e^{\frac{1}{2}s^\top s}\phi_p(\mathbf{z} - s)$, então

$$\begin{aligned} M_Z(s) &= E(e^{s^\top \mathbf{z}}) = 2\int_{\mathbb{R}} e^{\frac{1}{2}s^\top s}\phi_p(\mathbf{u} - s)\Phi_1(\boldsymbol{\lambda}\mathbf{u})d\mathbf{u} \\ &= 2e^{\frac{1}{2}s^\top s}\int_{\mathbb{R}} \phi_p(\mathbf{y})\Phi_1(\boldsymbol{\lambda}^\top(\mathbf{y} + s))d\mathbf{y} = 2e^{\frac{1}{2}s^\top s}\int_{\mathbb{R}}\int_{u \le \boldsymbol{\lambda}^\top s} \phi_p(\mathbf{y})\phi_1(u + \boldsymbol{\lambda}^\top \mathbf{y})d\mathbf{y}du \\ &= 2e^{\frac{1}{2}s^\top s}\int_{u \le \boldsymbol{\lambda}^\top s} E[\phi_1(u + \boldsymbol{\lambda}^\top \mathbf{Y})]du, \quad \text{em que } \ \mathbf{Y} \sim \mathrm{N}_p(0, \mathbb{I}_p) \end{aligned}$$

o resultado segue usando o Lema 1 que se encontra no Apêndice A. □

Como consequência imediata da proposição anterior, temos o seguinte Corolário

Corolário 1. *Seja* $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$*. Então sua fgm é dada por*

$$M_Y(\mathbf{s}) = 2e^{\mathbf{s}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{s}^\top \boldsymbol{\Sigma} s}\Phi_1(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^p. \tag{2.3}$$

Demonstração. A prova segue de $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}$ e $M_{\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}}(\mathbf{s}) = e^{\mathbf{s}^\top \boldsymbol{\mu}} M_Z(\boldsymbol{\Sigma}^{1/2}\mathbf{s})$. □

As proposições apresentadas a seguir são resultados relacionados com os momentos de um vetor aleatório skew-normal, cujas demonstrações podem ser obtidas a partir de sua fgm.

Proposição 3. *Seja* $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. *Então*

a) $E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$,

b) $E[\mathbf{Y}\mathbf{Y}^\top] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top$,

c) $Var(\mathbf{Y}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, *em que* $\Delta = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$.

A seguir apresentamos a classe das distribuições de misturas de escala skew-normal (MESN).

2.3 A Classe de Distribuições de Misturas de Escala Skew-Normal

A família de distribuições norml assimétrica (Azzalini, 1985), que contém como caso particular a distribuição normal, tem particular importância já que pode se adaptar à distribuições que estão em uma vizinhança de uma normal, isto é, que não tenham uma assimetria bem definida. Mesmo sendo atrativa, a skew-normal ainda não parece ser a mais adequada para análise de dados com a valores extremos, já que pode ser mostrado que a estimação dos parâmetros pode estar comprometida em virtude disso.

Ainda no contexto de modelos simétricos, a distribuição *t-student* torna-se uma alternativa à distribuição normal para lidar com valores extremos, já que apresenta caudas mais pesadas que a normal, podendo proporcionar ajustes mais robustos. Isso por que a *t-student* apresenta um atrativo adicional: um parâmetro extra que pode ser entendido como o responsável pela acomodação de valores extremos.

Neste sentido, a classe de distribuições de misturas de escala normal (MEN) pode ser vista como mais abrangente que o modelo *t-student*, propondo uma nova classe de distribuições mais robusta quanto a valores extremos. Portanto, assim como a classe MEN

procura tratar dados com valores extremos no contexto simétrico, a classe de distribuições propostas neste capítulo procura tratar dados com valores extremos no caso assimétrico. Sob essa motivação é que será apresentada a classe de modelos de misturas de escala skew-normal (MESN).
Neste capítulo será proposta a classe de distribuições MESN multivariada, sua representação estocástica e algumas de suas propriedades. Em seguida, serão apresentados alguns exemplos de distribuições pertencentes a essa classe. Ao final do capítulo será apresentado o algoritmo EM para estimação de máxima verossimilhança dos parâmetros em modelos pertencentes a essa classe.

2.4 A Distribuição MESN Multivariada

Nessa seção, nós definimos a distribuição de misturas de escala skew-normal e estudamos algumas de suas importantes propriedades, transformações lineares, momentos e distribuições marginais e condicionais.

Definição 2. *Um vetor aleatório p-dimensional* $\mathbf{Y}$ *é dito ter distribuição na classe MESN com parâmetro de locação* $\boldsymbol{\mu} \in \mathbb{R}^p$*, matriz de escala* $\boldsymbol{\Sigma}_{p\times p}$ *(positiva definida) e parâmetro de assimetria* $\boldsymbol{\lambda} \in \mathbb{R}^p$ *se sua fdp é dada por*

$$f(\mathbf{y}) = 2\int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})\Phi_1(\kappa^{-1/2}(u)A)dH(u), \tag{2.4}$$

em que $A = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$*,* U *uma variável aleatória positiva com fda* $H(\cdot|\boldsymbol{\nu})$ *e fdp* $h(\cdot|\boldsymbol{\nu})$*, e* $\kappa(\cdot)$ *uma função peso positiva bem definida. Usamos a notação* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *para representar um vetor aleatório com fdp dada em (2.4). Quando* $\boldsymbol{\mu} = \mathbf{0}$ *e* $\boldsymbol{\Sigma} = \mathbb{I}_p$ *temos a distribuição MESN padrão denotada por* $MESN_p(\boldsymbol{\lambda}; H)$*.*

Aqui, $\boldsymbol{\nu}$ é um parâmetro escalar ou vetorial indexando a distribuição do fator de escala U. Esse parâmetro adicional pode ser entendido por um fator de acomodação de valores extremos. Se a função de distribuição acumulada $H(u|\boldsymbol{\nu})$ converge fracamente para a distribuição de massa pontual em 1, então pode-se afirmar que a densidade em (2.4) converge para a densidade de um vetor aleatório com distribuição skew-normal. A prova desse resultado é similar àquela encontrada em Lange e Sinsheimer (1993). Quando $\boldsymbol{\lambda} = 0$, obtém-se a classe de distribuições de misturas de escala normal MEN representada pela função de distribução de probabilidade $f_0(\mathbf{y}) = \int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})dH(u|\boldsymbol{\nu})$. Além disso, quando $\kappa(u) = u^{-1}$ obtemos a distribuição normal/independente (NI) e quando $\kappa(u) = u^{-1}$ em (2.4), nós obtemos uma extensão da distribuição proposta por Lange e Sinsheimer (1993) que nós chamamos de skew-normal/independente (SNI). As Figuras 2, 3, 4 e 5 apresentam gráficos tridimensionais e gráficos de contorno das distribuições SN, ST, SSL e SCN bivariadas.

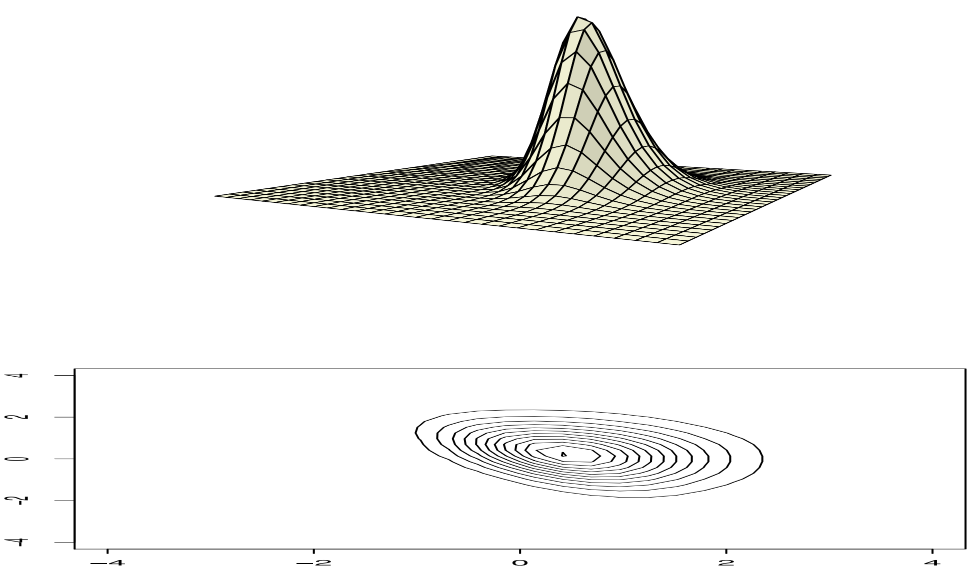

Figura 2: Gráficos da função densidade de probabilidade da skew-normal bivariada padrão, $\mathrm{SN}_2(\boldsymbol{\lambda})$, em que $\boldsymbol{\lambda} = (2, 1)^\top$

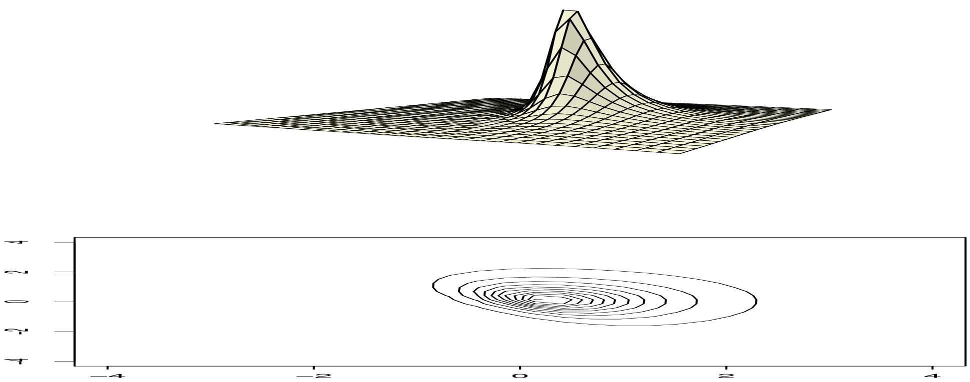

Figura 3: Gráficos da função densidade de probabilidade da skew-t bivariada padrão, $\mathrm{ST}_2(\boldsymbol{\lambda}, 2)$, em que $\boldsymbol{\lambda} = (2, 1)^\top$

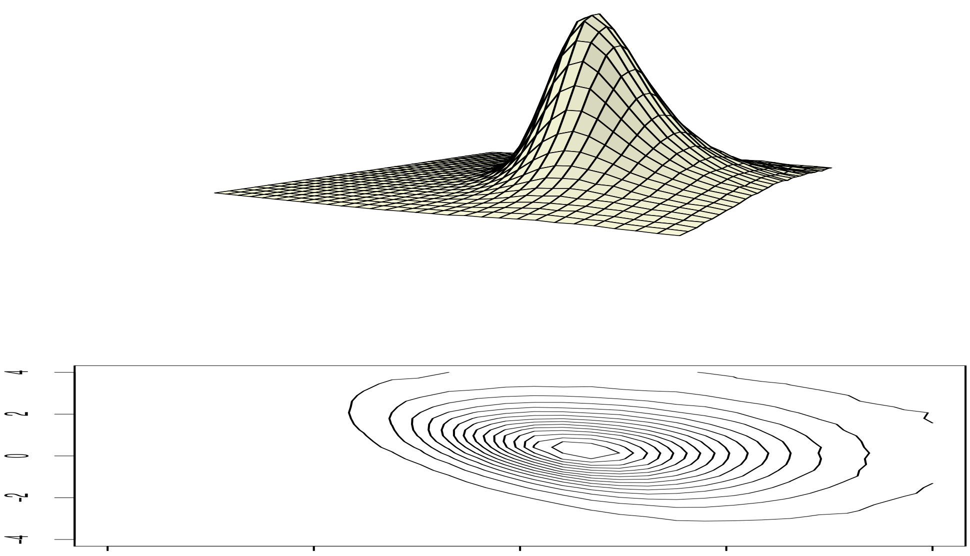

Figura 4: Gráficos da função densidade de probabilidade da skew-slash bivariada padrão, $\mathrm{SSL}_2(\boldsymbol{\lambda}, 1)$, em que $\boldsymbol{\lambda} = (2, 1)^\top$

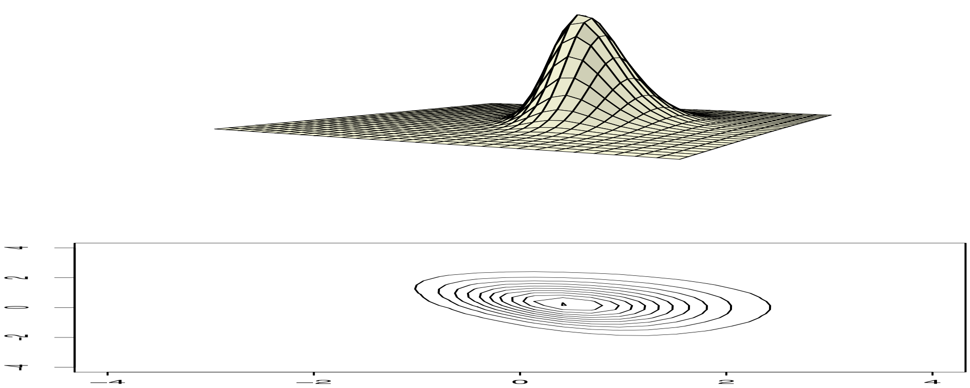

Figura 5: Gráficos da função densidade de probabilidade da skew-normal contaminada bivariada padrão, $\mathrm{SCN}_2(\boldsymbol{\lambda}, 0.5, 0.5)$, em que $\boldsymbol{\lambda} = (2, 1)^\top$

2.4.1 Representação Estocástica

A representação estocástica dada abaixo pode ser utilizada para gerar números pseudo aleatórios de um vetor aleatório $\mathbf{Y} \sim \mathrm{MESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ e também para estudar muitas de suas propriedades.

Proposição 4. *Um vetor aleatório* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *tem representação estocástica dada por*

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \kappa(U)^{1/2}\mathbf{Z}, \tag{2.5}$$

em que $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ *e* U *é uma variável aleatória positiva (com fda* H*) independente de* $\mathbf{Z}$*.*

Demonstração. A prova segue do fato que $\mathbf{Y}|U = u \sim \mathrm{SN}_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. $\square$

A seguinte proposição mostra outra forma de se representar estocasticamente um vetor aleatório com distribuição na classe MESN. Esse resultado será utilizado posteriormente para representar o modelo de misturas de escala skew-normal hierarquicamente. Tal representação é de extrema importância para a implementação do algoritmo EM de forma bastante geral.

Proposição 5. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Então sua representação estocástica pode ser dada por*

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Delta} T + \kappa^{1/2}(u)\boldsymbol{\Gamma}^{1/2}\mathbf{T}_1, \tag{2.6}$$

em que $T = \kappa^{1/2}(u)|T_0|$*,* $T_0 \sim N(0,1)$*, independente de* $\mathbf{T}_1 \sim N_p(\mathbf{0}, \mathbb{I}_p)$*,* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$*,* $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$*.*

Demonstração. A prova segue da Proposição 4 e da representação estocástica de um vetor aleatório com distribuição skew-normal. □

2.4.2 Propriedades

A próxima proposição apresenta a função geradora de momentos (fgm) de um vetor aleatório com distribuição MESN.

Proposição 6. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Então sua fgm é dada por*

$$M_{\mathbf{y}}(\mathbf{s}) = E[e^{\mathbf{s}^\top \mathbf{Y}}] = \int_0^\infty 2e^{\mathbf{s}^\top \boldsymbol{\mu} + \frac{1}{2}\kappa(u)\mathbf{s}^\top \boldsymbol{\Sigma}\mathbf{s}}\Phi_1(\kappa^{1/2}(u)\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}\mathbf{s})dH(u), \quad \mathbf{s} \in \mathbb{R}^p. \tag{2.7}$$

Demonstração. Da prova da Proposição 4 tem-se $\mathbf{Y}|U = u \sim \mathrm{SN}_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Agora, de propriedades conhecidas de esperança condicional, segue que $M_{\mathbf{y}}(\mathbf{s}) = E_U[E[e^{\mathbf{s}^\top \mathbf{Y}}|U]]$ e conclui-se a prova utilizando o Corolário 1. □

Na proposição seguinte, será derivado o vetor de médias e a matriz escala de um vetor aleatório com distribuição MESN. A prova segue da representação estocástica (2.6).

Proposição 7. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Então*

a) Se $E[\kappa^{1/2}(U)] < \infty$*, então* $E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}k_1\boldsymbol{\Delta}$ *,*

b) Se $E[\kappa(U)] < \infty$*, então* $Var[\mathbf{Y}] = k_2\boldsymbol{\Sigma} - \frac{2}{\pi}k_1^2\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$*,*

com $k_m = E[\kappa^{m/2}(U)]$*, m = 1, 2 e* $\boldsymbol{\Delta}$ *como em (2.6).*

Na Seção 2.6 apresentamos o algoritmo EM para modelos MESN, tendo isto em vista, a seguir, apresentamos o cálculo de alguns momentos condicionais importantes para a implementação desse algoritmo para a classe de distribuições MESN.

Proposição 8. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *e* $\mathbf{U} \sim H$ *o fator de mistura de escala. Então*

$$\kappa_r = \frac{2f_0(\mathbf{y})}{f(\mathbf{y})} E[\kappa^{-r}(U_y)\Phi(\kappa^{-1/2}(U_y)A)] \quad e \quad \tau_r = \frac{2f_0(\mathbf{y})}{f(\mathbf{y})} E[\kappa^{-r/2}(U_y)\phi(\kappa^{-1/2}(U_y)A)] \quad (2.8)$$

em que $\kappa_r = E[\kappa^{-r}(U)|\mathbf{y}]$ *e* $\tau_r = E[\kappa^{-r/2}(U)W_\Phi(\kappa^{-1/2}(U)A)|\mathbf{y}]$, *com* $W_\Phi(x) = \phi_1(x)/\Phi(x)$, $x \in \mathbb{R}$ *e com* ψ_0 *a fdp de* $Y_0 \sim MEN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$ *e* $U_y \stackrel{d}{=} U|Y_0$.

Demonstração. Veja Proposição 1 em Lachos et al (2009). $\square$

2.4.3 Distribuição Marginal

A proposição a seguir mostra que qualquer vetor aleatório com distribuição na classe MESN é invariante sob transformações lineares. Em particular, a distribuição marginal desse vetor permanece na classe MESN.

Proposição 9. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Então para qualquer vetor fixado* $\mathbf{b} \in \mathbb{R}^m$ *e uma matriz* $\mathbf{A} \in \mathbb{R}^{m \times p}$ *de posto completo,*

$$\mathbf{Z} = \mathbf{b} + \mathbf{A}\mathbf{Y} \sim MESN_p(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \boldsymbol{\lambda}^*; H), \quad (2.9)$$

com $\boldsymbol{\lambda}^* = \boldsymbol{\delta}^*/(1 - \boldsymbol{\delta}^{*\top}\boldsymbol{\delta}^*)^{1/2}$, $\boldsymbol{\delta}^* = (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$.

Além disso, se $m = p$ a matriz $\mathbf{A}$ é não-singular, então $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}$. Também, para qualquer $\mathbf{a} \in \mathbb{R}^p$,

$$\mathbf{a}^\top\mathbf{Y} \sim \mathrm{MESN}_p(\mathbf{a}^\top\boldsymbol{\mu}, \mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}, \lambda^*; H),$$

com $\lambda^* = \alpha/(1 - \alpha^2)^{1/2}$, $\alpha = \{\mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})\}^{-1/2}\mathbf{a}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$.

Demonstração. A prova desse resultado é obtida diretamente da Proposição 6, já que $M_{\mathbf{b}+\mathbf{AY}}(\mathbf{s}) = e^{\mathbf{s}^\top\mathbf{b}} M_{\mathbf{Y}}(\mathbf{A}^\top\mathbf{s})$. Quando $\mathbf{A}$ é uma matriz não singular, é fácil ver que $\boldsymbol{\delta}^* = \boldsymbol{\delta}$. $\square$

Pela Proposição 9, com $\mathbf{A} = [\mathbb{I}_{p_1}, \mathbf{0}_{p_2}]$, $p_1 + p_2 = p$, tem-se o seguinte resultado para um vetor aleatório MESN, relacionado com sua distribuição marginal.

Corolário 2. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *e* $\mathbf{Y}$ *particionado como* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *de dimensões* p_1 *e* p_2*, respectivamente. Além disso, sejam*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad e \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

as correspondentes partições de $\boldsymbol{\mu}$ *e* $\boldsymbol{\Sigma}$*, a distribuição marginal de* $\mathbf{Y}_1$ *é dada por*

$$f(\mathbf{y}_i|\boldsymbol{\theta}) = 2\int_0^\infty \phi_p(\mathbf{y}_i|\boldsymbol{\mu}_1, \kappa(u_i)\boldsymbol{\Sigma}_{11})\Phi_1(\kappa^{-1/2}(u_i)A_i)dH(u_i), \tag{2.10}$$

em que $A_i = (\boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}})^\top\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_1)$*,* U *uma variável aleatória positiva com fda* $H(\cdot|\boldsymbol{\nu})$ *e fdp* $h(\cdot|\boldsymbol{\nu})$*, e* $\kappa(\cdot)$ *uma função peso bem definida, isto é,* $\mathbf{Y}_1 \sim MESN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}}; H)$*,* $i = 1, 2$*, com*

$$\widetilde{\boldsymbol{\upsilon}} = \frac{\boldsymbol{\upsilon}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}},$$

$\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$*,* $\boldsymbol{\upsilon} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda} = (\boldsymbol{\upsilon}_1^\top, \boldsymbol{\upsilon}_2^\top)^\top$*.*

A seguir apresentamos uma proposição e um corolário para a obtenção da distribuição condicional e em seguida da esperança condicinal de $\mathbf{Y}_2$ dado $\mathbf{Y}_1$.

Proposição 10. *Sob a notação do Corolário 2 e tomando* $k(u) = u^{-1}$*, se* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *então a distribuição de* $\mathbf{Y}_2$ *dado* $\mathbf{Y}_1 = \mathbf{y}_1$ *e* $U = u$*, tem densidade dada por*

$$f(\mathbf{y}_2|\mathbf{y}_1, u) = \phi_{p_2}(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{22.1})\frac{\Phi_1(u^{1/2}\boldsymbol{\upsilon}^\top(\mathbf{y} - \boldsymbol{\mu}))}{\Phi_1(u^{1/2}\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}, \tag{2.11}$$

com $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$*. Além disso*

$$E[\mathbf{Y}_2|\mathbf{y}_1, u] = \boldsymbol{\mu}_{2.1} + u^{-1/2}\frac{\phi_1(u^{1/2}\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}{\Phi_1(u^{1/2}\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}\frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}}. \tag{2.12}$$

Demonstração. De fato, a densidade de $f(\mathbf{y}_2|\mathbf{y}_1, u) = f(\mathbf{y}|u)/f(\mathbf{y}_1|u)$, e (2.11) e seguem observando que $\mathbf{Y}|U = u \sim \mathrm{SN}_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$ e $\mathbf{Y}_1|U = u \sim \mathrm{SN}(\boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{\upsilon}})$. O resultado (2.12) segue do Lema 2 dado no Apêndice A, com $A_y = \boldsymbol{\upsilon}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1) - \boldsymbol{\upsilon}_2^\top\boldsymbol{\mu}_2$, $\mathbf{B} = \boldsymbol{\upsilon}_2$, $\boldsymbol{\mu} = \boldsymbol{\mu}_{2.1}$ e $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{22.1}$, o que conclui a demonstração. $\square$

Note que dado u, quando $\boldsymbol{\Sigma}_{21} = \mathbf{0}$ e $\boldsymbol{\lambda}_2 = \mathbf{0}$, é possível obter independência das componentes $\mathbf{Y}_1$ e $\mathbf{Y}_2$ de um vetor aleatório $\mathbf{Y}$ com distribuição MESN. O seguinte Corolário é um produto da Proposição 10, desde que $E[\mathbf{Y}_2|\mathbf{y}_1] = E_U[E[\mathbf{Y}_2|\mathbf{y}_1, U]|\mathbf{y}_1]$.

Proposição 11. *Considere a notação do Corolário 2 e tomando $k(u) = u^{-1}$. Se $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ então o primeiro momento de $\mathbf{Y}_2$ dado $\mathbf{Y}_1 = \mathbf{y}_1$ é dada por*

$$E[\mathbf{Y}_2|\mathbf{y}_1] = \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1 + \boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}} E[U^{-1/2} \frac{\phi_1(u^{1/2}\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}{\Phi_1(u^{1/2}\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}|\mathbf{y}_1], \tag{2.13}$$

com $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

A grande contribuição desse estudo é com certeza a utilização das técnicas mais recentes de influência local, mais precisamente as técnicas de Zhu e Lee (2001), e para nos ajudar com a utilização dessas técnicas segue a Proposição 12 e a partir dessa o Corólario 3.

Proposição 12. *Seja $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, então para qualquer função g, a distribuição de $g(\mathbf{Y} - \boldsymbol{\mu})$ não depende de $\boldsymbol{\lambda}$ e tem a mesma distribuição que $g(\mathbf{X} - \boldsymbol{\mu})$, em que $\mathbf{X} \sim MEN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. Um caso particular, se $\mathbf{W}$ é uma matriz $p \times p$ simétrica, então $(\mathbf{Y} - \boldsymbol{\mu})^\top \mathbf{W}(\mathbf{Y} - \boldsymbol{\mu})$ e $(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{W}(\mathbf{X} - \boldsymbol{\mu})$ são identicamente distribuídas.*

Demonstração. A prova segue do uso da Proposição 6 e um similar procedimento pode ser encontrado em Wang et al. (2004). □

Como um produto da Proposição 12, nós temos o interessante resultado que segue:

Corolário 3. *Seja $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. Então a forma quadrática*

$$d = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$$

tem a mesma distribuição que $d_0 = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$, em que $\mathbf{X} \sim MEN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$.

O resultado do Corolário 3 é interessante porque nos permite checar modelos na prática. Por outro lado, o Corolário 3 junto com o resultado encontrado em Lange e Sinsheimer (1993) nos permite obter o m-ésimo momento de d.

Corolário 4. *Seja* $\mathbf{Y} \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Então para qualquer* $m > 0$ *e tomando* $k(u) = u^{-1}$ *temos que*

$$E[d^m] = \frac{2^m \Gamma(m + p/2)}{\Gamma(p/2)} E[U^{-m}].$$

Esses resultados serão de grande utilidade para o desenvolvimento e melhor entendimento do trabalho. A seguir apresentamos alguns exemplos de distribuições da classe MESN e alguns de seus momentos com alguns cálculos tediosos porém não muito complicados.

2.5 Exemplos

Nas subseções que seguem serão apresentados três exemplos de distribuições classe de misturas de escala skew-normal (MESN), a skew-t (ST), a skew-slash (SSL) e a skew-normal contaminada (SCN). Nessas seções, após alguns resultados conhecidos nós definiremos as esperanças condicionais dessas distribuições e tomaremos a notação de κ_r para a esperança condicional $E[U^r|\mathbf{y}]$ e τ_r para a esperança condicional $E[U^r W_{\phi_1}(U^{1/2}A)|\mathbf{y}]$. Os resultados obtidos nessas seções serão de fundamental importância na implementação do algoritmo EM.

2.5.1 Distribuição skew-t Multivariada

A Distribuição skew-t multivariada (Branco e Dey, 2001; Azzalini e Capitanio, 2003) com ν graus de liberdade, $\mathrm{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ digamos, pode ser obtida do modelo de misturas de escalas skew-normal (2.4), tomando U distribuido como $Gamma(\nu/2, \nu/2)$, $\nu > 0$ e $\kappa(u) = 1/u$. A função densidade de probabilidade de $\mathbf{Y}$ é

$$f(\mathbf{y}) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T\left(\sqrt{\frac{\nu + p}{\nu + d}} A|0, 1, \nu + p\right), \quad \mathbf{y} \in \mathbb{R}^p, \tag{2.14}$$

em que $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ e $T(\cdot|\nu)$ denotam, respectivamente, a fdp e a fda da distribuição t-Student p-variada, e $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ a distância de Mahalanobis, que no capítulo 3 será explicada com mais detalhes. Um caso particular da distribuição skew-t é a distribuição Cauchy assimétrica, quando $\nu = 1$. Também, quando $\nu \uparrow \infty$, tem-se a distribuição skew-normal no limite.

Aplicações da distribuição skew-t em estimações robustas podem ser encontradas em Lin et al. (2007) e Azzalini e Genton (2008). Neste caso, da Proposição 7, a média e a matriz de covariâncias de $\mathbf{Y}$ são dadas por,

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\boldsymbol{\Delta}, \quad \nu > 1,$$

e

$$Var[\mathbf{Y}] = \frac{\nu}{\nu-2}\boldsymbol{\Sigma} - \frac{\nu}{\pi}\left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma(\frac{\nu}{2})}\right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top, \ \nu > 2.$$

em que $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$. A seguir, nós temos um importante resultado que será utilizado na implementação do algoritmo EM e para encontrarmos expressões de forma fechada do primeiro momento condicional dado na Proposição 11.

Proposição 13. *Seja* $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. *Então,*

$$\begin{aligned}
\kappa_r &= \frac{2^{r+1}\nu^{\nu/2}\Gamma(\frac{p+\nu+2r}{2})(d+\nu)^{\frac{-p+\nu+2r}{2}}}{f(\mathbf{y})\Gamma(\frac{\nu}{2})\sqrt{\pi^p}|\boldsymbol{\Sigma}|^{1/2}} T_1\left(\sqrt{\frac{p+\nu+2r}{d+\nu}}A|0,1,p+\nu+2r\right),\\
\tau_r &= \frac{2^{(r+1)/2}\nu^{\nu/2}\Gamma(\frac{p+\nu+2r}{2})(d+\nu+A^2)^{-\frac{p+\nu+2r}{2}}}{f(\mathbf{y})\Gamma(\frac{\nu}{2})\sqrt{\pi^{(p+1)}}|\boldsymbol{\Sigma}|^{1/2}}\frac{(\nu+d)^{(\nu+p)/2}}{(\nu+d+A^2)^{(\nu+p+r)/2}}.
\end{aligned}$$

em que $A = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$ *e* $W_{\Phi_1}(x) = \phi_1(x)/\Phi_1(x)$, $x \in \mathbb{R}$.

Demonstração. A prova segue do Lema 1 de Azzalini e Capitanio (2003), desde que $f(u|\mathbf{y}) = f(\mathbf{y}, u)/f(\mathbf{y})$, então

$$\kappa_r = \frac{2}{f(\mathbf{y})} = \int_0^\infty u^r\phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}A)G(\nu/2,\nu/2)du,$$

e

$$\tau_r = \frac{2}{f(\mathbf{y})} = \int_0^\infty u^r\phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\phi_1(u^{1/2}A)G(\nu/2,\nu/2)du,$$

em que $G(\nu/2,\nu/2)$ denota a função densidade de probabilidade da distribuição Gamma($\nu/2$, $\nu/2$). □

Para um vetor aleatótio $\mathbf{Y}$, particionado como $\mathbf{Y} \sim \mathrm{ST}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}, \widetilde{\boldsymbol{\upsilon}}, \nu)$. Deste modo, pela Proposição 11 nós temos o seguinte resultado:

Corolário 5. *Sob a notação da Proposição 11, se* $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. *Então,*

$$\begin{aligned} E[\mathbf{Y}_2|\mathbf{y}_1] &= \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1+\boldsymbol{\upsilon}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}} \times \\ & \frac{1}{f(\mathbf{y}_1)} \frac{\nu^{\nu^2}\Gamma(\frac{\nu+p_1-1}{2})}{\Gamma(\frac{\nu}{2}\sqrt{\pi^{(p_1+1)}}|\boldsymbol{\Sigma}_{11}|^{1/2})}(\nu + d_{y1} + (\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))^2)^{-\frac{\nu+p_1-1}{2}} \end{aligned}$$

em que $d_{y1} = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^1 (\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

2.5.2 Distribuição Skew-Slash Multivariada

Outra distribuição da classe MESN, chamada como skew-slash multivariada e denotada por $\mathrm{SSL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, é obtida quando a distribuição de U é $Beta(\nu, 1)$, $\nu > 0$ e $\kappa(u) = 1/u$. Sua fdp é dada por

$$f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1}\phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}A)du, \quad \mathbf{y} \in \mathbb{R}^p. \tag{2.15}$$

em que $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$.

A distribuiçaõ skew-slash se reduz à skew-normal quando $\nu \uparrow \infty$. Da Proposição 7, pode-se mostrar que

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

em que $\Delta = \Sigma^{1/2}\delta$. Como no caso skew-t nós temos os seguintes resultados:

Proposição 14. *Seja* $\mathbf{Y} \sim SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. *Então,*

$$\kappa_r = \frac{2^{\nu+r+1}\nu\Gamma(\frac{p+2\nu+2r}{2})}{P_1(\frac{p+2\nu+2r}{2}, \frac{d}{2})d^{-\frac{p+2\nu+2r}{2}}} f(\mathbf{y})\sqrt{\pi^p}|\boldsymbol{\Sigma}|^{1/2}E[\Phi(S^{1/2}A)],$$

em que $S_i \sim$ *Gamma*$(\frac{p+2\nu+2r}{2}, \frac{d}{2})\mathbb{I}_{(0,1)}$, *e*

$$\tau_r = \frac{2^{\nu+r+1/2}\nu\Gamma\left(\frac{p+2\nu+2r}{2}\right)}{f(\mathbf{y})\sqrt{\pi^{(p+1)}}|\boldsymbol{\Sigma}|^{1/2}}(d+A^2)^{-\frac{2\nu+p+2r}{2}}P_1(\frac{2\nu+p+2r}{2}, \frac{d+A^2}{2})$$

em que $P_x(a,b)$ *denota a fda da distribuição Gamma*(a,b) *avaliada em* x.

Corolário 6. *Sob a notação da Proposição 11, se* $\mathbf{Y} \sim SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. *Então,*

$$\begin{aligned} E[\mathbf{Y}_2|\mathbf{y}_1] &= \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{\sqrt{1+\boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}} \times \\ & \frac{2^\nu\nu}{f(\mathbf{y}_1)}\frac{\Gamma(\frac{p_1+2\nu-1}{2})(d_{y_1}+(\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))^2)^{-\frac{p_1+2\nu-1}{2}}}{\sqrt{\pi^{(p_1+1)}}|\boldsymbol{\Sigma}_{11}|^{1/2}} \times \\ & P_1(\frac{p_1+2\nu-1}{2}, \frac{d_{y_1}+(\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))^2}{2}), \end{aligned}$$

em que $d_{y_1} = (\mathbf{y}_1-\boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)$.

Aplicações da distribuição skew-slash podem ser encontradas em Wang e Genton (2006).

2.5.3 Distribuição Skew-Normal Contaminada Multivariada

A distribuição skew-normal contaminada é obtida quando o fator de mistura de escala U é uma variável aleatória discreta tomando um de dois estados. A função de probabilidade de U, dado o vetor de parâmetros $\boldsymbol{\nu} = (\nu_1, \nu_2)^\top$, é denotada por

$$h(u|\boldsymbol{\nu}) = \nu_1\mathbb{I}_{(u=\nu_2)} + (1-\nu_1)\mathbb{I}_{(u=1)}, \;\; 0 < \nu_1 < 1, \, 0 < \nu_2 \leq 1. \tag{2.16}$$

Para $\kappa(u) = u^{-1}$, segue que

$$f(\mathbf{y}) = 2\left\{\nu_1\phi_p(\mathbf{y}|\boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Sigma})\Phi(\nu_2^{1/2}A) + (1-\nu_1)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(A)\right\},$$

em que $A = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})$'.

Essa distribuição é denotada por $\mathrm{SCN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. O parâmetro ν_1 pode ser interpretado como a proporção de valores extremos ou valores aberrantes (*outliers*, do inglês), enquanto ν_2 pode ser interpretado como um fator de escala. Da Proposição 7, pode-se mostrar que

$$\begin{aligned} E[\mathbf{Y}] &= \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\left(\frac{\nu_1}{\nu_2^{1/2}} + 1 - \nu_1\right)\boldsymbol{\Delta}, \quad e \\ Var[\mathbf{Y}] &= \left(\frac{\nu_1}{\nu_2} + 1 - \nu_1\right)\boldsymbol{\Sigma} - \frac{2}{\pi}\left(\frac{\nu_1}{\nu_2^{1/2}} + 1 - \nu_1\right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \end{aligned}$$

em que $\Delta = \Sigma^{1/2}\delta$. A distribuição skew-normal contaminada se reduz a distribuição skew-normal quando $\nu_1 = \nu_2 = 1$.

Da Proposição 8 seguem os seguintes resultados:

Proposição 15. *Seja* $\mathbf{Y} \sim SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu_1, \nu_2)$. *Então,*

$$\begin{aligned} \kappa_r &= \frac{2}{f(\mathbf{y})}\left\{\nu_1\nu_2^r\phi_p(\mathbf{y}|\boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Sigma})\Phi(\nu_2^{1/2}A) + (1-\nu_1)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(A)\right\}, \\ \tau_r &= \frac{2}{f(\mathbf{y})}\left\{\nu_1\nu_2^{r/2}\phi_p(\mathbf{y}|\boldsymbol{\mu}, \nu_2^{-1}\boldsymbol{\Sigma})\phi_1(\nu_2^{1/2}A) + (1-\nu_1)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi_1(A)\right\}. \end{aligned}$$

e da mesma forma que para a skew-t e para a skew-slash nós temos o seguinte Corolário:

Corolário 7. *Sob a notação da Proposição 11, seja* $\mathbf{Y} \sim SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu_1, \nu_2)$. *Então,*

$$\begin{aligned} E[\mathbf{Y}_2|\mathbf{y}_1] = \; & \boldsymbol{\mu}_{2.1} + \frac{2\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}{f(\mathbf{y}_1)\sqrt{1+\boldsymbol{\upsilon}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\upsilon}_2}} \times \\ & [\nu_1\nu_2^{-1/2}\phi_{p_1}(\mathbf{y}_1|\boldsymbol{\mu}_1, \nu_2^{-1}\boldsymbol{\Sigma}_{11})\phi_1(\nu_2^{1/2}\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)) \\ & +(1-\nu_1)\phi_{p_1}(\mathbf{y}_1|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})\phi_1(\widetilde{\boldsymbol{\upsilon}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))] \end{aligned}$$

em que $d_{y_1} = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

A Figura 6 nos apresenta uma sobreposição das distribuições SN, ST, SSL e SCN tanto no caso de λ positivo quanto negativo, fica evidente que as distribuições ST, SSL e SCN possuem caudas mais pesadas do que a SN. No Capítulo 5, para as nossas aplicações, alguns conjuntos de dados serão ajustados as distribuições skew-normal, skew-t, skew-slash e skew-normal contaminada. Nas seções a seguir apresentamos o método de inferência pela máxima verossimilhança e a seguir apresentaremos o algoritmo EM.

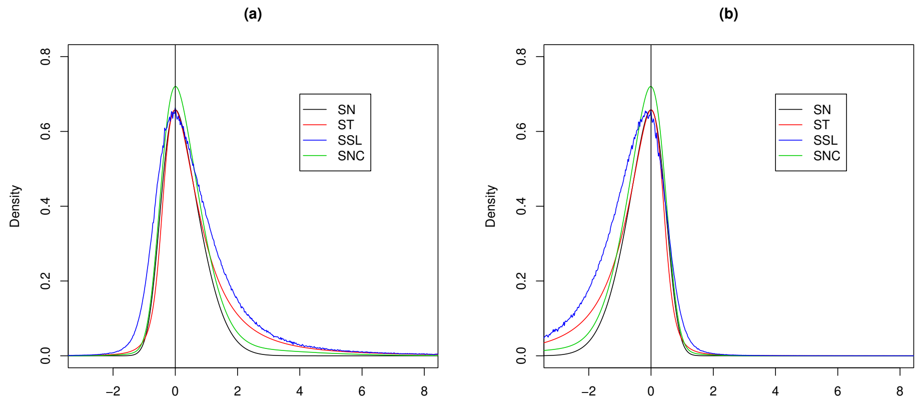

Figura 6: Função densidade de probabilidade de algumas distribuições da famíla MESN univariada padrão. (a) SN(3),ST$(3, 2)$,SCN$(3, 0.1, 0.1)$ e SSL$(3, 1)$ (b) SN(-3), ST$(-3, 2)$, SCN$(-3, 0.1, 0.1)$ e SSL$(-3, 1)$.

2.6 Estimação de Máxima Verossimilhança (EMV)

Encontrar os EMV de $\boldsymbol{\theta}$ por métodos convencionais é complicado, dado que as equações resultantes não tem forma fechada. Embora existam muitas metodologias para encontrar os EMV, aqui mostraremos como obter os EMV de $\boldsymbol{\theta}$ usando o algoritmo EM, o qual é fácil de implementar e computacionalmente conveniente.

A seguir será apresentada a representação hierárquica dos modelos na classe MESN, e a partir desta pode-se tratar o problema de estimação dos parâmetros via o algortítmo EM, considerando a abordagem por dados completos.

Utilizando a representação estocástica (2.6) e tomando $T_i \overset{iid}{\sim} HN_1(0, 1)$ (T_i distribuído com uma half-normal padrão univariada, média zero e variância um) o modelo MESN pode ser apresentado sob um ponto de vista com dados incompletos.

Proposição 16. *Considere a amostra* $\mathbf{Y}_i \sim MESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, H)$ *com* $i = 1, \ldots, n$. *Então*

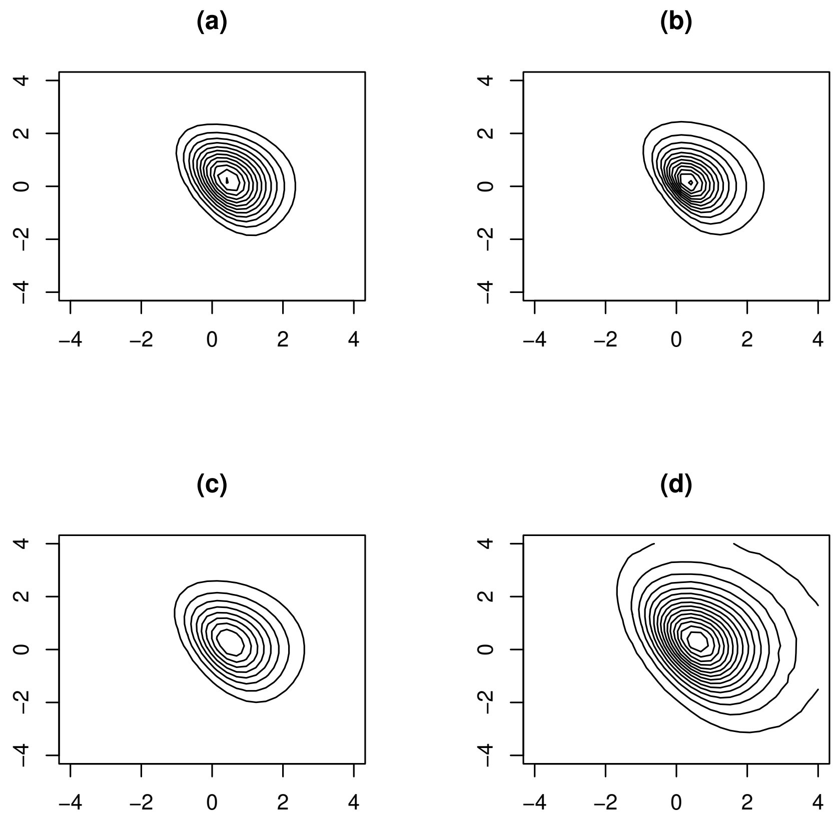

Figura 7: Contornos de algumas distribuições da classe MESN bivariada padrão. (a) $\mathrm{SN}_2(\boldsymbol{\lambda})$ (b) $\mathrm{ST}_2(\boldsymbol{\lambda}, 2)$ (c) $\mathrm{SCN}_2(\boldsymbol{\lambda}, 0.5, 0.5)$ (d) $\mathrm{SSL}_2(\boldsymbol{\lambda}, 1)$, em que $\boldsymbol{\lambda} = (2, 1)^\top$.

o modelo hierárquico para cada vetor aleatório MESN dessa amostra é dado por

$$\mathbf{Y}_i | T_i = t_i, U_i = u_i \overset{iid}{\sim} N_p(\boldsymbol{\mu} + \boldsymbol{\Delta} t_i, \kappa(u_i)\boldsymbol{\Gamma}), \tag{2.17}$$

$$T_i | U_i = u_i \overset{iid}{\sim} HN_1(0, \kappa(u_i)), \tag{2.18}$$

$$U_i \overset{iid}{\sim} H(u_i | \boldsymbol{\nu}), \tag{2.19}$$

em que $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$ *e* $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$. *Além disso, considerando* $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$, *a função de log-verossimilhança completa de* $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\nu})$, *com* $\boldsymbol{\alpha}$ *denotando o vetor com os elementos da matriz triangular superior* $\boldsymbol{\Sigma}$, *é dada por*

$$
\begin{aligned}
\ell_c(\boldsymbol{\theta}) &= C - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\kappa^{-1}(u_i)(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta}t_i)^{\top}\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta}t_i) \\
&\quad + \sum_{i=1}^{n}\log\left(h(u_i|\boldsymbol{\nu})\right) \\
&= C - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\left[\kappa^{-1}(u_i)(\mathbf{y}_i - \boldsymbol{\mu})^{\top}\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right. \\
&\quad \left. -2\kappa^{-1}(u_i)t_i\boldsymbol{\Delta}^{\top}\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) + \kappa^{-1}(u_i)t_i^2\boldsymbol{\Delta}^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}\right] \\
&\quad + \sum_{i=1}^{n}\log\left(h(u_i|\boldsymbol{\nu})\right), \qquad (2.20)
\end{aligned}
$$

com C que é uma constante que não dependente do vetor de parâmetros $\boldsymbol{\theta}$.

A estimação dos parâmetros do modelo para distribuições MESN pelo método da máxima verossímilhança, pode ser realizada via algoritmo EM (Dempester et al., 1977). O algoritmo é aplicado ao problema de estimação a partir de dados incompletos, aumentando o vetor de dados observados com a inclusão de variáveis latentes, não observaveis diretamente, de modo que a verossimilhança do vetor de dados completos simplifique as análises a serem envolvidas. O algoritmo procede iterativamente em duas etapas, E (do inglês, *Expectation*) e M (do inglês, *Maximization*). A inclusão desses dados não observáveis ao problema é tratada na etapa E, a qual consiste em tomar a esperança da log-verossimilhança completa condicional ao vetor de dados observados. Em seguida, no passo M, é realizada a maximização do resultado obtido no passo E com relação aos parâmetros do modelo, e em seguida igualamos a zero e isolamos o parâmetro de interesse. Muitas vezes não conseguimos obter formas fechadas e temos que recorrer a extensões do algoritmo EM que serão abordadas na próxima seção.

Para a estimação dos parâmetros na classe de distribuições MESN, pode-se obter a abordagem de dados incompletos através da representação hierárquica do modelo dado na Proposição 16. O vetor de dados observados é dado por $\mathbf{y} = (\mathbf{y}_1^{\top}, \ldots, \mathbf{y}_n^{\top})^{\top}$, e os dados aumentados por $\mathbf{u} = (u_1, \ldots, u_n)^{\top}$, $\mathbf{t} = (t_1, \ldots, t_n)^{\top}$. Portanto, na etapa E do algoritmo, toma-se o valor esperado da log-verossimilhança completa condicional à $\mathbf{y}$ e a $\boldsymbol{\theta}$ no seu estado corrente, dada em (2.20). Note que as seguintes quantidades deve ser obtidas

$$\begin{aligned}\widehat{u}_i &= \mathrm{E}\{\kappa^{-1}(U_i)|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\mathbf{y}_i\}, &(2.21)\\ \widehat{ut}_i &= \mathrm{E}\{\kappa^{-1}(U_i)T_i|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\mathbf{y}_i\}, &(2.22)\\ \widehat{ut^2}_i &= \mathrm{E}\{\kappa^{-1}(U_i)T_i^2|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\mathbf{y}_i\}. &(2.23)\end{aligned}$$

A primeira dessas quantidades pode ser obtida utilizando a Proposição 8. Para as demais quantidades, deve ser obter a distribuição condicional $T_i|\mathbf{Y}_i=\mathbf{y}_i, U_i=u_i$. Das equações (2.17) e (2.18) da representação hierárquica do modelo MESN mais o Lema 3 dado no Apêndice A tem-se que

$$\begin{aligned}2\phi_p(\mathbf{y}_i|\boldsymbol{\mu}+\boldsymbol{\Delta}t_i,\kappa(u_i)\boldsymbol{\Gamma})\times\phi_1(t_i|0,\kappa(u_i)) &= 2\phi_p(\mathbf{y}|\boldsymbol{\mu},\kappa(u_i)(\boldsymbol{\Gamma}+\boldsymbol{\Delta}\boldsymbol{\Delta}^\top))\\ &\times\ \phi_1(t_i|\mu_{T_i},\kappa(u_i)M_{T_i}^2)\end{aligned}$$

Portanto, $T_i|\mathbf{Y}_i=\mathbf{y}_i, U_i=u_i \sim HN_1(\mu_{T_i},\kappa(u_i)M_{T_i}^2)$ com $M_T^2=1/(1+\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})$ e $\mu_{T_i}=M_T^2\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})$.

Utilizando o resultado de (2.21) e algumas propriedades de esperança condicional, obtêm-se as seguintes expressões

$$\begin{aligned}\widehat{ut}_i &= \mathrm{E}\{\kappa^{-1}(U_i)T_i|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\mathbf{y}_i\}\\ &= \mathrm{E}_{U_i|\mathbf{y}_i}\{\mathrm{E}_{T_i|u_i,\mathbf{y}_i}[\kappa^{-1}(U_i)T_i]\}\\ &= \mathrm{E}_{U_i|\mathbf{y}_i}\{\kappa^{-1}(U_i)[\mu_{T_i}+W_\Phi\left(\frac{\kappa^{-1/2}(U_i)\mu_{T_i}}{M_{T_i}}\right)\kappa^{1/2}(U_i)M_{T_i}]\}\\ &= \mu_{T_i}\mathrm{E}_{U_i|\mathbf{y}_i}\{\kappa^{-1}(U_i)\}+M_{T_i}\mathrm{E}_{U_i|\mathbf{y}_i}\{W_\Phi\left(\frac{\kappa^{-1/2}(U_i)\mu_{T_i}}{M_{T_i}}\right)\kappa^{-1/2}(U_i)\}\\ &= \widehat{u}_i\widehat{\mu}_{T_i}+\widehat{M}_{T_i}\widehat{\tau}_i,\quad i=1,\ldots,n,\end{aligned}$$

em que a última igualdade utiliza o Lema 3 dado no Apêndice A para momentos de uma distribuição *half*-normal. Analogamente, tem-se que

$$\widehat{ut^2}_i=\widehat{u}_i\widehat{\mu}_{T_i}^2+\widehat{M}_{T_i}^2+\widehat{M}_{T_i}\widehat{\mu}_{T_i}\widehat{\tau}_i,$$

em que $\widehat{\tau}_i=E[U_i^{1/2}W_{\Phi_1}(\frac{U_i^{1/2}\widehat{\mu_{T_i}}}{\widehat{M}_{T_i}})|\widehat{\boldsymbol{\theta}},\mathbf{y}_i]$, $W_{\Phi_1}(x)=\phi_1(x)/\boldsymbol{\Phi}_1(x)$, $\widehat{M}_{T_i}^2=1/(1+\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\Delta}})$

e $\widehat{\mu}_{T_i} = \widehat{M}^2_{T_i}\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Gamma}}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}), \quad i = 1, \ldots, n.$

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

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta})|\mathbf{y}, \widehat{\boldsymbol{\theta}}] &= C - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n}\{(\mathbf{y}_i - \boldsymbol{\mu})^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\widehat{u}_i \\ &- 2\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\widehat{ut}_i + \boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}\widehat{ut^2}_i\} + \sum_{i=1}^{n}\mathrm{E}[\log(h(u_i|\boldsymbol{\nu}))]. \end{aligned}$$

A etapa M do algoritmo envolve a maximização em $\boldsymbol{\theta}$ da esperança condicional da log-verossimilhança completa acima. Assim como será mostrado posteriormente, obter as quantidades na etapa E para as distribuições skew-t e skew-normal contaminada não é uma tarefa árdua, já que as funções de distribuição de probabilidade definidas em (2.14) e (2.16) têm expressões fechadas, matematicamente atrativas e fáceis de serem implementadas. Entretanto esse não é o caso para a distribuição skew-slash. Neste caso, a expressão a ser maximizada envolve integrais complexas, sendo assim necessária a utilização de métodos alternativos para se obter uma resolução. A seguir estudaremos extensões do algoritmo EM que serão utilizadas como uma resolução do problema mencionado no parágrafo anterior.

2.6.1 O Algoritmo EM e suas Extensões

Na seção anterior após alguns cálculos chegamos ao valor esperado da log-verossimilhança completa, o que representa o passo E do algoritmo EM, e comentamos que quando se tem expressões intratáveis na etapa M do algoritmo, é comum recorrer a extensões do algoritmo EM. Essas extensões são os chamados algoritmos GEM (*generalized expectation maximization*), que em cada passo M aumenta a Q-função ao invés de maximiza-la. O algoritmo ECM (Meng e Rubin, 1993) é uma subclasse dos algoritmos GEM e em geral é mais utilizado do que o próprio algoritmo EM. Esse algoritmo substitui o passo M do algoritmo EM por uma sequência de passos de maximizações condicionais (CM - *conditional maximization*), ou passos CM. Por exemplo, vamos supor que temos uma distribuição qualquer com dois parâmetros a serem estimados, α e β por exemplo. Dado o passo E do algoritmo nós supomos conhecer $\beta^{(k)}$ então estimamos $\alpha^{(k+1)}$ dado que conhecemos

$\beta^{(k)}$ e em seguida fazemos o inverso, isto é, estimamos $\beta^{(k+1)}$ dado que $\alpha^{(k)}$ é conhecido e ficamos com um $\alpha^{(k+1)}$ em função de $\beta^{(k)}$ e vice-versa, assim em sucessivos passos.

O algoritmo ECME (Liu e Rubin, 1994) é uma extensão dos algoritmos EM e ECM. Esse novo algoritmo é computacionalmente mais rápido do que os anteriores. A sua diferença para o algoritmo ECM basicamente é a substituição de alguns dos passos CM's por passos de maxização condicional diretamente na atual função log-verossimilhança de dados observados.

Tanto o algoritmo EMC quanto o EMCE trazem o benefício de se trabalhar com funcões de baixas dimensões sujeitos a restrições ou condições. Qualquer problema que pode ser resolvido com o algoritmo EM ou com o algoritmo ECM também pode ser resolvido pelo ECME. A grande vantagem do uso do algoritmo ECME sobre os demais algoritmos citados é a sua velocidade computacional.

O algoritmo EM aplicado na função de log-verossimilhança dada em (2.20) é dado por

Passo E: Dado $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, obter $\widehat{u}_i$, $\widehat{ut}_i$ e $\widehat{ut^2}_i$ para $i = 1, \ldots, n$.

Passo M: Atualizar $\widehat{\boldsymbol{\theta}}$ maximizando $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E\{\ell_c(\boldsymbol{\theta})|\mathbf{y}, \widehat{\boldsymbol{\theta}}\}$ sobre $\boldsymbol{\theta}$, que resulta nas seguintes expressões com formas fechadas:

$$
\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \sum_{i=1}^{n}(\widehat{u}_i\mathbf{y}_i - \widehat{ut}_i\boldsymbol{\Delta})/(\sum_{i=1}^{n}\widehat{u}_i), \qquad (2.24)\\
\widehat{\boldsymbol{\Gamma}} &= \frac{1}{n}\sum_{i=1}^{n}[\widehat{u}_i(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top - 2\widehat{ut}_i\boldsymbol{\Delta}(\mathbf{y}_i - \boldsymbol{\mu})^\top 1 + \widehat{ut^2}_i\boldsymbol{\Delta}\boldsymbol{\Delta}^\top],\\
\widehat{\boldsymbol{\Delta}} &= \sum_{i=1}^{n}\widehat{ut}_i(\mathbf{y}_i - \boldsymbol{\mu})/\sum_{i=1}^{n}\widehat{ut^2}_i.
\end{aligned}
$$

A seguir aplicamos o algoritmo ECM para a função de log-verossmilhança dada em (2.20):

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

Passo CM: Atualizar $\widehat{\boldsymbol{\theta}}^{(k)}$ maximizando $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E\{\ell_c(\boldsymbol{\theta})|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\}$ sobre $\boldsymbol{\theta}$, que resulta nas seguintes formas fechadas para os parâmetros transformados $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ e $\boldsymbol{\Delta}$, respectivamente:

$$
\begin{aligned}
\widehat{\boldsymbol{\mu}}^{(k+1)} &= \sum_{i=1}^{n}(\widehat{\kappa}_i^{(k)}\mathbf{y}_i - \widehat{s}_{2i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)})/(\sum_{i=1}^{n}\widehat{\kappa}_i^{(k)}), \qquad (2.25)\\
\widehat{\boldsymbol{\Gamma}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\kappa}_i^{(k)}(\mathbf{y}_i-\widehat{\boldsymbol{\mu}}^{(k)})(\mathbf{y}_i-\widehat{\boldsymbol{\mu}}^{(k)})^\top - \widehat{s}_{2i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}(\mathbf{y}_i-\widehat{\boldsymbol{\mu}}^{(k)})^\top\\
&\quad - \widehat{s}_{2i}^{(k)}(\mathbf{y}_i-\widehat{\boldsymbol{\mu}}^{(k)})\widehat{\boldsymbol{\Delta}}^{\top(k)} + \widehat{s}_{3i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\Big],\\
\widehat{\boldsymbol{\Delta}}^{(k+1)} &= \sum_{i=1}^{n}\widehat{s}_{2i}^{(k)}(\mathbf{y}_i-\widehat{\boldsymbol{\mu}}^{(k)})/\sum_{i=1}^{n}\widehat{s}_{3i}^{(k)},
\end{aligned}
$$

E finalmente para a implementação do algoritmo ECME na log-verossimilhança dada em (2.20) é necessário apenas o acréscimo de mais um passo aos descritos no algoritmo ECM. O passo é descrito abaixo:

Passo CML: Atualizar $\boldsymbol{\nu}^{(k+1)}$ maximizando a função de verossimilhança marginal, obtendo

$$\boldsymbol{\nu}^{(k+1)} = arg\,max_{\boldsymbol{\nu}}\sum_{i=1}^{n} log(\psi(y_i|\boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}, \boldsymbol{\lambda}^{(k+1)}, \boldsymbol{\nu})), \qquad (2.26)$$

com $f(\mathbf{y}|\boldsymbol{\theta})$ dado em (2.4).

As iterações são repetidas até que alguma regra de convergência adequada seja satisfeita, por exemplo, se $||\boldsymbol{\Theta}^{(k+1)} - \boldsymbol{\Theta}^{(k)}||$ for suficientemente pequeno ou alguma distância envolvendo a log-verossimilhança, como $||\ell(\boldsymbol{\Theta}^{(k+1)}) - \ell(\boldsymbol{\Theta}^{(k)})||$. Veja Dias e Wedel (2004) no contexto de mistura de normais univariadas.

Os valores iniciais para o algoritmo podem ser obtidos utilizando o seguinte esquema: Para o caso skew-normal, utiliza-se como valores iniciais para o vetor de médias e matriz de covariâncias, o vetor de média amostral e matriz de covariâncias amostral, respectivamente. Para a j-ésima cordenada do vetor de assimetria, considere $\widehat{\rho}_j$ a assimetria amostral para a variável j. Então, $\lambda_j^{(0)} = 3 \times sign(\widehat{\rho}_j)$. As estimativas EM encontradas para esse caso são então passadas como valores inicias para o algoritmo considerando os outros elementos da classe MESN. Na prática, recomenda-se a utilização de diferentes valores iniciais para o algoritmo EM. Isto por que se existir mais de um máximo, pode-se determinar o global comparando seus valores de verossimilhança e comprovar a estabilidade da estimativa obtida. Detalhes sobre as vantagens e desvantagens da utilização dos algoritmos EM, ECM e ECME podem ser encontrados em Liu e Rubin, 1994.

No próximo capítulo apresentamos o modelo com erros nas variáveis multivariado, sua descrição, estrutura e aplicabilidades.

3 Modelo com Erros nas Variáveis Multivariado

3.1 Introdução

O problema de erro de medida está presente em diversas áreas, como por exemplo, geologia, química, física, biologia, entre outras. Em geral, está presente em laboratórios que necessitam quantificar substâncias (em especial, sua concentração) e para isso fazem uso de algum instrumento de medição (veja, Barnett, 1969, Theobald e Mallison, 1978, Shyr e Gleser, 1986, Bolfarine e Galea-Rojas, 1995 e Chapkevitch et al., 1996).

A cada ano que passa a comparação de instrumentos de medição tem se mostrado mais relevante. Como exemplo podemos citar um dos primeiros estudos nesta área, Grubbs (1948, 1973) apresenta uma classe particular de modelos utilizados para comparar instrumentos de medição conhecido como regressão linear com erros de medição. Barnett (1969) apresenta um exemplo onde quatro combinações instrumento-operador conhecidos para medir a capacidade vital num grupo de pacientes são avaliados. Leurgans (1980) compara dois métodos para medir a concentração da glicose no sangue e Jaech (1985) apresenta vários exemplos na área industrial. Christensen e Blackwood (1993), comparam cinco termopares, Galea-Rojas (1995) apresenta um estudo de inferência no modelo t de Student e Bedrick (2001) compara três métodos para medir sedimentos encontrados no solo.

Frequentemente a avaliação dos instrumentos é feita baseada em dados obtidos utilizando cada instrumento para medir uma característica comum em várias unidades experimentais. O modelo de calibração comparativa provavelmente é o caso particular mais utilizado de modelos com erro de medição. Este modelo específica os x_i's (variáves latentes) não são diretamente observáveis e no seu lugar observamos os X_i's.

Como uma introdução ao que veremos na próxima seção segue o modelo conhecido na literatura como modelo de regressão simples com erros na variável é definido pela seguinte relação

$$y_i = \alpha + \beta x_i + e_i, \tag{3.1}$$

e

$$X_i = x_i + u_i, \quad i = 1, \ldots, n, \tag{3.2}$$

em que $y_i = (y_1, \ldots, y_r)$ é a resposta para a i-ésima unidade experimental, $\mathbf{e_i} = (e_1, \ldots, e_r)$, é o erro de medição para a i-ésima unidade experimental, α e β são os parâmetros e em que é frequente considerar os erros e a variável latente com distribuição

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \overset{iid}{\sim} \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right), x_i \overset{iid}{\sim} \mathrm{N}_1(\mu_x, \sigma_x^2), \tag{3.3}$$

ambas independentes.

Sabemos que na prática existem várias situações onde a variável explanatória não pode ser observada diretamente, mas sim com erros. Entre outras situações, a inexatidão da medida pode ser resultado de uma opinião subjetiva, ou do uso de instrumentos de precisão limitada. Em estudos nutricionais, covariáveis relacionadas ao consumo de gordura saturada envolvem erros de medida consideráveis, em parte devido a uma tendência de algumas pessoas em omitir a descrição precisa de suas refeições (Carroll e Stefanski, 1995). Outra situação ocorre com o interesse em relacionar a produção de certo cereal com o nível de nitrogêneo disponível no solo (Fuller, 1987). Para termos o conhecimento do nível de nitrogêneo disponível no solo, é necessário amostrarmos o solo da unidade experimental e fazer análises laboratoriais na amostra selecionada. Como resultado da amostragem e da análise laboratorial, não temos os verdadeiros valores da covariável mas sim estimativas destas.

Os modelos com erros de medição podem ser estudados sob duas especificações diferentes:

1. As x_i's são constantes desconhecidas, chamadas de parâmetros incidentais.
2. As x_i's são variáveis aleatórias não-correlacionadas com média μ_x e variância ϕ_x e são não correlacionadas com o erro.

O modelo descrito no item 1 acima é denominado modelo funcional, neste caso o número de parâmetros cresce com o número de observações. O modelo descrito no item 2 acima é denominado modelo estrutural, neste modelo não é possível encontrar uma solução de máxima verossimilhança para os parâmetros do modelo, pois o mesmo não é identificável, no sentido que dois vetores de parâmetros diferentes podem dar origem a uma mesma distribuição. Do ponto de vista de aplicação, o modelo funcional é adequado quando temos interesse apenas nos indivíduos participantes do estudo, enquanto que o estrutural é adequado quando o interesse é generalizar os resultados do estudo para a população de onde vem os indivíduos envolvidos no estudo.

O objetivo principal desse capítulo é apresentar um estudo de inferência estatística sob o modelo com erros nas variáveis multivariado (MEM), onde é de interesse comparar vários instrumentos de medição da mesma quantidade desconhecida x em um grupo comum de n unidades experimentais, supondo que as observações obtidas seguem uma distribuição da família MESN e a característica de interesse x é medida na mesma escala.

3.2 Descrição do Modelo

Seja n o tamanho da amostra; X_i, o valor observado da covariável na unidade i; y_{ij}, o j-ésima resposta observada na unidade i e x_i, o valor (verdadeiro) não observado da covariável na unidade i, $i = 1, ..., n$ e $j = 1, ..., r$. Relatando essas variáveis nós postulamos os seguintes modelos de equações (veja, Barnett, 1969 e Shyr e Gleser, 1986),

$$X_i = x_i + u_i, \tag{3.4}$$

e

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta} x_i + \mathbf{e}_i, \tag{3.5}$$

em que $\mathbf{Y}_i = (y_{i1}, \ldots, y_{ir})^\top$ é o vetor de respostas para a i-ésima unidade experimental, $\mathbf{e}_i = (e_{i1}, \ldots, e_{ir})^\top$, é um vetor aleatório de erros de medição de dimensão r, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)^\top$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^\top$ são vetores de parâmetros de dimensão r. Seja $\boldsymbol{\epsilon}_i = (u_i, \mathbf{e_i}^\top)^\top$ e $\mathbf{Z}_i = (X_i, \mathbf{Y}_i{}^\top)^\top$. Então o modelo definido pelas equações (3.4)-(3.5) pode ser escrito na forma matricial a seguir

$$\mathbf{Z}_i = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i = \mathbf{a} + \mathbf{B}r_i, \tag{3.6}$$

em que $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top$ e $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top$ são vetores p x 1, com $p = r + 1$, e $\mathbf{B} = [\mathbf{b}|\mathbf{I}_p]$ uma matriz $p \times (p+1)$ e $\mathbf{r}_i = (x_i, \boldsymbol{\epsilon}_i^\top)^\top$, $i = 1, \ldots, n$. Assim utilizando (3.6), a distribuição de $\mathbf{Z}_i$ pode ser especificada uma vez sendo especificada a distribuição de $\mathbf{r}_i$, $i = 1, \ldots, n$.

Como em Lachos et al. (2008), o modelo estrutural com erros nas variáveis sob a classe das distribuições MESN é definida considerando que

$$\mathbf{r}_i = \begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \overset{iid}{\sim} \mathrm{MESN}_{p+1}\left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix}; H\right), \tag{3.7}$$

$i = 1, \ldots, n$, em que $D(\phi_x, \boldsymbol{\phi}) = diag(\phi_x, \phi_1, \ldots, \phi_p)^\top$, com $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$. Esse modelo está considerando que a população não é simetricamente distribuída. Por outro lado, os erros $\boldsymbol{\epsilon}_i$ estão relacionados aos erros de medição, de modo que é esperado que sejam simétricamente distribuídos. O parâmetro de assimetria λ_x incorpora assimetria na variável latente x_i e consequentemente nas quantidades observadas $\mathbf{Z}_i$, $i = 1, \ldots, n$, o qual será mostrado ter uma distribuição marginal MESN multivariada.

Tomando $T_i \overset{iid}{\sim} HN_1(0, 1)$, de (2.5) e da representação estocástica marginal de um vetor aleatório SN dado em Arellano-Valle et al. (2005a) e em (2.2), o modelo definido em (3.6) e (3.7) podem ser escritos hierarquicamente como

$$\mathbf{Z}_i | x_i, U_i = u_i \overset{ind}{\sim} N_p(\mathbf{a} + \mathbf{b}x_i, k(u_i)D(\boldsymbol{\phi})), \tag{3.8}$$

$$x_i | T_i = t_i, U_i = u_i \overset{ind}{\sim} \mathrm{N}_1(\mu_x + \tau_x t_i, k(u_i)\nu_x^2), \tag{3.9}$$

$$T_i | U_i = u_i \overset{iid}{\sim} \mathrm{HN}_1(0, k(u_i)), \tag{3.10}$$

$$U_i \overset{iid}{\sim} H(u_i, \boldsymbol{\nu}), \tag{3.11}$$

$i = 1, ...n$, todas independentes, em que $\nu_x^2 = \phi_x(1-\delta_x^2)$ e $\tau_x = \phi_x^{1/2}\delta_x$, com $\delta_x = \lambda_x/(1+\lambda^2)^{1/2}$. Se $\lambda_x = 0$, então o modelo assimétrico se reduz ao MEM simétrico considerando misturas de escala da distribuição normal (MEN-MEM). A densidade marginal de $\mathbf{Z}_i$ é dada por

$$f(\mathbf{z}_i|\boldsymbol{\theta}) = 2\int_0^\infty \phi_p(\mathbf{z}_i|\boldsymbol{\mu}, k(u_i)\boldsymbol{\Sigma})\Phi_1(k^{-1/2}(u_i)\overline{\boldsymbol{\lambda}}_x^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{z}_i-\boldsymbol{\mu}))dH(u_i). \tag{3.12}$$

Isto é, $\mathbf{Z}_i \overset{iid}{\sim} \text{MESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \overline{\boldsymbol{\lambda}}_x; H)$, em que

$$\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x,\ \boldsymbol{\Sigma} = \phi_x\mathbf{b}\mathbf{b}^\top + D(\boldsymbol{\phi}) \text{ e } \overline{\boldsymbol{\lambda}}_x = \frac{\lambda_x\phi_x\boldsymbol{\Sigma}^{-1/2}\mathbf{b}}{\sqrt{\phi_x+\lambda_x^2\Lambda_x}},$$

com $\Lambda_x = \phi_x/c$ e $c = 1+\phi_x\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$, $i = 1, ..., n$.

Daí a função log-verossimilhança para $\boldsymbol{\theta}$ dado a amostra observada $\mathbf{z} = (\mathbf{z}_1^\top, \ldots, \mathbf{z}_n^\top)^\top$ é dado por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}), \tag{3.13}$$

em que $\ell_i(\boldsymbol{\theta}) = log2 - \frac{p}{2}log2\pi - \frac{1}{2}log|\boldsymbol{\Sigma}| + log(K_i)$, com

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty k^{-p/2}(u)exp\{-\frac{1}{2}k^{-1}(u)d_i\Phi_1(k^{-1/2}(u)A_i)dH(u),$$

em que $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\overline{\boldsymbol{\lambda}}_x$ são como em (3.12), $d_i = (\mathbf{z}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_i-\boldsymbol{\mu})$ e $A_i = \overline{\boldsymbol{\lambda}}_x^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{z}_i-\boldsymbol{\mu}) = A_x a_i$, com

$$A_x = \frac{\lambda_x\Lambda_x}{\sqrt{\phi_x+\lambda_x^2\Lambda_x}} \quad \text{e} \quad a_i = (\mathbf{z}_i-\boldsymbol{\mu})^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}.$$

Devemos dar um destaque especial a estatística d_i, conhecida como a distância de Mahalanobis. Introduzida pelo matemático indiano Prasanta Chandra Mahalanobis em 1936, essa distância é baseada nas correlações entre variáveis com as quais distintos padrões podem ser identificados e analisados. Distingue-se da distância euclidiana já que tem em conta as correlações do conjunto de dados e é invariante à escala, ou seja, não depende da escala das medições.

Na prática, a distância de Mahalanobis, fornece uma estatística de diagnóstico útil para identificar unidades de amostras com observações aberrantes, especialmente no desenvolvimento de modelos baseados em regressão linear. Um ponto que tenha uma distância de Mahalanobis maior do que o resto da população amostral de pontos é dito ter maior alavancagem já que tem uma maior influência no declive ou nos coeficientes do modelo de regressão. A significância estatística da distância de Mahalanobis na detecção de valores atípicos multivariados pode ser avaliada por um teste chi-quadrado com k graus de liberdade. Para maiores detalhes, esse assunto é discutido em Pinheiro et al. (2001) e Osorio et al. (2009) para o caso simétrico e em Azzalini e Capitanio (1999) para o caso assimétrico.

Na seção a seguir nós discutimos o processo iterativo para estimação de parâmetros, baseado no algoritmo EM.

3.3 O Algoritmo de Estimação EM

O algoritmo EM para estimação de máxima verossimilhança do modelo MESN-MEM foi desenvolvido por Lachos et al. (2008) e nessa seção nós demonstramos como implementa-lo. A característica chave desse modelo é que ele pode ser formulado a partir de uma representação hierarquica flexível, dado em (3.8)-(3.11), o que ajuda nas derivações mais teóricas. Seja $\mathbf{z} = (\mathbf{z}_1{}^\top, ...\mathbf{z}_n{}^\top)^\top$, $\mathbf{x} = (x_1, ...x_n)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ e $\mathbf{t} = (t_1, \ldots, t_n)^\top$ e seja $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\alpha}^{(k)\top}, \boldsymbol{\beta}^{(k)\top}, \boldsymbol{\phi}^{(k)\top}, \mu_x^{(k)}, \phi_x^{(k)}, \lambda_x^{(k)})^\top$, denotam as estimativas de $\boldsymbol{\theta}$ na k-ésima iteração. Segue de (3.8)-(3.11), com $\nu_x^2 = \phi_x(1 - \delta_x^2)$, $\tau_x = \phi_x^{1/2}\delta_x$ e $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$ que a função log-verossimilhança completa associada com $\mathbf{z}_c = (\mathbf{z}, \mathbf{x}, \mathbf{t}, \mathbf{u})$ é da forma

$$\begin{aligned}\ell_c(\boldsymbol{\theta} \mid \mathbf{z}_c) \;=\; & -\frac{n}{2}log(|D(\boldsymbol{\phi})|) - \frac{1}{2}\sum_{i=1}^{n} k^{-1}(u_i)(\mathbf{z}_i - \mathbf{a} - \mathbf{b}x_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{z}_i - \mathbf{a} - \mathbf{b}x_i) \\ & -\frac{n}{2}log(\nu_x^2) - \frac{1}{2\nu_x^2}\sum_{i=1}^{n} k^{-1}(u_i)(x_i - \mu_x - \tau_x t_i)^2 + C. \qquad (3.14)\end{aligned}$$

em que C é uma constante que é independente do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, em que $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top)^\top$ e $\boldsymbol{\theta}_2 = (\mu_x, \phi_x, \lambda_x)^\top$. Seja $\boldsymbol{\theta}^{(k)}$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração. Dada a corrente estimativa de $\boldsymbol{\theta}^{(k)}$, o passo E calcula $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = E[\ell_c(\boldsymbol{\theta} \mid \mathbf{z}_c) \mid$

$\boldsymbol{\theta}^{(k)}, \mathbf{z}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$, em que $Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = Q_{1i}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) + Q_{2i}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$, com

$$\begin{aligned} Q_{1i}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) &= -\frac{1}{2}log(|D(\phi)|) - \frac{1}{2}[\widehat{u_i}^{(k)}(\mathbf{z}_i - \mathbf{a})^\top D^{-1}(\phi)(\mathbf{z}_i - \mathbf{a}) \\ &\quad -2\hat{ux}_i^{(k)}(\mathbf{z}_i - \mathbf{a})^\top D^{-1}(\phi)\mathbf{b} + \widehat{ux^2}_i^{(k)}\mathbf{b}^\top D^{-1}(\phi)\mathbf{b}] \end{aligned} \tag{3.15}$$

e

$$\begin{aligned} Q_{2i}(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) &= -\frac{1}{2}log(\nu_x^2) - \frac{1}{2\nu_x^2}[\widehat{ux_i{}^2}^{(k)} + \mu_x^2\widehat{u_i}^{(k)} + \tau_x^2\widehat{ut_i{}^2}^{(k)} \\ &\quad -2\widehat{ux_i}^{(k)}\mu_x - 2\widehat{utx_i}^{(k)}\tau_x + 2\mu_x\tau_x\widehat{ut_i}^{(k)}] \end{aligned} \tag{3.16}$$

e esses cálculos requerem as expressões de $\widehat{u_i}^{(k)} = E[k^{-1}(U_i) \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$, $\widehat{ut_i}^{(k)} = E[k^{-1}(U_i)t_i \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$, $\widehat{ut_i{}^2}^{(k)} = E[k^{-1}(U_i)t_i^2 \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$, $\widehat{ux_i}^{(k)} = E[k^{-1}(U_i)x_i \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$, $\widehat{ux_i{}^2}^{(k)} = E[k^{-1}(U_i)x_i^2 \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$ e $\widehat{utx_i}^{(k)} = E[k^{-1}(U_i)t_ix_i \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$, os quais podem ser facilmente avaliados por

$$\begin{aligned} \widehat{ut_i}^{(k)} &= \widehat{u_i}^{(k)}\widehat{\mu}_{Ti}^{(k)} + \widehat{M}_T^{(k)}\widehat{\eta_{1i}}_i^{(k)},\ \widehat{ut_i{}^2}^{(k)} = \widehat{u_i}^{(k)}\widehat{\mu}_{T_i}^{2(k)} + \widehat{M}_T^{2(k)} + \widehat{M}_T^{(k)}\widehat{\mu}_{T_i}^{(k)}\widehat{\eta_{1i}}_i^{(k)}, \\ \widehat{utx_i}^{(k)} &= \widehat{r}_i^{(k)}\ \widehat{ut_i}^{(k)} + \widehat{s}^{(k)}\ \widehat{ut_i{}^2}^{(k)}\widehat{ux_i}^{(k)} = \widehat{r}_i^{(k)}\ \widehat{u_i}^{(k)} + \widehat{s}^{(k)}\ \widehat{ut_i}^{(k)}, \\ \widehat{ux_i{}^2} &= \widehat{T_x}^{2(k)} + \widehat{r}_i^{2(k)}\ \widehat{u_i}^{(k)} + 2\widehat{r}_i^{(k)}\widehat{s}^{(k)}\ \widehat{ut_i}^{(k)}\ + \widehat{s}^{2(k)}\ \widehat{ut_i{}^2}^{(k)}, \end{aligned}$$

em que $\widehat{\eta_{1i}}^{(k)} = E[\kappa^{-1/2}(U_i)W_{\Phi_1}(\frac{\kappa^{-1/2}(U_i)\widehat{\mu}_{T_i}}{\widehat{M_T}})|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}_i]$, with $W_\Phi(\gamma) = \phi_1(\gamma)/\Phi(x)$, $\gamma \in \mathbb{R}$, $\widehat{M_T}^2 = [1 + \widehat{\tau_x}^2\widehat{\mathbf{b}}^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu_x}^2\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}\widehat{\mathbf{b}}]^{-1}$, $\widehat{\mu}_{T_i} = \widehat{\tau}_x\widehat{M}_T^2\widehat{\mathbf{b}}^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu_x}^2\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}(\mathbf{y}_i - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}\widehat{\mu_x})$, $\widehat{T_x}^2 = \widehat{\nu_x}^2[1 + \widehat{\nu_x}^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}}]^{-1}$, $\widehat{r}_i = \widehat{\mu}_x + \widehat{T_x}^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})(\mathbf{y}_i - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}\widehat{\mu_x})$ e $\widehat{s} = \widehat{\tau}_x(1 - \widehat{T_x}^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}})$, com todas essas quantidades avaliadas em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$.Em cada passo, as esperanças condicionais $\widehat{u_i}$ e $\widehat{\eta_{1i}}$ podem ser facilmente derivadas do resultado dado na Proposição 8. Para as distribuições skew-t e skew-normal contaminada da classe MESN nós temos expressões computacionalmente atrativas e essas podem ser facilmente implementadas. Entretanto, para o caso da skew-slash, pode ser empregada a integração de Monte Carlo, a qual equivale ao algoritmo conhecido como EM com integração de Monte Carlo.

Os passos CM, como em (2.25), então maximizam condicionalmente $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ com respeito a $\boldsymbol{\theta}$, obtendo uma nova estimativa de $\widehat{\boldsymbol{\theta}}^{(k+1)}$, como descrito abaixo:

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}}^{(k+1)} &= \overline{\mathbf{z}}_u^{(k)} - \overline{x}_u^{(k)}\widehat{\boldsymbol{\beta}}^{(k)}, \quad \widehat{\boldsymbol{\beta}}^{(k+1)} = \frac{\sum_{i=1}^{n} \widehat{ux_i}^{(k)}(\mathbf{z}_i - \overline{\mathbf{z}}_u^{(k)})}{\sum_{i=1}^{n} \widehat{ux_i^2}^{(k)} - n\overline{\widehat{u}^{(k)}}\overline{x}_u^{2(k)}}, \\
\widehat{\phi}_1^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{u_i}^{(k)}X_i^2 - 2\widehat{ux_i}^{(k)}X_i + \widehat{ux_i^2}^{(k)}), \\
\widehat{\phi}_{j+1}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{u_i}^{(k)}z_{ij}^2 + \widehat{u_i}^{(k)}\alpha_j^{2(k)} + \beta_j^{2(k)}\widehat{ux_i^2}^{(k)} - 2\widehat{u_i}^{(k)}\alpha_j^{(k)}z_{ij} - 2y_{ij}\beta_j^{(k)}\widehat{ux_i}^{(k)} \\
&\qquad +2\alpha_j^{(k)}\beta_j^{(k)}\widehat{ux_i}^{(k)}), \; j = 1, \ldots, r, \\
\widehat{\mu_x}^{(k+1)} &= \overline{x}_u^{(k)} - \widehat{\tau_x}^{(k)}\overline{t}_u^{(k)}, \quad \widehat{\nu_x}^{2(k)} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{ux_i^2}^{(k)} - \widehat{\mu_x}^{(k)}\widehat{ux_i}^{(k)}) - \widehat{\tau_x}^{(k)}\frac{1}{n}\sum_{i=1}^{n}\widehat{utx_i}^{(k)} \text{ e} \\
\widehat{\tau_x}^{(k+1)} &= \frac{\sum_{i=1}^{n}(\widehat{utx_i}^{(k)} - \overline{x}_u^{(k)}\widehat{ut_i}^{(k)})}{\sum_{i=1}^{n}(\widehat{t_i^2}^{(k)} - \overline{t}_u^{(k)}\widehat{ut_i}^{(k)})}, \\
\lambda_x^{(k+1)} &= \tau_x^{(k+1)}/\nu_x^{(k+1)} \text{ e } \phi_x^{(k+1)} = \tau_x^{2(k+1)} + \nu_x^{2(k+1)},
\end{aligned}
$$

em que, $\overline{\mathbf{z}}_u^{(k)} = \dfrac{\sum_{i=1}^{n}\widehat{u_i}^{(k)}\mathbf{z}_i}{\sum_{i=1}^{n}\widehat{u_i}^{(k)}}, \overline{x}_u^{(k)} = \dfrac{\sum_{i=1}^{n}\widehat{ux_i}^{(k)}}{\sum_{i=1}^{n}\widehat{u_i}^{(k)}}, \overline{t}_u^{(k)} = \dfrac{\sum_{i=1}^{n}\widehat{ut_i}^{(k)}}{\sum_{i=1}^{n}\widehat{u_i}^{(k)}}$ e $\overline{\widehat{u}^{(k)}} = \dfrac{1}{n}\sum_{i=1}^{n}\widehat{u_i}^{(k)}$.

As iterações do algoritmo acima são repetidas até a convergência apropriada que satisfaça a regra, por exemplo, de $||\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}||$ seja pequena suficiente. Uma crítica ao procedimento do tipo EM é que ele tende a ficar preso em modas locais. Um modo conveniente de se esquivar de tal limitação é tentar várias iterações EM com vários valores iniciais que representativos ao espaço parametrico. Se existirem várias modas, é possível encontrar a moda global comparando suas massas relativas e os valores da log-verossimilhança. Note que quando $\kappa(U_i) = 1$, $i = 1, \ldots, n$, (uma variável aleatória degenerada) o passo M se reduz as equações obtidas em Lachos, et al. (2005) sob a distribuição skew-normal e quando $\lambda_x = 0$ (ou $\tau_x = 0$) o passo M se reduz as equações obtidas por Bolfarine e Galea-Rojas (1995). Entretanto, quando $\kappa(U_i) = \frac{1}{U_i}, U_i \sim Gamma(\nu/2, \nu/2)$ e $\lambda_x = 0$, o passo M se reduz as equações obtidas por Bolfarine e Galea-Rojas (1996).

A seguir apresentamos um estudo sobre a utilização de variáveis latentes e da estimação das quantidades x_i via uma inferência Bayesiana empírica.

3.4 Estimação da Variável Latente

Um dos pontos iniciais de qualquer pesquisa cientifíca é a definição das variáveis de interesse. A definição *do que medir* e *como* medir está intrinsicamente relacionada aos objetivos de uma pesquisa. Em alguns casos, o *como* medir é um problema menor. Por exemplo, se um cientista deseja avaliar o efeito de um medicamento do controle de diabetes, ele pode medir a taxa de glicemia no sangue antes e depois do tratamento com o medicamento e, a partir daí tirar suas conclusões. No entanto, há situações em que o interesse da pesquisa não está ligado a variáveis tão concretas. Quando nos encontramos nesse tipo de situação é comum a utilização do que chamamos de variáveis latentes.

O modelo de variáveis latentes mede o relacionamento entre as variáveis latentes (ou fatores), e os indicadores. A forma funcional das equações estruturais é freqüentemente suposta como aditiva.

A implementação do modelo de variáveis latentes envolve quatro etapas. Na primeira, deve-se escolher quantas e quais variáveis latentes serão empregadas, sendo que essa escolha é baseada, principalmente, na experiência do pesquisador e na compreensão do problema em estudo (técnica amplamente utilizada para obtenção de prioris em inferência bayesiana). A segunda etapa é dedicada a identificação dos indicadores, para cada variável latente, sendo que essa escolha também é baseada na experiência dos pesquisadores em relação ao problema em estudo. A terceira etapa envolve a determinação das equações estruturais e das equações de medição para o modelo de variáveis latentes. Na quarta etapa, o desenvolvimento da função da máxima verossimilhança e a estimação dos parâmetros do modelo são realizados, conforme explicado anteriormente.

Para a obtenção de prioris em inferência Bayesiana em geral são utilizadas informações dadas por um especialista ou por um pesquisador experiente no assunto. Outro método utilizado é o de inferência bayesina empírica que consiste em obter uma priori a partir do próprio conjunto de dados.

Nessa seção, nós consideramos uma inferência bayesiana empírica para a variável latente, isto é utilizaremos prioris a partir dos próprios dados coletados, que é útil para estimar as quantidades x_i.

De Lachos et al. (2008), o estimador de x_i que minimiza o erro quadrático médio (EQM) é obtido através da média condicional de x_i dado $\mathbf{z}_i$ é dado por

$$\widehat{x_i} = E[x_i|\mathbf{z}_i] = \mu_x + \Lambda_x a_i + \frac{\lambda_x \Lambda_x}{\sqrt{1+\lambda_x^2 \Lambda_x}} \eta_{-1i}, \;\; i = 1, \ldots, n, \tag{3.17}$$

em que $\eta_{-1i} = E[k^{1/2}(U_i) W_\Phi(k^{1/2}(U_i)A_i)|\mathbf{z}_i]$. Expressões explícitas de η_{-1i} para as distribuições skew-t, skew-slash e skew-normal contaminada podem ser obtidas com o auxílio da Proposição 8. Na prática, o estimador de Bayes de x_i, denotado por $\widehat{x_i}$, pode ser obtido pela substituição do estimador de máxima verossimilhança $\boldsymbol{\theta}$ em (3.17).

A MESN-MEM dada em (3.4)-(3.6) consideradas duas fontes de variação, que podem gerar valores aberrantes em componentes de erro assim como na componente da variável latente (veja, por exemplo, Pinheiro et al., 2001 e Osorio et al., 2009). Deste modo, substituindo $\boldsymbol{\theta}$ em $x_i(\boldsymbol{\theta})$ com nossos estimadores atuais, nós obtemos a decomposição a seguir para a distância de Mahalanobis

$$\begin{aligned} d_i(\widehat{\boldsymbol{\theta}}) &= (\mathbf{z}_i - \widehat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{z}_i - \widehat{\boldsymbol{\mu}}) \\ &= \widehat{\mathbf{e}_i}^\top D^{-1}(\widehat{\boldsymbol{\phi}}) \widehat{\mathbf{e}_i} + \frac{1}{\widehat{\phi_x}} \widehat{\mu_{xi}^2} = \widehat{d_{ei}} + \widehat{d_{xi}}, \end{aligned} \tag{3.18}$$

em que $\widehat{\mathbf{e}_i} = \mathbf{z}_i - \widehat{\boldsymbol{\mu}} - \mathbf{b}\widehat{\mu_{xi}}$ são os resíduos estimados, com $\widehat{\mu_{xi}} = \widehat{\Lambda_x}\widehat{a_i}$. A equação (3.18) provê um modo simples para computar a distância de Mahalanobis. As distâncias estimadas d_i, d_{xi} e d_{ei} provém estatísticas dignósticas utéis para identificar observações com valores aberrantes.

Este resultado é interessante, pois permite avaliar os modelos estatísticos na prática.

Substituindo as estatísticas de máxima verossimilhança em $\widehat{\boldsymbol{\mu}}$ e $\widehat{\boldsymbol{\Sigma}}$ na distância de Mahalanobis d_i, podemos avaliar os ajustes dos modelos através da construção de envelopes (Montenegro et al., 2009a). Além disso através de gráficos da distância de Mahalanobis e considerando como *benchmark* o quantil υ da distribuição da forma quadrática d_i, podemos identificar valores aberrantes. Sob nossa aproximação, temos que $d_i \sim \chi^2_p$, para ao caso SN. Deste modo, nós podemos usar como pontos de corte o quantil $\upsilon = \chi^2_p(\xi)$, onde $0 < \xi < 1 <$, para identificar valores aberrantes. De Lachos et al. (2008) as propriedades relatadas para a distância de Mahalanobis e com o auxílio do Corolário 4 temos que $d_i \sim pF(p,\nu)$ para ST_1, $Pr(d_i \leq \upsilon) = Pr(\chi^2_p \leq \upsilon) - \frac{2^\nu\Gamma(\nu+p/2)}{\upsilon^\nu\Gamma(p/2)}Pr(\chi^2_{2\nu+p} \leq \upsilon)$ para a SSL_1 e $Pr(d_i \leq \upsilon) = \nu Pr(\chi^2_p \leq \gamma\upsilon) + (1-\nu)Pr(\chi^2_p \leq \upsilon)$ para a SCN.

No próximo capítulo aplicaremos os resultados aqui apresentados onde nós derivamos as medidas de influência local, dado o estimador de máxima verossimilhança $\widehat{\boldsymbol{\theta}}$ para obtermos a matriz Hessiana que será de grande utilidade para prosseguirmos nosso estudo.

3.5 A Matriz de Informação Observada

De (3.13) e a notação usada em (3.12), e depois de algumas manipulações algébricas nós encontramos que a função de log-verossimilhança pode ser escrita como:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{3.19}$$

em que $\ell_i(\boldsymbol{\theta}) = log2 - \frac{p}{2}log2\pi - \frac{1}{2}log|\boldsymbol{\Sigma}| + log(K_i)$,$i = 1,\ldots,n$. Desta forma, as segundas derivadas de $\ell(\boldsymbol{\theta})$ com respeito a $\boldsymbol{\theta}$ é dada por

$$\mathbf{L} = \sum_{i=1}^{n} \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = -\frac{n}{2}\frac{\partial^2 log|\Sigma|}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \sum_{i=1}^{n}\frac{1}{K_i^2}\frac{\partial K_i}{\partial\boldsymbol{\theta}}\frac{\partial K_i}{\partial\boldsymbol{\theta}^\top} + \sum_{i=1}^{n}\frac{1}{K_i}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},$$

em que

$$\frac{\partial K_i}{\partial\boldsymbol{\theta}} = I_i^\phi(\frac{p+1}{2})\frac{\partial A_i}{\partial\boldsymbol{\theta}} - \frac{1}{2}I_i^\phi(\frac{p+2}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}},$$

e

$$\begin{aligned}
\frac{\partial^2 K_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{1}{4} I_i^\phi(\frac{p+4}{2}) \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} - \frac{1}{2} I_1^\phi(\frac{p+2}{2}) \frac{\partial^2 d_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\
&-\frac{1}{2} I_i^\phi(\frac{p+3}{2})(\frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^T} + \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top}) - I_i^\phi(\frac{p+3}{2}) A_i \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top} \\
&+ I_i^\phi(\frac{p+1}{2}) \frac{\partial^2 A_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},
\end{aligned}$$

com,

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

e

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

Note que nós também podemos escrever $K_i = I_i^\Phi(\frac{p}{2})$. Da substituição direta de $H(u_i)$ nas integrais acima, imediatamente obtemos os seguintes resultados para as respectivas distribuições consideradas:

- *skew-t.*

$$\begin{aligned}
I_i^\Phi(w) &= \frac{2^w \nu^{\nu/2} \Gamma(w + \nu/2)}{\Gamma(\nu/2)(\nu + d_i)^{\nu/2+w}} T_1\left(\frac{A_i}{(d_i+\nu)^{1/2}} \sqrt{2w+\nu} | 0, 1, 2w+\nu\right) \quad \text{e} \\
I_i^\phi(w) &= \frac{2^w \nu^{\nu/2}}{\sqrt{2\pi}\Gamma(\nu/2)} (\frac{1}{d_i + A_i^2 + \nu})^{\frac{\nu+2w}{2}} \Gamma(\frac{\nu+2w}{2}).
\end{aligned}$$

- *skew-slash.*

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

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

- *skew-normal contaminada.*

$$
\begin{aligned}
I_i^{\Phi}(w) &= \sqrt{2\pi}\{\nu\gamma^{w-1/2}\phi_1(d_i|0,\frac{1}{\gamma})\Phi(\gamma^{1/2}A_i)+(1-\nu)\phi_1(d_i|0,1)\Phi(A_i)\} \quad \text{e} \\
I_i^{\phi}(w) &= \nu\gamma^{w-1/2}\phi_1(d_i+A_i^2|0,\frac{1}{\gamma})+(1-\nu)\phi_1(d_i+A_i^2).
\end{aligned}
$$

As derivadas de log $|\mathbf{\Sigma}|$, d_i e A_i envolvem manipulações algébricas tediosas mas não complicadas e o resultado final pode ser encontrado no Apêndice B. Intervalos de confiança assintóticos e testes de estimação de máxima verossimilhaça podem ser obtidos usando essa matriz, isto é, se $\mathbf{J} = -\mathbf{L}$ denota a matriz de informação observada da log-verossimilhança marginal $\ell(\boldsymbol{\theta})$ da MESN-MEM, então intervalos de confiança assintóticos e testes de hipóteses para os parêmetros de $\boldsymbol{\theta}$ são obtidos assumindo que o estimador de máxima verossimilhança tem aproximadamente uma distribuição $N_{3p+1}(\boldsymbol{\theta}, \mathbf{J}^{-1})$. Na prática, $\mathbf{J}$ é usualmente desconhecida e tem que ser substituida pelo estimador de máxima verossimilhança $\widehat{\mathbf{J}}$, que é, a matriz $\widehat{\mathbf{J}}$ avaliada nos estimadores de máxima verossimilhança $\hat{\boldsymbol{\theta}}$. Falando de uma forma mais geral, para modelos como os da Proposição 10, a matriz de informação pode ser derivada dos resultados aqui apresentados.

4 *Influência Local*

4.1 Introdução

Modelos estatísticos que são aplicados para analisar as características essenciais de um conjunto de dados são quase sempre descrições aproximadas dos mais complicados processos. Numa análise estatística é importante levar em consideração a estabilidade dos resultados obtidos e das inferências de um modo geral, com relação a possíveis perturbações nos dados ou no modelo estatístico.

Enquanto as análises de resíduos são feitas para investigar problemas com o modelo ajustado, uma análise diagnóstico é feita assumindo o modelo como correto, e investigando a robustez das conclusões com pequenas perturbações nos dados.

Em situações práticas é de interesse fazer a análise dos dados com o intuito de verificar se alguma observação em particular controla propriedades importantes na estimação dos parâmetros de interesse. Assim, é importante fazer um estudo de detecção de observações atípicas e avaliação da exclusão de algumas dessas observações na estimação dos parâmetros do modelo. Esse problema tem sido estudado no contexto de modelo de regressão linear por muitos autores, dentre os quais citamos Cook e Weisberg (1982), Cook (1986) e Chatterjje e Hadi (1988).

Um método mais geral para avaliar a influência de determinadas observações sob certo esquema de perturbação é proposto por Cook (1986), conhecido na literatura como *método de influência local*, o qual permite avaliar a influência que pequenas perturbações podem exercer sobre as componentes do modelo, tais como estimativa dos parâmetros, e outros resultados da análise. Apesar da abrangente aplicabilidade do método proposto por Cook (1986), em alguns casos torna-se difícil ou as vezes intratável aplicá-lo diretamente em modelos de interesse prático, sendo necessário o uso de métodos alternativos. Poon e Poon (1999) propõe um método de influência local baseado na curvatura normal conforme, que é uma medida padronizada da curvatura normal utilizada por Cook (1986). Já Loynes (2001) propõe uma nova medida de influência local que chama de "propagação de curvatura" (sendo essa uma tradução para *spreed of curvature*) baseado num estudo probabilístico da curvatura normal.

Na literatura, alguns trabalhos percursores em influência e diagnóstico são dados para modelos com erros na variável e foram desenvolvidos por Kelly (1984) que propõe um procedimento de diagnóstico baseado em uma função de influência. Outro método foi proposto por Tanakaet et al. (1991) que usam a função de influência de Hampel para desenvolver métodos de análise de influência de uma observação individual em estruturas de covariância. Abdullah (1995) apresenta vários métodos para detectar observações influentes no modelo de regressão funcional com erros nas variáveis. Entretanto essas medidas de diagnóstico avaliam o efeito dos parâmetros de regressão estimados depois de excluir uma única observação do conjunto de dados.

Inspirados pelo poder e pela amplitude da aplicabilidade do algoritmo EM, Zhu e Lee (2001) desenvolveram uma alternativa interessante à proposta de Cook (1986) e as demais metodologias até o momento, baseada na função de verossimilhança completa resultante da implementação do algoritmo-EM (função Q-afastamento) e permite uma rápida aplicação do método, especialmente em modelos complicados.

Antes de nos aprofundarmos no estudo de influência local seguimos com uma breve descrição sobre as diferenças entre robustez com relação a valores aberrantes e diagnósticos, que desempenham papel fundamental para a melhor compreensão do leitor nesse trabalho.

4.2 Robustez com Relação Valores Aberrantes e Diagnósticos

Um problema comum em análises de dados é a presença de pontos aberrantes. Entende-se por ponto aberrante (ou atípico) por uma observação que apresenta um comportamento atípico em relação ao restante dos dados.

Quando se tem uma única variável, o valor aberrante caracteriza-se por assumir um valor muito mais alto ou muito mais baixo que os demais.

Há várias causas possíveis para a ocorrência de valores aberrantes:

a) Erros de medida;

b) Erros de transcrição ou digitação;

c) Erro ao considerar uma unidade amostral que não pertence à população de interesse, por exemplo, num estudo com portadores de uma determinada doença, observações de pacientes não portadores, erroneamente incluídos no estudo por erro de diagnóstico, podem resultar em valores aberrantes;

d) Variabilidade normal dos dados.

Com exceção do item (d), todas as demais causas estão relacionadas a erros e, consequentemente, a identificações de valores aberrantes nessas situações exige a sua retirada da amostra final.

A escola de robustez com relação a valores aberrantes e a escola de diagnósticos (veja, por exemplo, Besley e Welsch (1980), Cook e Weisberg (1982), e Cook (1986)) consideram o problema de valores aberrantes de duas perspectivas diferentes. Talvez seja por isso que, até recentemente, temos poucas interações entre ambos os campos de pesquisa. A escola de diagnósticos tenta desenvolver procedimentos para identificar observações influentes ou atípicas. Depois dessas observações serem detectadas é possível que a correção seja feita através da atribuição de pesos menores a essas observações ou até mesmo as descartando por completo. A utilização desses métodos resultam em resultados finais mais robustos.

O objetivo principal da escola de diagnósticos, contudo, é detectar, isto é, identificar observações aberrantes e influentes. Numerosas dificuldades podem surgir na etapa de identificação de valores aberrantes. A mais notória é o "*masking effect*" (veja, por exemplo, Rousseeuw e Van Zomeren (1990)), em uma possível tradução livre para o português seria o efeito de mascarar. Se existem vários valores aberrantes agrupados em uma região do espaço amostral longe dos demais dados, a maior parte dos métodos de detecção de valores aberrantes não robustos falha em identificar essas observações como valores aberrantes. Em outras palavras, os valores aberrantes mascaram uns aos outros. A escola de estatísticas robustas, em contraste, tenta desenvolver procedimentos que são insensíveis a observações anômalas. Seu objetivo principal é "proteger" (veja Huber (1991)). Procedimentos de estatísticas robustas automaticamente incluem a identificação de valores aberrantes na etapa de estimação. As observações discordantes podem, freqüentemente, ser identificadas em uma segunda etapa como um subproduto das primeiras análises. Isso facilita a tarefa para o pesquisador. Além disso, Procedimentos de estatísticas robustas são concebidos de tal forma que eles podem lidar com o "*masking effect*" descrito acima. Comumente é dito que as aproximações mencionadas acima não precisam ser mutuamente exclusivas, porém Fieller(1993) contesta essa afirmação.

Assim concluimos que em primeiro lugar, métodos robustos para valores aberrantes podem fornecer ferramentas excelentes para a identificação de observações aberrantes e/ou influentes (veja Rousseeuw e Van Zomeren (1990) e Fung (1993)). Em segundo lugar, cada estatístico, incluindo os da escola de estatísticas robustas, pode estar interessado em saber se existem quaisquer observações que não se ajustam dentro de um modelo e se existe alguma razão particular para isso ocorrer no ajuste. Isso, em geral, leva para um melhor entendimento do fenômeno que é estudado. Isto é, não notamos que alguma das escolas tenha se sobreposto à outra. As duas abordagens podem ser consideradas como complementares (veja, por exemplo, Huber (1991) e Davies e Gather (1993)).

Na Seção 4.3 é apresentado uma breve introdução do método de influência local proposto por Cook(1986) e na Seção 4.4 a aproximação proposta por Zhu e Lee (2001).

4.3 A Influência Local de Cook

Dentre os métodos mais utilizadas, na prática, para medir influência em modelos de regressão, a análise da influência local sugerida por Cook (1986) é com certeza uma das mais destacadas. Nessa seção veremos uma breve introdução dessa metodologia que na seção seguinte será de grande utilidade para um melhor entendimento do método de Zhu e Lee (2001).

Sabemos que a detecção de observações influentes é um passo importante na análise de dados. Existem várias alternativas para avaliar a influência dos dados e/ou modelo pertubado sobre estimadores dos parâmetros. Por exemplo, ver Cook e Weisberg (1982) e Chatterjje e Hadi (1988).

Eliminação de observações é comum quando pretendemos avaliar o efeito de uma observação sob o processo e estimação. A análise de influência global a partir do efeito de observações é avaliado por eliminar esta observação do conjunto de dados. Num método mais geral, a influência local proposta por Cook (1986) é baseada em avaliar a mudança nos resultados da análise quando incorporamos "pequenas pertubações" ao modelo. Dentro desse contexto, podemos perturbar a variável resposta, a variável explicativa, a matriz de covariância, etc. A abordagem inicial é baseada na análise do *likehood displacement*, que é usualmente traduzido como "afastamento da verossimilhança" (ver Cook e Weisberg, 1982 e Cook, 1986).

$$LD(\boldsymbol{\omega}) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\},$$

em que $\ell(\boldsymbol{\theta})$ é a função de de log-verossimilhança do modelo postulado, $\boldsymbol{\Theta} \subseteq \mathbb{R}^k$ vetor de parâmetros desconhecidos e $k = 3p + 1$ número de parâmetros, $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ é a log-verossimilhança do modelo pertubado, $\boldsymbol{\omega}$ é um vetor de perturbações de dimensões $q \times 1$ em um subconjunto aberto $\Omega \subset \mathbb{R}^q$, com $q < k$. Quando avaliamos $\boldsymbol{\omega}$ em $\boldsymbol{\omega}_0$ para $\boldsymbol{\omega}_0 \subset \Omega$ então temos os dados não perturbados. $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotam respectivamente, os estimadores de máxima verossimilhança (EMV) baseados em $\ell(\boldsymbol{\theta})$ e $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. A ideia de influência local (Cook, 1986) consiste em caracterizar o comportamento de $LD(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$. O procedimento consiste em selecionar uma direção unitária $\mathbf{d}$, $||\mathbf{d}|| = 1$, e então, considerar o gráfico de $LD(\boldsymbol{\omega}_0 + a\mathbf{d})$ contra $a \in \mathbb{R}$. Este gráfico é chamado de linha projetada. Note que $LD(\boldsymbol{\omega}_0) = 0$, ou seja, $LD(\boldsymbol{\omega}_0 + a\mathbf{d})$ tem um mínimo local em $a = 0$. Cada linha ajustada pode ser caracterizada pela curvatura normal $C_d(\boldsymbol{\theta})$ em torno de $a = 0$. A sugestãó é considerar a direção $\mathbf{d}_{max}$ correspodente à maior curvatura $C_{dmax}(\boldsymbol{\theta})$. O gráfico de índice de $\mathbf{d}_{max}$ pode mostrar aquelas observações que sob pequenas perturbações exercem uma notável influência sobre $LD(\boldsymbol{\omega})$. Cook (1986) mostrou que a curvatura normal na direção de $\mathbf{d}$ toma a forma

$$C_1(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta} \mathbf{d}|, \tag{4.1}$$

em que $-\mathbf{L} = \dfrac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$ é a matriz de informação de Fisher observada para o modelo $(\boldsymbol{\omega} = \boldsymbol{\omega}_0)$ e $\boldsymbol{\Delta}$ é a matriz de dimensão k x q com elementos

$$\Delta_{rs} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_r \partial \omega_s}, \tag{4.2}$$

avaliados em $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $r = 1, ..., k$ e $s = 1, ..., q$. Portanto, a maximização de (4.1) é equivalente a encontrar o maior autovalor absoluto C_{dmax} da matriz $\mathbf{B} = -\boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta}$ e $\mathbf{d}_{max}$ é o autovalor normalizado correspondente.

Desta forma, $\mathbf{d}_{max}$ pode ser utilizado como uma ferramenta útil na análise de disgnóstico. O gráfico de elementos de $|\mathbf{d}_{max}|$ pode mostrar qual o tipo de perturbação tem a maior influência em $LD(\boldsymbol{\omega})$ em torno de $\boldsymbol{\omega}_0$. Cook (1986) propõe examinar as componentes de $\mathbf{d}_{max}$, independente do valor de C_{dmax}, uma vez que pode indicar observações que são conjuntamente influentes.

Em algumas situações é de interesse avaliar a influência sobre um subconjunto $\boldsymbol{\theta}_1$ de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. Por exemplo, podemos ter interesse em $\boldsymbol{\theta}_1 = (\mu_x, \boldsymbol{\beta}^\top)^\top$ ou em $\boldsymbol{\theta}_1 = \lambda_x$. Em tais situações, a curvatura na direção $\mathbf{d}$ é dada por

$$C_1(\boldsymbol{\theta}_1) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^\top (\mathbf{L}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta} \mathbf{d}|,$$

em que

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{22}^{-1} \end{pmatrix},$$

e $\mathbf{L}_{22}$ é obtido na partição de $\mathbf{L}$, segundo a partição de $\boldsymbol{\theta}$. O autovetor $\mathbf{d}_{max}$ corresponde ao maior autovalor absoluto da matriz $\mathbf{B} = \boldsymbol{\Delta}^\top (\mathbf{L}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}$.

Escobar e Meeker (1992) (ver também Verbeke e Molenberghs, 2000) consideram que uma outra direção é dada por $\mathbf{d} = \mathbf{e}_{in}$, um vetor de dimensão n x 1 de zeros com um na i-ésima posição. Neste caso a curvatura normal, chamada de influência local total do indivíduo i, é dada por $C_i = 2|\mathbf{e}_{in}^\top \mathbf{B}\mathbf{e}_{in}| = 2|b_{ii}|$ em que b_{ii} é o i-ésimo elemento diagonal de $\mathbf{B}$, $i = 1, ..., n$.

Com finalidade de comparar a influência local e global, utilizamos uma das medidas mais utilizadas para avaliar a influência de uma observação, via eliminação, é a distância de Cook (D_i) (Cook e Weisberg, 1982) e o afastamento da verossimilhança (LD_i) (Cook, 1977), definidas, respectivamente, por

$$D_i = \frac{(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)})^\top (-\mathbf{L})(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)})}{k} \tag{4.3}$$

$$LD_i = 2[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{(i)})], \tag{4.4}$$

em que $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{(i)}$ denotam, respectivamente, a estimativa deo vetor $\boldsymbol{\theta}$ com todos os dados da amostra e com a eliminação da observação i, $i = 1, ..., n$. Veja Zhao e Lee (1998) para mais detalhes.

Como já comentamos nas seções anteriores muitas vezes se torna impossível a utilização do método e nesse sentido Zhu e Lee (2001) apresentam uma metodologia para resolver essa "falha" encontrada na influência local. Essa nova metodologia é apresentada na próxima seção.

4.4 Método de Influência Local

Existem basicamente duas abordagens para detectar observações que influenciam seriamente nos resultados de uma análise estatística. A primeira abordagem consiste na eliminação de casos em que o impacto na estimação devido à eliminação de uma observação é diretamente disponível por algumas medidas, tais como, o afastamento pela verossimilhança e a distância de Cook (1977). A segunda abordagem é a de influência local (Cook, 1986) em que esse impacto é estudado através de diferentes esquemas de perturbações. Uma alternativa interessante à proposta de Cook (1986) e útil nos casos em que torna-se difícil ou impossível aplicar diretamente os métodos apresentados por Cook (1977, 1986), foi proposto por Zhu e Lee (2001) e detalhamos a seguir.

Considerando o vetor de perturbação $\omega = (\omega_1, ..., \omega_g)^\top$ variando em uma região aberta $\boldsymbol{\Omega} \in \mathbb{R}^g$. Seja $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \mathbf{Y}_c)$, $\boldsymbol{\theta} \in \mathbb{R}^h$, a função de log-verossimilhança dos dados completos do modelo perturbado. Assumimos que existe um vetor de não perturbação $\boldsymbol{\omega}_0$ tal que $\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega}_0 \mid \mathbf{Y}_c) = \ell_c(\boldsymbol{\theta} \mid \mathbf{Y}_c)$, para todo $\boldsymbol{\theta}$. Denotemos $\widehat{\boldsymbol{\theta}}(\omega)$ como a função que maximiza $Q(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \mathbf{Y}_c) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}]$. O gráfico de influência é definido como $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$, onde $f_Q(\boldsymbol{\omega})$ é a função Q-afastamento definida como

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

É fácil observar que valores mais próximos de zero indicam que os dados são robustos a perturbação aplicada, quanto mais distante de zero maior o influência da perturbação aplicada.

Segundo a ideia de Cook (1986) e Zhu e Lee (2001), a curvatura normal $C_{f_Q,\mathbf{d}}$ de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$ na direção de algum vetor unitário $\mathbf{d}$ pode ser resumida pelo comportamento local da função Q-afastamento. Pode ser mostrado que (Zhu e Lee, 2001)

$$C_{f_{Q,\mathbf{d}}} = -2\mathbf{d}^\top \ddot{Q}_{\boldsymbol{\omega}_0}\mathbf{d} \text{ e } -\ddot{Q}_{\boldsymbol{\omega}_0} = \Delta_{\boldsymbol{\omega}_0}^\top\{-\ddot{Q}_\theta(\widehat{\boldsymbol{\theta}})\}^{-1}\Delta_{\boldsymbol{\omega}_0},$$

em que $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \dfrac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ e $\Delta_{\boldsymbol{\omega}} = \dfrac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}$.

Como em Cook (1986), a expressão $-\ddot{Q}_{\boldsymbol{\omega}_0}$ é importante para detectar observações localmente influentes. Uma descrição de $-\ddot{Q}_{\boldsymbol{\omega}_0}$, é dada por sua decomposição espectral

$$-2\ddot{Q}_{\boldsymbol{\omega}_0} = \sum_{m=1}^{p} \lambda_m \mathbf{e}_m \mathbf{e}_m^{'},$$

em que $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_1, \mathbf{e}_p)$ são os pares de autovalores e autovetores da matriz $-\ddot{Q}_{\boldsymbol{\omega}_0}$ com $\lambda_1 \geq \ldots \geq \lambda_p, \lambda_{p+1} = \ldots = \lambda_q = 0$ e $\mathbf{e}_1, \ldots, \mathbf{e}_p$ são elementos associados à base ortonormal. Assim, podemos expressar

$$C_{f_{Q,\mathbf{u}_l}} = \sum_{m=1}^{p} \lambda_m \mathbf{e}_{ml}^2,$$

em que $\mathbf{e}_{ml}$ é o l-ésimo componente de $\mathbf{e}_m \in \mathbb{R}^q$ e $\mathbf{u}_l$ é um vetor coluna em $\mathbb{R}^q$ com o l-ésimo componente igual a um, e todos os outros iguais a zero.

A estimativa dos casos influentes pode ser baseada na inspeção visual de $\{M(0)_l, l = 1, ..., g\}$ plotado contra o índice l. Pode-se obter $M(0)_l$ via

$$B_{f_{Q,\mathbf{u}_l}} = \frac{-2\mathbf{u}_l^\top \ddot{Q}\boldsymbol{\omega}_0 \mathbf{u}_l}{tr[-2\ddot{Q}\boldsymbol{\omega}_0]},$$

em que $\mathbf{u}_l$ é um vetor coluna como definido anteriormente. Duas importantes propriedades de $B_{f_{Q,\mathbf{u}_l}}$ são dadas pelo seguinte teorema.

Teorema 4.1

a) Invariância sob reparametrização de θ: se ψ: $\Theta \to \Psi$ é diferenciável com uma Jacobiana não singular, então $C_{f_{Q,\mathbf{u}_l}}$ e $B_{f_{Q,\mathbf{u}_l}}$ são invariantes com respeito a qualquer reparametrização $\psi = \psi(\theta)$.

b) Invariância sob reparametrização conformal de ω: seja $\phi = \phi(\omega)$ uma reparametrização conformal de ω. Então $B_{f_{Q,\mathbf{u}_l}}$ em qualquer direção de uma unidade em ω^0 é invariante com respeito a reparametrização conformal $\phi = \phi(\omega)$.

O Teorema acima nos apresenta duas importantes propriedades que $C_{f_{Q,\mathbf{u}_l}}$ e $B_{f_{Q,\mathbf{u}_l}}$ são invariantes com respeito a qualquer reparametrização de θ. Daí, os resultados obtidos da reparametrização conformal de ω, $B_{f_{Q,\mathbf{u}_l}}$ continuam invariantes.

Lee e Xu (2004) propõem usar $1/g + c * SM(0)$ como um *benchmark* para estabelecer o l-ésimo caso como influente, onde $c*$ é uma constante selecionada. Dependendo da aplicação real, $c*$ pode ser selecionada arbitrariamente e $SM(0)$ é o desvio padrão de $\{M(0)_l, l = 1, ..., g\}$. Veja Montenegro et al. (2009a) para mais detalhes.

Na próxima seção é apresentado o método para a obtenção da matriz hessiana, matriz essa vital para o estudo da influência local pelo método de Zhu e Lee.

4.5 A Matriz Hessiana $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$

Para obter a medida diagnóstico para a influência local de um esquema de perturbação local, é necessário computar $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$, onde $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ é o vetor de parâmetros. Daí, a matriz Hessiana é dada por $\ddot{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \ddot{Q}_i(\boldsymbol{\theta})$ com

$$\ddot{Q}_i(\boldsymbol{\theta}) = -\frac{\partial^2 Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_{1i}(\boldsymbol{\theta}_1) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_{2i}(\boldsymbol{\theta}_2) \end{pmatrix},$$

em que $\ddot{Q}_{1i}(\boldsymbol{\theta}_1) = -\partial^2 Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\top$ e $\ddot{Q}_{2i}(\boldsymbol{\theta}_2) = -\partial^2 Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top$. Portanto a matriz hessiana tem elementos dados por (veja Magnus e Neudecker, 1988)

$$\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top} &= -\widehat{u}_i \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi})\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top} &= -\widehat{ux}_i \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi})\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= -\widehat{ux^2}_i \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi})\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\alpha}^\top} &= \left(-\widehat{u}_i D(\mathbf{z}_i - \mathbf{a}) + \widehat{ux}_i D(\mathbf{b})\right) D^{-2}(\boldsymbol{\phi})\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\beta}^\top} &= \Big(-\widehat{ux}_i D(\mathbf{z}_i - \mathbf{a}) + \widehat{ux^2}_i D(\mathbf{b})\Big) D^{-2}(\boldsymbol{\phi})\mathbb{I}_{(p)}^\top, \\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} &= \big[\frac{1}{2}D(\boldsymbol{\phi}) - \widehat{u}_i D^2(\mathbf{z}_i - \mathbf{a}) + 2\widehat{ux}_i D(\mathbf{z}_i - \mathbf{a})D(\mathbf{b}) - \widehat{ux^2}_i D^2(\mathbf{b})\big] D^{-3}(\boldsymbol{\phi}), \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\mu_x} &= -\frac{1}{\nu_x^2}\widehat{u}_i, \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\phi_x} &= \frac{\lambda_x(1+\lambda_x^2)^{1/2}}{2\phi_x^{3/2}}\widehat{ut}_i + \frac{(1+\lambda_x^2)}{\phi_x^2}B_{1i}, \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\lambda_x} &= -\frac{2\lambda_x}{\phi_x}B_{1i} - \frac{(1+2\lambda_x^2)}{\phi_x^{1/2}(1+\lambda_x^2)^{1/2}}\widehat{ut}_i, \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\phi_x} &= \frac{1}{2\phi_x^2} - \frac{3\lambda_x(1+\lambda_x^2)^{1/2}}{4\phi_x^{5/2}}B_{2i} - \frac{(1+\lambda_x^2)}{\phi_x^3}B_{3i}, \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\lambda_x\partial\phi_x} &= \frac{(1+2\lambda_x^2)}{2\phi_x^{3/2}(1+\lambda_x^2)^{1/2}}B_{2i} + \frac{\lambda_x}{\phi_x^2}B_{3i}, \\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2, \boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\lambda_x\partial\lambda_x} &= \frac{(1-\lambda_x^2)}{(1+\lambda_x^2)^2} - \frac{1}{\phi_x}B_{3i} - \widehat{ut^2}_i - \frac{\lambda_x(3+2\lambda_x^2)}{\phi_x^{1/2}(1+\lambda_x^2)^{3/2}}B_{2i},
\end{aligned}$$

em que $\mathbb{I}_{(p)} = [\mathbf{0}_q, \mathbf{I}_q]$, tal que $\mathbf{0}_q = (0, \ldots, 0)^\top$ é um vetor $q \times 1$ com $q = 1 - p$ e $\mathbf{I}_q$ é matriz identidade $q \times q$, $B_{1i} = \widehat{u}_i \mu_x - \widehat{ux}_i$, $B_{2i} = \widehat{ut}_i \mu_x - \widehat{utx}_i$ e $B_{3i} = \widehat{ux^2}_i + \widehat{u}_i \mu_x^2 - 2\widehat{ux}_i \mu_x$.

De posse desses resultados podemos seguir para a próxima seção que é dedicada a dois esquemas de pertubação, ponderação de casos e perturbação na variável resposta.

4.6 Esquemas de Perturbação

Nessa seção são considerados dois diferentes esquemas de perturbação para o modelo básico definido em (3.6)-(3.7): *ponderação de casos* para detectar observações com contribuição marcante na log-verossimilhança e que pode exercer alta influência nas estimativas de máxima verossimilhança e *medida de perturbação para um instrumento* feito nos valores observados dos instrumentos usados no estudo, o qual seus próprios valores preditos podem indicar observações com larga influência.

4.6.1 Perturbação de Ponderação de Casos

Primeiro considere a seguinte atribuição arbitrária de pesos para o valor esperado dos dados da função log-verossimilhança completa (função Q-afastamento perturbada), os quais podem capturar saídas em direções gerais. Note que, para $w_i = 0$ e $w_j = 1, j \neq i$ a i-ésima unidade experimental é excluída da função log-verossimilhança de dados completos. Além disso, é possível mostrar que a influência local para esse esquema de perturbação é equivalente para o método de supressão (Osorio, 2006). A função Q-afastamento perturbada é representada pela forma

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \widehat{\boldsymbol{\theta}}) &= E[\ell_c(\boldsymbol{\theta}, \boldsymbol{\omega} | \mathbf{z}_c)] = \sum_{i=1}^{n} w_i E[\ell_i(\boldsymbol{\theta} | \mathbf{z}_i)] \\
&= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1, \omega_i | \widehat{\boldsymbol{\theta}}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\theta}_2, \omega_i | \widehat{\boldsymbol{\theta}}),
\end{aligned}
$$

em que $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ um vetor $n \times 1$ com $\boldsymbol{\omega}_0 = \mathbf{1}_n$, $Q_{1i}(\boldsymbol{\theta}_1, \omega_i \hat{\boldsymbol{\theta}}) = \omega_i Q_{1i}(\boldsymbol{\theta}_1 | \widehat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\theta}_2, \omega_i \widehat{\boldsymbol{\theta}}) = \omega_i Q_{2i}(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}})$ com $Q_{1i}(\boldsymbol{\theta}_1, \omega_i \widehat{\boldsymbol{\theta}})$ e $Q_{2i}(\boldsymbol{\theta}_2, \omega_i \widehat{\boldsymbol{\theta}})$ são como apresentados em (3.15) e (3.16), respectivamente. Note que essa perturbação geral pode ser decomposta em duas outras sub-perturbações definidas por i) $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1,\omega_i|\widehat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}})$ e ii) $Q(\boldsymbol{\theta},\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}) + Q_{2i}(\boldsymbol{\theta}_2,\omega_i|\widehat{\boldsymbol{\theta}})$. Sob esse esquema de perturbação a matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ tem elementos que são apresentados a seguir:

$$
\begin{aligned}
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\alpha}\partial\omega_i} &= \widehat{u}_i\mathbb{I}_{(p)}D^{-1}(\boldsymbol{\phi})(\mathbf{z}_i-\mathbf{a}) - \widehat{ux}_i\mathbb{I}_{(p)}D^{-1}(\boldsymbol{\phi})\mathbf{b},\\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} &= \widehat{ux}_i\mathbb{I}_{(p)}D^{-1}(\boldsymbol{\phi})(\mathbf{z}_i-\mathbf{a}) - \widehat{ux^2}_i\mathbb{I}_{(p)}D^{-1}(\boldsymbol{\phi})\mathbf{b},\\
\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1,\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\phi}\partial\omega_i} &= \frac{1}{2}\big[-D(\boldsymbol{\phi}) + \widehat{u}_iD^2(\mathbf{z}_i-\mathbf{a}) - 2\widehat{ux}_iD(\mathbf{z}_i-\mathbf{a})D(\mathbf{b}) + \widehat{ux^2}_iD^2(\mathbf{b})\big]D^{-2}(\boldsymbol{\phi})\mathbf{1}_p,\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\mu_x\partial\omega_i} &= -\frac{1}{\nu_x^2}(\widehat{u}_i\mu_x - \widehat{ux}_i + \widehat{ut}_i\tau_x),\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\phi_x\partial\omega_i} &= -\frac{1}{2\phi_x} + \frac{\lambda_x(1+\lambda_x^2)^{1/2}}{2\phi_x^{3/2}}B_{2i} + \frac{1+\lambda_x^2}{2\phi_x^2}B_{3i},\\
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2,\boldsymbol{\omega}_0|\widehat{\boldsymbol{\theta}})}{\partial\lambda_x\partial\omega_i} &= \frac{\lambda_x}{(1+\lambda_x^2)} - \frac{\lambda_x}{\phi_x}B_{3i} - \lambda_x\widehat{ut^2}_i - \frac{(1+2\lambda_x^2)}{\phi_x^{1/2}(1+\lambda_x^2)^{1/2}}B_{2i}.
\end{aligned}
$$

em que $\mathbb{I}_{(p)} = [\mathbf{0}_q,\mathbf{I}_q]$, tal que $\mathbf{0}_q = (0,\ldots,0)^\top$ é um vetor $q\times 1$ e $\mathbf{I}_q$ é matriz identidade $q\times q$ com $q = 1-p$, $B_{1i} = \widehat{u}_i\mu_x - \widehat{ux}_i$, $B_{2i} = \widehat{ut}_i\mu_x - \widehat{utx}_i$ e $B_{3i} = \widehat{ux^2}_i + \widehat{u}_i\mu_x^2 - 2\widehat{ux}_i\mu_x$.

4.6.2 Perturbação na Variável Resposta

Nesse caso as medições são obtidas quando um instrumento é modificado considerando esquemas de perturbações aditivas e multiplicativas. Supondo que as medidas do instrumento $m(=1,...,p)$ são escolhidas para serem perturbadas, então o modelo perturbado é dado por

$$\mathbf{Z}_{mi}(\boldsymbol{\omega}) = \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i,$$

com $\boldsymbol{\omega} = (\omega_1,...,\omega_n)^\top$ e

(1) Perturbação aditiva: $\mathbf{Z}_{mi}(\omega_i) = \mathbf{Z}_i + \omega_i\mathbf{e}_m$, onde $\mathbf{e}_m$ é um vetor nulo p-dimensional com um na m-ésima posição. Nesse caso $\boldsymbol{\omega}_0 = \mathbf{0}$, e

(2) Perturbação multiplicativa: $\mathbf{Z}_{mi}(\omega_i) = \mathbf{Z}_i \odot \mathbf{1}_m(\omega_i)$, onde $\mathbf{1}_m(\omega_i)$ p-dimensional denotando um vetor tendo uns na m-ésima componente substituída por ω_i e $\odot$ denota o produto de Hadamard que diz que se duas matrizes do mesmo tipo, $\mathbf{A}_{n\times p}$, $\mathbf{B}_{n\times p}$, então a matriz $\mathbf{C}_{n\times p}$ cujo elemento genérico é dado pelo produto dos correspondentes elementos

de $\mathbf{A}$ e $\mathbf{B}$: $\mathbf{C}_{n\times p} = \mathbf{A}_{n\times p} \odot \mathbf{B}_{n\times p} \Longleftrightarrow a_{ij} \cdot b_{ij}$. Aqui $\boldsymbol{\omega}_0 = \mathbf{1}$.

Para ambos os casos, a função log-verossimilhança perturbada para dados completos é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\theta}_1, \omega_i|\widehat{\theta}) + Q_{2i}(\boldsymbol{\theta}_2|\widehat{\theta}),$$

em que $Q_{1i}(\boldsymbol{\theta}_1, \omega_i|\widehat{\theta})$ é dado como na equação (14), substituindo $\mathbf{z}_{mi}(\omega_i)$ por $\mathbf{z}_i$. Sob esse esquema de perturbação segue a matriz $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$

$$\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0 \mid \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial w_i} = \widehat{u_i} \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi}) \mathbf{p}_{mi},$$

$$\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0 \mid \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial w_i} = \widehat{ux_i} \mathbb{I}_{(p)} D^{-1}(\boldsymbol{\phi}) \mathbf{p}_{mi},$$

$$\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\omega}_0 \mid \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\phi} \partial w_i} = (\widehat{u_i} D(\boldsymbol{z_i} - \mathbf{a}) - \widehat{ux_i} D(\mathbf{b})) D^{-2}(\boldsymbol{\phi}) \mathbf{p}_{mi},$$

em que $\mathbb{I}_{(p)} = [\mathbf{0}_q, \mathbb{I}_q]$, tal que $\mathbf{0}_q = (0, ..., 0)^\top$ é um vetor $q \times 1$ e $\mathbb{I}_q$ é uma matriz identidade $q \times q$ com $q = 1 - p$, $\mathbf{p}_{mi} = \mathbf{e}_m$ sob o caso aditivo e $\mathbf{p}_{mi} = \mathbf{0}_m(z_{mi})$ sob o caso multiplicativo, com $\mathbf{0}_m(z_{mi})$ um vetor p-dimensional nulo com z_{mi} na m-ésima posição, $m = 1, \ldots, p$ e $i = 1, \ldots, n$. Para ilustrar, vamos supor que temos 2 instrumentos onde obtemos 4 medidas de cada um, desta forma temos a matriz que segue:

$$\mathbf{Z}_{2,4}(\boldsymbol{\omega}) = \begin{pmatrix} 23 & 20 \\ 24 & 22 \\ 27 & 25 \\ 19 & 17 \end{pmatrix},$$

inicialmente vamos perturbar o instrumento 1, desta forma temos sob o caso aditivo $\mathbf{p}_{1,1} = (1\ 0)$, $\mathbf{p}_{1,2} = (1\ 0)$, $\mathbf{p}_{1,3} = (1\ 0)$ e $\mathbf{p}_{1,4} = (1\ 0)$, enquanto sob o caso multiplicativo temos $\mathbf{p}_{1,1} = (23\ 0)$, $\mathbf{p}_{1,2} = (24\ 0)$, $\mathbf{p}_{1,3} = (27\ 0)$ e $\mathbf{p}_{1,4} = (19\ 0)$. De forma similar perturbando o instrumento 2 sob o caso aditivo temos $\mathbf{p}_{2,1} = (0\ 1)$, $\mathbf{p}_{2,2} = (0\ 1)$, $\mathbf{p}_{2,3} = (0\ 1)$ e $\mathbf{p}_{2,4} = (0\ 1)$, enquanto sob o caso multiplicativo temos $\mathbf{p}_{2,1} = (0\ 20)$, $\mathbf{p}_{2,2} = (0\ 22)$, $\mathbf{p}_{2,3} = (0\ 25)$ e $\mathbf{p}_{2,4} = (0\ 17)$.

O próximo capítulo é dedicado a ilustrar os resultados de influência local para dois conjuntos de dados conhecidos na literatura.

5 *Aplicação*

5.1 Introdução

Neste capítulo apresentamos duas aplicações com dados reais já utilizados anteriormente na literatura. O que diferencia esse estudo dos outros já realizados na área é a utilização do método de influência local proposto por Zhu e Lee (2001) para análise de diagnóstico. Serão aplicados os métodos de perturbação desenvolvidos no capítulo anterior.

5.2 Conjunto de Dados de Chipkevitch et al. (1996)

A estimação do tamanho testicular em adolescentes é importante para a avaliação do desenvolvimento normal da puberdade, para diagnóstico de distúrbios e para a avaliação de efeitos de doenças relacionadas. A medição clínica, do tamanho testicular na puberdade, é um método mais efetivo e avaliável para a estimação da função testicular do que outros métodos existentes, tais como espermogramas ou biopsia, os quais podem ser agressivos especialmente em adolescentes.

O estudo foi realizado com todos os pacientes vistos na divisão da medicina de adolescentes do Hospital da Criança Darcy Vargas, onde uma consulta de saúde com ultra-som foi feita para investigar a suspeita de patologia escrotal durante o ano de 1993.

Os pacientes com uma grande hidrocele (derrame líquido entre as túnicas da vaginal do testículo) ou outras condições que poderiam prejudicar as medições clínicas dos testículos foram excluídos do estudo. Acredita-se que varicocele (tumor produzido pela dilatação

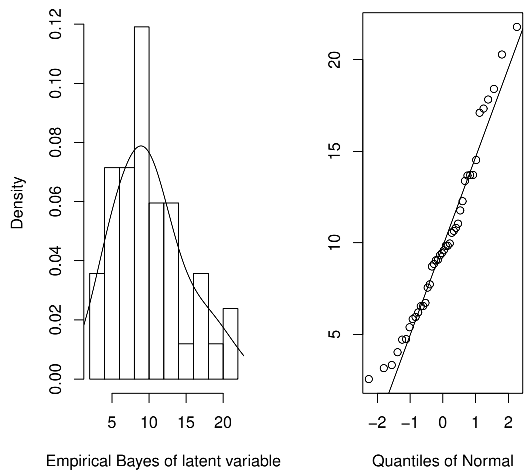

Figura 8: Histograma e Q-Q plots normal das estimativas empíricas de Bayes na variável latente.

Tabela 1: EMV dos quatro modelos ajustados no conjunto de dados de Chipkevitch. Erros quadrados (SE) são estimados assintoticamente baseados no erro padrão da matriz de informação observada.

	SN		ST		SSL		SCN	
Parâmetro	Estimativa	SE	Estimativa	SE	Estimativa	SE	Estimativa	SE
α_2	0,1022	0,5655	0,0431	0,4944	0,1203	0,5326	0,1559	0,5018
α_3	-0,0097	0,6216	-0,2467	0,5265	-0,1746	0,5788	-0,1527	0,5229
α_4	0,0482	0,6277	0,1119	0,5679	0,1032	0,6005	0,1464	0,5644
α_5	1,5390	0,6337	1,5453	0,5612	1,5800	0,5986	1,6404	0,5730
β_2	0,8838	0,0509	0,8990	0,0514	0,8889	0,0518	0,8911	0,0511
β_3	0,9495	0,0559	0,9866	0,0566	0,9755	0,0579	0,9782	0,0547
β_4	1,1419	0,0565	1,1536	0,0587	1,1466	0,0583	1,1540	0,0574
β_5	1,0826	0,0570	1,0956	0,0579	1,0862	0,0579	1,0885	0,0584
ϕ_1	1,3385	0,3714	0,9291	0,2890	0,8363	0,2512	0,7467	0,2245
ϕ_2	1,3284	0,3480	0,9539	0,2836	0,8427	0,2384	0,7972	0,2230
ϕ_3	1,6736	0,4322	0,9028	0,2961	0,8952	0,2868	0,7294	0,2106
ϕ_4	1,1578	0,3710	0,9482	0,3160	0,8003	0,2586	0,7845	0,2539
ϕ_5	1,4105	0,3994	1,0198	0,3378	0,9109	0,2810	0,9119	0,2762
μ_x	3,9957	1,3959	4,1696	1,0965	4,2065	1,3518	4,2778	1,3815
ϕ_x	59,2790	21,5489	38,1464	14,3678	34,9551	13,6922	30,9967	13,5730
λ_x	4,7826	4,7895	3,4308	2,4547	3,7395	3,2961	3,2417	2,8696
ν	-	-	6,0000	-	3,0000	-	0,3000	-
γ	-	-	-	-	-	-	0,3000	-
log-likelihood	-422,5881		-416,7961		-419,3554		-415,9791	
AIC	877,1762		867,5922		872,7108		867,9582	
BIC	904,9789		897,1326		902,2512		899,2363	
HQ	866,2716		856,0061		861,1247		855,6905	

varicosa das veias do cordão espérmático) não prejudica a medição com o paciente deitado de costas.

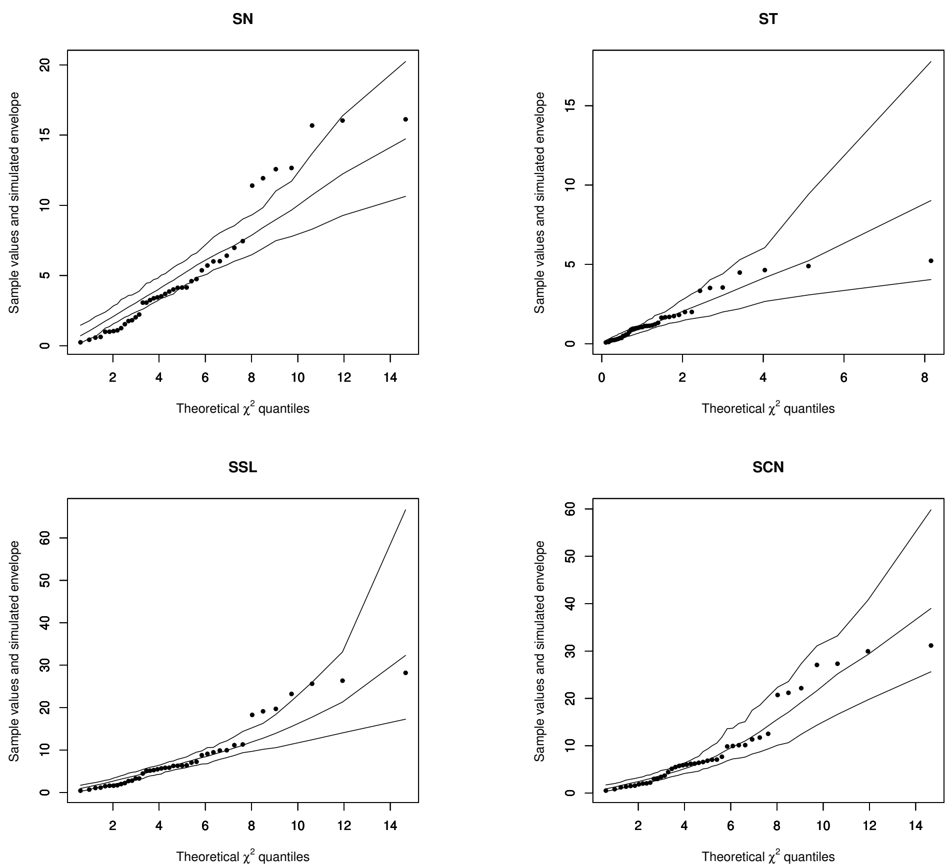

Figura 9: Conjunto de dados de Chipkevitch - Envelopes simulados.

Tabela 2: Conjunto de dados de Chipkevitch. Medida C_{fQ} de Zhu e Lee.

Distribuição	Ponderação de Pesos	Aditivo (Instrumento I)	Multiplicativo (Instrumento IV)
SN	6,1869	2,5646	758,8755
ST	2,9470	3,1364	564,2995
SSL	3,7971	2,7113	626,0928
SCN	3,1219	3,0951	563,8815

A amostra inclui 42 adolescentes ente 11 e 17 anos de idade (a idade média foi de 13, 9 e o desvio padrão 1,3).

A seguir ajustamos esses dados as distribuiuções SN, ST, SSL e SCN multivariadas. Cinco diferentes técnicas foram utilizadas: ultrassom(US), método gráfico proposto pelos autores (I), medida dimensional (II), orquidômetro prader (III), e orquidômetro anel (IV) com o ultrassom (US) assumido ser o instrumento de referência. Esse conjunto de dados foi

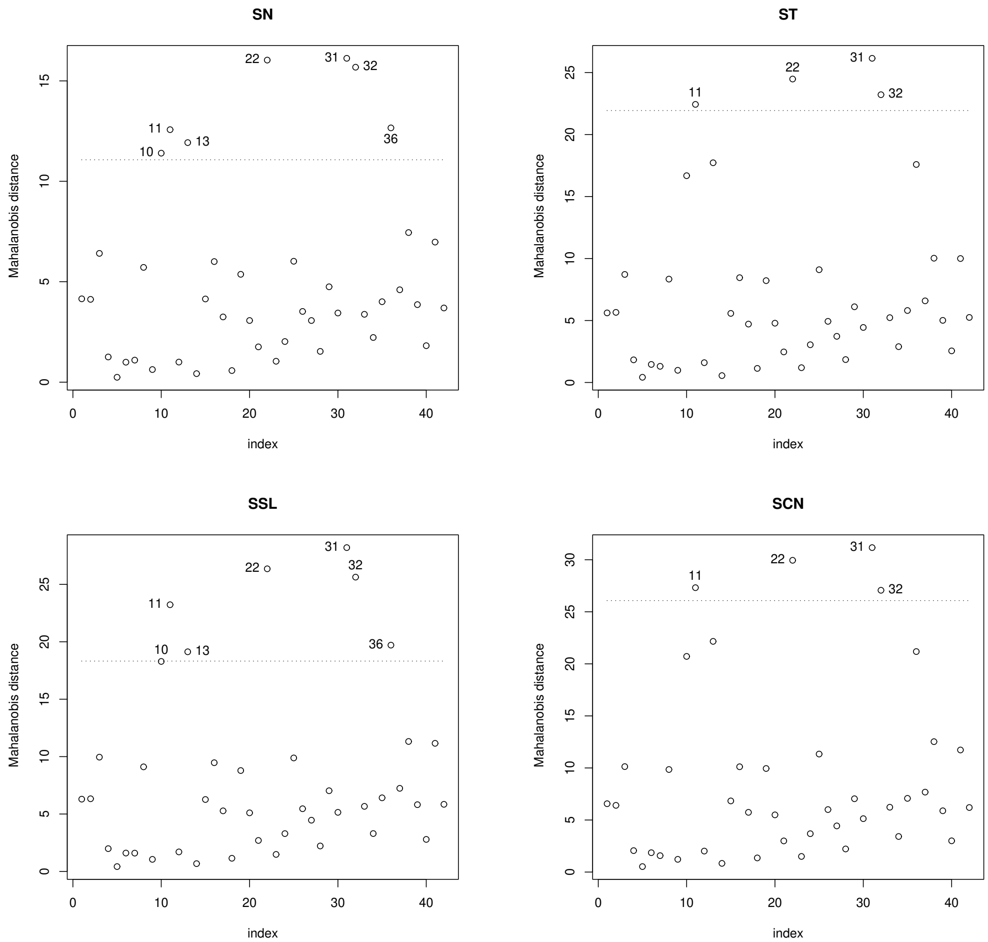

Figura 10: Gráfico de dispersão para o Índice e a distância de Mahalanobis para os quatro modelos ajustados.

estudado em Lachos et al. (2008) sob MESN-MEM. Nossa intenção nesse trabalho é utilizar novamente esse conjunto de dados para os modelos da classe MESN-MEM mas agora com a abordagem da aproximação de influência local de Zhu e Lee (2001). Ordenando os dados para verificar a existência de assimetria na resposta latente (x_i), primeiramente ajustamos os dados ao modelo normal MEM. Figura 8 mostra o histograma correspondente e o Q-Q plot das estimativas de Bayes empírica de x_i em que a variável latente é positivamente assimétrica e dessa forma um modelo Gaussiano pode não prover um bom ajuste. Desta forma pareceria apropriado ajustar uma distribuição MESN para x_i e distribuição MEN para $\boldsymbol{\epsilon}_i$, tal como definido em (3.6) e (3.7). Os valores iniciais para o al-

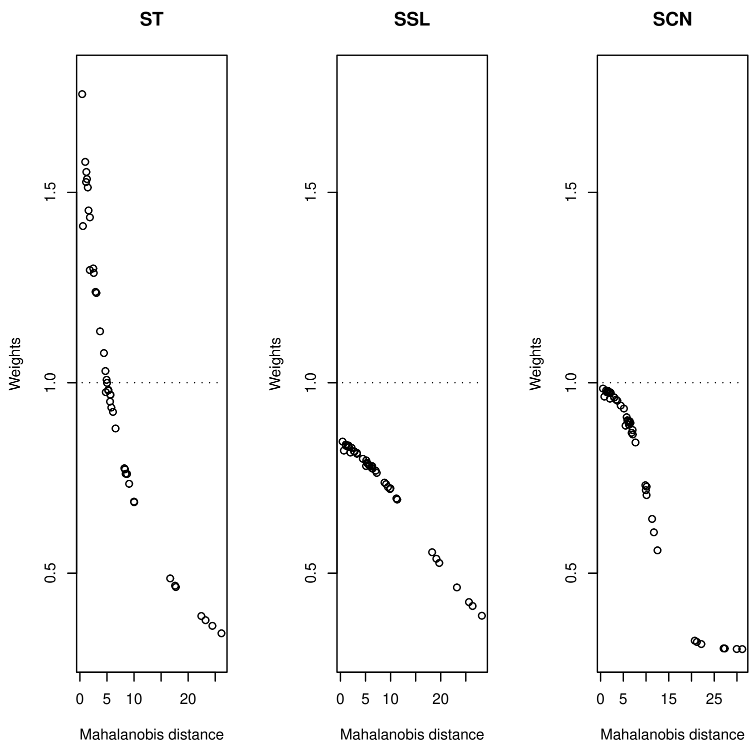

Figura 11: u_i estimado vs. distância de Mahalanobis para os modelos ST, SSL e SCN.

goritmo EM são os obtidos para a distribuição SN em Lachos et al. (2008), os parâmetros ν e γ são obtidos segundo maximização da log-verossimilhança e inspeção visual dos gráficos obtidos (log-verossimilhança x parâmetros). Estimativas de máxima verossimilhança e erro padrão todos calculados e a matriz de informação para a MESN-MEM são dadas na Tabela 1. O critério AIC indica que a distribuição MESN com caudas pesadas apresenta um melhor ajuste que o modelo SN-MEM, sugerindo uma não normalidade dos dados. Substituindo as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ na distância de Mahalanobis d_i, na Figura 9, nós apresentamos envelopes simulados (linhas representam o quinto percentil, a média e o nonagésimo-quinto percentil de 100 pontos simulados para cada observação). Esses gráficos cruzados demonstram uma forte evidência que as distribuições assimétricas com caudas pesadas, em particular a distribuição ST, resultam um melhor ajuste para o conjunto de dados que a distribuição skew-normal.

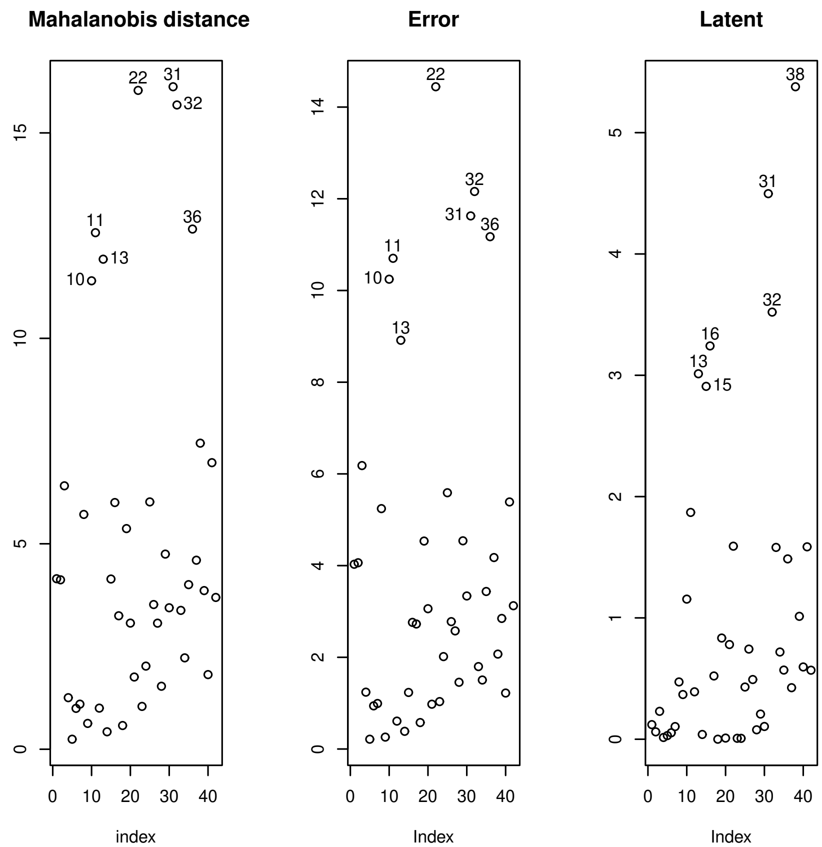

Figura 12: $d^2_{\mathbf{e}i}$ (Erro) e d^2_{xi} (Latente) estimados no ajuste skew–normal do conjunto de dados de Chipkevitch.

Ordenamos os dados para detectar observações aberrantes e para isso consideramos a distância de Mahalanobis. A Figura 10 mostra tais distâncias para os quatro modelos ajustados. As linhas de ponto de corte correspondem ao quantil $\xi = 0,95$. Nós podemos ver nessas figuras que as observações 11, 22, 31 e 32 são detectadas como possíveis valores aberrantes nós casos ST e SCN, essas mais as observações 10, 13 e 36 são detectados como possíveis valores aberrantes nos casos SN e SSL. Do algoritmo EM, os pesos estimados $(\widehat{u}_i, i = 1, ..., 42,)$ para essas observações são as menores para todos modelos MESN com caudas pesadas (as 4 últimas observações de todas as distribuições na Figura 11, sendo 11, 32, 22 e 31 para a ST e a SSL e 32, 11, 22 e 31 para a SCN), confirmando os aspectos de robustez das estimativas de máxima verossimilhança versus observações aberrantes dos modelos SMSN com caudas pesadas. Para o caso skew-normal, $\widehat{u}_i$ é inversamente

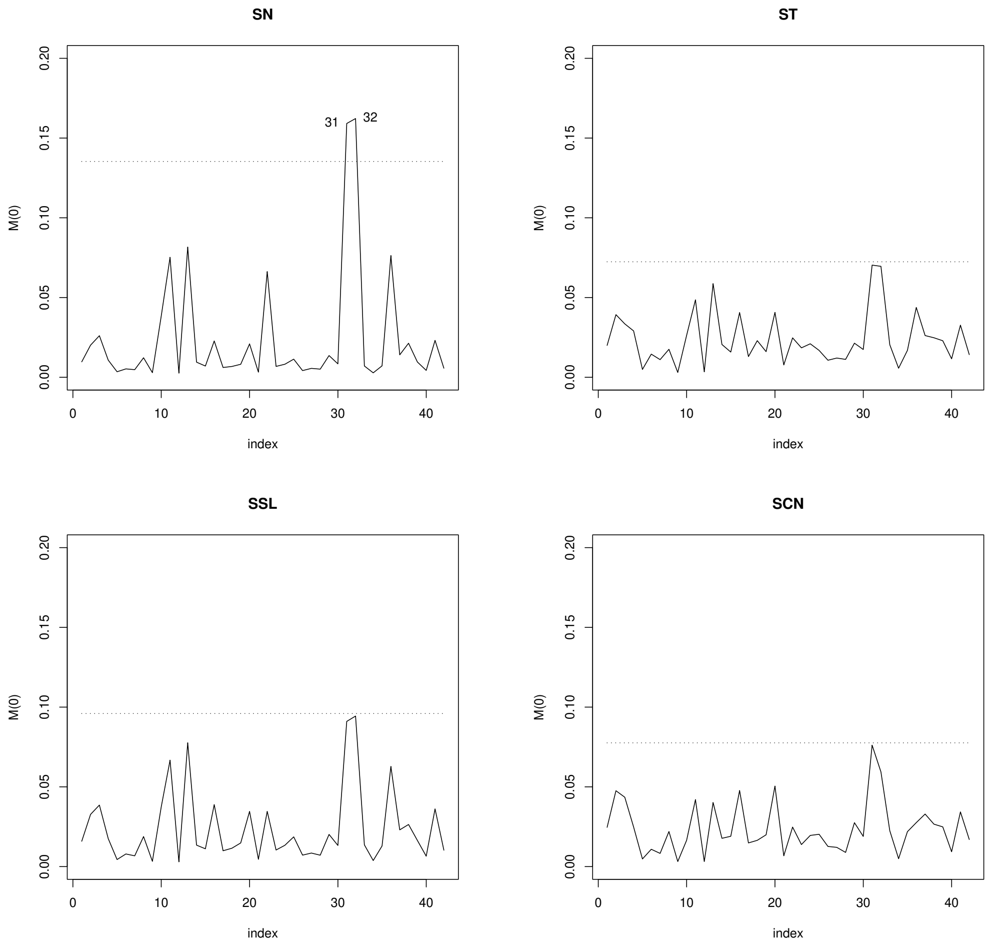

Figura 13: Conjunto de dados de Chipkevitch - Gráficos de Índice de M(0) sob a perturbação da ponderação de casos para os quatro modelos ajustados. As linhas horizontais delimitam o *benchmark* de Lee e Xu (2004) para M(0) com $c* = 3$.

proporcional a distância de Mahalanobis. Deste modo, d_i maior implica em menor $\widehat{u_i}$, o estimador de $\boldsymbol{\theta}$ tende dar menor peso para observações atípicas no senso da distância de Mahalanobis. Para mais detalhes, veja Ghosh et al. (2008).

A distância estimada $d_{\mathbf{e}i}$(Erro) e d_{xi}(Latente), definidas em (3.18), leva a úteis estatísticas diagnósticos para identificar observações aberrantes. A Figura 12 apresenta estatísticas diagnóstico para SN-MEM. Observações 10, 11, 13, 22, 31, 32 e 36 apresentam grande valor de $d_{\mathbf{e}i}$ sugerindo valores aberrantes com relação ao erro ($\mathbf{e}$-*outlier*). Entretanto, observações 13, 31 e 32 apresentam grandes valores de d_{xi} sugerindo ob-

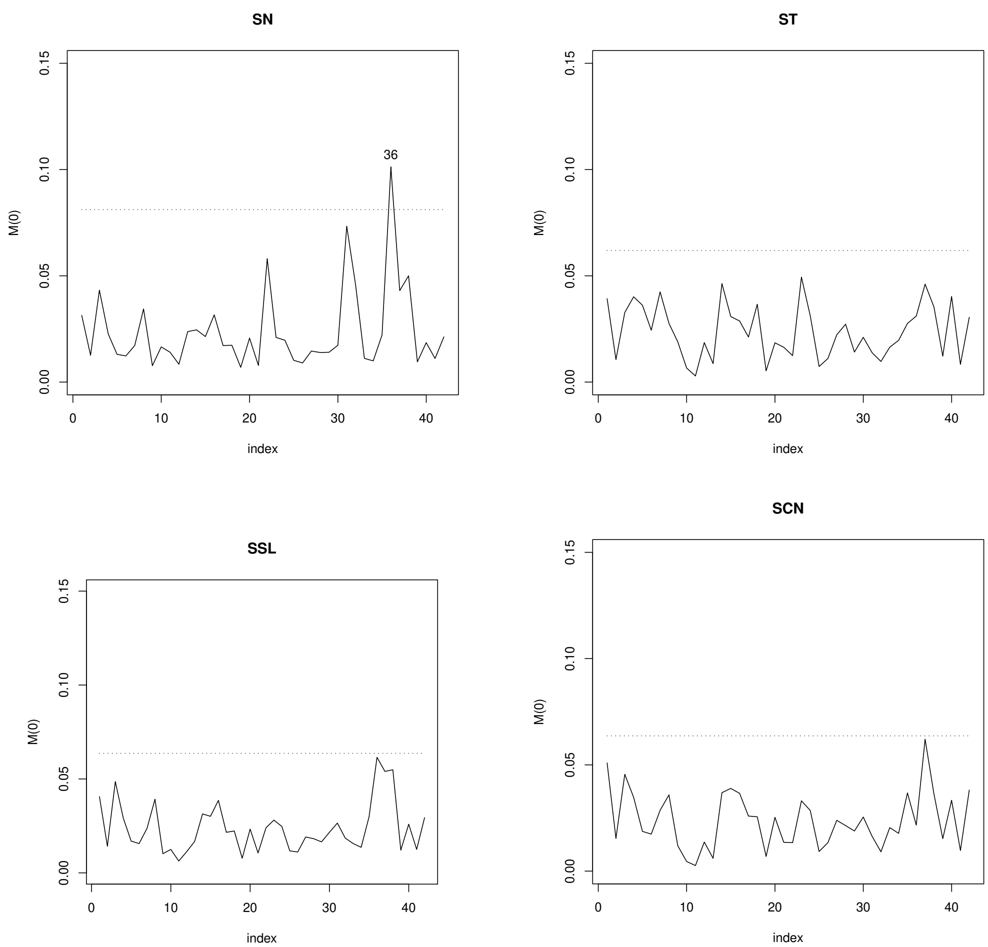

Figura 14: Conjunto de dados de Chipkevitch - Gráficos de Índice de M(0) sob o caso de perturbação aditiva para a medição do instrumento I para os quatro modelos ajustados. As linhas horizontais delimitam o *benchmark* de Lee e Xu (2004) para M(0) com $c* = 3$.

servações aberrantes. O gráfico d_{xi} da alguma indicação de que as observações 16, 15 e 38 são possivelmente valores aberrantes com relação a variável latente (x-*outlier*), o qual não pode ser concluído da Figura 10.

Nós identificamos observações influentes para o conjunto de dados de Chipkevitch usando $M(0)$ de uma curvatura padrão B_i obtida considerando o esquema de perturbação do caso de pesos e medida de perturbação de um instrumento: casos aditivo e multiplicativo. A análise de influência desenvolvida aqui é baseada no algoritmo EM apresentada em Lachos et al. (2008). Aqui nós conduzimos um estudo de influência local de Zhu e

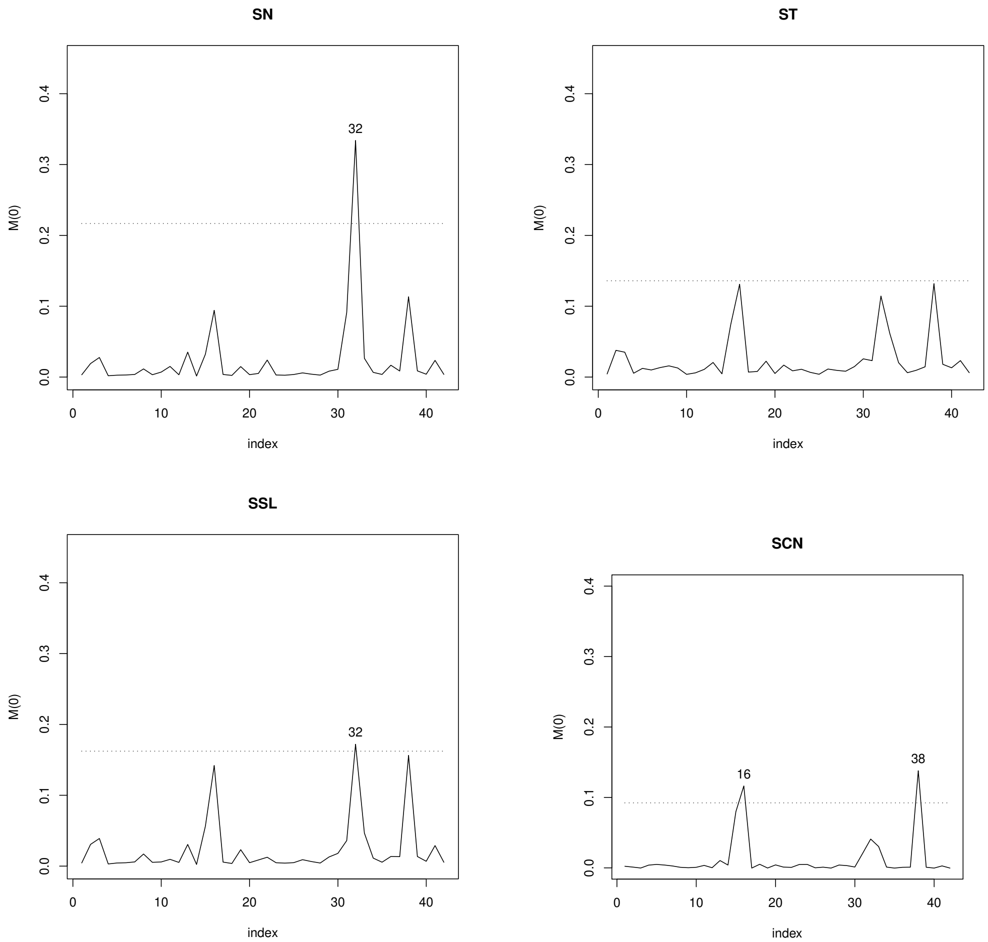

Figura 15: Conjunto de dados de Chipkevitch - Gráficos de Índice de M(0) sob o caso de perturbação multiplicativa para a medição do instrumento IV para os quatro modelos ajustados. As linhas horizontais delimitam o *benchmark* Lee e Xu (20040 para M(0) com $c* = 3.5$.

Lee (2001) com enfoque de interesse em $\boldsymbol{\theta}$. As Figuras 13, 14 e 15 apresentam os gráficos índice de $M(0)$ para os modelos sob a distribuição MESN. Da Figura 13 é notado que sob SN-MEM as observações 31 e 32, que apareceram como possíveis valores aberrantes na Figura 10, são identificadas como influentes.

Nós examinamos agora o efeito das medidas de perturbação utilizando os instrumentos I e IV, definimos que o instrumento I foi utilizado no método I, o II no método I, III no étodo III, o IV no método IV e o US no método US. Os instrumentos I e IV são escolhidos porque eles apresentam maiores valores de $C_{fQ,\mathbf{dmax}}$. Os valores de $C_{fQ,\mathbf{dmax}}$

para perturbações aditivas das medidas obtidas pelo instrumento I e as perturbações multiplicativas para as medidas obtidas pelo instrumento IV bem como $C_{fQ,\mathbf{dmax}}$ para o caso de ponderação de pesos são encontradas na Tabela 2. As Figuras 14 e 15 ilustram o gráfico índice para perturbação de medidas de um instrumento. Usando esse esquema de perturbação, nós podemos examinar a influência nas medidas obtidas pelos instrumentos I e IV, sob os casos aditivos e multiplicativos, respectivamente.

Na Figura 14 é observado alguma influência da medida da observação 36 quando o instrumento I é perturbado sob uma SN-MEM.

Na Figura 15 é observado alguma influência da medida da observação 32 quando o instrumento IV é perturbado sob uma SN-MEM e sob uma SSL-MEM. Como esperado, a influência de tais observações é reduzida quando nós consideramos distribuições com caudas mais pesadas que a distribuição skew-normal. Entretanto, nesse caso de perturbação, a observação 16 é eminente sob SCN-MEM. Para esse conjunto de dados a ST-MEM acomoda um pouco melhor as observações influentes.

Como sugerido por Lu e Song, nós usamos os quantis TRC e MRC, para mostrar o impacto das observações influentes detectadas as quais são definidas, respectivamente, por

$$TRC = \sum_{j=1}^{n_p} |\frac{\widehat{\theta}_j - \widehat{\theta_{[i]j}}}{\widehat{\theta}_j}| \text{ e } MRC = max_{j=1,\ldots,n_p}|\frac{\widehat{\theta}_j - \widehat{\theta_{[i]j}}}{\widehat{\theta}_j}|,$$

onde n_p é a dimensão de θ e o subscrito $[i]$ significa o estimador de máxima verossimilhança de θ com a i-ésima observação, y_i, excluída. A comparação dessas medidas, baseado em diferentes modelos, com as observações 31 e 32 mais influentes excluídas, são dadas na Tabela 3. Note que as maiores mudanças são obtidas sob a distribuição SN. Como esperado, os resultados indicam que os estimadores de máxima verossimilhança são menos sensíveis a presença de observações influentes quando nós usamos distribuições com caudas mais pesadas que a SN.

Tabela 3: Conjunto de dados de Chipkevitch. Comparação das mudanças relativas nos estimadores de máxima verossimilhança em termos de TRC e MRC para os quatro modelos MESN selecionados.

Observação	Distribuição	TRC	MRC
31 e 32	SN	53,2386	49,3093
	ST	2,5758	1,2128
	SSL	3,0932	2,0785
	SCN	2,4882	1,8232

5.3 Conjunto de Dados de Barnett (1969)

A seguir desenvolvemos um estudo idêntico ao anterior com o conjunto de dados de Barnett (1969), relativos a um estudo médico para avaliar a qualidade de dois instrumentos para medir a capacidade vital pulmonar de um grupo de pacientes. Um instrumento é de tipo padrão e outro novo, portátil e mais fácil de operar. Além disso, é comparado o uso por um operador especializado e por um não especializado. Os instrumentos foram comparados em um grupo comum de 72 pacientes, quando operados por um operador especializado e por um não especializado. Assim temos as seguintes combinações, (I) instrumento padrão e operador especializado, (II) instrumento padrão e operador não especializado, (III) instrumento novo e operador especializado, (IV) instrumento novo e operador não especializado. Assumimos a combinação (I) como a de referência nesse estudo.

Como na seção anterior ajustamos os dados as distribuiuções SN, ST, SSL e SCN multivariadas. Os valores iniciais para o algoritmo EM com relação a distribuição SN são os obtidos para a distribuição Normal em Lachos et al. (2009), e para as demais distribuições é utilizado os valores obtidos pelo algoritmo para a distribuição skew-normal. Os parâmetros ν e γ são obtidos segundo maximização da log-verossimilhança e inspeção visual dos gráficos obtidos (log-verossimilhança $\times$ parâmetros). Ordenando os dados para verificar a existência de assimetria na resposta latente (x_i), primeiramente ajustamos os dados ao modelo normal MEM. A Figura 16 mostra o histograma correspondente e o Q-Q plot das estimativas de Bayes empírica de x_i em que a variável latente é um pouco assimétrica e dessa forma um modelo Gaussiano pode não prover um bom ajuste.

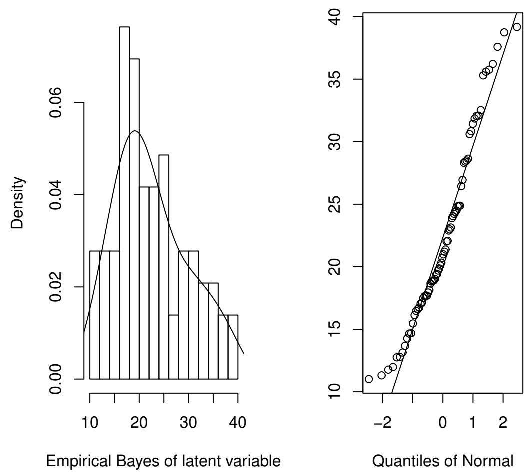

Figura 16: Histograma e Q-Q plots normal das estimativas empíricas de Bayes na variável latente.

Tabela 4: EMV dos quatro modelos ajustados no conjunto de dados de Barnett. Erros quadrados (SE) são estimados assintoticamente baseados no erro padrão da matriz de informação observada.

	SN		ST		SSL		SCN	
Parâmetro	Estimativa	SE	Estimativa	SE	Estimativa	SE	Estimativa	SE
α_2	-2,0730	1,0474	-1,6757	0,9646	-1,7860	0,9931	-1,7711	0,9792
α_3	-5,2700	1,2353	-4,8412	1,1936	-5,0060	1,2118	-4,8388	1,1938
α_4	-4,3682	1,2449	-3,9940	1,2122	-4,1231	1,2316	-4,0077	1,2037
β_2	1,0609	0,0444	1,0455	0,0422	1,0489	0,0429	1,0511	0,0429
β_3	1,1912	0,0524	1,1755	0,0523	1,1800	0,0526	1,1781	0,0522
β_4	1,1304	0,0528	1,1207	0,0529	1,1226	0,0533	1,1233	0,0524
ϕ_1	5,0400	0,9962	4,0142	0,8922	3,2863	0,7239	2,3749	0,4915
ϕ_2	1,7957	0,5800	1,3477	0,4973	1,1131	0,4065	0,8192	0,2802
ϕ_3	3,0426	0,8102	2,7485	0,7544	2,2091	0,5958	1,5073	0,4165
ϕ_4	3,9340	0,8959	3,4681	0,8660	2,8079	0,6820	1,9253	0,4624
μ_x	13,0568	1,2591	12,9233	1,4427	12,9850	1,3658	12,9299	1,4284
ϕ_x	141,8255	34,0546	126,3277	35,1883	100,2414	26,4007	71,9769	19,8276
λ_x	5,1778	2,9387	4,9102	3,2819	4,9742	3,0917	4,9439	3,2947
ν	-	-	12,0000	-	3,0000	-	0,7000	-
γ	-	-	-	-	-	-	0,4000	-
log-likelihood	-733,5978		-732,3871		-733,2834		-731,2196	
AIC	1493,2		1492,8		1494,6		1492,4	
BIC	1522,8		1524,6		1526,4		1526,6	
HQ	1486,1		1485,1		1486,9		1484,2	

Analisando o histograma para as medidas observadas dos instrumentos de referência vemos na Figura 16, que os dados suportam o uso de uma distribuição assimétrica com caudas pesadas. Sendo assim, deve acreditamos ser apropriado ajustar uma distribuição MESN para x_i e uma distribuição MEN para $\boldsymbol{\epsilon}_i$, tal como definido em (3.6) e (3.7). Podemos

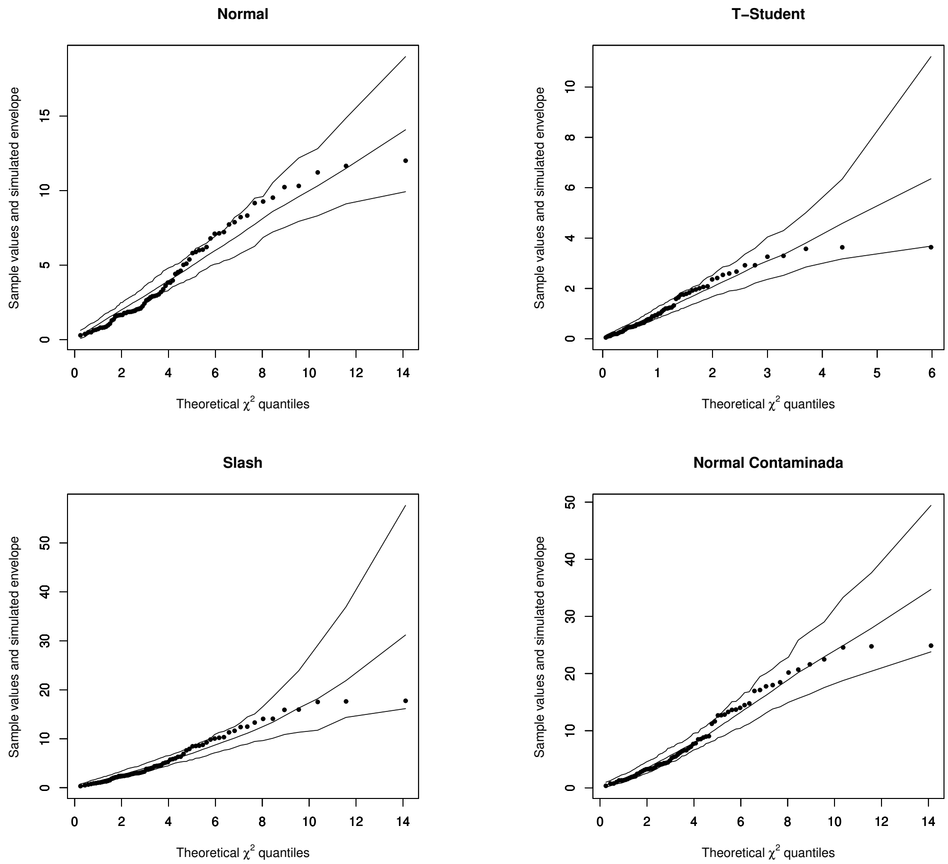

Figura 17: Conjunto de dados de Barnett - Envelopes simulados (distribuições simétricas $\boldsymbol{\lambda} = 0$).

Tabela 5: Conjunto de dados de Barnett. Log-verossimilhança e critérios AIC, BIC e HQ para distribuições simétricas ($\lambda_x = 0$).

	Normal	T-Student	Slash	Normal Contaminada
log-likelihood	-738,1865	-736,9632	-738,0118	-735,4502
AIC	1500,4	1499,9	1502,0	1498,9
BIC	1527,7	1529,5	1531,6	1530,8
HQ	1493,8	1492,8	1494,9	1491,2

ver na Tabela 4 as estimativas de máxima verossimilhança e o erro padrão para cada uma claculada através da matriz de informação para a MESN-MEM. Substituindo as estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ na distância de Mahalanobis d_i, nas Figuras 17 e 18, nós apresentamos envelopes simulados (linhas representam o quinto percentil, a

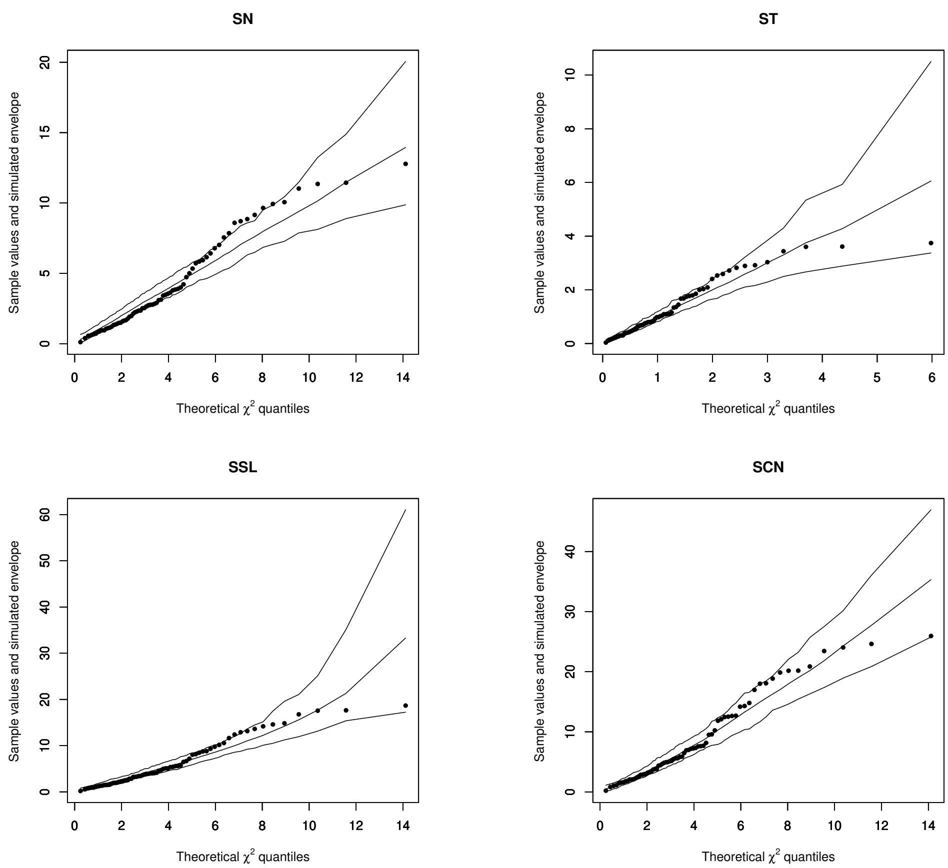

Figura 18: Conjunto de dados de Barnett - Envelopes simulados.

Tabela 6: Conjunto de dados de Barnett. Resultados da estatística da razão de verossimilhanças para testar a hipótese de interesse H_0.

Modelo	Valor	P-Valor
Normal/SN-MEM	9,1774	0,0024
t-Student/ST-MEM	9,1522	0,0025
Slash/SSL-MEM	9,4568	0,0021
Normal Contaminada/SCN-MEM	8,4612	0,0036

média e o nonagésimo-quinto percentil de 100 pontos simulados para cada observação). Esses gráficos cruzados demonstram alguma evidência de que as distribuições assimétricas com caudas mais pesadas podem resultar num melhor ajuste para o conjunto de dados do que a distribuição skew-normal. Para validar nossa hipótese apresentamos a Tabela 5 que nos mostra que todas as distribuições simétricas obtém resultados inferiores nos critérios AIC, BIC e HQ do que nos casos assimétricos, que podem ser observados na Tabela 4. Em vias gerais estamos buscando uma forma de validar a relevância do uso de modelos assimétricos para esse conjunto de dados. Uma forma simples de buscar evidências a favor de nossa hipótese é através de um teste de hipóteses, onde nossa hipótese nula (H_0) é que $\lambda_x = 0$, isto é, um modelo assimétrico se adequa melhor do que um simétrico. Segundo Montenegro et al. (2009b) a estatística razão de verossimilhança $LR = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\} < \chi^2_{(0.95,1)}$. No nosso caso específico $\ell(\widehat{\boldsymbol{\theta}})$ é o modelo irrestrito (assimétrico), isto é, com $\lambda_x \neq 0$ enquanto $\ell(\boldsymbol{\theta})$ é o modelo restrito (simétrico), $\lambda_x = 0$. Na tabela 6 observamos que existem evidências que nos levam a acreditar que o uso de modelos assimétricos para esse conjunto de dados seja realmente o mais adequado.

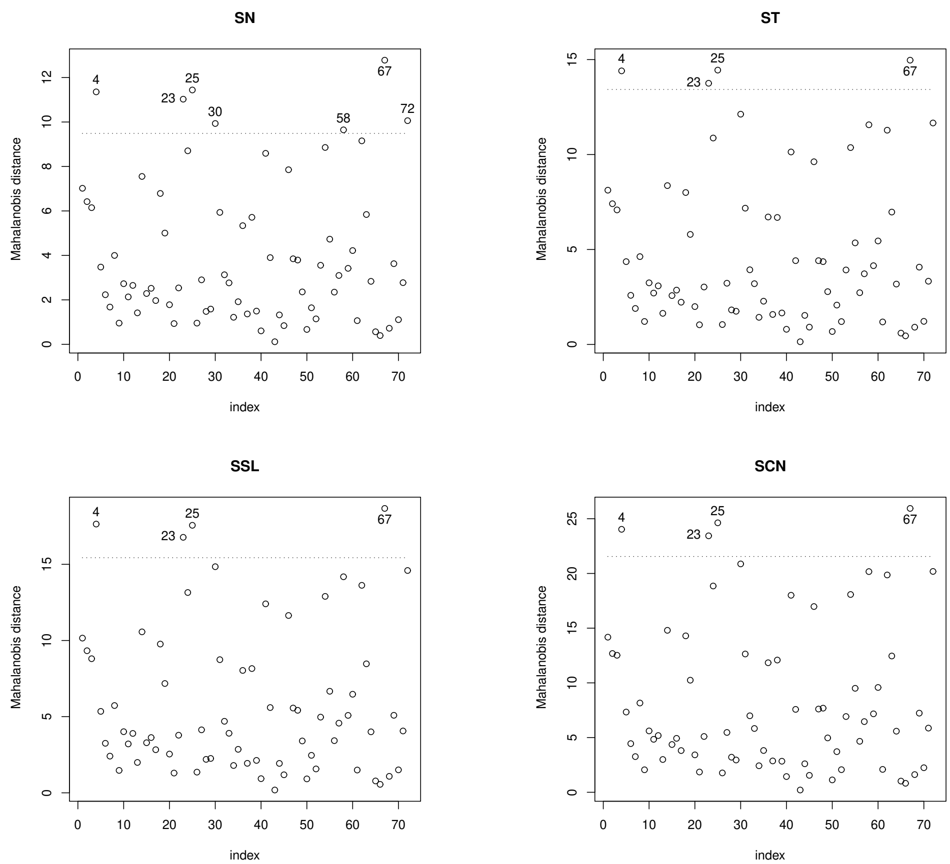

Figura 19: Gráfico de dispersão para o Índice e a distância de Mahalanobis para os quatro modelos ajustados.

Ordenamos os dados para detectar observações atípicas e para isso consideramos a

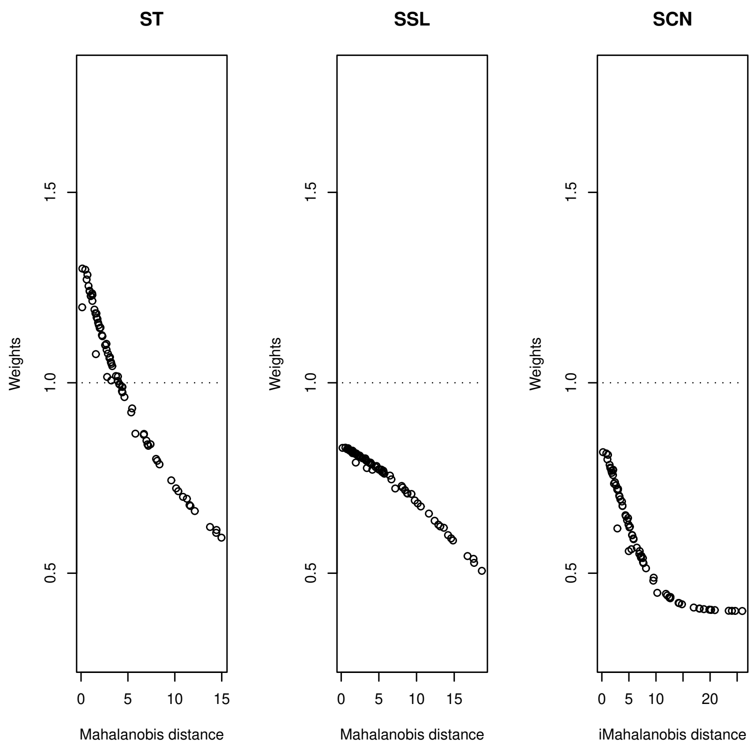

Figura 20: u_i estimado vs. distância de Mahalanobis para os modelos ST, SSL e SCN.

Tabela 7: Conjunto de dados de Barnett. Medida C_{fQ} de Zhu e Lee.

Distribuição	Ponderação de Pesos	Aditivo (Instrumento II)	Multiplicativo (Instrumento III)
SN	4,5761	1,4144	770,1751
ST	3,3740	1,6471	674,5077
SSL	3,7279	1,4886	698,6358
SCN	3,5861	1,6321	690,6927

distância de Mahalanobis. A Figura 19 mostra tais distâncias para os quatro modelos ajustados. As linhas de ponto de corte correspondem ao quantil $\xi = 0,95$. Nós podemos ver nessas figuras que as observações 4, 23, 25 e 67 são detectadas como possíveis valores aberrantes nós casos ST, SSL e SCN, essas e mais as observações 30, 58 e 72 são detectados como possíveis valores aberrantes no caso SN. Do algoritmo EM, os pesos estimados ($\widehat{u_i}, i = 1, ..., 42,$) para essas observações são as menores para todos modelos MESN com caudas pesadas (estão entre as 7 últimas observações de todas as distribuições na Figura 20, sendo 58, 72, 30, 23, 4, 25 e 67 para a ST e a SCN e 58, 72, 30, 23, 25, 4 e 67 para

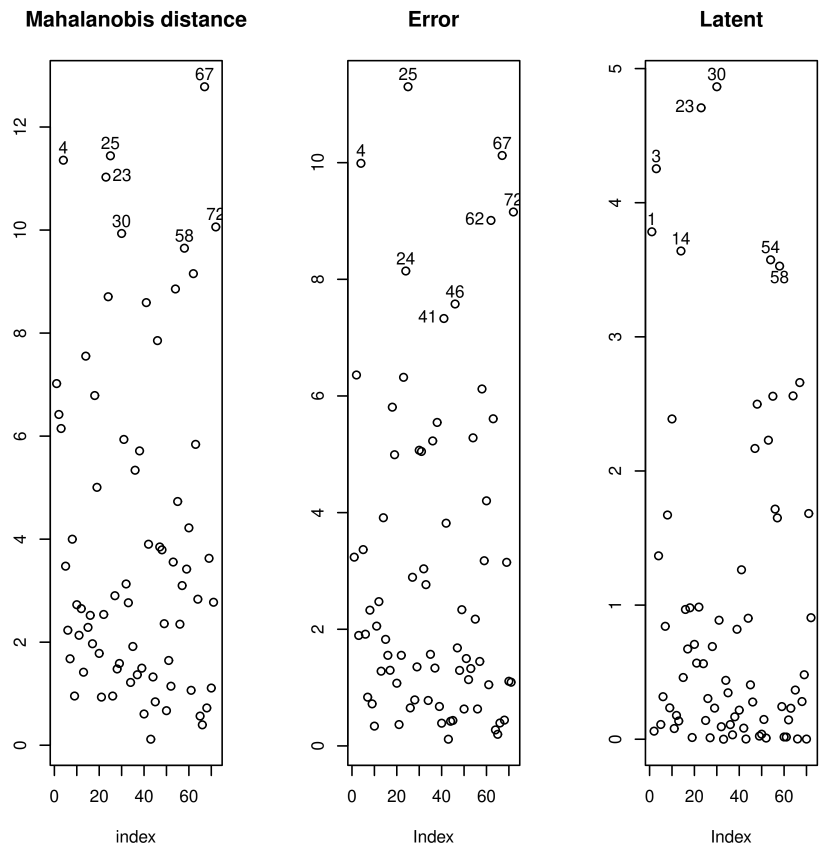

Figura 21: $d^2_{\mathbf{e}i}$ (Erro) e d^2_{xi} (Latente) estimados no ajuste skew–normal do conjunto de dados de Barnett.

a SSL), confirmando os aspectos de robustez das estimativas de máxima verossimilhança versus observações aberrantes dos modelos SMSN com caudas pesadas. Para o caso skew-normal, $\widehat{u_i}$ é inversamente proporcional a distância de Mahalanobis. Deste modo, d_i maior implica em menor $\widehat{u_i}$, o estimador de $\boldsymbol{\theta}$ tende dar menor peso para observações atípicas no senso da distância de Mahalanobis.

A distância estimada $d_{\mathbf{e}i}$(Erro) e d_{xi}(Latente), definidas em (3.18), leva a úteis estatísticas diagnósticos para identificar observações aberrantes. A Figura 21 apresenta estatísticas diagnóstico para SN-MEM. Observações 4, 24, 25, 41, 46, 62, 67 e 72 apresentam grande valor de $d_{\mathbf{e}i}$ sugerindo valores aberrantes com relação ao erro ($\mathbf{e}$-*outlier*). O gráfico d_{xi} da alguma indicação de que as observações 1, 3, 14, 23, 30, 54 e 58 são possivelmente valores aberrantes com relação a variável latente (x-*outlier*). A Figura 22

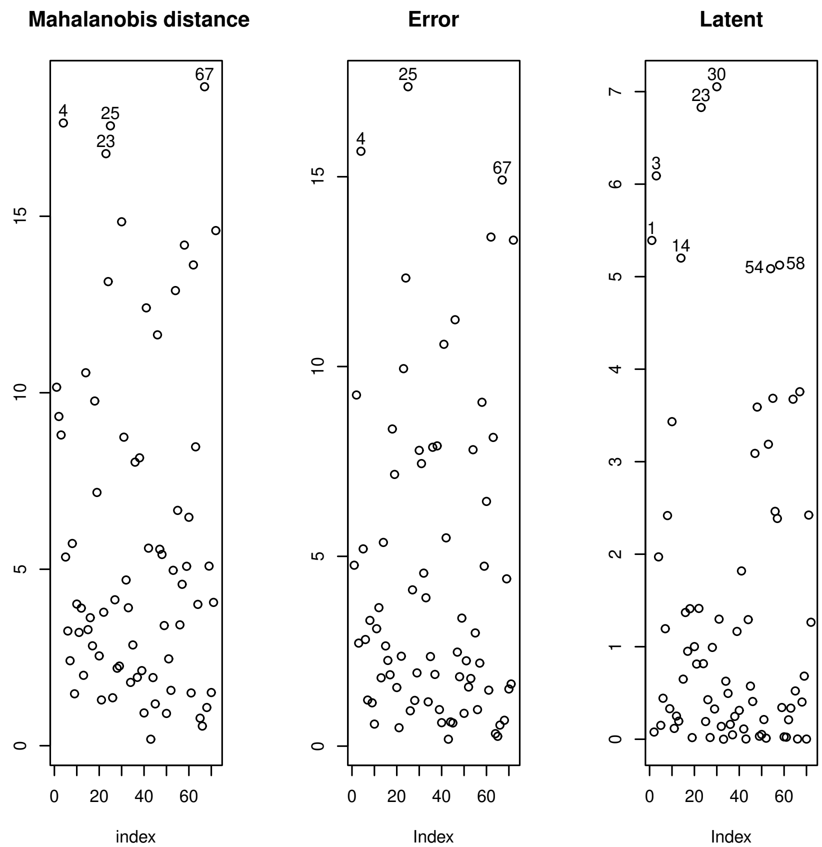

Figura 22: $d^2_{\mathbf{e}i}$ (Erro) e d^2_{xi} (Latente) estimados no ajuste skew–slash do conjunto de dados de Barnett.

apresenta estatísticas diagnóstico para SSL-MEM. Observações 4, 25 e 67 apresentam grande valor de $d_{\mathbf{e}i}$ sugerindo valores aberrantes com relação ao erro ($\mathbf{e}$-*outlier*). O gráfico d_{xi} da alguma indicação de que as observações 1, 3, 14, 23, 30, 54 e 58 são possivelmente valores aberrantes, idem ao caso sob a distribuição SN-MEM, com relação a variável latente (x-*outlier*).

Nós identificamos observações influentes para o conjunto de dados de Barnett de forma similar a descrita na seção anterior. As Figuras 23, 24 e 25 apresentam os gráficos índice de $M(0)$ para o modelo selecionado MESN. Observamos na Figura 23 que a observação 23, que apareceram como possíveis valores aberrantes na Figura 19, é identificada como influente para todos os modelos, mas aparentemente os dados se comportam melhor sob uma distribuição ST-MEM ou SSL-MEM.

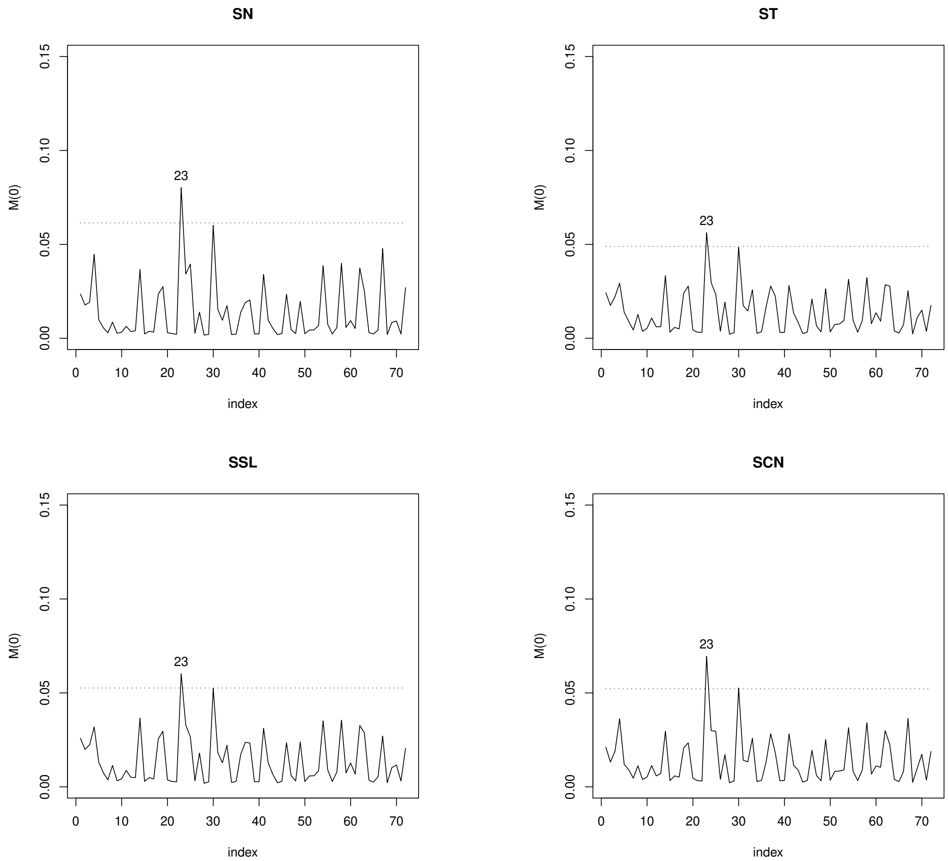

Figura 23: Conjunto de dados de Barnett - Gráficos de Índice de M(0) sob a perturbação da ponderação de casos para os quatro modelos ajustados. As linhas horizontais delimitam o *benchmark* de Lee e Xu (2004) para M(0) com $c* = 3$.

A seguir vamos examinar o efeito das medidas de perturbação utilizando os instrumentos II e III. Os instrumentos II e III são escolhidos porque eles apresentam maiores valores de $C_{fQ,\mathbf{dmax}}$. Os valores de $C_{fQ,\mathbf{dmax}}$ para perturbações aditivas das medidas obtidas pelo instrumento II e as perturbações multiplicativas para as medidas obtidas pelo instrumento III bem como $C_{fQ,\mathbf{dmax}}$ para o caso de ponderação de pesos são encontradas na Tabela 7. As Figuras 24 e 25 ilustram o gráfico índice para perturbação de medidas de um instrumento. Usando esse esquema de perturbação, nós podemos examinar a influência nas medidas obtidas pelos instrumentos II e III, sob os casos aditivos e multiplicativos, respectivamente, de forma análoga a feita na seção anterior.

Na Figura 24 é observado alguma influência da observação 23 quando o instrumento

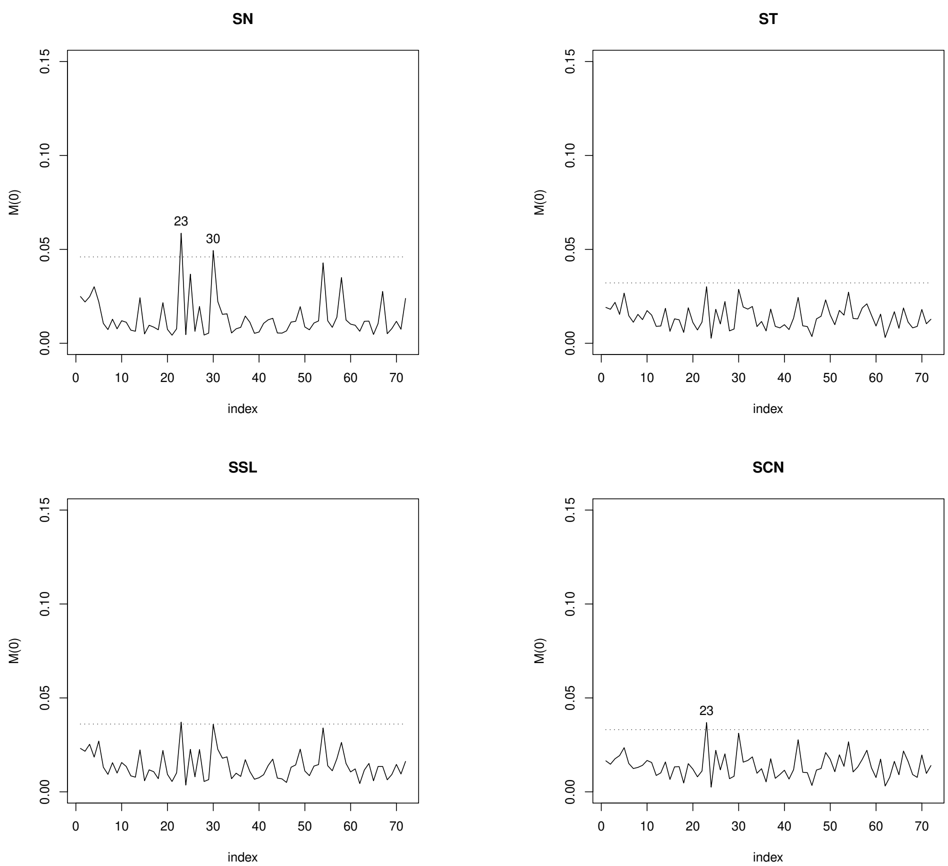

Figura 24: Conjunto de dados de Barnett - Gráficos de Índice de M(0) sob o caso de perturbação aditiva para a medição do instrumento II para os quatro modelos ajustados. As linhas horizontais delimitam o *benchmark* de Lee e Xu (2004) para M(0) com $c* = 3$.

II é perturbado sob uma SCN-MEM e das observações 23 e 30 sob uma SN-MEM.

Na Figura 25 é observado alguma influência das observações 1 e 3 quando o instrumento III é perturbado sob uma ST-MEM e SSL-MEM. Exceto esse caso da perturbação multiplicativa onde os modelos sob a ST-MEM e SSL-MEM apresentam pontos influentes e a SN-MEM não, os outros resultados foram dentro do esperado que a influência de observações atípicas sejam reduzidas quando nós consideramos distribuições com caudas mais pesadas que a distribuição skew-normal. Para esse conjunto de dados a SCN-MEM acomoda um pouco melhor as observações influentes.

As medidas TRC e MRC foram calculadas com as observações 23 e 30 excluídas e

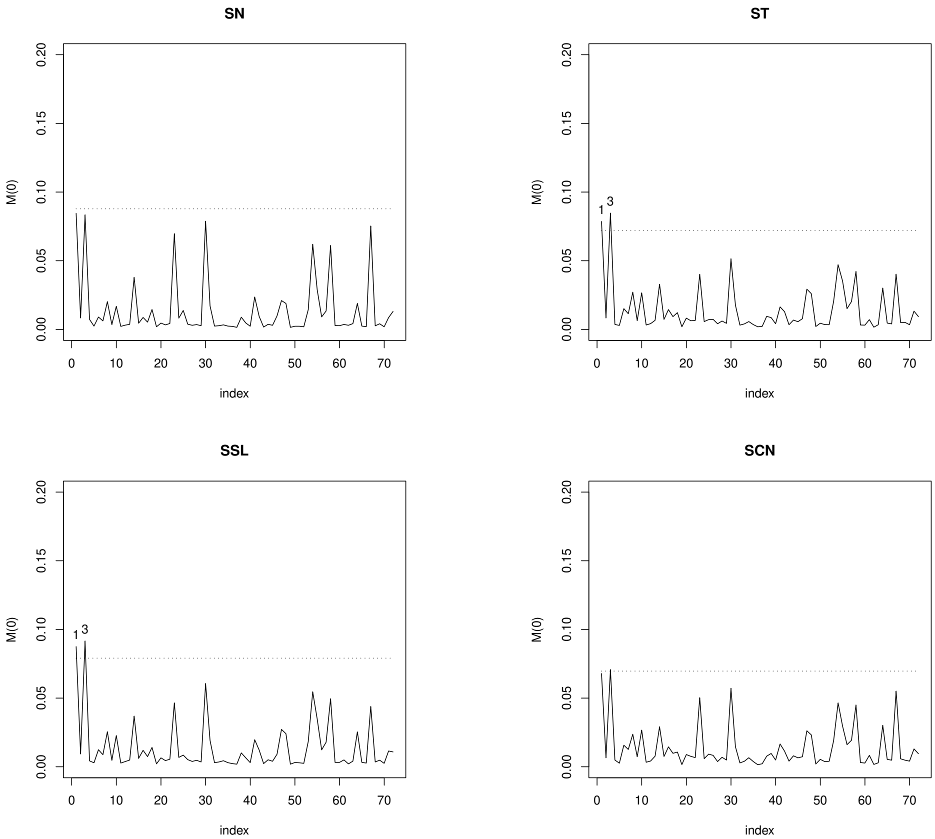

Figura 25: Conjunto de dados de Barnett - Gráficos de Índice de M(0) sob o caso de perturbação multiplicativa para a medição do instrumento II para os quatro modelos ajustados. As linhas horizontais delimitam o *benchmark* Lee e Xu (20040 para M(0) com $c* = 3.5$.

Tabela 8: Conjunto de dados de Barnett. Comparação das mudanças relativas nos estimadores de máxima verossimilhança em termos de TRC e MRC para os quatro modelos MESN selecionados.

Observação	Distribuição	TRC	MRC
23 e 30	SN	0,9777	0,1807
	ST	0,8648	0,1691
	SSL	0,9005	0,1800
	SCN	0,9431	0,1656

os resultados são apresentados na Tabela 8. Note que apesar de pequenas as maiores mudanças são obtidas sob a distribuição SN. Como esperado, os resultados indicam que os estimadores de máxima verossimilhança são menos sensíveis a presença de observações influentes quando nós usamos distribuições com caudas mais pesadas que a SN, apesar de nesse caso ganho ser muito pequeno quando comparado com o da seção anterior.

6 *Considerações Finais*

6.1 Conclusões

Nessa dissertação discutimos a utilização de distribuições assimétricas como uma alternativa quando os modelos simétricos se mostram pouco eficientes na acomodação de observações atípicas. Foi apresentado no Capítulo 2 a classe de distribuições de misturas de escala skew-normal proposta por Branco e Dey (2001) e mais recentemente discutida por Ghosh et al. (2008).

No Capítulo 3, foi apresentado uma generalização do modelo de calibração comparativa conhecida como modelos de erros de medição e em seguida apresentamos o modelo estrutural com erros nas variáveis sob a classe das distribuições MESN, como definido em Lachos et al. (2008). Além disso, uma breve introdução a variáveis latentes e uma extensão do algoritmo EM para a classe MESN-MEM foram apresentadas.

A metodologia desenvolvida por Zhu e Lee (2001) tem um enfoque especial no Capítulo 4 onde é apresentada com detalhes. Outras metodologias como a de Cook (1986) e a de Poon e Poon (1999) foram descritas para um melhor entendimento de influência local. Também foram descritos alguns esquemas de perturbação.

Na aplicação podemos validar a utilidade de modelos assimétricos com cauda mais pesadas quando um modelo simétrico não é satisfatório. Podemos avaliar através de envelopes simulados e de critérios de seleção de modelos, AIC, BIC, HQ, a adequação ou não de um determinado modelo. Para cada esquema de perturbação podemos analisar através de uma visualização gráfica que em geral modelos com caudas mais pesadas acomodam melhor as observações atípicas, isso porque uma das características da classe de distribuição MESN é que elas podem atribuir naturalmente pesos diferentes para cada observação e consequentemente controlar a influência da observação no processo de estimação, Zeller (2009).

Em resumo, o estudo de diagnóstico de influência local é uma das ferramentas para avaliar o impacto da estimação na presença de pequenas perturbações nos dados ou no modelo. Permitindo, dessa forma, detectar observações influentes que possam causar mudanças significativas nas conclusões estatísticas. Com a utilização de modelos que acomodem bem as observações mais extremas e a utilização de métodos confiáveis, é de se esperar que se consiga identificar os dados que realmente possam influenciar na estimação dos parâmetros.

Nesse trabalho utilizamos o pacote estatístico R 2.9.0 e o Matlab R2007a.

6.2 Perspectivas Futuras

Algumas linhas de pesquisa podem ser levadas em consideração:

1. Apresentar um estudo de estimação e diagnóstico para outros membros da família MESN.

2. Estudar com mais detalhes o caso mais simples (para dois instrumentos de medição), que corresponde ao modelo de regressão linear simples com erros nas variáveis estrutural sob a classe das distribuições MESN.

Apêndice A - Lemas

Neste apêndice apresentamos os lemas citados nas Seções 2.2.2, 2.4.3 e 2.6, que são utéis para um melhor entendimento do texto. A seguir o lema utilizado na Seção 2.2.2 para demonstração da Proposição 8.

Lema 1. *Seja* $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*. Então, para algum vetor fixo* $\mathbf{a}$ *de dimensão* k *e alguma matriz fixa* $\mathbf{B}_{k\times n}$*, temos*

$$
\begin{aligned}
E[\Phi_k(\mathbf{a}+\mathbf{B}\mathbf{Y}|\boldsymbol{\eta},\boldsymbol{\Omega})] &= \Phi_k(\mathbf{a}|\boldsymbol{\eta}-\mathbf{B}\boldsymbol{\mu},\boldsymbol{\Omega}+\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top) \qquad (A.1)\\
E[\phi_k(\mathbf{a}+\mathbf{B}\mathbf{Y}|\boldsymbol{\eta},\boldsymbol{\Omega})] &= \phi_k(\mathbf{a}|\boldsymbol{\eta}-\mathbf{B}\boldsymbol{\mu},\boldsymbol{\Omega}+\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top) \qquad (A.2)
\end{aligned}
$$

O próximo lema é citado na Seção 2.4.3 para obtermos a demonstração da Proposição 10.

Lema 2. *Seja* $\mathbf{Y} \in \mathbb{R}^p$ *um vetor aleatório com a seguinte função distribuição de probabildade*

$$
f(\mathbf{y}|u) = k^{-1}(u)\phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}A_y + u^{1/2}\mathbf{B}^\top\mathbf{y}),
$$

oem que u *é uma constante positiva,* $A_y \in \mathbb{R}$*,* $\mathbf{B}$ *qualquer vetor p-dimensional fixado e* $k(u) = \Phi_1(u^{1/2}\frac{A_y+\mathbf{B}\boldsymbol{\mu}}{\sqrt{1+\mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B}}})$ *é uma constante padronizada. Então,*

$$
E[\mathbf{Y}|u] = \mathbf{y} + u^{-1/2}\frac{\boldsymbol{\Sigma}\mathbf{B}}{\sqrt{1+\mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B}}}W_{\Phi_1}(u^{1/2}\frac{A+\mathbf{B}^\top\boldsymbol{\mu}}{\sqrt{1+\mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B}}})
$$

Demonstração. Primeiro observe, pelo uso do Lema 2 de Arellano-Valle et al. (2005), que

$$
\begin{aligned}
E[\mathbf{Y}|u] &= k^{-1}\int_{\mathbb{R}}\int_0^\infty \mathbf{y}\phi_1(t|u^{1/2}A_y + u^{1/2}\mathbf{B}^\top\mathbf{y}, 1)\phi(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})dtd\mathbf{y}\\
&= k^{-1}\int_0^\infty \phi_1(t|u^{1/2}A_y + u^{1/2}\mathbf{B}^\top\mathbf{y}, 1+\mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B})E_{\mathbf{Y}|t}[\mathbf{Y}]dt,
\end{aligned}
$$

em que $\mathbf{Y}|t \sim N_p(\boldsymbol{\mu} - \boldsymbol{\Lambda}\mathbf{B}(A + \mathbf{B}^\top\boldsymbol{\mu}) + u^{-1/2}\Lambda\mathbf{B}t, u^{-1}\boldsymbol{\Lambda})$, com $\boldsymbol{\Lambda} = (\boldsymbol{\Sigma}^{-1} + \mathbf{B}\mathbf{B}^\top)^{-1}$, e a demonstração segue pelo uso das propriedades conhecidas da distribuição normal truncada (Veja Johnson et al., 1994).

Os lemas a seguir são utilizados na Seção 2.6 com o intuito de auxiliar o leitor em pontos especifícos da estimação de máxima verossimilhança.

Lema 3. *Seja* $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *e* $\mathbf{X} \sim N_q(\boldsymbol{\eta}, \boldsymbol{\Omega})$. *Então*

$$\begin{aligned}\phi_p(\mathbf{y}|\boldsymbol{\mu} + \mathbf{A}\mathbf{x}, \boldsymbol{\Sigma})\phi_q(\mathbf{x}|\boldsymbol{\eta}, \boldsymbol{\Omega}) &= \phi_p(\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\eta}, \Sigma + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top) \\ &\times \phi_q(\mathbf{x}|\boldsymbol{\eta} + \Lambda\mathbf{A}^\top\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{\eta}), \Lambda)\end{aligned}$$

em que $\Lambda = (\boldsymbol{\Omega}^{-1} + \mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{A})$

Demonstração. Fazendo $\mathbf{z} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{A}\boldsymbol{\eta}$ e $\mathbf{w} = \mathbf{x} - \boldsymbol{\eta}$ a prova segue do fato que

$$\begin{aligned}(\mathbf{z} - \mathbf{A}\mathbf{w})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{A}\mathbf{w}) + \mathbf{w}^\top\boldsymbol{\Omega}^{-1}\mathbf{w} &= \mathbf{z}(\Sigma + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top)^{-1}\mathbf{z} \\ &+ (\mathbf{w} - \Lambda\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{z})^\top\Lambda^{-1}(\mathbf{w} - \Lambda\mathbf{A}^\top\boldsymbol{\Sigma}^{-1}\mathbf{z})\end{aligned}$$

a prova segue também por notar que $|\boldsymbol{\Sigma} + \mathbf{A}\Omega\mathbf{A}^\top||\Lambda| = |\boldsymbol{\Sigma}||\boldsymbol{\Omega}|$.

O seguinte lema é uma propriedade da distribuição half-normal (veja Johnson et al. 1994, Seção 10.1).

Lema 4. *Seja* $X \sim N(\eta, \tau^2)$. *Então, para qualquer constante real* a *segue que*

$$\begin{aligned}E[X|X > a] &= \eta + \frac{\phi_1(\frac{a-\eta}{\tau})}{1 - \Phi_1(\frac{a-\eta}{\tau})}\tau \\ E[X^2|X > a] &= \eta^2 + \tau^2 + \frac{\phi_1(\frac{a-\eta}{\tau})}{1 - \Phi_1(\frac{a-\eta}{\tau})}(\eta + a)\tau\end{aligned}$$

Apêndice B - A Matriz de Informação Observada

Neste apêndice apresentamos a matriz de informação observada da Seção 3.5.

Considere as seguintes notações:

$\mathbf{A}_i = (\mathbf{Y}_i - \boldsymbol{\alpha} - 2q_i\boldsymbol{\beta})^\top D^{-1}(\boldsymbol{\psi})$, $\mathbf{M} = D^{-1}(\boldsymbol{\phi})\mathbf{b}\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})$,

$\mathbf{M}_1 = D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi})$, $\boldsymbol{\psi} = (\phi_2, ...\phi_p)^\top$, $q_i = \mu_x + c^{-1}\phi_x a_i$,

$\mathbf{W}_{2i} = \mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x$, $\mathbb{I}_{(p)} = [\mathbf{0}, \mathbf{I}_{(p-1)}]$ de dimensão $(p-1)$ x p e $D_1(\boldsymbol{\beta}) = [\mathbf{0}, D(\boldsymbol{\beta})]$ de dimensão p x r.

Considerando as notações acima, derivando em forma vetorial e após algumas manipulações algébricas, temos que

$\bullet A_i$

De $A_i = A_x a_i$ segue que $\dfrac{\partial A_i}{\partial \boldsymbol{\gamma}} = [A_x \dfrac{\partial a_i}{\partial \boldsymbol{\gamma}} + a_i \dfrac{\partial A_x}{\partial \boldsymbol{\gamma}}]$ e assim

$$\frac{\partial^2 A_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = [\frac{\partial A_x}{\partial \boldsymbol{\gamma}} \frac{\partial a_i}{\partial \boldsymbol{\tau}^\top} + A_x \frac{\partial^2 a_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} + \frac{\partial a_i}{\partial \boldsymbol{\gamma}} \frac{\partial A_x}{\partial \boldsymbol{\tau}^\top} + a_i \frac{\partial^2 A_x}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}]$$

com $\boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi_x, \boldsymbol{\phi}, \lambda_x$, $A_x = \lambda_x \Lambda_x/(\phi_x + \lambda_x^2 \Lambda_x)^{1/2}$, $\Lambda_x = \phi_x/c$, $a_i = \mathbf{X}_i{}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$, $\mathbf{X}_i = \mathbf{Z}_i - \mathbf{a} - \mathbf{b}\mu_x$, $c = 1 + \phi_x \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$, $i = 1, ..., n$. Usando os resultados de Nel

(1980) relacionados para derivação de vetores segue que,

$$
\begin{aligned}
\frac{\partial A_x}{\partial \boldsymbol{\gamma}} &= 0, \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \frac{\partial A_x}{\partial \boldsymbol{\beta}} = -\frac{(2c+\lambda_x^2)}{\lambda_x^2} {A_x}^3 D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \frac{\partial A_x}{\partial \lambda_x} = \frac{\phi_x}{\Lambda_x^2 \lambda_x^3} {A_x}^3,\\
\frac{\partial A_x}{\partial \boldsymbol{\phi}} &= \frac{(2c+\lambda_x^2)}{2\lambda_x^2} {A_x}^3 D(\mathbf{b}) D^{-2}(\boldsymbol{\phi})\mathbf{b}, \frac{\partial A_x}{\partial \phi_x} = \frac{(2c+\lambda_x^2-c^2)}{2\phi_x^2\lambda_x^2} {A_x}^3,\\
\frac{\partial a_i}{\partial \mu_x} &= -\mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}, \frac{\partial a_i}{\partial \boldsymbol{\alpha}} = -D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \frac{\partial a_i}{\partial \boldsymbol{\beta}} = D^{-1}(\boldsymbol{\psi})\mathbf{W}_{2i} - \mu_x D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial a_i}{\partial \phi_x} &= 0, \frac{\partial a_i}{\partial \boldsymbol{\phi}} = -D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{X}_i, \frac{\partial a_i}{\partial \lambda_x} = 0,\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= -(4\frac{\phi_x}{\lambda_x^2} {A_x}^3 - \frac{3(2c+\lambda_x^2)^2}{\lambda_x^4} {A_x}^5)\mathbf{M}_1 - \frac{2c+\lambda_x^2}{\lambda_x^2} {A_x}^3 D^{-1}(\boldsymbol{\psi}),\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial \phi_x} &= -[\frac{2(c-1)}{\lambda_x^2\phi_x} {A_x}^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^{-2})}{2\lambda_x^4\phi_x^2} {A_x}^5] D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial \boldsymbol{\phi}^\top} &= [2\frac{\phi_x}{\lambda_x^2} {A_x}^3 - \frac{3(2c+\lambda_x^2)^2}{2\lambda_x^4} {A_x}^5] D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\mathbf{b}^T D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\\
&+ \frac{2c+\lambda_x^2}{\lambda_x^2} {A_x}^3 \mathbb{I}_{(p)} D(\mathbf{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\beta}\partial \lambda_x} &= \frac{\phi_x {A_x}^3}{\lambda_x^5 \Lambda_x^2}(-3{A_x}^2(2c+\lambda_x^2) + 4\lambda_x^2\Lambda_x) D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta},\\
\frac{\partial^2 A_x}{\partial \phi_x\partial \phi_x} &= -\frac{\lambda_x^2+1}{\lambda_x^2\phi_x^3} {A_x}^3 + \frac{3(2c+\lambda_x^2-c^2)^2}{4\lambda_x^4\phi_x^4} {A_x}^5,\\
\frac{\partial^2 A_x}{\partial \phi_x\partial \boldsymbol{\phi}^\top} &= -[\frac{(c-1)}{\lambda_x^2\phi_x} {A_x}^3 + \frac{3(2c+\lambda_x^2)(2c+\lambda_x^2-c^2)}{4\lambda_x^4\phi_x^2} {A_x}^5]\mathbf{b}^\top D(\mathbf{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 A_x}{\partial \phi_x\partial \lambda_x} &= \frac{c-2}{\lambda_x^3\Lambda_x\phi_x} {A_x}^3 + \frac{3(2c+\lambda_x^2-c^2)}{2\lambda_x^5\Lambda_x^2\phi_x} {A_x}^5,\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}^\top} &= [-\frac{\phi_x}{\lambda_x^2} {A_x}^3 + \frac{3(2c+\lambda_x^2)^2}{4\lambda_x^4} {A_x}^5] D(\mathbf{b})D^{-1}(\boldsymbol{\phi})\mathbf{M}D^{-1}(\boldsymbol{\phi})D(\mathbf{b})\\
&- \frac{2c+\lambda_x^2}{\lambda_x^2} D^2(\mathbf{b})D^{-3}(\boldsymbol{\phi}) {A_x}^3,\\
\frac{\partial^2 A_x}{\partial \boldsymbol{\phi}\partial \lambda_x} &= \frac{\phi_x {A_x}^3}{2\lambda_x^5\Lambda_x^2}[3{A_x}^2(2c+\lambda_x^2) - 4\lambda_x^2\Lambda_x] D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{b},\\
\frac{\partial^2 A_x}{\partial \lambda_x\partial \lambda_x} &= -\frac{3\phi_x}{\lambda_x^4\Lambda_x^2} {A_x}^3 + \frac{3\phi_x^2}{\lambda_x^6\Lambda_x^4} {A_x}^5, \frac{\partial^2 a_i}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}^\top} = 2D(\mathbf{X}_i)D(\mathbf{b})D^{-3}(\boldsymbol{\phi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\tau}^\top} &= 0, \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \phi_x, \lambda_x, \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}, \phi_x, \lambda_x|\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\tau}^\top} &= 0, \boldsymbol{\gamma} = \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\tau} = \phi_x, \lambda_x, \frac{\partial^2 a_i}{\partial \mu_x\partial \boldsymbol{\beta}^\top} = -2\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi}),\\
\frac{\partial^2 a_i}{\partial \mu_x\partial \boldsymbol{\phi}^\top} &= \mathbf{b}^\top D(\mathbf{b})D^{-2}(\boldsymbol{\phi}), \frac{\partial^2 a_i}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\beta}^\top} = -D^{-1}(\boldsymbol{\psi}), \frac{\partial^2 a_i}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\phi}^\top} = \mathbb{I}_{(p)} D(\mathbf{b})D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 a_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= -2\mu_x D^{-1}(\boldsymbol{\psi}), \frac{\partial^2 a_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\phi}^\top} = -\mathbb{I}_{(p)} D(\mathbf{Z}_i - \boldsymbol{\alpha} - 2\mu_x\mathbf{b})D^{-2}(\boldsymbol{\phi}),
\end{aligned}
$$

$\bullet d_i$

Para $d_i = \mathbf{X}_i{}^\top \mathbf{\Sigma}^{-1}\mathbf{X}_i$, com $d_{i\boldsymbol{\gamma}} = \frac{\partial d_i}{\partial \boldsymbol{\gamma}}$, e $d_{i\boldsymbol{\gamma}\boldsymbol{\gamma}^\top} = \frac{\partial^2 d_i}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\tau}^\top}$, disto segue que

$d_{i\mu_x} = -2\mathbf{b}^\top\mathbf{\Sigma}^{-1}\mathbf{X}_i$, $d_{i\boldsymbol{\alpha}} = -2\mathbb{I}_{(p)}\mathbf{\Sigma}^{-1}\mathbf{X}_i$,

$d_{i\boldsymbol{\beta}} = -2q_i D^{-1}(\boldsymbol{\psi})\mathbf{W}_{2i} + 2c^{-1}\phi_x a_i q_i D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}$, $d_{i\phi_x} = -c^{-2}a_i{}^2$, $d_{i\lambda_x} = 0$,

$d_{i\boldsymbol{\phi}} = -D^{-2}(\boldsymbol{\phi})D(\mathbf{X}_i)\mathbf{X}_i + 2c^{-1}\phi_x a_i D^{-2}(\boldsymbol{\phi})D(\mathbf{b})\mathbf{X}_i - c^{-2}\phi_x^2 a_i{}^2 D^{-2}(\boldsymbol{\phi})D(\mathbf{b})\mathbf{b}$,

$d_{i\mu_x\mu_x} = 2\mathbf{b}^\top\mathbf{\Sigma}^{-1}\mathbf{b}$, $d_{i\mu_x\boldsymbol{\alpha}^\top} = 2\mathbf{b}^\top\mathbf{\Sigma}^{-1}\mathbb{I}_{(p)}{}^\top$, $d_{i\mu_x\boldsymbol{\beta}^\top} = -2c^{-1}\mathbf{A}_i$,

$d_{i\mu_x\phi_x} = 2\frac{(c-1)}{c^2\phi_x}a_i$, $d_{i\mu_x\boldsymbol{\phi}^\top} = 2c^{-1}\mathbf{X}_i\mathbf{\Sigma}^{-1}D^{-1}(\boldsymbol{\phi})D(\mathbf{b})$, $d_{i\boldsymbol{\alpha}\boldsymbol{\alpha}^\top} = 2\mathbb{I}_{(p)}\mathbf{\Sigma}^{-1}\mathbb{I}_{(p)}^\top$,

$d_{i\boldsymbol{\alpha}\boldsymbol{\beta}^\top} = 2q_i[D^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_x\mathbf{M}_1] + 2c^{-1}\phi_x D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}(\mathbf{Y}_i - \boldsymbol{\alpha})^\top D^{-1}(\boldsymbol{\psi})$,

$d_{i\boldsymbol{\alpha}\phi_x} = 2c^{-2}a_i D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}$, $d_{i\boldsymbol{\alpha}\boldsymbol{\phi}^\top} = 2\mathbb{I}_{(p)}\mathbf{\Sigma}^{-1}D^{-1}(\boldsymbol{\psi})[D(\mathbf{X}_i) - c^{-1}\phi_x a_i D(\mathbf{b})]$,

$$\begin{aligned}
d_{i\boldsymbol{\beta}\boldsymbol{\beta}^\top} = {} & 4\frac{\phi_x^2}{c^2}a_i[D^{-1}(\boldsymbol{\psi})(\mathbf{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)\boldsymbol{\beta}^\top D^{-1}(\boldsymbol{\psi})\\
& +D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}(\mathbf{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)^\top D^{-1}(\boldsymbol{\psi})]\\
& -2c^{-1}\phi_x D^{-1}(\boldsymbol{\psi})(\mathbf{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)(\mathbf{Y}_i - \boldsymbol{\alpha} - 2\boldsymbol{\beta}\mu_x)^\top D^{-1}(\boldsymbol{\psi})\\
& +2\mu_x(q_i + c^{-1}\phi_x a_i)D^{-1}(\boldsymbol{\psi}) + 2\frac{\phi_x^2}{c^2}a_i{}^2[D^{-1}(\boldsymbol{\psi}) - 4\frac{\phi_x}{c}\mathbf{M}_1],
\end{aligned}$$

$d_{i\boldsymbol{\beta}\phi_x} = -2c^{-2}a_i\mathbf{A}_i{}^\top$, $d_{i\phi_x\phi_x} = 2\frac{c^{-3}}{\phi_x}(c-1)a_i{}^2$,

$d_{i\boldsymbol{\beta}\boldsymbol{\phi}^\top} = 2[q_i\mathbb{I}_{(p)}D(\mathbf{Z}_i - \boldsymbol{\alpha} - q_i\mathbf{b}) + c^{-1}\phi_x\{\mathbf{A}_i{}^\top(\mathbf{Z}_i - \boldsymbol{\alpha} - \mathbf{b}q_i)^\top D(\mathbf{b})]D^{-2}(\boldsymbol{\phi})$,

$d_{i\phi_x\boldsymbol{\phi}^\top} = (-2c^3\phi_x a_i{}^2 D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{b} + 2c^{-2}a_i D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\mathbf{X}_i)^\top$,

$$\begin{aligned}
d_{i\boldsymbol{\phi}\boldsymbol{\phi}^\top} = {} & 2D^{-3}(\boldsymbol{\phi})D^2(\mathbf{X}_i) - 4c^{-1}\phi_x a_i D^{-3}(\boldsymbol{\phi})D(\mathbf{b})D(\mathbf{X}_i)\\
& -2c^{-1}\phi_x D^{-2}(\boldsymbol{\phi})D(\mathbf{b})\mathbf{X}_i\mathbf{X}_i{}^\top D(\mathbf{b})D^{-2}(\boldsymbol{\phi})\\
& +2c^{-2}\phi_x^2 D^{-2}(\boldsymbol{\phi})D(\mathbf{b})\mathbf{X}_i\mathbf{X}_i{}^\top\mathbf{M}D^{-1}(\boldsymbol{\phi})D(\mathbf{b})\\
& +2c^{-2}\phi_x^2 a_i{}^2 D^{-3}(\boldsymbol{\phi})D^2(\mathbf{b}) - 2c^{-3}\phi_x^3 a_i{}^2 D^{-1}(\boldsymbol{\phi})D(\mathbf{b})\mathbf{M}D(\mathbf{b})D^{-1}(\boldsymbol{\phi})\\
& +2c^{-2}\phi_x^2 D^{-1}(\boldsymbol{\phi})D(\mathbf{b})\mathbf{M}\mathbf{X}_i\mathbf{X}_i{}^\top D(\mathbf{b})D^{-2}(\boldsymbol{\phi}).
\end{aligned}$$

$\bullet log(|\boldsymbol{\Sigma}|)$

$$
\begin{aligned}
\frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \boldsymbol{\tau} \partial \boldsymbol{\gamma}^\top} &= 0, \boldsymbol{\tau} = \mu_x, \boldsymbol{\alpha}, \lambda_x | \boldsymbol{\gamma} = \mu_x, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi_x, \boldsymbol{\phi}, \lambda_x,\\
\frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \boldsymbol{\beta} \partial \phi_x} &= 2c^{-2} D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}, \frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \phi_x \partial \phi_x} = \frac{1}{c^2 \phi_x^2}(c-1)^2,\\
\frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\phi}^\top} &= -2c^{-1}\phi_x [D_1(\boldsymbol{\beta}) - c^{-1}\phi_x D^{-1}(\boldsymbol{\psi})\boldsymbol{\beta}\mathbf{b}^\top D(\mathbf{b})] D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= 2c^{-1}\phi_x [D^{-1}(\boldsymbol{\psi}) - 2c^{-1}\phi_x \mathbf{M}_1], \frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \phi_x \partial \boldsymbol{\phi}^\top} = -c^{-2}\mathbf{b}^\top D(\mathbf{b}) D^{-2}(\boldsymbol{\phi}),\\
\frac{\partial^2 log(|\boldsymbol{\Sigma}|)}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= -D^{-2}(\boldsymbol{\phi}) - c^{-2}\phi_x^2 D(\mathbf{b}) D^{-1}(\boldsymbol{\phi}) \mathbf{M} D^{-1}(\boldsymbol{\phi}) D(\mathbf{b}) + 2c^{-1}\phi_x D^2(\mathbf{b}) D^{-3}(\boldsymbol{\phi}).
\end{aligned}
$$

Apêndice C - Implementação do Algoritmo EM

Nesse apêndice mostramos o programa escrito em Matlab R2007a, para calcular os estimadores de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$ usando o algoritmo EM, conforme descrito na Seção 3.3, para qualquer conjunto de dados.

```
function y=EMVcalSNI(dados,nu,Ind,alfa,beta,phi,mux,phix,lambdax)
%dados é o conjunto de dados
%Ind é uma indicados da distribuição, em que 1:SN, 2:ST, 3:SSL e 4:SCN
%nu, alfa, beta phi, mux, phix e lambdax são os parâmetros a serem estimados
%encontra os EMV no modelo com erros nas variáveis multivariado
n=size(dados,1); %número de observações
p=size(dados,2); %número de instrumentos
u=[alfa' beta' phi' mux phix lambdax nu']'; %valores iniciais
criterio=norm(u);
k=0;
uj=zeros(n,1);
utj=zeros(n,1);
ut2j=zeros(n,1);
uxj=zeros(n,1);
ux2j=zeros(n,1);
uxtj=zeros(n,1);
xj=zeros(n,1);

while criterio> 0.0001
```

```
k=k+1
deltax=lambdax/sqrt(1+lambdax ^ 2);
w=sqrt(phix)*deltax;v=phix/(1+lambdax ^ 2);
a=[0 alfa']'; b=[1 beta']';
Dphi=diag(phi);
Sigma=Dphi+phix*b*b';
med=a+b*mux;
Mt=(1+w ^ 2*b'*(Dphi+v*b*b') ^ (-1)*b) ^ (-1);
Tx=(v ^ (-1)+b'*Dphi ^ (-1)*b) ^ (-1);
s=w*(1-Tx*b'*inv(Dphi)*b);
c=1+phix*b'*inv(Dphi)*b;
Lambdax=phix/c;
for j=1:n
z=(dados(j,:))';
d(j)=(z-med)'*inv(Sigma)*(z-med);
metj=Mt*w*b'*(Dphi+v*b*b') ^ (-1)*(z-a-b*mux);
Ajj=metj/sqrt(Mt);
ai=b'*inv(Dphi)*(z-med);
muxi=Lambdax*ai;

%Etapa E - Calculando o Estimador de Bayes

if Ind==1

uj(j,1)=1;
esper=normpdf(Ajj,0,1)/normcdf(Ajj,0,1);
xi(j,1)=mux+muxi+esper*Lambdax*lambdax/sqrt(phix+lambdax ^ 2*Lambdax);

end

if Ind==2
```

```
esper=2*(nu/2)^(nu/2)*gamma((p+nu+1)/2)/(gamma(nu/2)*sqrt(2*pi)^(p+1)
*det(Sigma)^(1/2)*((d(j)+nu+Ajj^2)/2)^((p+1+nu)/2)
*veroST1(z,med,Sigma,Ajj,nu));
uj(j,1)=2*(nu/2)^(nu/2)*gamma((p+nu+2)/2)*tcdf(sqrt((p+nu+2)/(d(j)+nu))
*Ajj,p+nu+2)/(gamma(nu/2)*sqrt(2*pi)^p*det(Sigma)^(1/2)
*veroST1(z,med,Sigma,Ajj,nu)*((d(j)+nu)/2)^((p+2+nu)/2));
esper2=(nu)^(nu/2)*gamma((p+nu-1)/2)/(gamma(nu/2)*sqrt(pi)^(p+1)
*det(Sigma)^(1/2)*((d(j)+nu+Ajj^2))^((p-1+nu)/2)
veroST1(z,med,Sigma,Ajj,nu));
xi(j,1)=mux+muxi+esper2*Lambdax*lambdax/sqrt(phix+lambdax^2*Lambdax);

end

if Ind==3

faux1=@(x)exp(-0.5*d(j)*x).*x.^(p/2+nu).*normcdf(sqrt(x).*Ajj,0,1);
faux2=@(x)exp(-0.5*d(j)*x).*x.^(p/2+nu-1).*normcdf(sqrt(x).*Ajj,0,1);
faux3=@(x)exp(-0.5*(d(j)+Ajj^2)*x).*x.^(p/2+nu-1/2);
faux4=@(x)exp(-0.5*(d(j)+Ajj^2)*x).*x.^(p/2+nu-3/2);
uj(j,1)=quad(faux1,0,1)/quad(faux2,0,1);
esper=(1/sqrt(2*pi))*(quad(faux3,0,1)/quad(faux2,0,1));
esper2=(1/sqrt(2*pi))*(quad(faux4,0,1)/quad(faux2,0,1));
xi(j,1)=mux+muxi+esper2*Lambdax*lambdax/sqrt(phix+lambdax^2*Lambdax);

end

if Ind==4

fy=2*(nu(1)*mvnpdf(z',med',Sigma/nu(2))*normcdf(nu(2)^(1/2)*Ajj,0,1)+(1-nu(1))
*mvnpdf(z',med',Sigma)*normcdf(Ajj,0,1));
```

```
uj(j,1)=2*(nu(1)*nu(2)*mvnpdf(z',med',Sigma/nu(2))*normcdf(nu(2)^(1/2)*Ajj,0,1)
+(1-nu(1))*mvnpdf(z',med',Sigma)*normcdf(Ajj,0,1))/fy;
esper=2*(nu(1)*nu(2)^(1/2)*mvnpdf(z',med',Sigma/nu(2))*normpdf(nu(2)^(1/2)
*Ajj,0,1)+(1-nu(1))*mvnpdf(z',med',Sigma)*normpdf(Ajj,0,1))/fy;
esper2=2*(nu(1)*nu(2)^(-1/2)*mvnpdf(z',med',Sigma/nu(2))*normpdf(nu(2)^(1/2)
*Ajj,0,1)+(1-nu(1))*mvnpdf(z',med',Sigma)*normpdf(Ajj,0,1))/fy;
xi(j,1)=mux+muxi+esper2*Lambdax*lambdax/sqrt(phix+lambdax^2*Lambdax);

end

utj(j,1)=uj(j,1)*metj+sqrt(Mt)*esper;
ut2j(j,1)=Mt+uj(j,1)*metj^2+sqrt(Mt)*metj*esper;
r=mux+Tx*b'*inv(Dphi)*(z-a-b*mux);
uxj(j,1)=r*uj(j,1)+s*utj(j,1);
ux2j(j,1)=Tx+r^2*uj(j,1)+2*r*s*utj(j,1)+s^2*ut2j(j,1);
uxtj(j,1)=r*utj(j,1)+s*ut2j(j,1);

end

suma1=0;
suma2=0;
suma4=0;
suma5=0;
suma6=0;

for j=1:n
suma1=suma1+uj(j)*dados(j,2:p)'-beta*uxj(j,1);
suma2=suma2+(dados(j,2:p)'-alfa)*uxj(j,1);
suma4=suma4+uj(j)*(dados(j,:)'-a).*(dados(j,:)'-a)-2*(dados(j,:)'-a).*b*uxj(j,1)+ux2j(j,1)*b.*
suma5=suma5+ux2j(j,1)+uj(j)*mux^2+w^2*ut2j(j,1)-2*uxj(j,1) *mux-2*w*uxtj(j,1)
+2*mux*w*utj(j,1);
```

```
suma6=suma6+uxtj(j,1)-mux*utj(j,1);

end

alfa=suma1/sum(uj);
beta=suma2/sum(ux2j);
phi=suma4/n;
v=suma5/n;
w=suma6/sum(ut2j);
% w=0; % caso normal
mux=(sum(uxj)-w*sum(utj))/sum(uj);
lambdax=w/sqrt(v);
phix=v+w^2;

u=[alfa' beta' phi' mux phix lambdax nu']';
param(k,:)=u';

if k>2
criterio=norm(param(k,:)-param(k-1,:));
end

end
y.theta=u;
```

Apêndice D - Critérios de Seleção de Modelos

Na análise de regressão geralmente não se consegue identificar um único modelo candidato a representar a relação existente entre variável resposta e variáveis independentes. Pra se decidir qual é o modelo mais apropriado dentre um conjunto de modelos candidatos são utilizados os modelos de seleção. A ideia básica dos critérios de seleção de modelos consiste em incluir um termo penalizador para a inclusão de novos regressores, pois cada regressor a mais significa um grau de liberdade a menos. dentre os critérios mais conhecidos e utilizados estão o AIC e BIC. Neste apêndice apresentamos o cálculo dos critérios AIC, BIC e HQ utilizados nas Seções 5.3 e 5.2 para aplicação aos conjunos de dados de Chipkevitch e Barnett respectivamente.

1. Critério de Informação de Akaike (AIC)

$$AIC = -2\ell(\widehat{\boldsymbol{\theta}}) + 2p$$

2. Critério de Informação Bayesiano (BIC)

$$BIC = -2\ell(\widehat{\boldsymbol{\theta}}) + p\ log(n)$$

3. Critério de Hannan-Quinn (HQ)

$$BIC = -2\ell(\widehat{\boldsymbol{\theta}}) + p\ log(log(n))$$

em que $\ell(\widehat{\boldsymbol{\theta}})$ = verossimilhança obtida via algoritmo EM, n = número de observações e p = número de parâmetros.

Apêndice E - Conjuntos de Dados e ν ótimo

Neste apêndice apresentamos os conjuntos de dados utilizados nas Seções 5.2 (Tabela 9) e 5.3 (Tabela 10). Além disso a Figura 26 nos mostra um gráfico da log-verossimilhança *versus* ν e γ ótimos para o conjunto de dados de Chipkevitch e a Figura 27 nos mostra um gráfico semelhante para o conjunto de dados de Barnett.

Tabela 9: Conjunto de Dados de Chipkevitch.

	Instrumentos				
Observação	US	I	II	III	IV
1	5,0	7,5	5,9	8,0	9,0
2	5,7	5,0	4,8	6,0	10,0
3	7,4	5,0	6,8	9,0	12,0
4	2,6	3,5	3,1	4,0	4,0
5	5,7	5,0	5,0	6,0	7,0
6	6,1	5,0	4,4	7,0	8,0
7	6,2	5,0	6,0	8,0	9,0
8	10,4	10,0	8,8	10,0	10,0
9	9,1	7,5	7,9	10,0	11,0
10	14,8	10,0	13,0	12,0	15,0
11	16,4	12,5	10,3	17,5	17,5
12	9,6	7,5	8,2	10,0	11,0
13	15,7	15,0	19,8	20,0	20,0
14	3,0	2,0	2,0	3,0	4,0
15	16,4	15,0	17,3	20,0	20,0
16	17,6	15,0	17,3	20,0	22,5
17	10,0	7,5	7,9	12,0	12,0
18	4,1	3,5	4,4	4,0	6,0
19	12,7	10,0	11,4	12,0	12,0
20	2,7	3,5	4,1	2,5	6,0

	Instrumentos				
Observação	US	I	II	III	IV
21	10,2	10,0	11,1	12,0	13,5
22	16,5	10,0	15,3	15,0	15,0
23	4,5	3,5	3,9	6,0	7,0
24	5,6	5,0	4,5	4,5	6,0
25	11,0	7,5	9,7	9,0	11,0
26	9,2	10,0	11,3	12,0	13,5
27	8,5	7,5	8,8	12,0	12,0
28	5,4	5,0	6,1	8,0	8,0
29	6,7	7,5	7,2	10,0	8,0
30	5,3	5,0	5,9	8,0	10,0
31	20,0	20,0	16,3	25,0	22,5
32	18,8	15,0	16,3	20,0	25,0
33	13,9	12,5	12,2	15,0	17,5
34	9,4	10,0	10,3	12,0	13,5
35	9,1	7,5	10,8	12,0	12,0
36	14,1	15,0	13,0	13,5	15,0
37	9,3	10,0	8,4	10,0	10,0
38	20,9	20,0	22,1	25,0	25,0
39	11,5	10,0	10,6	15,0	13,5
40	9,7	10,0	9,7	11,0	12,0
41	13,7	12,5	11,6	17,5	15,0
42	8,9	10,0	8,1	12,0	12,0

Tabela 10: Conjunto de Dados de Barnett.

	Instrumentos			
Observação	I	II	III	IV
1	34,5	35,3	40,3	37,20
2	13,1	13,2	16,1	16,00
3	38,2	37,2	41,5	37,00
4	21,1	28,8	27,4	25,20
5	18,6	14,2	15,4	16,90
6	19,4	17,8	20,2	18,00
7	23,6	22,6	24,3	23,50
8	28,8	29,2	26,5	28,60
9	19,8	17,2	18,0	16,60
10	31,2	31,8	32,5	30,40
11	17,6	16,3	13,9	12,00
12	14,8	17,6	17,0	16,40

	Instrumentos			
Observação	I	II	III	IV
13	18,4	16,6	14,0	16,50
14	35,8	34,8	36,8	39,60
15	18,8	20,0	20,9	20,70
16	24,0	23,2	25,5	24,80
17	22,2	21,2	22,9	22,70
18	25,4	25,0	26,2	19,60
19	9,2	12,0	6,4	10,30
20	22,4	21,6	23,0	23,00
21	22,4	21,3	20,3	21,40
22	22,6	25,1	24,0	24,50
23	38,6	41,8	39,8	36,80
24	27,8	21,0	18,9	20,00
25	22,2	14,0	18,4	13,60
26	18,8	18,2	19,0	18,40
27	9,4	9,6	10,6	10,00
28	24,8	22,2	21,5	21,50
29	16,6	17,8	17,6	18,00
30	40,4	41,8	40,0	37,70
31	25,4	25,6	20,8	22,50
32	17,8	17,0	13,9	12,00
33	12,8	13,0	8,0	11,30
34	19,4	20,6	20,3	18,80
35	17,6	20,0	18,6	18,60
36	20,4	16,6	14,7	11,60
37	10,6	10,0	8,5	6,00
38	20,0	18,0	12,7	17,00
39	22,8	22,8	23,8	23,50
40	19,4	18,0	16,7	15,80
41	25,8	27,0	28,5	21,10
42	14,0	14,4	16,8	14,80
43	12,6	11,0	10,0	10,30
44	23,2	24,2	23,6	23,60
45	20,0	19,4	19,8	19,80
46	24,0	19,0	14,7	17,40
47	28,8	29,8	32,4	31,40
48	34,2	31,5	32,0	32,00
49	10,0	11,3	6,5	8,40
50	14,0	14,0	13,5	13,80
51	18,8	17,1	16,0	13,50
52	12,8	12,6	11,6	13,30
53	31,2	30,0	31,1	32,50
54	37,7	33,4	39,0	37,00
55	34,2	32,2	31,2	32,90

	Instrumentos			
Observação	I	II	III	IV
56	27,4	28,8	28,5	28,80
57	28,4	29,2	27,1	27,50
58	38,0	37,4	34,4	34,00
59	21,0	16,8	16,5	19,30
60	18,2	14,0	10,6	10,50
61	14,0	13,2	13,5	11,00
62	22,0	16,8	16,4	11,10
63	19,4	19,0	18,2	12,70
64	32,6	32,0	32,5	32,70
65	19,6	19,4	18,9	19,20
66	13,2	12,6	11,4	10,00
67	28,4	30,6	36,5	35,10
68	20,6	18,4	17,2	17,80
69	22,0	19,7	19,0	22,70
70	12,6	11,5	8,6	11,50
71	30,4	28,4	28,5	26,70
72	21,4	21,8	25,6	27,20

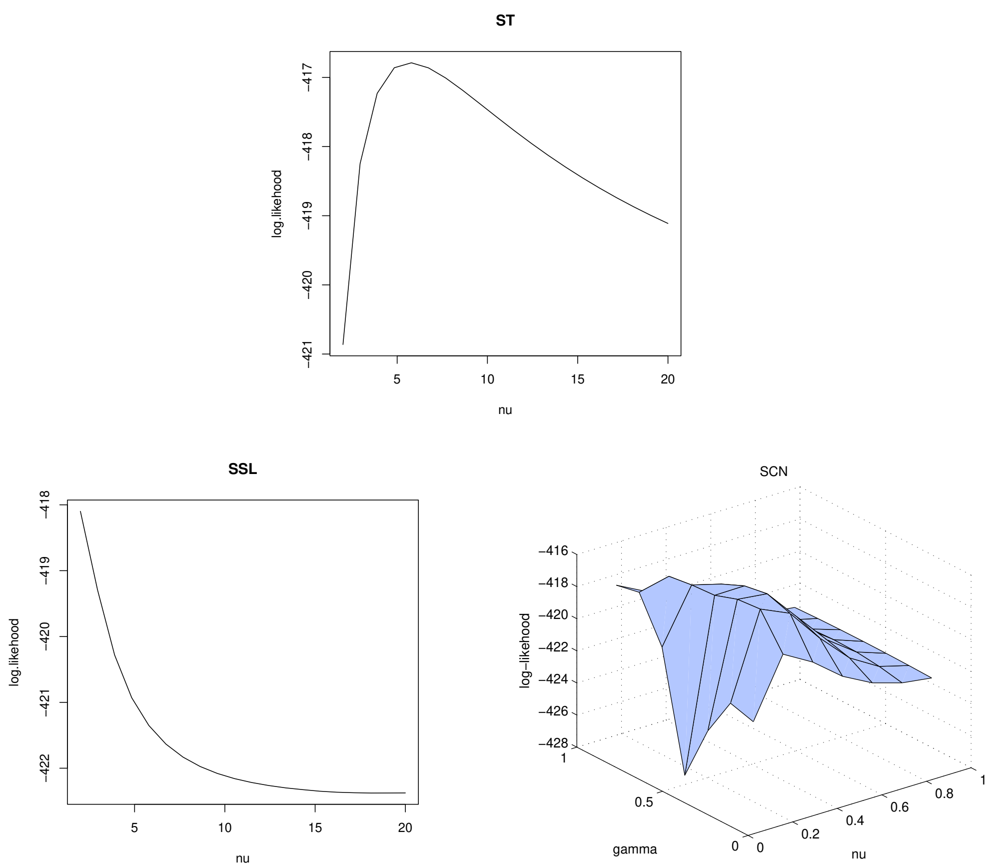

Figura 26: Conjunto de dados de Chipkevitch - Gráfico da log-verossimilhança *versus* ν e γ ótimos.

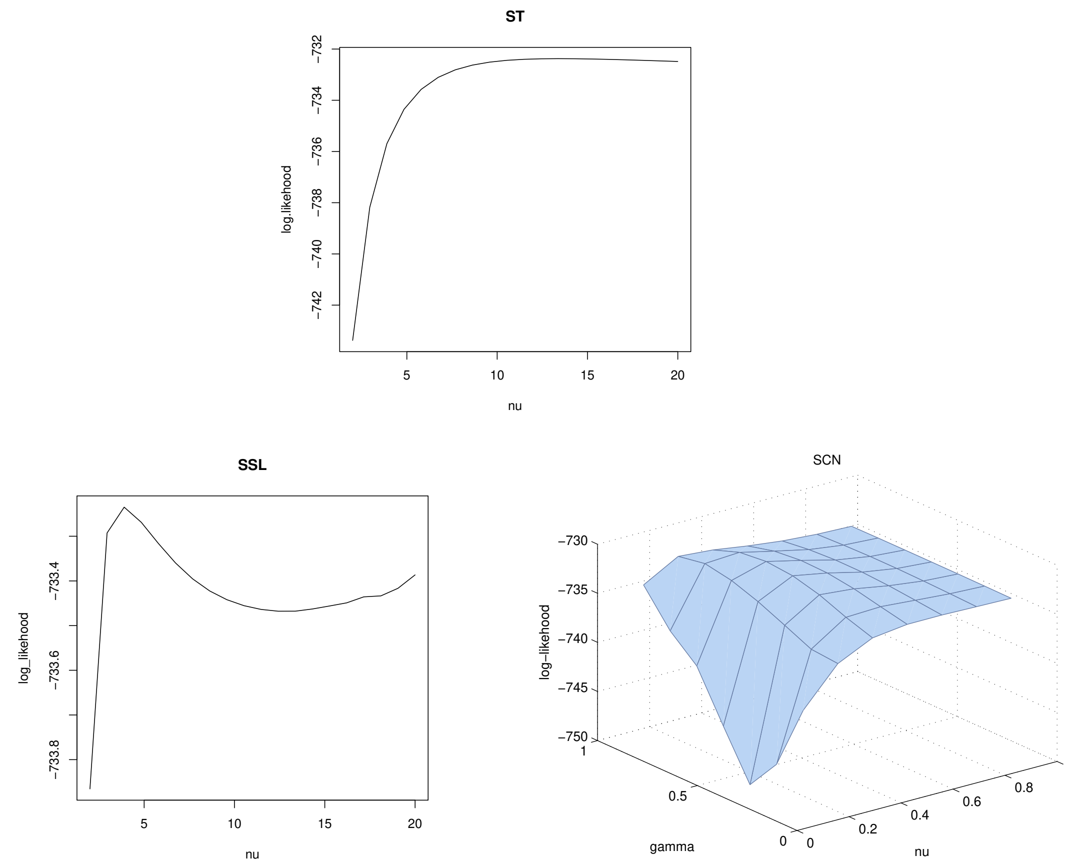

Figura 27: Conjunto de dados de Barnett - Gráfico da log-verossimilhança *versus* ν e γ ótimos.

Referências

[1] Abdullah, M.B.(1995). Detection of influential observations in functional errors in variables model. *Communication in Statistics - Theory and Methods,* **24**(6), 1585-1595.

[2] Arellano–Valle, R.B., Bolfarine, H. & Lachos, V.H. (2005a). Skew-normal linear mixed models. *Journal of Data Science,* **3**, 415-438.

[3] Arellano–Valle, R.B., Ozan, S., Bolfarine, H. & Lachos, V.H. (2005b). Skew-normal measurement error model. *Journal of Multivariate Analysis,* **98**, 265-281.

[4] Arellano–Valle, R.B. & Genton, M.G. (2005). On fundamental skew distrbutions. *Journal of Multivariate Analysis,* **96**, 93-116.

[5] Azzalini, A. (1985). A class of distributions which includes the normal ones . *Scandinavian Journal of Statistics*, **12**, 171-178.

[6] Azzalini, A. & Dalla–Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

[7] Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society,* B, **61**, 579-602.

[8] Azzalini, A. & Capitanio, A. (2003). Distributions generated by pertubation of simmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society,* B, **65**, 367-389.

[9] Azzalini, A. & Genton, M. G. (2008). Robust Likehood methods based on skew-t and related distributions. *International Statistical Review,* B, **76**, 106-129.

[10] Barnett, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics,* **25**, 129-142.

[11] Bedrick, E.J. (2001). An efficient scores test for comparing several measuring devices. *Journal of Quality Technology,* **33**, **1**, 96-103.

[12] Besley, E. Kuh & R.E. Welsch (1980). Regression diagnostics. Wiley. New York.

[13] Bolfarine, H. & Galea–Rojas, M. (1995). Maximum likelihood estimation of simultaneous pairwise linear structural relationships. *Biometrical Journal,* **37**, 673-689.

[14] Bolfarine, H. & Galea–Rojas, M. (1996). On structural comparative calibration under t-model. *Computational Statistics,* **11**, 63-85.

[15] Branco, M. & Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis,* **79**, 93-113.

[16] Carroll, R.J. & Stefanski, L.A. (1995). *Measurement error in nonlinear models.* Chapman and Hall, New York.

[17] Chatterjje, S. & Hadi, A.S. (1988). *Sensitivity analysis in linear regression.* John Wiley, New York.

[18] Cheng, C.L. & Van Ness, J.W. (1999). *Statistical regression with measurement error.* Arnold, London.

[19] Chipkevitch, E., Nishimura, R., Tu, D. & Galea–Rojas, M. (1996). Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods. *Journal of Urology,* **156**, 2050-2053.

[20] Christensen, R. & Blackwood, L. (1993). Tests for precision and accuracy of multiple measuring devices. *Technometrics,* **35**, 411-420.

[21] Cook, R.D. (1977). Detection of influential observations in linear regresion. *Technometrics,* **19**, 15-18.

[22] Cook, R.D. & Weisberg, S. (1982). *Residuals and influence in regression.* Chapman and Hall, London.

[23] Cook, R.D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society,* B, **48**, 133-169.

[24] Davies, P.L. & Gather, U. (1993). The identification of multiple outliers (with discussion). *Journal of the American Statistical Association,* **88**, 782-801.

[25] De Castro, M., Lachos, V.H. & Galea–Rojas, M. (2008). Heteroscedastic skew-normal measurement error models. *Communications in statistics - Theory and methods.* Submetido.

[26] 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,* B, **39**, 1-22.

[27] Dias, J.G. & Wendel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic gaussian mixture likehoods. *Statistics and Computing,* **14**, 323-332.

[28] Escobar, E. & Meeker, W. (1992). Assessing influence in regression analysis with censured data. *Biometrics,* **48**, 507-528.

[29] Fieller, N. (1993). Comment on "The identification of multiple outliers" by L. Davies and U. Gather. *Journal of American Statistical Association,* **88**, 794-795.

[30] Fuller, W.A. (1987). *Measurement error models.* Wiley, New York.

[31] Fung, W.K. (1993). Unmasking outliers and leverage points: A confirmation. *Journal of American Statistical Association,* **88**, 515-519.

[32] Galea–Rojas, M. (1995). Calibração estrutural e funcional. Tese de doutorado, Departamento de Estatística. IME-USP.

[33] Ghosh, P., Lachos, V.H. and Vilca, L.F. (2008). Skew-normal/independent distributions, with applications. RT-IMECC 02-07, IMECC-UNICAMP.

[34] Grubbs, F.E. (1948). On estimating precision of mesuring instruments nd product variability. *Journal of the American Statistical Association,* **43**, 243-264.

[35] Grubbs, F.E. (1973). Errors of measurement precision, accuracy and the statistical comparison of measuring instruments. *Technometrics,* **15**, 53-66.

[36] Henze, N.A. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal Statistics,* **13**, 271-275.

[37] Huber, P.J. (1991). Between robusteness and diagnosics. Em *Directions in Robust Statistics and Diagnostics,* W. Stahel and S. Weisberg (eds.), IMA, **33**, 121-130. Springer Verlag, New York.

[38] Jaech, J.L. (1985). Statistical analysis of measurement errors. *Exxon Monographs.* John Wiley, New York.

[39] Johnson, N. L., Kotz S. & Balakrishnan, N. (1994). *Continous univariate distributions.* John Wiley, New York.

[40] Kelly, G. (1984). The influence function in the errors in variables problem. *Annals of Statistics,* **12**, 87-100.

[41] Lachos, V.H. (2002). Inferência e Diagnóstico em Modelos de Grubbs. Dissertação de mestrado, Departamento de Estatística, IMECC-UNICAMP.

[42] Lachos, V.H. (2004). Modelos Lineares Mistos Assimétricos. Tese de doutorado, Departamento de Estatística, IME-USP.

[43] 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 2005-12, IME-USP.

[44] Lachos, V.H, Vilca, L.F. & Galea–Rojas, M. (2007a). Influence diagnostics for the Grubbs' model. *Statistical Papers*, **48**, 419-436.

[45] Lachos, V.H, Bolfarine, H. & Vilca, L.F. (2007b). Robust Multivariate Measurement Error Model with Skew-Normal. RT-IMECC 27-07, IMECC-UNICAMP.

[46] Lachos, V.H., Bolfarine, H. & Vilca, L.F. (2008). Robust multivariate measurement error models with scale mixture of skew-normal distribtion. *Statistics*, sob revisão.

[47] Lachos, V.H., Vilca, L.F., Bolfarine, H. & Ghosh, P. (2009). Multivariate measurement error models based on scale mixtures of the skew-normal distributions. *Statistics,* **doi: 10.1080/02331880903236926.**

[48] Lange, K.L. & Sinsheimer. (1993). Normal/independent distributions and their applications in robust regression . *Journal of Computational and Graphical Statistics,* **2**, 175-198.

[49] Leurgans, S. (1980). Evaluation laboratiry Measurement Techniques. *In Biostatistics casebol.* John Wiley, New York.

[50] Lin, T.I., Lee, J.C. & Yen, S.Y. (2007). Finite mixture modelling using the skew-normal distribution. *Statistica Sinica*, **17**, 909-927.

[51] Liu, C. & Rubin, D.B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **81**, 663-648.

[52] Loynes, R.M. (2001). A new measure in local influence. *Journal of Statistical Planning and Inference,* **92**, 47-53.

[53] Lu, B. & Song, X.Y. (2006). Local influence of multivariate probit latent variable models. *Journal of Multivariate Analysis*, **97**, 1783-1798.

[54] Magnus, J.R. & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics.* Wiley, New York.

[55] Meng, X.L., & Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, **80**, 267-278.

[56] Montenegro, L.C., Bolfarine, H. & Lachos, V.H. (2009a). Influence diagnostics for a skew extension of the Grubbs measurement error model. *Communication in Statistics - Simulation and Computation,* **38**(4), 667-681.

[57] Montenegro, L.C., Bolfarine, H. & Lachos, V.H. (2009b). Local influence analysis of skew-normal linear mixed models.*Communication Statistics - Theory and Methods,* **38**, 484-496.

[58] Nel, D.G. (1980). On matrix differetiation in statistics.*Soth African Statistical Journal*, **14**, 137-193.

[59] Osorio, F.S. (2006). Diagnóstico de influência em modelos elípticos com efeitos mistos. Tese de doutorado, Departamento de Estatística, IME-USP.

[60] Osorio, F.S., Paula, G.A. & Galea, M. (2009). On estimation and influence diagnostics for the Grubbs' model under heavy-tailed distributions. *Computational Statistics and Data Analysis,* **53** 1249-1263.

[61] Pinheiro, J.C., Liu, C.H. & Wu, Y.N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics*, **10**, 249-276.

[62] Poon W.Y. & Poon, Y.S. (1999). Conformal normal curvature and assessment of local influence. *Journal of Royal Statistical Society,* B, **61**, 51-61.

[63] Rousseeuw, P.J. & Van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association,* **85**, 633-639.

[64] Shyr, I. & Gleser, L. (1986). Inference about comparative precision in linear structural relationships. *Journal of Statistical Planning and Inference,* **14**, 339-358.

[65] Solano - Dávila, O.L. (2000). Influência local em modelos de calibração comparativa. Dissertação de Mestrado, Departamento de Estatística, IMECC-UNICAMP.

[66] Tanaka, Y., Watadoni, S. & Moon, S.H. (1991). Influence in covariance structure analysis with an application to confirmatory fator analysis.*Communication in Statistic - Theory and Methods,* **20**(12), 3805-3821. 339-358.

[67] Theobald, C.M. & Mallison, J.R. (1978). Comparative calibration, linear structural relationship and congeneric measurements. *Biometrics,* **34**, 35-45.

[68] Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* Springer-Verlag, New York.

[69] Wang, J., Boyer, J. & Genton, M.G. (2004). A skew-simmetric representation of multivariate distributions. *Statistica Sinica,* **14**, 1259-1270.

[70] Wang, J. & Genton, M.G. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference,* **136**, 209-220.

[71] Zeller, C.B. (2009). Distribuições misturas de escala skew-normal: Estimação e diagnóstico em modelos lineares. Tese de doutorado, Departamento de Estatística, IMECC-UNICAMP.

[72] Zhao, Y. & Lee, A. (1998). Influence diagnostics for simultaneous equations models. *Australian and New Zealand Journal Statistics,* **40**, 345-357.

[73] Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society,* B, **63**, 111-126.