UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

THAÍS CASTELO BRANCO MONHO

Modelo de Regressão Normal Inversa Gaussiana com Erros nas Variáveis

Campinas

2019

Thaís Castelo Branco Monho

Modelo de Regressão Normal Inversa Gaussiana com Erros nas Variáveis

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

Orientador: Filidor Edilfonso Vilca Labra

Este exemplar corresponde à versão final da Dissertação defendida pela aluna Thaís Castelo Branco Monho e orientada pelo Prof. Dr. Filidor Edilfonso Vilca Labra.

Campinas

2019

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M748m Monho, Thaís Castelo Branco, 1991-
Modelo de regressão normal inversa gaussiana com erros nas variáveis / Thaís Castelo Branco Monho. – Campinas, SP : [s.n.], 2019.

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

1. Modelos com erros nas variáveis. 2. Modelos de regressão (Estatística). 3. Algoritmos de esperança-maximização. 4. Distribuições bivariadas (Estatística). I. Vilca Labra, Filidor Edilfonso, 1964-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Measurement error regression under a normal inverse gaussian model
Palavras-chave em inglês:
Errors-in-variables models
Regression models (Statistics)
Expectation-maximization algorithms
Bivariate distributions (Statistics)
Área de concentração: Estatística
Titulação: Mestra em Estatística
Banca examinadora:
Filidor Edilfonso Vilca Labra [Orientador]
Larissa Avila Matos
Alexandre Galvão Patriota
Data de defesa: 23-07-2019
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-4788-9129
- Currículo Lattes do autor: http://lattes.cnpq.br/4504173240581437

Dissertação de Mestrado defendida em 23 de julho de 2019 e aprovada pela banca examinadora composta pelos Profs. Drs.

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

Prof(a). Dr(a). LARISSA AVILA MATOS

Prof(a). Dr(a). ALEXANDRE GALVÃO PATRIOTA

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Este trabalho é dedicado a Deus que
iluminou todo meu caminho durante esta caminhada.

Agradecimentos

Primeiro de tudo, gostaria de agradecer a Deus e a Nossa Senhora por me guiarem, iluminarem e por me darem forças para seguir em frente com os meus objetivos.

À minha família, agradeço todo o incentivo, paciência, amor e compreensão pelas minhas faltas e momentos de afastamento.

Ao meu namorado Guilherme Antunes, agradeço todo o seu amor, carinho, admiração e apoio ao longo do mestrado.

Ao professor Filidor, agradeço pela confiança, compreensão, apoio e pela excepcional orientação durante a elaboração deste trabalho.

Aos professores do Departamento de Estatística, agradeço pelos ensinamentos concedidos.

Aos colegas de Mestrado, agradeço pela ajuda, apoio, amizade e companheirismo.

Ao CREMESP, agradeço pelo incentivo, compreensão e auxílio durante o mestrado.

Por fim, a todos aqueles que contribuíram, direta ou indiretamente, para a realização desta dissertação, o meu sincero agradecimento.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Resumo

As distribuições multivariadas de caudas pesadas têm encontrado muitas aplicações em estatística, onde a distribuição normal tem sido usada. A distribuição Normal Inversa Gaussianal (NIG) é uma mistura variância-média, de uma normal multivariada com uma distribuição univariada Inversa Gaussiana (IG). O principal objetivo deste trabalho é o estudo dos modelos estruturais com erros nas variáveis, sob a distribuição NIG simétrica multivariada. Assume-se que as variáveis observadas seguem uma distribuição NIG bivariada simétrica. Devido à complexidade da função de verossimilhança, a estimação dos parâmetros do modelo por maximização direta é extremamente difícil. Para superar este problema, propomos um algoritmo EM rápido e preciso, para obter as estimativa de máxima verossimilhança dos parâmetros. Discutimos aspectos da robustez do uso da distribuição NIG, que inclui a aplicação do método de detecção de observações atípicas no modelo com erros nas variáveis sob a distribuição NIG simétrica. Estudos de simulação e aplicações de um conjunto de dados reais, são fornecidos para ilustrar a metodologia proposta.

Palavras-chave:Distribuições de caudas pesadas; Distribuições hiperbólicas generalizadas; Distribuição Normal Inversa Gaussiana; Modelo com erros nas variáveis.

Abstract

The heavy-tailed multivariate distributions have found several applications in statistics, where the normal distribution has been used. The Normal Inverse Gaussian (NIG) distribution is a recent variance-mean mixture of a multivariate Gaussian with a univariate Inverse Gaussian (IG) distribution. The main goal of this work is the study of the structural errors-in-variables models, under the symmetric Multivariate Normal Inverse Gaussian (NIG) distribution. It is assumed that the observed variables follow a symmetric bivariate NIG distribution. Due to the complexity of the likelihood function, the estimation of the model parameters by direct maximization is exceedingly difficult. To overcome this problem, we propose a fast and accurate EM algorithm for the maximum likelihood estimation. We discuss aspects of robustness of the use of the NIG distribution, which includes the application of method to detect outlying observations in the error-in-variables model under the symmetric NIG distribution. Simulation studies and applications to real data set are given to illustrate the proposed methodology.

Keywords: Heavy-tailed distributions; Generalized hyperbolic distributions; Normal inverse gaussian Distribution; Errors-in-variables models.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

Fdp	Função densidade de probabilidade
Fda	Função de distribuição acumulada
GH	Distribuição hiperbólica generalizada
GIG	Distribuição inversa gaussiana generalizada
IG	Distribuição inversa gaussiana
NIG	Distribuição normal inversa gaussiana
NIGS	Distribuição normal inversa gaussiana simétrica

Sumário

1 Introdução

Em 1963, Benoit Mandelbrot notou que o comportamento dos log-retornos de ativos financeiros apresentavam caudas mais pesadas do que a distribuição normal. Assim, um novo tipo de distribuição, chamada hiperbólica generalizada, provou ser útil no tratamento de dados financeiros. O que tornou as distribuições hiperbólicas tão populares foi o fato dos seus parâmetros serem suficientemente flexíveis para se adequarem aos mais variados conjuntos de dados e contextos. Uma subfamília da classe da distribuições hiperbólicas, é a classe de distribuições normal inversa gaussiana (NIG), que é uma distribuição normal com mistura variância-média, em que a densidade da mistura é a distribuição inversa gaussiana (IG). Uma das razões pelas quais a distribuição NIG é tão popular, é que seus quatro parâmetros a tornam flexível o suficiente para ajustar bem diferentes conjuntos de dados e consequentemente a torna potencialmente útil em diferentes contextos.

Esta distribuição tem sido aplicada em áreas como economia, física, biologia, agronomia entre outros. A sua aplicação na área financeira ou econômica surge em primeiro lugar através de Eberlein e Keller (1995). Em seu estudo, utilizaram esta família de distribuições para ajustar dados de ações da bolsa alemã. Um trabalho mais abrangente é elaborado por Prause (1999), no qual surge a aplicação das distribuições hiperbólicas generalizadas para ajustar dados financeiros de ativos da bolsa alemã e índices americanos.

A distribuição hiperbólica, introduzida por Barndorff-Nielsen (1978), possui a vantagem de ter caudas mais pesadas e de permitir grande concentração das observações em torno da média ou do valor central. Estas características são verificadas empiricamente em aplicações de retornos dos ativos financeiros. O bom ajuste aos dados da distribuição hiperbólica, motiva o uso desta distribuição em outras áreas onde não tem sido usada com muita frequência.

A distribuição NIG simétrica é uma distribuição que possui caudas mais pesadas do que a distribuição normal, e que apresenta um valor de curtose maior que 3, medida esta que caracteriza o achatamento da curva da função de distribuição. Uma distribuição é dita ser leptocúrtica, quando o seu valor de curtose é maior do que 3, já quando o valor da curtose é igual a 3 temos uma distribuição mesocúrtica, que é o caso da distribuição normal, e quando a curtose é menor do que 3, temos uma distribuição platicúrtica. Sendo assim, valores de curtose maiores do que 3, nos dão indícios de que uma distribuição NIG seria mais apropriada do que uma distribuição normal, para o ajuste dos dados.

As distribuições com caudas pesadas possuem maior probabilidade na área das caudas quando comparadas à distribuição normal. Para dados do campo financeiro, como acontece com os retornos de ativos, temos que suas distribuições, em grande parte dos casos, são assimétricas com caudas mais pesadas do que uma distribuição normal. Logo, utilizar uma distribuição normal não seria uma boa escolha, deste modo, a princípio foi sugerido o uso de distribuições de Pareto, mas essas distribuições apresentavam caudas demasiadamente altas. Assim, as distribuições hiperbólicas generalizadas surgiram como uma alternativa no tratamento dos dados financeiros, pois seus parâmetros são suficientemente flexíveis para se adequarem aos mais variados conjuntos de dados, especialmente a distribuição NIG, por ser mais simples de ser aplicada.

É bem conhecido, que as distribuições elípticas simétricas, que é o caso da distribuição NIG multivariada, têm sido usadas no lugar da distribuição normal por serem mais robustas, já as distribuições elípticas assimétricas têm sido usadas no lugar da distribuição skew-normal proposta por Azzalini (1985). Ambas classes de distribuições possuem aplicações em um grande números de áreas, tanto em áreas teóricas como aplicadas (na teoria de distribuições e em modelos de regressão).

1.1 Distribuição normal com mistura variância-média

A família de distribuições normais com mistura variância-média tem a importante habilidade de gerar inúmeras formas de distribuições flexíveis. Esta família possui alguns casos bastante relevantes de distribuições assimétricas, como é o caso da distribuição normal inversa gaussiana (NIG), que será o foco de estudo neste trabalho.

Em inúmeras aplicações as caudas da distribuição normal são menos pesadas do que o desejado, e suas formas também não são adequadas para dados altamente assimétricos ou muito concentrados, próximos de valores centrais. Desta maneira, nestes casos a utilização de distribuições normais com misturas variância-média é a mais indicada.

Se um vetor aleatório p-dimensional $\mathbf{Z}$ segue uma distribuição normal, então dizemos que o vetor aleatório p-dimensional $\mathbf{X}$ tem distribuição normal com mistura variância-média se pode ser representado por

$$\mathbf{X} = \boldsymbol{\mu} + U\boldsymbol{\beta} + \sqrt{U}\mathbf{Z}, \tag{1.1}$$

em que $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, e U é uma variável aleatória univariada positiva independente de $\mathbf{Z}$. Nesta representação, $\boldsymbol{\beta} \in \mathbb{R}^p$ é o parâmetro de assimetria, e $\boldsymbol{\mu}$ é o parâmetro de locação. Se a variância de U for finita, então a média e variância de $\mathbf{X}$ são dadas por

$$E[\mathbf{X}] = \mu + \boldsymbol{\beta}E[U] \quad \text{e} \quad Var[\mathbf{X}] = E[U]\boldsymbol{\Sigma} + Var[U]\boldsymbol{\beta}\boldsymbol{\beta}^\top.$$

Algumas observações sobre a distribuição em (1.1):

1. Quando $\boldsymbol{\beta} = \mathbf{0}$, temos a família de distribuições de escala mistura normal (SMN), em que temos como casos particulares as distribuições normal, t-Student, normal contaminada, slash, entre outras distribuições simétricas;

2. Quando a variável aleatória U segue uma distribuição inversa gaussiana, $U \sim IG(\chi, \psi)$, temos como caso particular a distribuição normal inversa gaussiana;

3. Dado $U = u$, a distribuição condicional de $\mathbf{X}$ segue uma distribuição normal, i.e. $\mathbf{X}|U = u \sim N_p(\boldsymbol{\mu} + u\boldsymbol{\beta}, u\boldsymbol{\Sigma})$. Sob o modelo condicional, o procedimento para estimação será mais simples, e o algoritmo EM parece ser uma boa opção como método de estimação dos parâmetros do modelo;

4. Para $\boldsymbol{\beta} \neq \mathbf{0}$, a distribuição é assimétrica, e assim é natural pensar em compará-la com as distribuições assimétricas propostas por Azzalini (1985), por exemplo. As distribuições assimétricas de Azzalini, têm encontrado aplicações em diversas áreas, como por exemplo modelos de regressão. Então, é natural pensar no uso da distribuição em (1.1) em modelos de regressão.

1.2 Modelo de regressão com erros nas variáveis

Em geral, os modelos de regressão linear são ajustados sobre a hipótese de que as covariáveis são medidas sem erro, considerando-as como efeito fixo. Porém, em muitas situações práticas, essas covariáveis apresentam efeitos aleatórios, uma vez que fatores do acaso podem interferir no processo de medição. Nesse caso, fatores aleatórios, como correntes de ar ou vibrações, introduziriam erros nas mensurações. Uma alternativa para evitar essa alteração seria repetir inúmeras vezes a medição, permitindo avaliar a incerteza estatística nas covariáveis.

Situações desse tipo ocorrem na engenharia, em que a redução dos "erros", no contexto mais amplo da palavra, é tratada especificamente pela metrologia. Em síntese, essa ciência analisa os aspectos relativos às medições e calibrações de instrumentos, utilizando-se metodologias estatísticas. Tais medições, em conjunto com o uso de uma técnica estatística inadequada certamente, resultam em índices de qualidade incoerentes. Para modelar estas situações, incluindo no modelo os erros de medições ou erros no processo, o modelo de regressão com erros nas variáveis, proposto por Fuller (1987) contorna estes problemas. O modelo proposto por Fuller é definido sob a classe das distribuições normais. Devido à presença de observações atípicas, generalizações têm sido consideradas sobre a classe das distribuições simétricas elípticas, que contêm como casos particulares as distribuições normal, t-Student, exponencial potência, normal contaminada, slash, entre outras.

A inferência nos modelos lineares sob a classe das distribuições elípticas tem sido objeto de estudo por muitos autores, como por exemplo, em Fang e Zhang (1990) e Arellano-

Valle (1994). Em modelos lineares com erros nas variáveis na classe das distribuições elípticas, existem numerosos trabalhos, como por exemplo Lachos et al.(2009) e Patriota e Bolfarine (2009). A inferência nestes modelos é importante pelas suas aplicações em diferentes áreas, como por exemplo, nas ciências médicas ou ciências sociais. Apesar dos constantes avanços tecnológicos terem tornado cada vez mais precisos os procedimentos de mensuração, não é realista supor que algumas variáveis no modelo sejam medidas sem erros e o mais comum é não ter acesso aos seus verdadeiros valores. O modelo linear com erros nas variáveis mencionado acima é definido a seguir.

Considere a relação linear,

$$Y_i = \alpha + \beta\ x_i + e_i,\ i = 1, \ldots, n,$$

com Y_i sendo a variável resposta da i-ésima observação e x_i a variável preditora da i-ésima observação. Porém, como muitas vezes x_i pode não condizer com seu valor real é considerado medido com erro de acordo com a equação

$$X_i = x_i + u_i,$$

em que X_i é um estimador não viesado do verdadeiro valor de x_i. Temos que α representa o intercepto e β a inclinação, sendo ambos parâmetros desconhecidos. Os erros e_i e u_i são variáveis aleatórias com média 0 e variância ϕ_e e ϕ_u, respectivamente, os quais são independentemente distribuídos, para $i = 1, ..., n$.

O modelo de regressão linear com erros nas variáveis pode ser visto como um modelo usado para a comparação de dois instrumentos de medições, este estudo é conhecido como calibração comparativa (Barnett, 1969), onde se comparam vários instrumentos ou métodos de medições. A necessidade de se comparar instrumentos ou métodos de medições tem aparecido com frequência em diversas áreas, por exemplo: Grubbs (1948,1973) compara três cronômetros; Barnett (1969) apresenta um exemplo em que quatro combinadores, instrumento-operador concebido para medir a capacidade vital em um grupo de pacientes são avaliados e Jaech (1985) apresentou vários exemplos na área industrial. Christensen e Blackwood (1993) comparam cinco termopares; Bolfarine e Galea-Rojas (1995) apresentam um estudo de inferência no modelo t-Student; Bedrick (2001) compara três métodos para medir sedimentos de solo. Outros trabalhos sob a classe das distribuições simétricas podem ser encontrados em Arellano-Valle (1994), e para trabalhos mais recentes veja resultados em Patriota (2010) e Patriota e Bolfarine (2009).

Em muitas aplicações, por exemplo, na área industrial e agricultura, é assumido que todos os instrumentos medem a característica de interesse na mesma escala, aqui o modelo de Grubbs parece ser mais adequado, veja também, Christensen e Blackwood (1993).

1.3 Objetivo do trabalho

O objetivo deste trabalho é apresentar o modelos de regressão linear com erros nas variáveis sob a classe das distribuições NIG simétrica. Esta discussão tem como base, principalmente os trabalhos desenvolvidos por Arellano-Valle e Bolfarine (1996), baseados em distribuições simétricas, tais como as distribuições t-Student, normal contaminada e slash, em geral baseada nas distribuições elípticas (Galea et al., 2000). A distribuição NIG possui uma representação estocástica que facilita a implementação de um algoritmo EM para obter as estimativas de máxima verossimilhança dos parâmetros do modelo e de aplicações de análise de diagnóstico. Para a análise de dados, consideramos alguns resultados de análise de diagnóstico no modelo de regressão linear com erros nas variáveis utilizado os métodos de Zhu et al. (2001). De uma forma geral, podemos enumerar os objetivos desse trabalho como:

1. Apresentar a distribuição NIG multivariada. Apresentar propriedades da distribuição NIG multivariada em direção do uso no modelo de regressão linear com erros nas variáveis,

2. Discutir alguns aspectos de inferência, em que a estimação de máxima verossimilhança será feita via um algoritmo EM,

3. Desenvolver um estudo de análise de diagnóstico no modelo de regressão linear com erros nas variáveis, eliminação de casos e ponderação de casos.

1.4 Organização do Trabalho

Este trabalho consta de 6 capítulos. No Capítulo 2, apresentamos uma revisão resumida da distribuição hiperbólica, apresentando as propriedades que serão usadas nos capítulos subsequentes.

No Capítulo 3, apresentamos o modelo de regressão linear com erros nas variáveis baseado na distribuição NIG simétrica. Estudos de estimação de máxima verossimilhança via algoritmo EM são discutidos. A estimação dos parâmetros do modelo via algoritmo EM sob algumas restrições serão discutidas, com o intuito de realizar alguns testes de hipóteses. Estudos de simulações são desenvolvidos para avaliar o desempenho dos algoritmos propostos e para estudar o comportamento da estatística de teste.

No Capítulo 4, apresentamos uma discussão sobre a identificação de observações atípicas, exclusão de casos e métodos de influência local propostos por Cook (1986). No Capítulo 5, apresentamos aplicações a dois conjuntos de dados reais, e finalmente no Capítulo 6, apresentamos um resumo e conclusões dos resultado obtidos e de pesquisas futuras.

2 Distribuição Normal Inversa Gaussiana

2.1 Introdução

A distribuição normal inversa gaussiana (NIG) é um caso especial da distribuição hiperbólica generalizada (GH), introduzida por Barndorff-Nielsen (1978). Uma das razões que faz a distribuição NIG ser tão popular, é que seus quatro parâmetros a tornam flexível o suficiente para ajustar bem diferentes conjuntos de dados e consequentemente a torna potencialmente útil em diferentes contextos.

A distribuição NIG simétrica, que é um caso particular da distribuição normal com mistura variância-média com a densidade da mistura sendo a distribuição inversa gaussiana (IG), é uma distribuição que possui caudas mais pesadas do que a distribuição normal, apresentando então valores de curtose maiores do que de uma distribuição normal. Sendo assim, esta distribuição é considerada mais indicada para analisar dados mais leptocúrticos.

2.2 Distribuição normal com mistura variância-média

A família de distribuições normais com mistura variância-média tem a importante habilidade de gerar inúmeras formas de distribuições flexíveis. Esta família possui alguns casos bastante relevantes de distribuições assimétricas, como é o caso da distribuição hiperbólica generalizada (GH), que tem como casos particulares outras inúmeras distribuições importantes como, por exemplo, a normal inversa gaussiana (NIG).

Seja Z é uma variável aleatória com distribuição normal univariada e X uma variável aleatória seguindo uma distribuição normal com mistura variância-média, então X pode ser definida como

$$X = \mu + U\beta + \sqrt{U}Z, \tag{2.1}$$

em que μ $\beta \in \mathbb{R}$ e $Z \sim N(0, \sigma^2)$. β é o parâmetro que representa assimetria, μ é o parâmetro de locação e U é a variável aleatória positiva que representa o peso e é independente de Z, com função distribuição acumulada (fda) $H(.)$. Sua fda pode ser obtida por meio de

$$F(x) = \int_0^\infty \Phi((x - \mu - u\beta)u^{-1/2}, \sigma^2)dH(u), \qquad x \in \mathbb{R}, \tag{2.2}$$

em que $\Phi(., \sigma^2)$ é a fda de uma distribuição $N(0, \sigma^2)$. Quando $H(.)$ é absolutamente contínua, com função distribuição (fdp) $h(u)$, podemos obter a fdp de X como sendo

$$f(x) = \int_0^\infty u^{-1/2}\phi((x - \mu - u\beta)u^{-1/2}, \sigma^2)h(u)du, \qquad x \in \mathbb{R}, \tag{2.3}$$

em que $\phi(., \sigma^2)$ é a fdp de uma distribuição $N(0, \sigma^2)$. Dado U com variância finita, temos que

$$E[X] = \mu + \beta E[U] \quad \text{e} \quad Var[X] = E[U]\sigma^2 + Var[U]\beta^2.$$

Quando $\beta = 0$ e U^{-1} segue uma distribuição gamma com parâmetros $(\nu/2, \nu/2)$, ou seja, U segue uma distribuição inversa gamma com parâmetros $(\nu/2, \nu/2)$, temos que X segue a conhecida distribuição t-Student com ν graus de liberdade (Kotz e Nadarajah, 2004). Já quando $\beta \neq 0$ e U seguindo uma distribuição inversa gaussiana generalizada (GIG), resultando assim em uma distribuição GH (Barndorff-Nielsen, 1997), a partir da qual podemos gerar outras distribuições.

2.3 Distribuição gaussiana inversa generalizada

A distribuição inversa gaussiana generalizada (GIG) foi proposta inicialmente por Étienne Halphen (1941) para analisar a frequência dos fluxos de rios, porém devido a sua forma complexa envolvendo funções de Bessel e funções exponenciais permaneceu em esquecimento por vários anos. Essa distribuição foi redescoberta e popularizada por Barndorff-Nielsen (1978) e posteriormente, por sua influência, foi muito estudada por Jørgensen (1982).

Uma variável aleatória positiva tem distribuição GIG se a sua fdp é dada por

$$h(u) = \left(\frac{\psi}{\chi}\right)^{\lambda/2} \frac{u^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{u} + \psi u\right)\right\} \mathbb{1}_{\{u>0\}}(u), \tag{2.4}$$

em que $-\infty < \lambda < \infty$, $\psi > 0$, $\gamma > 0$ e K_λ é a função de Bessel modificada de terceiro tipo avaliada em u e indexada por λ, o qual é definida por

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} \exp\left\{-\frac{x}{2}\left(y + \frac{1}{y}\right)\right\} dy,$$

o gráfico da função $K_\lambda(.)$ pode ser observado na Figura 1 para alguns valores de λ. Esta distribuição é denotada por $U \sim GIG(\psi, \chi, \lambda)$ e os parâmetros satisfazem as seguintes relações

$$\begin{aligned} \chi \geqslant 0, \psi > 0 \quad &\text{se} \quad \lambda > 0, \\ \chi > 0, \psi \geqslant 0 \quad &\text{se} \quad \lambda < 0, \\ \chi > 0, \psi > 0 \quad &\text{se} \quad \lambda = 0. \end{aligned}$$

Algumas propriedades da função de Bessel modificada de terceiro tipo serão necessárias no decorrer do trabalho. Suas propriedades mais importantes são:

1. $K_\lambda(x) = K_{-\lambda}(x)$,

2. $K_{\lambda+1}(x) = \dfrac{2\lambda}{x} K_\lambda(x) + K_{\lambda-1}(x)$,

3. $K_{r+1/2}(x) = \sqrt{\dfrac{\pi}{2x}} \exp(-x) \displaystyle\sum_{k=0}^{r} \dfrac{(r+k)!(2x)^{-k}}{(r-k)!k!}$ $r = 0, 1, 2, ...$, um caso particular é, para $r = 0$ temos $K_{1/2}(x) = \sqrt{\dfrac{\pi}{2x}} \exp(-x)$,

4. $K'_{\lambda-1}(x) = \dfrac{d}{dx} K_{\lambda-1}(x) = -\dfrac{1}{2}(K_{\lambda-1}(x) + K_{\lambda+1}(x))$.

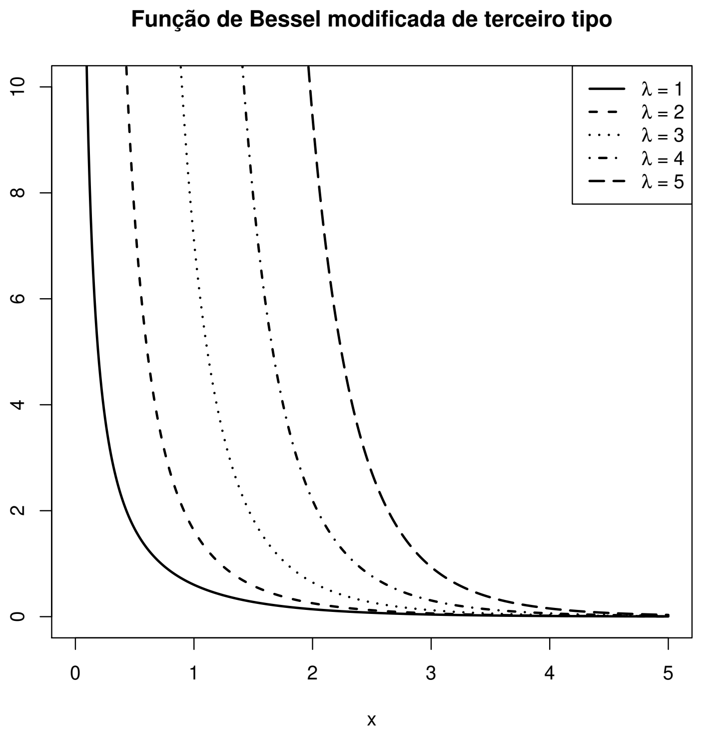


Figura 1 – Função de Bessel de terceiro tipo $K_\lambda(x)$.

2.3.1 Subclasses

A distribuição Inversa Gaussiana Generalizada possui algumas subclasses muito conhecidas, como por exemplo,

- **Distribuição gama**
  Se $\chi = 0$, $\psi > 0$ e $\lambda > 0$, temos uma distribuição *Gama*$(\lambda, \psi/2)$, com função densidade dada por

$$h(u) = \left(\frac{\psi}{2}\right)^{\lambda} \frac{u^{\lambda-1}}{\Gamma(\lambda)} \exp\left\{-\frac{1}{2}\psi u\right\} \mathbb{1}_{\{u>0\}}(u),$$

- **Distribuição inversa gama**
  Se $\chi > 0$, $\gamma = 0$ e $\lambda < 0$, temos uma distribuição $IGama(\lambda, \chi/2)$, com função densidade dada por

$$h(u) = \left(\frac{2}{\chi}\right)^{\lambda} \frac{u^{\lambda-1}}{\Gamma(-\lambda)} \exp\left\{-\frac{\chi}{2u}\right\} \mathbb{1}_{\{u>0\}}(u),$$

- **Distribuição inversa gaussiana**
  Se $\lambda = -1/2$, temos uma distribuição $IG(\chi, \psi)$, que será aprofundada adiante, com função dada por

$$h(u) = \sqrt{\frac{\chi}{2\pi}} \exp(\sqrt{\psi\chi}) u^{-3/2} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{u} + \psi u\right)\right\} \mathbb{1}_{\{u>0\}}(u).$$

.

2.3.2 Propriedade dos momentos

Os momentos da distribuição GIG serão extremamente necessários no decorrer do trabalho, principalmente no algoritmo EM. Temos que o m-ésimo momento é dado por

$$E[U^m] = \left(\frac{\chi}{\psi}\right)^{m/2} \frac{K_{\lambda+m}(\sqrt{\chi\psi}),}{K_{\lambda}(\sqrt{\chi\psi})}, \tag{2.5}$$

e desta propriedade, segue que

$$E[U] = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}, \tag{2.6}$$

$$E[U^{-1}] = \left(\frac{\psi}{\chi}\right)^{1/2} \frac{K_{\lambda-1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})} = \left(\frac{\psi}{\chi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})} - \frac{2\lambda}{\chi}, \tag{2.7}$$

$$Var[U] = \left(\frac{\chi}{\psi}\right) \left[\frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})} + \left(\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}\right)^2\right]. \tag{2.8}$$

As esperanças (2.6) e (2.7) serão utilizadas no próximo capítulo para a estimação dos parâmetros do modelo via algoritmo EM.

2.3.3 Caracterização dos parâmetros

Esta seção é destinada para analisar graficamente o comportamento da fdp da distribuição GIG para diferentes valores dos parâmetros. Na Figura 2, temos os gráficos da densidade em (2.4) para alguns valores de λ fixando $\chi = \psi = 1$, quanto menor o parâmetro λ mais assimétrica será a distribuição.

Já nas Figuras 3 e 5 fixamos $\psi = 1$ para $\lambda = 0$ e $\lambda = 0.5$, respectivamente, variando os valores de χ, e assim podemos notar que com o aumento do valor de χ o gráfico se torna menos assimétrico e nas Figuras 4 e 6 fixamos $\chi = 1$ para $\lambda = 0$ e $\lambda = 0.5$,

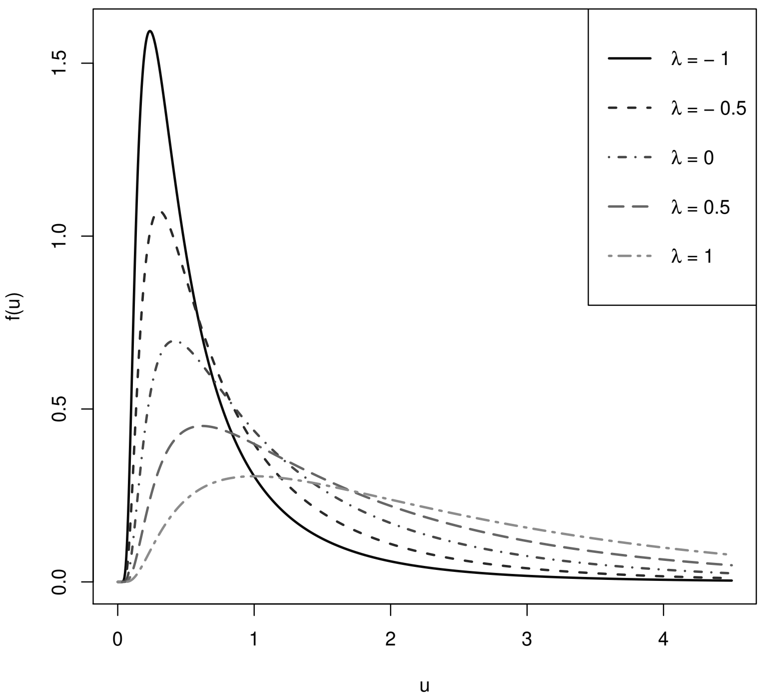


Figura 2 – Fdp da GIG para diferentes valores de λ, fixando $\chi = \psi = 1$.

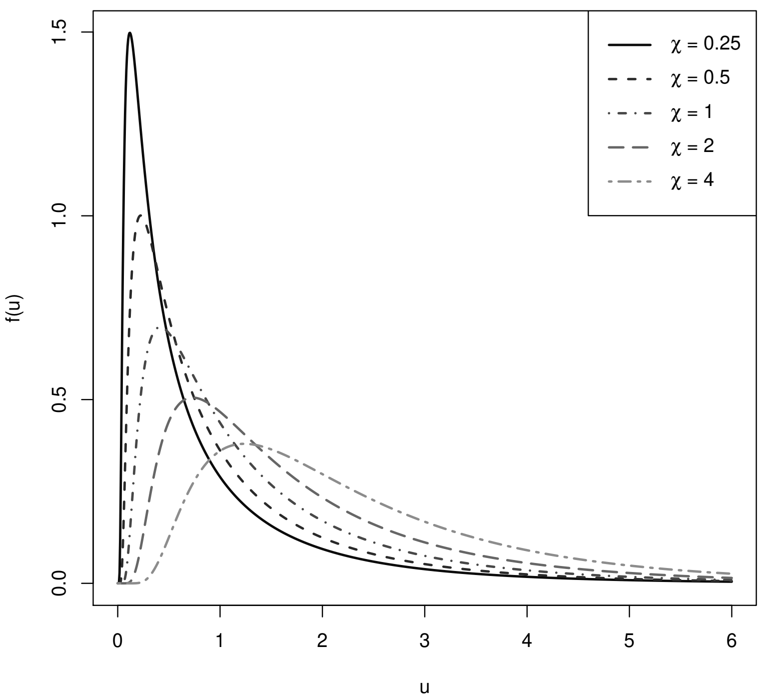


Figura 3 – Fdp da GIG para diferentes valores de χ, fixando $\psi = 1$ e $\lambda = 0$.

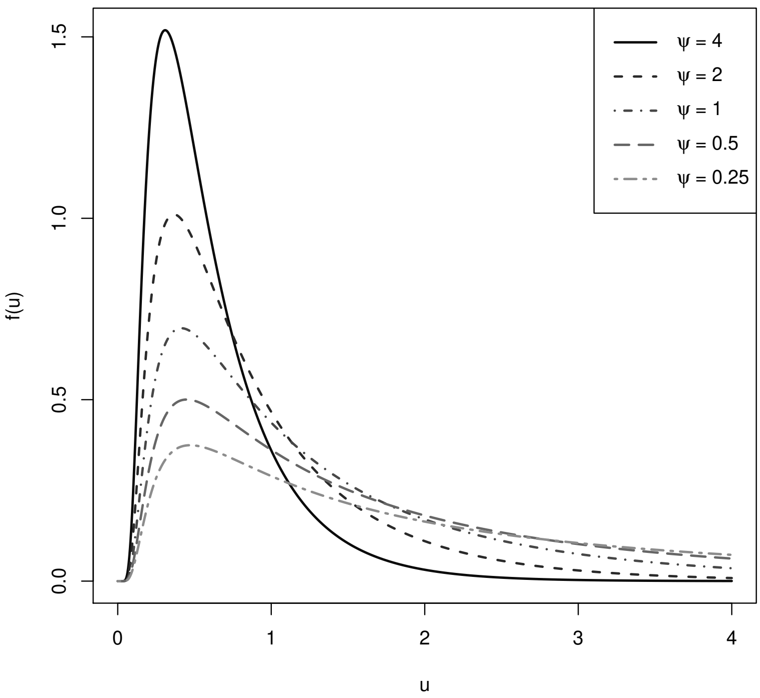


Figura 4 – Fdp da GIG para diferentes valores de ψ, fixando $\chi = 1$ e $\lambda = 0$.

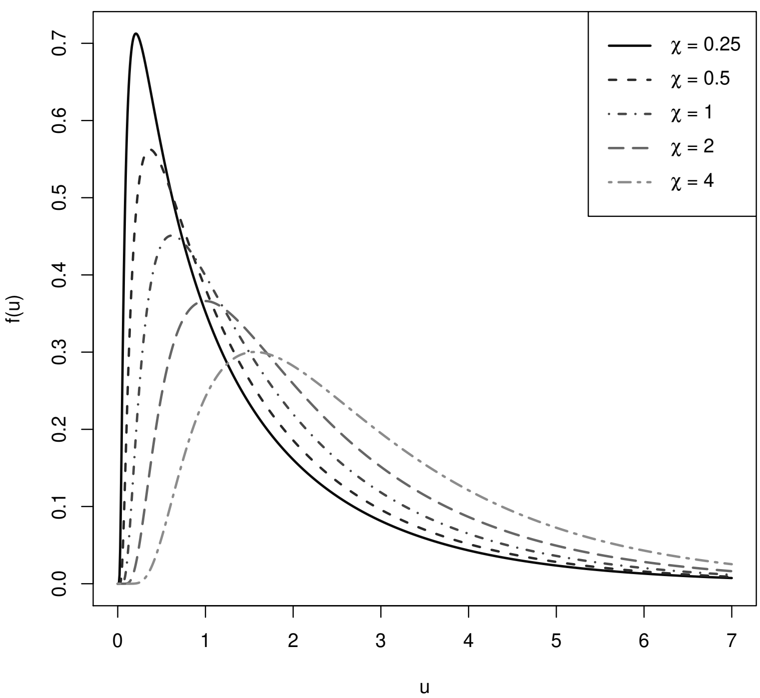


Figura 5 – Fdp da GIG para diferentes valores de χ, fixando $\psi = 1$ e $\lambda = 0.5$.

respectivamente, variando ψ e podemos ver que quanto maior o valor do parâmetro ψ mais assimétrica será a distribuição.

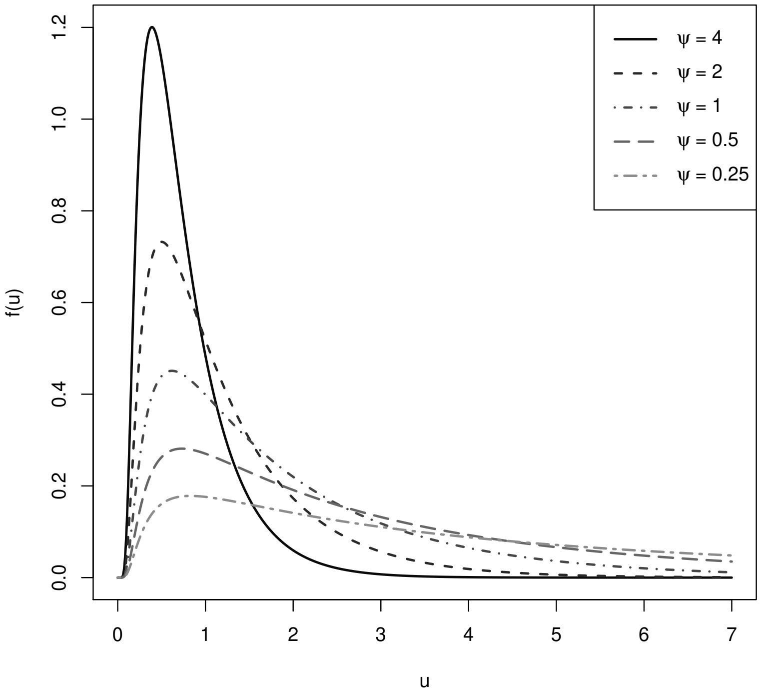


Figura 6 – Fdp da GIG para diferentes valores de ψ, fixando $\chi = 1$ e $\lambda = 0.5$.

2.3.4 Distribuição inversa gaussiana

Temos um dos casos mais conhecidos da distribuição GIG quando temos $\lambda = -1/2$, que corresponde a uma distribuição inversa gaussiana (IG). Existem diversas parametrizações possíveis para esta distribuição, segundo Johnson et al. (1994) temos que $U \sim IG(\mu, \lambda)$ se a função densidade de probabilidade é dada por

$$h_1(u; \mu, \lambda) = \left(\frac{\lambda}{2\pi u^3}\right)^{1/2} \exp\left\{-\frac{\lambda(u-\mu)^2}{2u\mu^2}\right\} \mathbb{1}_{\{u>0\}}(u), \tag{2.9}$$

com média $\mu > 0$ e $\lambda \in \mathbb{R}_+$. Mas reparametrizando (2.9) com $\mu = \delta/\gamma$ e $\lambda = \delta^2$ obtemos as seguintes possíveis funções densidade de probabilidade da distribuição IG.

$$h_2(u) = \left(\frac{\delta^2}{2\pi u^3}\right)^{1/2} \exp\left\{-\frac{\delta^2}{2(\delta/\gamma)^2 u}\left(u - \frac{\delta}{\gamma}\right)\right\}, \tag{2.10}$$

$$h_3(u) = \left(\frac{\delta^2}{2\pi u^3}\right)^{1/2} \exp\left\{-\frac{\gamma^2}{2u}\left(u^2 - 2u\frac{\delta}{\gamma} + \frac{\delta^2}{\gamma^2}\right)\right\}, \tag{2.11}$$

$$h_4(u) = \left(\frac{\delta^2}{2\pi u^3}\right)^{1/2} \exp\left\{-\frac{\gamma^2 u}{2} + \delta\gamma - \frac{\delta^2}{2u}\right\}, \tag{2.12}$$

$$h_5(u) = \left(\frac{\delta^2}{2\pi u^3}\right)^{1/2} \exp\left\{\delta\gamma - \frac{(\delta^2/u + \gamma^2 u)}{2}\right\}, \tag{2.13}$$

com $\delta \in \mathbb{R}_+$ e $\gamma \in \mathbb{R}_+$.

Substituindo, $\delta^2 = \chi$ e $\gamma^2 = \psi$ na parametrização $h_5(u)$, temos a parametrização mencionada na seção anterior (2.4) da distribuição GIG quando $\lambda = -0.5$, assim $U \sim IG(\chi, \psi)$ é dada por

$$h(u; \chi, \psi) = \sqrt{\frac{\chi}{2\pi}} \exp(\sqrt{\psi\chi}) u^{-3/2} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{u} + \psi u\right)\right\} \mathbb{1}_{\{u>0\}}(u). \tag{2.14}$$

A seguir, serão apresentadas algumas propriedades importantes sobre a distribuição inversa gaussiana, que serão úteis com o decorrer do trabalho.

Seja $U \sim IG(\chi, \psi)$, temos que:

1. A média é dada por

$$E[U] = \left(\frac{\chi}{\psi}\right)^{1/2}.$$

2. A variância é dada por

$$Var[U] = \left(\frac{\chi}{\psi^3}\right)^{1/2}.$$

3. A moda é dada por

$$Mo[U] = \left(\frac{-3}{2\psi}\right) + \sqrt{\frac{\chi}{\psi}}\left(1 + \frac{9}{4}\frac{1}{\chi\psi}\right)^{1/2}.$$

E algumas relações importantes da distribuição inversa gaussiana podem ser vistas abaixo.

1.

$$\left(\frac{\psi(U - \sqrt{\chi/\psi})^2}{U}\right) \sim {\chi^2}_{(1)},$$

2. Para c uma constante real positiva,

$$cU \sim IG(\sqrt{c\chi}, \sqrt{\psi/c}).$$

Por meio da distribuição acumulada da normal padrão podemos representar a função de distribuição acumulada de U. Sendo assim, se $U \sim IG(\chi, \psi)$, temos que a função distribuição acumulada é dada por

$$F(u;\chi,\psi) = \Phi\left\{\left(\frac{u}{\chi}\right)^{-1/2}\left(\frac{u\sqrt{\psi}-\sqrt{\chi}}{\sqrt{\chi}}\right)\right\} + \exp\{2\sqrt{\chi\psi}\}\Phi\left\{\left(\frac{1}{\sqrt{\chi\psi}}\right)^{-1/2}\left(\frac{u\sqrt{\psi}-\sqrt{\chi}}{\sqrt{\chi}}\right)\right\},$$

em que $u > 0$ e $\Phi(.)$ é a função distribuição acumulada (fda) da distribuição normal padrão, expressão deduzida por Shuster (1968) e Chhikara e Folks (1974). A seguir temos algumas relações importantes da distribuição IG.

Através das figuras a seguir conseguimos analisar como os diferentes parâmetros interferem graficamente na distribuição IG. Na Figura 7, fixando o parâmetro $\psi = 1$ temos que o aumento no parâmetro χ leva a diminuição da assimetria da distribuição e na Figura 8, fixando o parâmetro $\chi = 1$ temos que o aumento do parâmetro ψ leva ao aumento da simetria da distribuição.

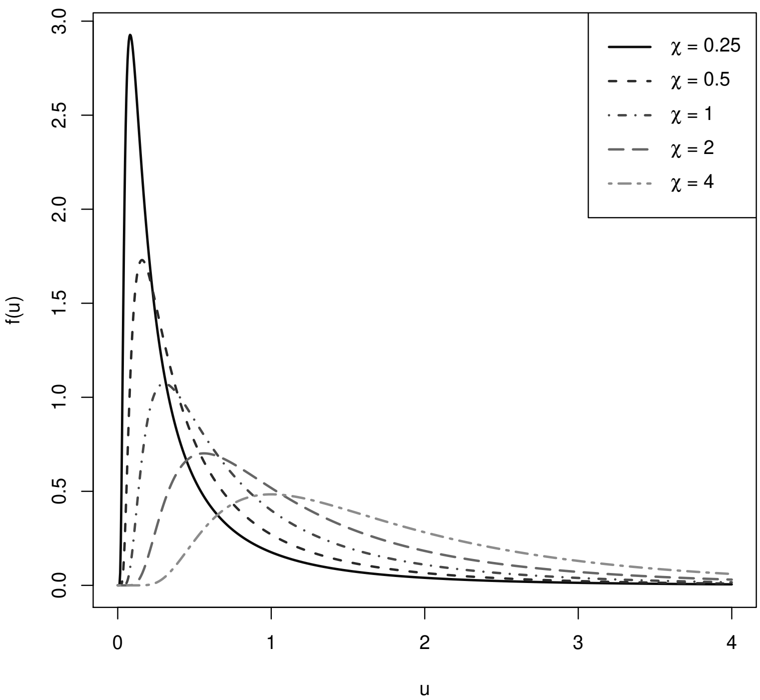


Figura 7 – Fdp da IG para diferentes valores de χ, fixando $\psi = 1$.

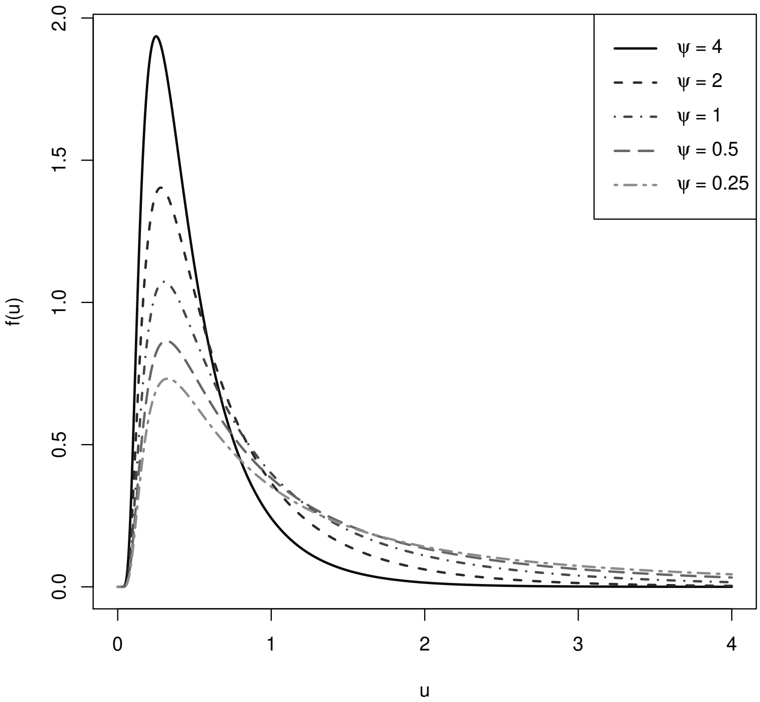


Figura 8 – Fdp da IG para diferentes valores de ψ, fixando $\chi = 1$.

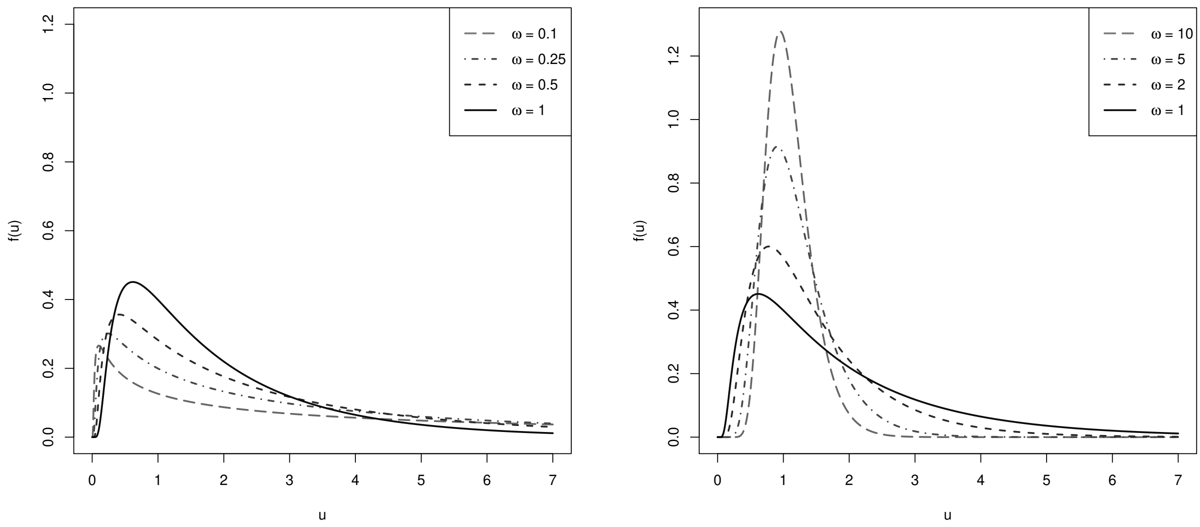
2.4 Reparametrização

Para o decorrer do trabalho será utilizada uma reparametrização da densidade dada em (2.4), fixando $\psi = \omega\eta$ e $\chi = \omega/\eta$ obtemos uma nova parametrização para a distribuição GIG com fdp dada por,

$$h(u) = \frac{(u/\eta)^{\lambda-1}}{2\eta K_\lambda(\omega)} \exp\left\{-\frac{\omega}{2}\left(\frac{u}{\eta} + \frac{\eta}{u}\right)\right\} \mathbb{1}_{\{u>0\}}(u), \tag{2.15}$$

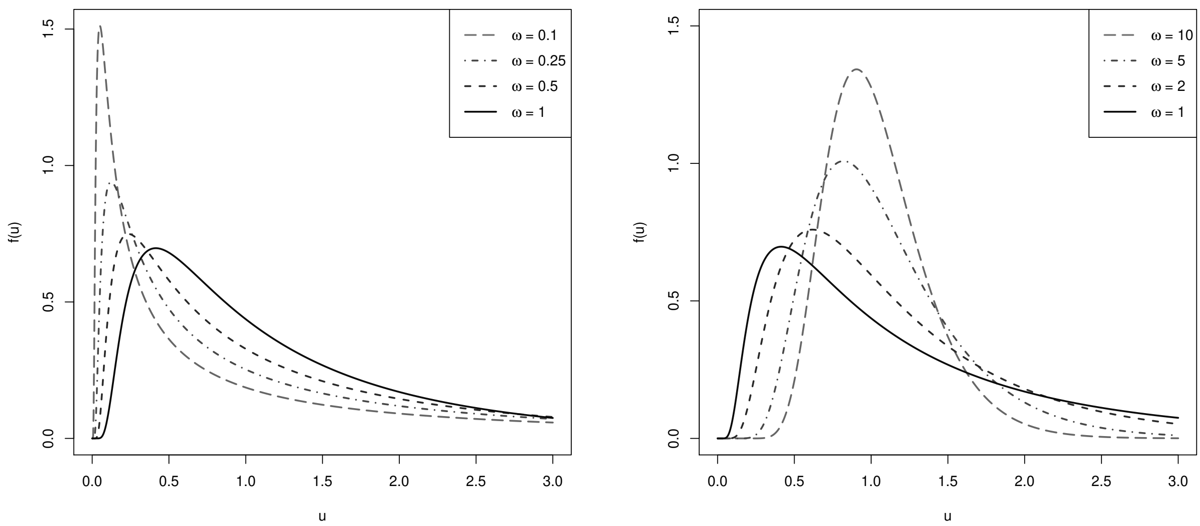
em que $\eta > 0$ é o parâmetro escala, $\omega > 0$ é um parâmetro de concentração, e λ é o parâmetro de indexação. Deste modo, quando denotamos $U \sim GIG(\omega, \eta, \lambda)$ temos que U segue uma GIG com a parametrização em (2.15).

Nas Figuras 9 e 10, apresentamos a distribuição GIG sob a parametrização (2.15) para $\eta = 1$, $\lambda = 0$ e $\lambda = 0.5$ e variando os valores de ω, podemos notar que a sua diminuição leva ao achatamento do gráfico da GIG.



(a) $\omega \leqslant 1$ (b) $\omega \geqslant 1$

Figura 9 – Fdp da GIG para diferentes valores de ω, fixando $\eta = 1$ e $\lambda = 0.5$.



(a) $\omega \leqslant 1$ (b) $\omega \geqslant 1$

Figura 10 – Fdp da GIG para diferentes valores de ω, fixando $\eta = 1$ e $\lambda = 0$.

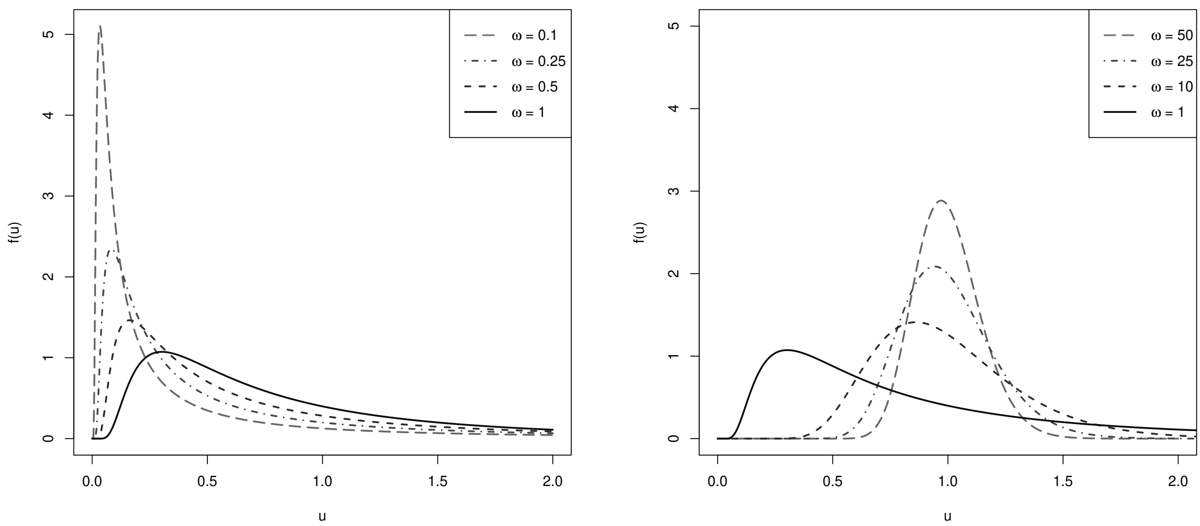
Considerando uma distribuição GIG com $\lambda = -1/2$, ou seja, uma distribuição IG, temos que a fdp obtida a partir da parametrização (2.15) será dada por

$$h(u) = \frac{(u/\eta)^{-3/2}}{2\eta K_{1/2}(\omega)} \exp\left\{-\frac{\omega}{2}\left(\frac{u}{\eta} + \frac{\eta}{u}\right)\right\} \mathbb{1}_{\{u>0\}}(u), \tag{2.16}$$

mais especificamente, será utilizada a IG com $\eta = 1$ e $\chi = \psi = \omega$, ou seja, $U \sim IG(\omega, \omega)$ com densidade dada por

$$h(u) = \frac{u^{-3/2}}{2K_{1/2}(\omega)} \exp\left\{-\frac{1}{2}\left(\omega u + \frac{\omega}{u}\right)\right\} \mathbb{1}_{\{u>0\}}(u). \tag{2.17}$$

Pela Figura 11, podemos observar o comportamento da distribuição IG, em que com o aumento do parâmetro ω a distribuição tem uma diminuição na simetria, sendo assim, quanto menor for o ω mais assimétrica será a distribuição IG.



(a) $\omega \leqslant 2$ (b) $\omega \geqslant 2$

Figura 11 – Fdp da IG para diferentes valores de ω, fixando $\eta = 1$.

2.5 Distribuição hiperbólica generalizada

As distribuições com caudas pesadas tiveram início com o economista Pareto (1986) e mais estudadas depois pelo matemático francês Paul Pierre Lévy (1925). No início estas distribuições eram estudadas apenas do ponto de vista teórico, mas hoje podem ser aplicadas a diversas áreas, como é o caso da matemática financeira. As distribuições com caudas pesadas possuem maior probabilidade na área das caudas, quando comparadas à distribuição normal com mesma média e mesma variância.

Em dados do campo financeiro, como acontece com os retornos de ações, suas distribuições, em grande parte dos casos, são assimétricas e possuem caudas mais pesadas do que uma distribuição normal. Logo, utilizar uma distribuição normal não seria uma boa escolha, deste modo, a princípio foi sugerido o uso de distribuições de Pareto por Mandelbrot (1963), mas essas distribuições apresentam caudas demasiadamente pesadas. Assim, as distribuições hiperbólicas surgiram como uma alternativa no tratamento de

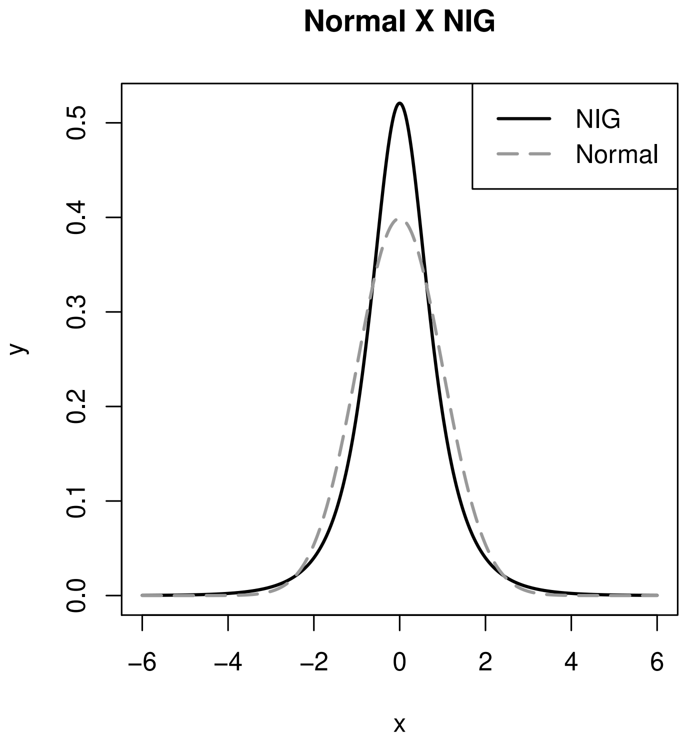


Figura 12 – Comparação entre $NIG(1, 0, 1, 0)$ e $N(0, 1)$.

dados financeiros, pois seus parâmetros são suficientemente flexíveis para se adequarem aos mais variados conjuntos de dados. Sendo mais tarde generalizadas e obtendo portanto, as distribuições GH.

Como podemos ver pela Figura 12, quando comparamos a distribuição normal com a distribuição NIG simétrica, que é um caso especial da distribuição GH, temos que esta distribuição se ajusta melhor a dados leptocúrticos do que a distribuição normal, como temos muitas vezes em dados financeiros. Além da área financeira, as distribuições hiperbólicas generalizadas são aplicadas às mais diversas áreas de conhecimento, como por exemplo, a física, biologia, agronomia, entre outros.

2.5.1 Caso univariado

A classe de distribuições GH tem grande destaque na literatura de modelagem financeira, particularmente no caso univariado. Um motivo importante para este destaque é a sua ligação com os processos de Lévy, isto é, processos com incrementos independentes e estacionários, como o movimento browniano, que são usados para modelar processos de preços em tempo contínuo.

Uma variável aleatória X tem distribuição GH se for definida como uma variável aleatória gerada pela combinação de uma variável U com uma variável com distribuição normal Z de acordo com a seguinte equação

$$X = \mu + U\beta + \sqrt{U}Z, \tag{2.18}$$

em que $Z \sim N(0, \sigma^2)$, e $U \sim GIG(\lambda, \chi, \psi)$ independente de Z. A sua função densidade de probabilidade foi proposta por McNeil et al.(2005), que pode ser expressa por

$$g(x) = \left[\frac{\chi + d(x)}{\psi + \beta^2/\sigma^2}\right]^{(\lambda-1/2)/2} \frac{(\psi/\chi)^{\lambda/2} K_{\lambda-1/2}(\sqrt{[\psi + \beta^2/\sigma^2][\chi + d(x)]})}{(\sqrt{2\pi\sigma^2} K_\lambda(\sqrt{\psi\chi})\exp[-\beta^2/\sigma^2(x-\mu)]} \mathbb{1}_{\{x>0\}}(x), \tag{2.19}$$

em que $d(x) = (x-\mu)^2/\sigma^2$ e $\boldsymbol{\theta} = (\lambda, \chi, \psi, \mu, \sigma, \beta)$ é o vetor dos parâmetros do modelo, com $\chi, \psi > 0$, $\mu, \beta, \lambda, \sigma \in \mathbb{R}$ e $\sigma > 0$,

Existe um conjunto grande de parametrizações alternativas para a distribuição GH na literatura, sendo mais comum encontrar essa distribuição com a seguinte reparametrização. O parâmetro de assimetria β é substituído pelo parâmetros β' e os parâmetros não negativos χ e ψ são substituído pelos parâmetros não negativos δ e α de acordo com

$$2^{\text{a}} \text{ parametrização: } \beta' = \frac{\beta}{\sigma^2}, \quad \delta = \sqrt{\chi} \quad e \quad \alpha = \sqrt{\psi + \frac{\beta^2}{\sigma^2}}.$$

Estes parâmetros devem satisfazer as seguintes restrições

$$\begin{aligned} \delta &\geqslant 0 \;\text{ e }\; \alpha^2 > \beta'^2\sigma^2 \;\text{ se }\; \lambda > 0, \\ \delta &> 0 \;\text{ e }\; \alpha^2 > \beta'^2\sigma^2 \;\text{ se }\; \lambda = 0, \\ \delta &> 0 \;\text{ e }\; \alpha^2 \geqslant \beta'^2\sigma^2 \;\text{ se }\; \lambda < 0. \end{aligned}$$

Blæsild (1981) usa essa parametrização para mostrar que as distribuições GH formam uma classe fechada de distribuições sob transformações lineares e condicionais. Entretanto, esta parametrização tem o problema de que os parâmetros α e δ são geralmente invariantes sob qualquer uma destas operações.

As diferentes parametrizações permitem desenvolver diferentes significados aos parâmetros utilizados. Abaixo seguem algumas das parametrizações conhecidas, considerando $\sigma = 1$:

$$\begin{aligned} &3^{\text{a}} \text{ parametrização:} \quad \zeta = \delta\sqrt{\alpha^2 - \beta'^2},\ \rho = \beta'/\alpha, \\ &4^{\text{a}} \text{ parametrização:} \quad \xi = (1+\zeta)^{-1/1},\ \bar{\chi} = \xi\rho, \\ &5^{\text{a}} \text{ parametrização:} \quad \bar{\alpha} = \alpha\delta,\ \bar{\beta'} = \beta'\delta. \end{aligned}$$

Os parâmetros da distribuição hiperbólica generalizada podem ser interpretados da seguinte forma: μ e δ representam respectivamente, a localização e a escala, β representa a assimetria e α representa a forma. Aumentando ξ ou diminuindo ζ ou $\overline{\alpha}$ temos um aumento no achatamento, representando a curtose. O parâmetro λ determina o peso das caudas da distribuição, como também define as diferentes subclasses das distribuições GH.

Outra parametrização foi proposta por Barndorff-Nielsen e Stelzer (2011), que representa as distribuições hiperbólicas generalizadas da seguinte forma,

$$g(x) = \frac{\bar{\gamma}\bar{\alpha}^{1/2-\lambda}}{\sqrt{2\pi}\delta K_\lambda(\bar{\gamma})} \left(1+\theta^2\right)^{\lambda/2-1/4} K_{\lambda-1/2}(\bar{\alpha}\sqrt{1+\theta^2})\exp\{\beta(x-\mu)\}\mathbb{1}_{\{x>0\}}(x), \qquad (2.20)$$

em que

$$\gamma = \sqrt{\alpha^2-\beta^2},\ \bar{\delta\alpha},\ \bar{\beta} = \delta\beta,\ \bar{\gamma} = \delta\gamma \ \text{ e } \ \theta = \frac{x-\mu}{\delta}. \qquad (2.21)$$

2.5.1.1 Caracterização dos parâmetros

Nesta seção podemos observar o efeito de cada um dos parâmetros da distribuição hiperbólica generalizada quando consideramos a fdp dada em (2.19), podemos ter mais informações consultando Wang (2015).

Quando $\beta = \bar{\beta} = \beta' = \rho = \bar{\chi} = 0$ obtemos uma distribuição GH simétrica, já quando temos que $\beta > 0$ observamos uma assimetria a esquerda e se $\beta < 0$ temos uma assimetria a direita e seu efeito pode ser visto na Figura 15. Quando maior é o valor do parâmetro β mais evidente será a assimetria. O efeito de μ é bastante intuitivo, já que é o parâmetro de locação e se efeito pode ser visto na Figura 13. O parâmetro χ interfere no achatamento da densidade, sendo que um aumento de χ torna a densidade mais achatado, que podemos ver na Figura 14.

Temos também que o menor valor do parâmetro ψ resulta num aumento do achatamento da densidade. Fixando todos os outros parâmetros, com um aumento de ψ, temos um aumento da variância, que pode ser analisada na Figura 16.

2.5.1.2 Algumas propriedades

A distribuição GH possui algumas propriedades importantes que serão úteis, como por exemplo, seja $X \sim GH(\lambda, \chi, \psi, \mu, \sigma, \beta)$ e $U \sim GIG(\lambda, \chi, \psi)$ com representação como em (2.19). Então,

1. A média de X é dada por

$$E[X] = \mu + E[U]\beta,$$

2. A variância de Xé dada por

$$Var[X] = Var[U]\beta^2 + E[U]\sigma^2.$$

Blaesild (1981) provou que a distribuição GH é fechada sob transformações lineares, como pode ser visto no seguinte teorema.

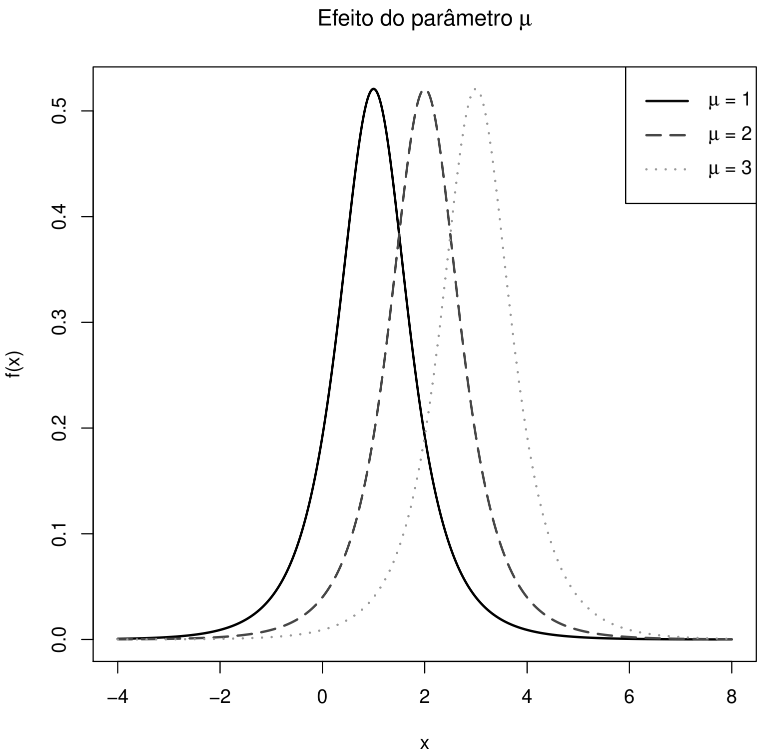


Figura 13 – Comparação do efeito do parâmetro μ com $\mu = 1, 2$ e 3, fixando $\lambda = -0.5$, $\psi = \chi = \sigma = 1$ e $\beta = 0$.

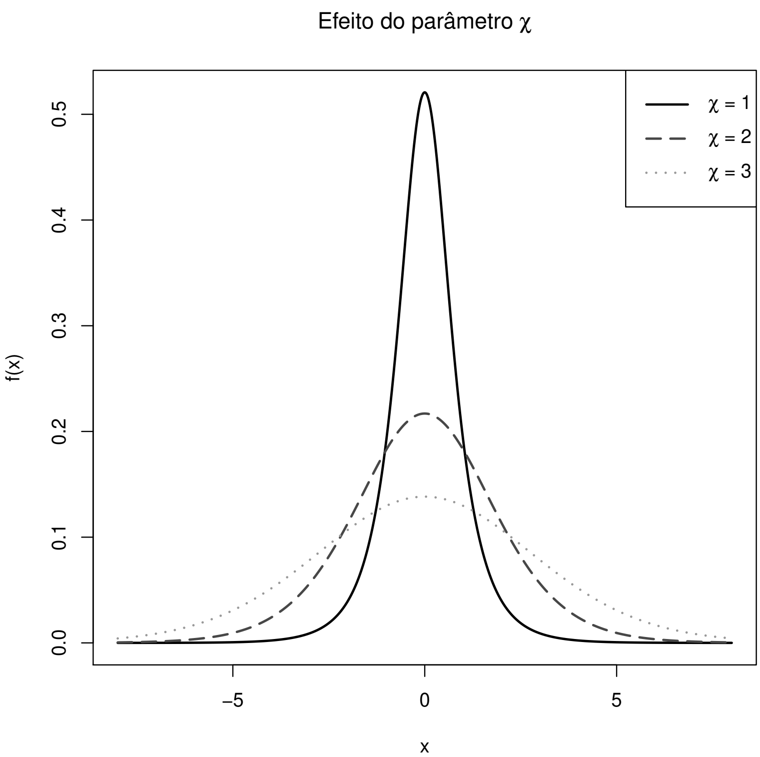
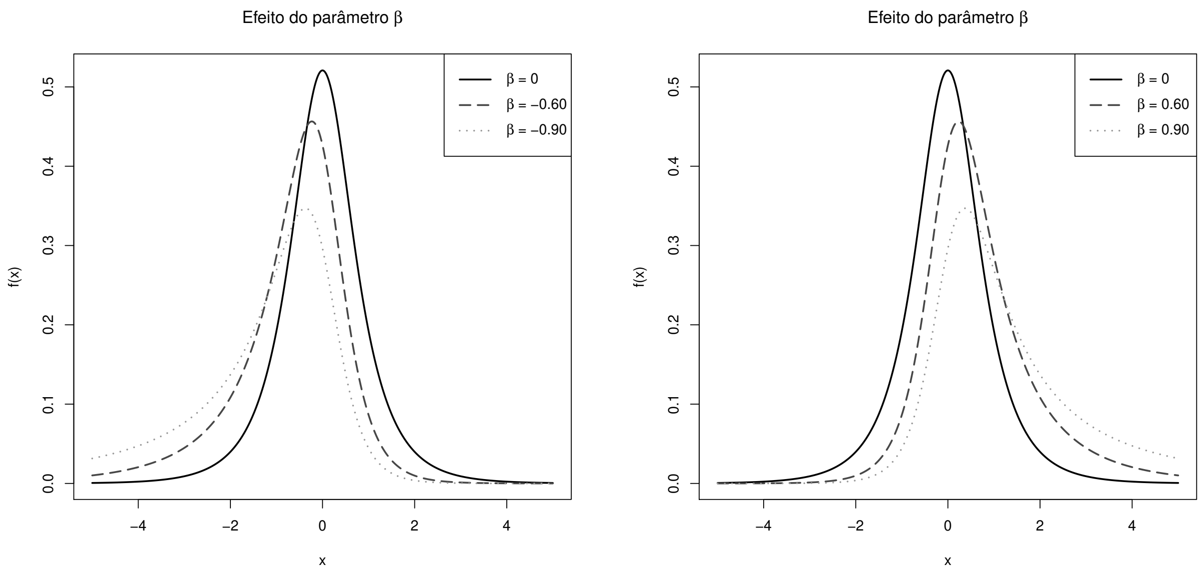


Figura 14 – Comparação do efeito do parâmetro χ com $\chi = 1, 2$ e 3, fixando $\lambda = -0.5$, $\psi = \sigma = 1$ e $\beta = \mu = 0$.



(a) Para $\beta = 0, -0.6$ e -0.9. (b) Para $\beta = 0, 0.6$ e 0.9.

Figura 15 – Comparação do efeito do parâmetro β com $\beta = -0.9, -0.6, 0, 0.6$ e 0.9, fixando $\lambda = -0.5$, $\alpha = \chi = \sigma = 1$ e $\mu = 0$.

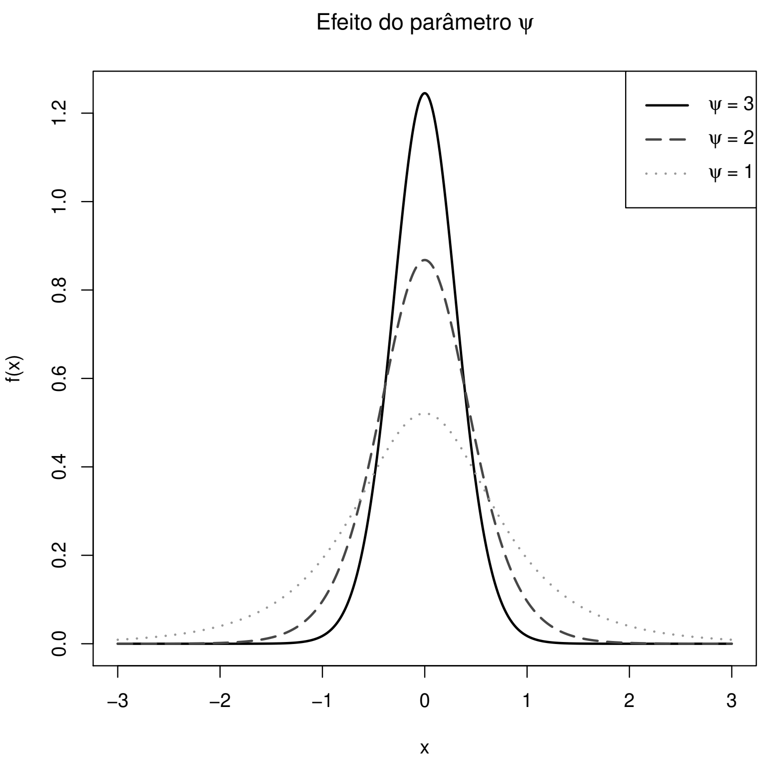


Figura 16 – Comparação do efeito do parâmetro ψ com $\psi = 1, 2$ e 3, fixando $\lambda = -0.5$, $\delta = \sigma = 1$ e $\beta = \mu = 0$.

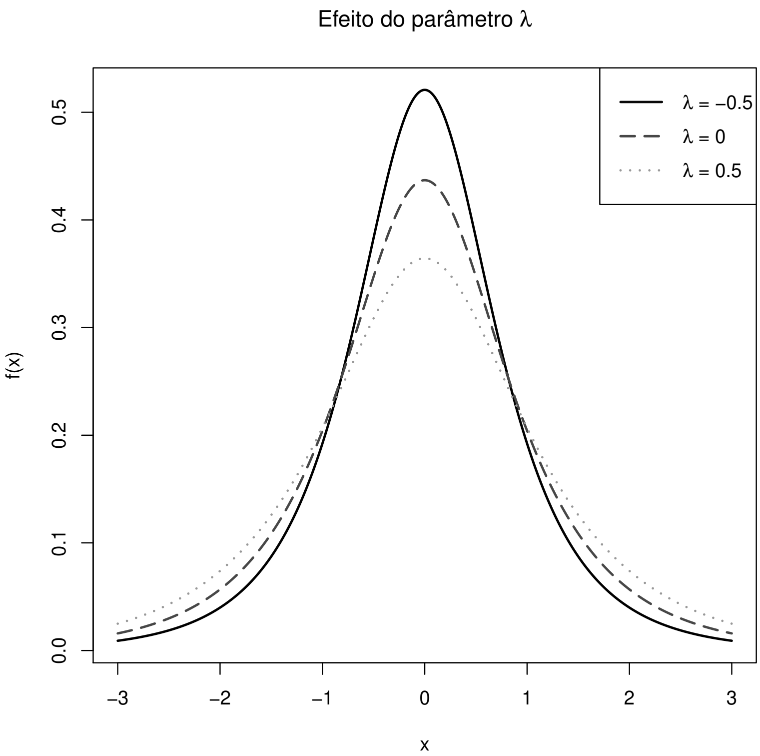


Figura 17 – Comparação do efeito do parâmetro λ com $\lambda = -0.5, 0$ e 0.5, fixando $\psi = \chi = \sigma = 1$ e $\beta = 0$ e $\mu = 0$.

Teorema 2.1. *Seja $X \sim GH(\lambda, \chi, \psi, \mu, \sigma, \beta)$. Então a transformação linear $Y = aX + b$, em que $a, b \in \mathbb{R}$, Y também é uma distribuição hiperbólica generalizada $Y \sim GH(\lambda, \chi, \psi, a\mu + b, a^2\sigma^2, a\beta)$.*

2.5.1.3 Casos especiais

A família da distribuição GH é extremamente flexível e contém muitos casos especiais conhecidos, como por exemplo,

- **Distribuição t-Student**

  Se $\lambda = -\nu/2$, $\psi = \beta = 0$, $\sigma = 1$ e $\chi = \nu$, temos uma distribuição $GH(-\nu/2, \nu, 0, \mu, 0) = t_{(\nu)}$, com fdp dada por

$$g(x) = \frac{\Gamma\left(\dfrac{\mu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\dfrac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

- **Distribuição hiperbólica**

Se $\lambda = 1$ temos uma distribuição Hiperbólica $HYP(\chi, \psi, \sigma, \mu, \beta)$, com fdp dada por

$$g(x) = \left[\frac{\chi + d(x)}{\psi + \beta^2/\sigma^2}\right]^{1/4} \frac{(\psi/\chi)^{1/2} K_{1/2}(\sqrt{[\psi + \beta^2/\sigma^2][\chi + d(x)]})}{(\sqrt{2\pi\sigma^2} K_1(\sqrt{\psi\chi}) \exp[-\beta^2/\sigma^2(x-\mu)]},$$

com $x, \mu \in \mathbb{R}$ e $d(x) = \dfrac{(x-\mu)^2}{\sigma^2}$.

- **Distribuição normal inversa gaussiana**

Se $\lambda = -1/2$, temos uma distribuição $NIG(\chi, \psi, \sigma, \mu, \beta)$, que será aprofundada adiante, com função densidade dada por

$$g(x) = \left[\frac{\chi + d(x)}{\psi + \beta^2/\sigma^2}\right]^{-1/2} \frac{(\psi/\chi)^{-1/4} K_1(\sqrt{[\psi + \beta^2/\sigma^2][\chi + d(x)]})}{(\sqrt{2\pi\sigma^2} K_{1/2}(\sqrt{\psi\chi}) \exp[-\beta^2/\sigma^2(x-\mu)]}$$

com $x, \mu \in \mathbb{R}$ e $d(x) = \dfrac{(x-\mu)^2}{\sigma^2}$.

A distribuição normal inversa gaussiana simétrica (NIGS) é obtida quando o parâmetro de assimetria é igual a zero ($\beta = 0$). Assim, temos que a distribuição NIGS é definida pela relação

$$X = \mu + \sqrt{U} Z,$$

em que $Z \sim N(0, \sigma^2)$ e U são independentes, e U segue uma distribuição IG, definida em (2.14). A distribuição NIGS será denotada por $X \sim NIGS(\mu, \sigma; \chi, \psi)$. Usando a função geradora de momentos (Hanssen e Oigard, 2001), temos o seguinte teorema:

Teorema 2.2. *Seja $X \sim NIGS(\mu, \sigma; \chi, \psi)$, com χ, $\psi > 0$. Então,*

1. *A média é dada por*

$$E[X] = \mu$$

2. *A variância é dada por*

$$Var[X] = E[U]\sigma^2$$

Quando $\psi \to 0$, obtemos a distribuição Cauchy como um caso especial da distribuição NIGS (Lillestøl, 2000). A distribuição normal também pode ser obtida como um caso limite da distribuição NIGS quando $\psi \to \infty$.

2.5.2 Caso multivariado

Um vetor aleatório p-variado X tem distribuição GH se for definido como um vetor aleatório gerado pela combinação de uma variável aleatória positiva U com um vetor aleatório normal p-variado $\mathbf{Z}$ de acordo com a seguinte equação

$$\mathbf{X} = \boldsymbol{\mu} + U\boldsymbol{\beta} + \sqrt{U}\mathbf{Z}, \tag{2.22}$$

em que $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, e $U \sim GIG(\psi, \chi, \lambda)$ independente de $\mathbf{Z}$.

A sua fdp foi derivada por McNeil et al.(2005), como

$$f(\mathbf{x}) = \left[\frac{\chi + d(\mathbf{x})}{\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}\right]^{(\lambda - p/2)/2} \times \frac{(\psi/\chi)^{\lambda/2} K_{\lambda - p/2}(\sqrt{[\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}][\chi + d(\mathbf{x})]})}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2} K_\lambda(\sqrt{\psi\chi})\exp[(\boldsymbol{\mu} - \mathbf{x})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}]}, \tag{2.23}$$

em que $d(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ é a distância de Mahalanobis e $\boldsymbol{\theta} = (\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta})$ é o vetor dos parâmetros do modelo. Usamos a notação $\mathbf{X} \sim GH_p(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta})$ para indicar que X segue a distribuição em (2.23), e assumido que $|\boldsymbol{\Sigma}| = 1$ para garantir a identificabilidade do modelo, para mais informações ver Browne et al (2015).

Podemos ver facilmente que $\mathbf{X}|U = u \sim N_p(\boldsymbol{\mu} + u\boldsymbol{\beta}, u\boldsymbol{\Sigma})$. Assim, pelo teorema de Bayes a distribuição condicional de U dado $\mathbf{X} = \mathbf{x}$, pode ser dada por

$$\begin{aligned} f(u|\mathbf{x}) &= \frac{f(\mathbf{x}|u)h(u)}{f(\mathbf{x})} \\ &= \left(\frac{\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}{\chi + d(\mathbf{x})}\right)^{(\lambda - p/2)/2} \times \frac{u^{\lambda - p/2 - 1}}{2K_{\lambda - p/2}(\sqrt{[\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}][\chi + d(\mathbf{x})]})} \\ &\times \exp\left\{-\frac{1}{2}\left(\frac{(\chi + d(\mathbf{x}))}{u} + (\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})u\right)\right\}, \end{aligned}$$

e assim, $U|\mathbf{X} = \mathbf{x} \sim GIG(\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}, \chi + d(\mathbf{x}), \lambda - p/2)$.

De acordo com Browne et al (2015), seja $\mathbf{X} \sim GH_p(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta})$ e $U \sim GIG((\lambda, \chi, \psi)$. Então,

1. A média é dada por

$$E[\mathbf{X}] = \boldsymbol{\mu} + E[U]\boldsymbol{\beta},$$

2. A variância é dada por

$$\begin{aligned} Var[\mathbf{X}] &= E[Var(\mathbf{X}|U)] + Var[E(\mathbf{X}|U)] \\ &= Var[U]\boldsymbol{\beta}\boldsymbol{\beta}' + E[U]\boldsymbol{\Sigma}. \end{aligned}$$

A função geradora de momentos da distribuição GH é facilmente calculada, uma vez que a função geradora de momento da mistura é conhecida. Assim, ela é dada por

$$\begin{aligned} M_{GH}(\boldsymbol{t}) &= E[E[\exp\{\mathbf{t}'\mathbf{X}\}|U]] = e^{\boldsymbol{t}'\boldsymbol{\mu}} E[\exp\{U(\boldsymbol{t}'\boldsymbol{\beta} + 1/2\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})\}] \\ &= e^{\boldsymbol{t}'\boldsymbol{\mu}} \left(\frac{\psi}{\psi - 2\boldsymbol{t}'\boldsymbol{\beta} - \boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}\right)^{\lambda/2} \frac{K_\lambda(\sqrt{\psi(\chi - 2\boldsymbol{t}'\boldsymbol{\beta} - \boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})})}{K_\lambda(\sqrt{\chi\psi})}, \quad \chi \geqslant 2\boldsymbol{t}'\boldsymbol{\beta} + \boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}. \end{aligned}$$

Como provado por Blaesild (1981) temos que a distribuição GH é fechada sob transformações lineares. Assim, segue que

Teorema 2.3. *Seja* $\mathbf{X} \sim GH_p(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta})$*, então a transformação linear* $\mathbf{Y} = \mathbf{AX}+\mathbf{b}$*, em que* $\mathbf{b} \in \mathbb{R}^k$ *e* $\mathbf{A} \in \mathbb{R}^{kxd}$*,* $\mathbf{Y}$ *também é uma distribuição hiperbólica generalizada* $\mathbf{Y} \sim GH_p(\lambda, \chi, \psi, \mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}', \mathbf{A}\boldsymbol{\beta})$*.*

Considerando o parâmetro de assimetria igual a zero ($\boldsymbol{\beta} = \mathbf{0}$), temos a distribuição GH multivariada simétrica, com $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ e $U \sim GIG(\lambda, \chi, \psi)$. Assim, temos que $\mathbf{X} = \boldsymbol{\mu} + \sqrt{U}\mathbf{Z}$ e sua função densidade de probabilidade É dada por

$$f(\mathbf{x}) = \left[\frac{\chi + d(\mathbf{x})}{\psi}\right]^{(\lambda-p/2)/2} \times \frac{(\psi/\chi)^{\lambda/2} K_{\lambda-p/2}(\sqrt{\psi[\chi + d(\mathbf{x})]})}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2} K_\lambda(\sqrt{\psi\chi})}, \qquad (2.24)$$

em que $d(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ e $\boldsymbol{\theta} = (\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Serão apresentados a seguir alguns casos especiais da distribuição GH simétrica, em que serão utilizados algumas propriedades assintóticas da função de Bessel K_λ que podem ser encontrados em Abramowitz e Stegun (1972).

- **Distribuição t-Student**
  A distribuição $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ é obtida considerando $\lambda < 0$ fixo e $\psi \to 0$, assim temos que
  $$\lim_{\psi \to 0} f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2} \frac{\Gamma(p/2 - \lambda)}{(\pi\chi)^{p/2}\Gamma(-\lambda)} (1 + d(\mathbf{x})/\chi)^{\lambda - p/2}.$$
  Desta maneira, a densidade da distribuição t-Student com ν graus de liberdade é obtida considerando $\nu = \chi = -2\lambda$,

- **Distribuição normal contaminada**
  Denotada por $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu, \eta)$, com $0 \leqslant \nu \leqslant 1$ e $0 < \eta < 1$, esta distribuição será obtida através da mistura de $GH_p(\boldsymbol{\mu}, (1/\eta)\boldsymbol{\Sigma}, \lambda, \chi, \psi)$ e $GH_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \chi, \psi)$ com fdp dadas por f_1 e f_2, respectivamente. Desta forma, considere a fdp dada por
  $$f(\mathbf{x}) = \nu f_1(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}/\eta, \lambda, \chi, \psi) + (1 - \nu) f_2(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \chi, \psi),$$
  ASSIM TOMANDO $\psi \to \infty$ e $\chi \to \infty$, tal que $\chi/\psi \to 1$, temos a distribuição normal contaminada multivariada, com fdp dada por
  $$\lim_{(\psi,\chi)\to(\infty,\infty)} f(\mathbf{x}) = \nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}/\eta) + (1 - \nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

- **Distribuição slash**
  Denotada por $SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, a fdp é obtida através do vetor aleatório $\mathbf{X} = \boldsymbol{\mu} + k(U)^{1/2}\mathbf{Z}$, em que $k(u) = e^{-u}$ e $U \sim GIG(1, 0, \psi)$, ou seja, $U \sim Gama(1, \psi/2)$, dessa maneira, temos que
  $$\lim_{\chi \to 0} f(\mathbf{x}) = f(\mathbf{x}) = (\psi/2) \int_0^1 u^{(\psi/2)-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, e^{-u}\boldsymbol{\Sigma}) du.$$

Assim, obtemos $SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ com $\nu = \psi/2$.

- **Distribuição normal inversa gaussiana**
  Seja $\mathbf{X}$ um vetor aleatório com distribuição $NIG_p(\chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\beta})$, a sua função densidade de probabilidade é dada por

$$f(\mathbf{x}) = \left[\frac{\chi + d(\mathbf{x})}{\psi + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}\right]^{-(p+1)/4} \frac{(\psi/\chi)^{-1/4} K_{-1/2-p/2}(\sqrt{[\psi + \boldsymbol{\beta}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}][\chi + d(\mathbf{x})]})}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2} K_{1/2}(\sqrt{\psi\chi}) \exp[(\boldsymbol{\mu} - \mathbf{x})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}]}, \tag{2.25}$$

em que $d(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$.

Um caso especial da distribuição NIG é a distribuição normal inversa gaussiana simétrica, que obtemos quando $\boldsymbol{\beta} = \mathbf{0}$. Assim, um vetor aleatório p-dimensional $\mathbf{X}$ tem distribuição NIGS multivariada como uma mistura de um vetor aleatório normal p-dimensional $\mathbf{Z}$ com U seguindo distribuição IG com parâmetros χ, ψ, de acordo com a seguinte equação

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{U}\mathbf{Z}, \tag{2.26}$$

em que $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, e $U \sim IG(\psi, \chi)$ independente de $\mathbf{Z}$. Assim, sua função densidade de probabilidade é dada por

$$f(\mathbf{x}) = \left[\frac{\chi + d(\mathbf{x})}{\psi}\right]^{-(p+1)/4} \frac{(\psi/\chi)^{-1/4} K_{-1/2-p/2}(\sqrt{\psi(\chi + d(\mathbf{x}))})}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2} K_{1/2}(\sqrt{\psi\chi})}. \tag{2.27}$$

Usamos a notação $\mathbf{X} \sim NIGS_p(\chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ para indicar que X segue a distribuição em (2.27).

A seguir apresentamos a função geradora de momentos (fgm) e a função característica da distribuição NIGS. Então, seja $\mathbf{X} \sim NIGS_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \chi, \psi)$, com χ, $\psi > 0$. Então,

1. A função característica de $\mathbf{X}$ é dada por

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \left(\frac{\sqrt{\psi}}{\sqrt{\psi + \mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t}}}\right)^{-1/2} \frac{K_{1/2}(\sqrt{\chi(\psi + \mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t})})}{K_{1/2}(\sqrt{\chi\psi})}$$

2. A fgm de $\mathbf{X}$ é dada por

$$M_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \left(\frac{\sqrt{\psi}}{\sqrt{\psi - \mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t}}}\right)^{-1/2} \frac{K_{1/2}(\sqrt{\chi(\psi - \mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t})})}{K_{1/2}(\sqrt{\chi\psi})},$$

com $\mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t} < \psi$, e $\mathbf{t} \in \mathbb{R}^p$. A prova e outras propriedades podem ser vistas em Zeller (2009).

A partir dessas propriedades podemos encontrar os momentos de $\mathbf{X}$ através da fgm de $\mathbf{X}$ que é dada por

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu}} M_{IG}(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}/2), \ \mathbf{t} \in \mathbb{R}^p, \tag{2.28}$$

em que

$$M_{IG}(s) = \left(\frac{\sqrt{\psi}}{\sqrt{\psi - 2s}} \right)^{-1/2} \frac{K_{1/2}(\sqrt{\chi(\psi - 2s)})}{K_{1/2}(\sqrt{\chi\psi})} \tag{2.29}$$

é a fgm de $U \sim IG(\chi, \psi)$.

Sejam U e $\mathbf{Z}$ independentes, podemos obter o vetor de médias e a matriz de covariâncias de um vetor aleatório com distribuição NIGS. Assim, seja $\mathbf{X} \sim NIGS_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \chi, \psi)$, com χ, $\psi > 0$. Então,

1. se $E[U^{1/2}] < \infty$ temos que $E[\mathbf{X}]$ existe e é dado por
$$E[\mathbf{X}] = \boldsymbol{\mu},$$

2. se $E[U] < \infty$ temos que o segundo momento $E[\mathbf{X}]$ existe e é dado por
$$Var[\mathbf{X}] = Var[U]\boldsymbol{\Sigma}.$$

Um caso especial da distribuição NIGS multivariada que será utilizado neste trabalho é a distribuição NIGS bivariada. Dizemos que $\mathbf{X} = (X_1, X_2)^\top$ segue uma distribuição NIGS bivariada padrão se $\mathbf{X} = \sqrt{U}\mathbf{Z}$, em que $\mathbf{Z} \sim N_2(0, \boldsymbol{\Sigma})$ e $U \sim IG(\chi, \psi)$, com $\mathbf{Z}_0$ e U independentes. Esta distribuição será denotada por $\mathbf{X} \sim NIGS_2(0, \boldsymbol{\Sigma}, \chi, \psi)$.

Note que $\mathbf{X}$, dado $U = u$, tem uma distribuição normal bivariada $N_2(\mathbf{0}, u\boldsymbol{\Sigma})$, em que a fdp de $\mathbf{X}$ pode ser obtida da seguinte forma

$$\begin{aligned}
f(\mathbf{x}) &= \int_0^\infty f(\mathbf{x}|u)h(u)du \\
&= \int_0^\infty \frac{1}{2\pi u\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2u}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}\right\} \frac{\sqrt{\chi}u^{-3/2}}{\sqrt{2\pi}} \exp(\sqrt{\chi\psi}) \exp\left\{-\frac{1}{2}\left(\frac{\chi}{u} + \psi u\right)\right\} du \\
&= \frac{\sqrt{\chi}}{(2\pi)^{3/2}\sqrt{1-\rho^2}} \exp(\sqrt{\chi\psi}) \int_0^\infty u^{-5/2} \exp\left\{-\frac{1}{2}\left(\frac{\chi + \mathbf{x}^T\Sigma^{-1}\mathbf{x}}{u} + \psi u\right)\right\} du \\
&= \frac{\sqrt{\chi}}{(2\pi)^{3/2}\sqrt{1-\rho^2}} \exp(\sqrt{\chi\psi}) \int_0^\infty u^{-5/2} \exp\left\{-\frac{1}{2}\left(\frac{\chi + d(\mathbf{x})}{u} + \psi u\right)\right\} du,
\end{aligned}$$

em que $d(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}$. Assim, transformando $u = \dfrac{\sqrt{d(\mathbf{x}) + \chi}}{\sqrt{\psi}} y$, temos

$$
\begin{aligned}
f(\mathbf{x}) &= \frac{\sqrt{\chi}}{(2\pi)^{3/2}\sqrt{1-\rho^2}}\exp(\sqrt{\chi\psi}) \int_0^\infty \left(\frac{\sqrt{d(\mathbf{x})+\chi}}{\sqrt{\psi}}y\right)^{-5/2} \times \\
&\times \exp\left\{-\frac{1}{2}\left(\frac{\chi + d(\mathbf{x})}{\left(\frac{\sqrt{d(\mathbf{x})+\chi}}{\sqrt{\psi}}y\right)} + \psi\left(\frac{\sqrt{d(\mathbf{x})+\chi}}{\sqrt{\psi}}y\right)\right)\right\}\left(\frac{\sqrt{d(\mathbf{x})+\chi}}{\sqrt{\psi}}y\right) dy \\
&= \frac{\sqrt{\chi}}{(2\pi)^{3/2}\sqrt{1-\rho^2}}\exp(\sqrt{\chi\psi})\left(\frac{\sqrt{d(\mathbf{x})+\chi}}{\sqrt{\psi}}y\right)^{-3/2} \times \\
&\times \int_0^\infty y^{-5/2}\exp\left\{-\frac{1}{2}\sqrt{\psi(d(\mathbf{x})+\chi)}\left(\frac{1}{y}+y\right)\right\} dy \\
&= \frac{\sqrt{\chi}}{2\sqrt{1-\rho^2}}\left(\frac{\sqrt{\psi}}{\pi\sqrt{d(\mathbf{x})+\chi}}\right)^{3/2} \exp(\sqrt{\chi\psi})K_{3/2}(\sqrt{\psi(d(\mathbf{x})+\chi)}).
\end{aligned}
$$

Assim, a fdp de $\mathbf{X} \sim NIGS_2(\mathbf{0}, \mathbf{\Sigma}, \chi, \psi)$ é dada por

$$
f(\mathbf{x}) = \frac{\sqrt{\chi}}{2\sqrt{1-\rho^2}}\left(\frac{\sqrt{\psi}}{\pi\sqrt{d(\mathbf{x})+\chi}}\right)^{3/2} \exp(\sqrt{\chi\psi})K_{3/2}(\sqrt{\psi(d(\mathbf{x})+\chi)}), \tag{2.30}
$$

em que $d(\mathbf{x})$ é a distância de Mahalanobis já definida anteriormente.

2.6 Discussão sobre reparametrização

Nesta seção, será discutida a parametrização utilizada no decorrer do trabalho. Na literatura, é assumido que $|\mathbf{\Sigma}| = 1$ para garantir a identificabilidade do modelo, como esta condição pode ser muito restritiva para as aplicações, recentemente algumas outras condições foram propostas para tornar esse modelo mais flexível, como por exemplo o que foi proposto por Browne et al (2015) que é utilizar $U \sim GIG(\omega, 1, \lambda)$. Fixando $\psi = \omega\eta$ e $\chi = \omega/\eta$, ou então $\omega = \sqrt{\chi\psi}$ e $\eta = \sqrt{\psi/\chi}$, obtendo uma nova parametrização para a distribuição GIG com fdp dada por,

$$
h(u) = \frac{(u/\eta)^{\lambda-1}}{2\eta K_\lambda(\omega)}\exp\left\{-\frac{\omega}{2}\left(\frac{u}{\eta}+\frac{\eta}{u}\right)\right\}, \; u > 0, \tag{2.31}
$$

em que $\eta > 0$ é o parâmetro escala, $\omega > 0$ é um parâmetro de concentração, e λ é o parâmetro de indexação. Sob esta parametrização a fdp da distribuição GH multivariada será dada por

$$
f(\mathbf{x}) = \left[\frac{\omega + d}{\omega + \boldsymbol{\beta}^\top\mathbf{\Sigma}^{-1}\boldsymbol{\beta}}\right]^{(\lambda-p/2)/2} \frac{K_{\lambda-p/2}(\sqrt{[\omega+\boldsymbol{\beta}^\top\mathbf{\Sigma}^{-1}\boldsymbol{\beta}][\omega+d]})}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}K_\lambda(\omega)\exp[-(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}\boldsymbol{\beta}]}. \tag{2.32}
$$

Quando consideramos uma distribuição GIG com $\lambda = -1/2$ temos uma distribuição IG que a fdp em (2.31), será dada por

$$h(u) = \frac{(u/\eta)^{-3/2}}{2\eta K_{1/2}(\omega)} \exp\left\{-\frac{\omega}{2}\left(\frac{u}{\eta} + \frac{\eta}{u}\right)\right\}, \tag{2.33}$$

mais especificamente, será utilizada a distribuição IG com $\eta = 1$ e $\chi = \psi = \omega$, ou seja, $U \sim IG(\omega, \omega)$ com fdp dada por

$$h(u) = \frac{u^{-3/2}}{2K_{1/2}(\omega)} \exp\left\{-\frac{1}{2}\left(\omega u + \frac{\omega}{u}\right)\right\}. \tag{2.34}$$

Da mesma maneira que ocorre com a distribuição GH multivariada, a condição $|\boldsymbol{\Sigma}| = 1$ é muito restritiva quando consideramos uma distribuição NIG multivariada. Uma proposta para contornar esta restrição é utilizar $U \sim IG(\omega, \omega)$, definida em (2.34). Sob esta parametrização a fdp da NIG multivariada será dada por

$$f(\mathbf{x}) = \left[\frac{\omega + d}{\omega + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}\right]^{-(p+1)/4} \frac{K_{-1/2-p/2}(\sqrt{[\omega + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}][\omega + d]})}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2} K_{1/2}(\omega) \exp[-(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}]}. \tag{2.35}$$

Quando analisamos o parâmetro ω no caso univariado, considerando $\sigma = 1$, $\beta = 0$, $\mu = 0$ e $\lambda = -1/2$, podemos comparar seu efeito com a distribuição normal, em que pode ser observado que quanto menor for o valor de ω mais leptocúrtica será a distribuição, e quanto maior for seu valor, mais a distribuição se assemelha a uma distribuição normal, que pode ser visto na Figura 18 com $\omega = 20$. Isto ocorre devido ao fato da distribuição normal ser um caso limite da distribuição NIGS quando $\omega \to \infty$.

Em especial, neste trabalho será utilizada uma distribuição NIGS bivarida, em que sua fdp é um caso especial da densidade dada em (2.35). Assim, temos que a fdp da distribuição NIGS bivariada que será utilizada no trabalho é dada por

$$f(\mathbf{x}) = \left[\frac{\omega + d(\mathbf{x})}{\omega}\right]^{-3/4} \frac{K_1(\sqrt{\omega[\omega + d(\mathbf{x})]})}{(2\pi)|\boldsymbol{\Sigma}|^{1/2} K_{1/2}(\omega)}, \tag{2.36}$$

em que $d(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$.

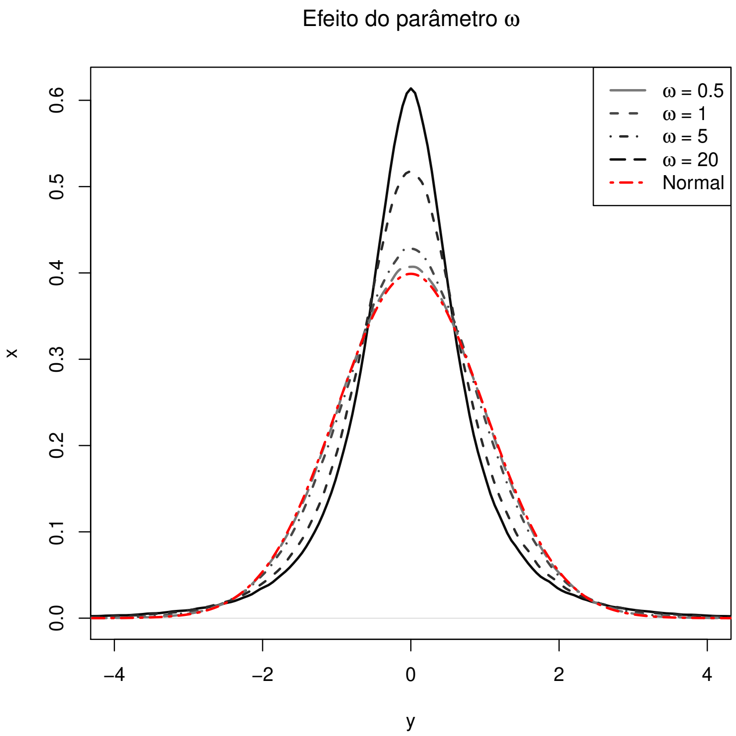


Figura 18 – Comparação do efeito do parâmetro ω da NIG com a distribuição normal, considerando $\omega = 0.5, 1, 5$ e 10, fixando $\beta = 0$, $\mu = 0$ e $\lambda = -1/2$.

3 Modelo de Regressão com Erros nas Variáveis

Em geral, os modelos de regressão linear são ajustados sob a hipótese de que as covariáveis são medidas sem erro. Porém, em muitas situações na prática, essas covariáveis não podem ser medidas diretamente. Nesse caso, fatores aleatórios, como correntes de ar ou vibrações, introduziriam erros nas mensurações. Uma alternativa para contornar este problema seria repetir a medição, permitindo avaliar a incerteza estatística no resultado final.

Situações desse tipo ocorrem na engenharia, em que a redução dos "erros", no contexto mais amplo da palavra, é tratada especificamente pela ciência denominada por metrologia. Em síntese, essa ciência analisa os aspectos relativos às medições e calibrações de instrumentos, utilizando-se metodologias estatísticas. Tais medições, em conjunto com o uso de uma técnica estatística inadequada certamente resultam em índices de qualidade incoerentes. Para modelar estas situações, incluindo no modelo os erros de medições ou erros no processo, o modelo de regressão com erros nas variáveis, proposta por Fuller (1987) vem a contornar estes problemas.

Os modelos de regressão com erros nas variáveis são uma extensão dos modelos de regressão usuais, em que duas variáveis possuem uma relação linear entre si do tipo

$$y_i = \alpha + \beta x_i,$$

porém nem y e nem x são observados diretamente devido à presença de erros na medição das covariáveis. Assim, considerar um modelo de regressão linear simples afetaria a precisão dos parâmetros do modelo, desta maneira consideram-se os valores observados de X e Y como

$$\begin{aligned} Y_i &= y_i + e_i \\ X_i &= x_i + u_i, \end{aligned}$$

para $i = 1, ..., n$.

Os modelos com erros nas variáveis podem ser classificados em funcional e estrutural. No modelo funcional, temos as suposições de que e_i e u_i são variáveis aleatórias independentes, com média zero e variâncias ϕ_e e ϕ_u respectivamente e ϕ_u e os valores de x_i não observados como constantes. Assim, temos que o número de parâmetros cresce de acordo com o número de observações, pois os valores de x_i são parâmetros chamados de incidentais, e são associados à i-ésima observação. Desta maneira, temos que os parâmetros do modelo são dados por $\alpha, \beta, \phi_e, \phi_u$ e x_i, para $i = 1, \ldots, n$, totalizando $n + 4$ parâmetros.

Já no modelo estrutural temos que x_i é considerado variável aleatória, com média μ_x e variância ϕ_x. Assim, temos que o modelo estrutural possui 6 parâmetros dados por $\alpha, \beta, \mu_x, \phi_e, \phi_u$ e ϕ_x.

Neste trabalho será considerado o modelo de regressão com erros nas variáveis estrutural, ele considera uma variável aleatória bivariada desconhecida (x_i, y_i), que representa os valores reais, sendo $y_i = \alpha + \beta x_i$, .

$$\begin{aligned} Y_i &= \alpha + \beta x_i + e_i \\ X_i &= x_i + u_i, \; i = 1, ..., n, \end{aligned} \tag{3.1}$$

em que α representa o intercepto e β a inclinação, sendo ambos parâmetros desconhecidos. Os erros e_i e u_i são variáveis aleatórias com média 0 e variância ϕ_e e ϕ_u respectivamente e independentemente distribuídos, para $i = 1, ..., n$.

Para a utilização deste modelo é necessário levar em consideração algumas condições de identificabilidade (Fuller, 1987), que são dadas por

(i) A variância ϕ_u ou a variância ϕ_e é conhecida;

(ii) A razão das variâncias $\lambda = \dfrac{\phi_u}{\phi_e}$ é conhecida;

(iii) O coeficiente de atenuação $k_x = \dfrac{\phi_x}{\phi_x + \phi_e}$ é conhecido, em que ϕ_x é a variância de x_i, $i = 1, ..., n$;

(iv) O intercepto α é conhecido.

Qualquer uma das suposições citadas reduz o número de parâmetros tornando o modelo identificável, para este trabalho será considerado como condição de identificabilidade a razão de variâncias conhecida com $\lambda = 1$.

Mas diferente dos modelos com erros nas variáveis mais comuns será utilizado ao invés de uma distribuição Normal, uma distribuição Normal Inversa Gaussiana, que é indicada para casos em que temos indícios que os dados seguem uma distribuição leptocúrtica. Iremos comparar também o caso em que ω é parâmetro desconhecido do modelo e quando ele é pré-fixado.

3.1 Modelo

Nesta parte do trabalho iremos assumir um modelo com erros de medição estrutural baseado na Distribuição Normal Inversa Gaussiana. Antes de especificar o modelo sob a distribuição NIG simétrica, vamos representar o modelo em (3.1) em forma matricial. Seja $\boldsymbol{\epsilon}_i = (e_i, u_i)^\top$ e $\boldsymbol{Z}_i = (Y_i, X_i)^\top$, o modelo pode ser escrito na forma

$$\begin{aligned}\mathbf{Z}_i &= \mathbf{a} + \mathbf{b}x_i + \boldsymbol{\epsilon}_i \\ &= \mathbf{a} + \mathbf{B}\mathbf{r}_i,\end{aligned} \tag{3.2}$$

em que $\mathbf{a} = (\alpha, 0)^\top$, $\mathbf{b} = (\beta, 1)^\top$, $\mathbf{B} = (\mathbf{b}, \mathbf{1}_2)$ e o vetor $\mathbf{r}_i = (x_i, \boldsymbol{\epsilon}_i{}^\top)^\top = (x_i, e_i, u_i)^\top$. Assim, assumimos que $(x_i, e_i, u_i)^\top$ são independentemente distribuídos com distribuição

$$\mathbf{r}_i = \begin{pmatrix} x_i \\ e_i \\ u_i \end{pmatrix} \stackrel{iid}{\sim} NIGS_3 \left[\begin{pmatrix} \mu_x \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \phi_x & 0 & 0 \\ 0 & \phi_e & 0 \\ 0 & 0 & \phi_u \end{pmatrix}, \quad \omega \right], \; i = 1, \ldots, n.$$

Assumindo que $\phi_e = \phi_u = \phi$ e utilizando as propriedades de linearidade, tem-se que a distribuição marginal de $\mathbf{Z}_i = (Y_i, X_i)^\top$ é uma NIGS dada por

$$Z_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim NIGS_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \omega),$$

com fdp dada em (2.36), em que

$$\boldsymbol{\mu} = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \beta^2\phi_x + \phi & \beta\phi_x \\ \beta\phi_x & \phi_x + \phi \end{pmatrix}, \; i = 1, ..., n.$$

As expressões acima podem ser escritas em forma matricial como

$$\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x; \quad \boldsymbol{\Sigma} = \phi_x \mathbf{b}\mathbf{b}^\top + \phi\mathbf{I}_2,$$

em que $\mathbf{a}$ e $\mathbf{b}$ são como em (3.2) e $\mathbf{I}_2$ é a matriz identidade de ordem 2×2.

Seja $\mathbf{z}_1, \ldots, \mathbf{z}_n$ a amostra observada, então a função de log-verossimilhança para os parâmetros α, β, μ_x, ϕ_x, ϕ e ω pode ser escrita como $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$, em que

$$\ell_i(\theta) = -\frac{3}{4}[\log(\omega + d(\mathbf{z}_i)) - \log\omega] + \log K_1(\sqrt{\omega(\omega + d(\mathbf{z}_i))}\,) - \log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \log K_{1/2}(\omega),$$

com $d(\mathbf{z}_i) = (\mathbf{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$. Pode-se ver que a função de log-verossimilhança observada envolve expressões complexas, como por exemplo o termo $K_1(.)$, que pode tornar difícil trabalhar diretamente com a função $\ell(\boldsymbol{\theta})$ para encontra os estimadores de máxima verossimilhança (EMV) dos parâmetros do modelo. Desta maneira, será utilizado um algoritmo EM que será descrito a diante.

3.2 Estimação dos Parâmetros

Como mencionado anteriormente, é trabalhoso encontrar os EMV diretamente da função $\ell(\boldsymbol{\theta})$. Assim, nesta seção será apresentada a estimação dos parâmetros utilizando o algoritmo EM (Dempster et al., 1977) para diferentes situações, como por exemplo, quando comparamos o parâmetro ω fixo ou parte do modelo, como também para alguns casos especias que são utilizados para realizar teste de hipóteses.

3.2.1 Algoritmo EM considerando ω fixo no modelo

A seguir, vamos discutir um procedimento para obtenção dos EMV dos parâmetros do modelo, especificamente vamos utilizar o algoritmo EM, sendo primeiramente considerado o parâmetro ω fixo no modelo. Uma característica importante deste modelo é que ele pode ser formulado como representação hierárquica que é extremamente útil para futuros cálculos. A partir da representação estocástica, dada em (2.26) segue que através da representação hierárquica do modelo sob a distribuição NIGS, temos que

$$\begin{aligned} \mathbf{Z}_i|x_i, U_i = u_i &\sim N_2(\mathbf{a} + \mathbf{b}x_i, u_i\phi\mathbf{I}_2), \\ x_i|U_i = u_i &\sim N_1(\mu_x, u_i\phi_x), \\ U_i &\overset{iid}{\sim} IG(\omega, \omega) \end{aligned} \tag{3.3}$$

para $i = 1, ..., n$ independentes. A partir do teorema de Bayes temos a seguinte propriedade que será importante para o desenvolvimento teórico do algoritmo EM. Desta maneira, temos que

$$\begin{aligned} f_{\mathbf{z}_i,x_i,u_i}(\mathbf{z}_i, x_i, u_i) &= f_{\mathbf{z}_i|x_i,u_i}(\mathbf{z}_i|x_i, u_i) f_{x_i,u_i}(x_i, u_i) \\ &= f_{\mathbf{z}_i|x_i,u_i}(\mathbf{z}_i|x_i, u_i) f_{x_i|u_i}(x_i|u_i) f_{u_i}(u_i). \end{aligned}$$

Seja $\mathbf{z} = (\mathbf{z}_1^\top, ..., \mathbf{z}_n^\top)^\top$, $\mathbf{x} = (x_1, ..., x_n)^\top$, $\mathbf{u} = (u_1, ..., u_n)^\top$ e $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\alpha}}^{(k)}, \hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{\mu}_x}^{(k)}, \hat{\boldsymbol{\phi}_x}^{(k)}, \hat{\boldsymbol{\phi}}^{(k)})^\top$ os estimadores dos parâmetros na k-ésima iteração. Assim de (3.3) temos que a função de log-verossimilhança completa associada a $\mathbf{z}_c = (\mathbf{z}, \mathbf{x}, \mathbf{u})$ é dada por

$$\begin{aligned} \ell_c(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\mathbf{z}_c) &= \log(\prod_{i=1}^{n} f_{z_i,x_i,u_i}(z_i, x_i, u_i)) \\ &= \log(\prod_{i=1}^{n} f_{z_i|x_i,u_i}(z_i|x_i, u_i) f_{x_i|u_i}(x_i|u_i) f_{u_i}(u_i) \\ &= \sum_{i=1}^{n} \log f_{z_i|x_i,u_i}(z_i|x_i, u_i) + \sum_{i=1}^{n} \log f_{x_i|u_i}(x_i|u_i) + \sum_{i=1}^{n} \log f_{u_i}(u_i), \end{aligned} \tag{3.4}$$

em que os termos da função de log-verossimilhança completa são dadas por

$$\begin{aligned} \log(f_{\mathbf{z}_i|x_i,u_i}(z_i|x_i, u_i)) &= -\frac{1}{2u_i\phi}[(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)^\top(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)] \\ &\quad - \log(\phi) - \log(u_i) - \log(2\pi) \\ &= -\frac{1}{2u_i\phi}[(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)^\top(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)] - \log(\phi) + C_1, \end{aligned} \tag{3.5}$$

$$\begin{aligned} \log(f_{x_i|u_i}(x_i|u_i)) &= -\frac{1}{2\phi_x u_i}(x_i - \mu_x)^2 - \frac{1}{2}\log(\phi_x) - \frac{1}{2}\log(u_i) - \frac{1}{2}\log(2\pi) \\ &= -\frac{1}{2\phi_x u_i}(x_i - \mu_x)^2 - \frac{1}{2}\log(\phi_x) + C_2, \end{aligned} \tag{3.6}$$

$$\log(f_{u_i}(u_i)) = \frac{1}{2}\log(\omega) - \frac{1}{2}\log(2\pi) - \frac{3}{2}\log(u_i) + \omega - \frac{1}{2}\omega(u_i + 1/u_i), \tag{3.7}$$

em que C_1 e C_2 são constantes independentes dos parâmetros do modelo. Neste caso, $\log(f_{u_i}(u_i))$ é considerado constante devido ω ser parâmetro pré-fixado do modelo.

Desta maneira, a partir de (3.5)-(3.7) temos que a função de log-verossimilhança completa do modelo é dada por

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{z}_c) =& - n\log(\phi) - \frac{1}{2}\sum_{i=1}^{n}(u_i\phi)^{-1}[(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)^\top(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)] \\ & - \frac{n}{2}\log(\phi_x) - \frac{1}{2\phi_x}\sum_{i=1}^{n} u_i^{-1}(x_i - \mu_x)^2 + C,\end{aligned}$$

com Csendo uma constante independente do vetor dos parâmetros $\boldsymbol{\theta}$. Considere $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}, \hat{\mu_x}^{(k)}, \hat{\phi_x}^{(k)}, \hat{\phi}^{(k)})$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração.

Após manipulações algébricas, a esperança condicional da função de log-verossimilhança completa é dada por

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}).$$

Utilizando propriedades conhecidas da esperança condicional, como

$$E_{\mathbf{x},\mathbf{u}}[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}] = E_{\mathbf{u}}[E_{\mathbf{x}}[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{u}, \mathbf{z}]|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}],$$

entre outras encontradas em (Lachos et al., 2009), podemos obter a esperança condicional da função de log-verossimilhança completa, que será dada por

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(\alpha, \beta, \phi|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}),$$

em que

$$\begin{aligned}Q_{1i}(\alpha, \beta, \phi|\hat{\boldsymbol{\theta}}^{(k)}) =& - \log(\phi) - \frac{1}{2\phi}\hat{q}_i^{(k)}(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\hat{x_i}^{(k)})^\top(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\hat{x_i}^{(k)}) \\ & - \frac{1}{2\phi}\hat{\Lambda}_x^{(k)}\mathbf{b}^\top\mathbf{b}, \\ Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}) =& - \log(\phi_x) - \frac{1}{2\phi_x}(\hat{\Lambda}_x^{(k)} + \hat{q}_i^{(k)}(\hat{x_i}^{(k)} - \hat{\mu_x}^{(k)})^2),\end{aligned}$$

com

$$\begin{aligned}\hat{x}_i^{(k)} &= \mathrm{E}[x_i|u_i, \mathbf{z}, \hat{\boldsymbol{\theta}}^{(k)}] = \hat{\mu_x}^{(k)} + \frac{\hat{\Lambda}_x^{(k)}}{\hat{\phi}^{(k)}}\hat{\mathbf{b}}^{\top(k)}(\mathbf{z}_i - \hat{\mathbf{a}}^{(k)} - \hat{\mathbf{b}}^{(k)}\hat{\mu_x}^{(k)}), \\ \hat{x}_i^{2(k)} &= \mathrm{E}[x_i^2|u_i, \mathbf{z}, \hat{\boldsymbol{\theta}}^{(k)}] = \frac{\hat{\Lambda}_x^{(k)}}{\hat{q}_i^{(k)}} + (\hat{x}_i^{(k)})^2, \\ c^{(k)} &= 1 + \frac{\hat{\phi_x}^{(k)}}{\hat{\phi}^{(k)}}\hat{\mathbf{b}}^{\top(k)}\hat{\mathbf{b}}^{(k)}.\end{aligned} \tag{3.8}$$

Para encontrar Λ_x, através de (3.3) obtemos que

$$f_{\mathbf{z}_i,x_i,u_i}(z_i,x_i,u_i) = f(\mathbf{z}_i)f(u_i|\mathbf{z}_i)f(x_i|u_i,\mathbf{z}_i).$$

Assim, utilizando o Lema 3 em Arellano-Valle et al. (2005a), obtemos de acordo com Lachos et al. (2009), que

$$x_i|u_i,\mathbf{z}_i \sim N_p\left(\mu_x + \Lambda_x\mathbf{b}^\top D^{-1}(\phi)(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\mu_x),\ \frac{\Lambda_x}{q_i}\right),$$

desta maneira, temos que $\hat{\Lambda}_x^{(k)} = \hat{q}_i^{(k)}\mathrm{Var}[x_i|u_i,\mathbf{z},\hat{\boldsymbol{\theta}}^{(k)}] = \hat{\phi_x}^{(k)}/c^{(k)}$.

De (2.24), temos que $U_i|\hat{\boldsymbol{\theta}}^{(k)},\mathbf{Z}_i = \mathbf{z}_i \sim \mathrm{GIG}(\omega,\omega + d(\mathbf{z}_i),\lambda = -3/2)$, em que $d(\mathbf{z}_i) = (\mathbf{z}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$ é a distância de Mahalanobis, sendo assim teremos que

$$\begin{aligned}\hat{q}_i^{(k)} &= \mathrm{E}[U_i^{-1}|\hat{\boldsymbol{\theta}}^{(k)},\mathbf{Z}_i = \mathbf{z}_i] \\ &= \frac{3}{\omega + d(\hat{\mathbf{z}}_i)^{(k)}} + \sqrt{\frac{\omega}{\omega + d(\hat{\mathbf{z}}_i)^{(k)}}}\frac{K_{1/2}(\sqrt{\omega(\omega + d(\hat{\mathbf{z}}_i)^{(k)}}))}{K_{3/2}(\sqrt{\omega(\omega + d(\hat{\mathbf{z}}_i)^{(k)}}))}.\end{aligned}$$

Desta maneira, temos o seguinte algoritmo EM:

Passo E: Dado $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, calcule $\hat{q}_i^{(k)}$, $\hat{x}_i^{(k)}$, $\hat{x}_i^{2(k)}$ para $i = 1,...,n$ a partir de $d(\hat{\mathbf{z}}_i)^{(k)}$

Passo M: Atualize $\hat{\boldsymbol{\theta}}^{(k+1)}$ maximizando $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, o que leva às seguintes expressões fechadas:

$$\begin{aligned}\hat{\alpha}^{(k+1)} &= \frac{\sum_{i=1}^n \hat{q}_i^{(k)}(Y_i - \hat{\beta}^{(k)}\hat{x_i}^{(k)})}{\sum_{i=1}^n \hat{q}_i^{(k)}}, \\ \hat{\beta}^{(k+1)} &= \frac{\sum_{i=1}^n (Y_i - \hat{\alpha}^{(k)})\hat{q}_i^{(k)}\hat{x_i}^{(k)}}{\sum_{i=1}^n (\hat{\Lambda}_x^{(k)} + \hat{q}_i^{(k)}\hat{x}_i^{2(k)})}, \\ \hat{\mu_x}^{(k+1)} &= \frac{\sum_{i=1}^n \hat{q}_i^{(k)}\hat{x_i}^{(k)}}{\sum_{i=1}^n \hat{q}_i^{(k)}}, \\ \hat{\phi_x}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n \hat{\Lambda}_x^{(k)} + \hat{q}_i^{(k)}(\hat{x}_i^{2(k)} - 2\hat{\mu_x}^{(k)}\hat{x_i}^{(k)} + (\hat{\mu_x}^{(k)})^2), \\ \hat{\phi}^{(k+1)} &= \frac{1}{2n}\sum_{i=1}^n \hat{q}_i^{(k)}[X_i^2 - 2X_i\hat{x_i}^{(k)} + \hat{x}_i^{2(k)} + Y_i^2 - \hat{\alpha}^{(k)}Y_i - 2\hat{\beta}^{(k)}Y_i\hat{x_i}^{(k)} \\ &\quad + 2\hat{\alpha}^{(k)}\hat{\beta}^{(k)}\hat{x_i}^{(k)} + (\hat{\alpha}^{(k)})^2 + (\hat{\beta}^{(k)})^2\hat{x}_i^{2(k)}].\end{aligned}$$

Os passos são repetidos até a convergência do algoritmo EM, uma regra de convergência utilizada é $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|$ suficientemente pequeno. O algoritmo EM é um algoritmo relativamente simples e não custoso computacionalmente, sendo uma vantagem de sua utilização.

Valor Inicial: Neste caso, temos que o valor de ω já é previamente fixado, porém para utilizar o algoritmo EM é necessária uma aproximação inicial dos valores dos parâmetros desconhecidos, já que serão necessários no cálculo da esperança condicional do passo E, portanto foram utilizados como valores iniciais os estimadores pelo método dos momentos para razão de variâncias conhecida ($\lambda = 1$), que neste caso, também são estimadores de máxima verossimilhança ajustados pelos graus de liberdade considerando uma distribuição Normal, devido a simetria estes valores já são suficientes. Logo, os valores iniciais escolhidos serão

$$\begin{aligned}
\hat{\mu_x}^{(0)} &= \overline{X}, \\
\hat{\beta}^{(0)} &= \frac{S_{YY} - S_{XX} + \sqrt{(S_{YY} - S_{XX})^2 + 4S_{XY}^2}}{2S_{XY}}, \\
\hat{\alpha}^{(0)} &= \overline{Y} - \hat{\beta}^{(0)}\overline{X}, \\
\hat{\phi_x}^{(0)} &= \frac{S_{YY}}{\hat{\beta}^{(0)}}, \\
\hat{\phi}^{(0)} &= S_{XX} - \hat{\phi_x}^{(0)},
\end{aligned}$$

em que S_{XY}, S_{XX} e S_{YY} são os estimadores de máxima verossimilhança para as variâncias e covariâncias, ajustadas pelo grau de liberdade. Assim, temos que

$$\begin{aligned}
S_{XY} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \\
S_{XX} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\
S_{YY} &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.
\end{aligned}$$

Caso $S_{XY} = 0$ e $S_{YY} < S_{XX}$, então $\hat{\beta}^{(0)} = 0$. Mas, se $S_{XY} = 0$ e $S_Y > S_X$, teremos que $\hat{\beta}^{(0)}$ será indefinido.

3.2.2 Algoritmo EM considerando ω parâmetro

Já nesta parte do trabalho iremos assumir um modelo com erros de medição estrutural com ω desconhecido, sendo este mais complexo, dado que será estimado um parâmetro a mais com relação ao modelo anterior. Temos neste caso que $\boldsymbol{\theta} = (\alpha, \beta, \mu_x, \phi_x, \phi, \omega)$ é o vetor de parâmetros a ser estimado.

A partir de (3.5)-(3.7), temos que neste modelo a função de log-verossimilhança completa será dada por

$$\begin{aligned}\ell_c(\boldsymbol{\theta}|\mathbf{z}_c) =& -n\log(\phi) - \frac{1}{2}\sum_{i=1}^{n}(u_i\phi)^{-1}[(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)^\top(\mathbf{z}_i - \mathbf{a} + \mathbf{b}x_i)] \\ & - \frac{n}{2}\log(\phi_x) - \frac{1}{2\phi_x}\sum_{i=1}^{n} u_i^{-1}(x_i - \mu_x)^2 + \frac{n}{2}\log(\omega) + n\omega \\ & - \frac{1}{2}\omega\sum_{i=1}^{n}(u_i + 1/u_i) + C,\end{aligned} \tag{3.9}$$

com C sendo uma constante independente do vetor dos parâmetros $\boldsymbol{\theta}$. Considere $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}, \hat{\mu_x}^{(k)}, \hat{\phi_x}^{(k)}, \hat{\phi}^{(k)}, \hat{\omega}^{(k)})$ a estimativa de $\boldsymbol{\theta}$ na k-ésima iteração. A esperança condicional da função de log-verossimilhança completa é dada por

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}),$$

e utilizando a propriedade da esperança condicional,

$$E_{\mathbf{x},\mathbf{u}}[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}] = E_{\mathbf{u}}[E_{\mathbf{x}}[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{u}, \mathbf{z}]|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}],$$

temos que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(\alpha, \beta, \phi|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{3i}(\omega|\hat{\boldsymbol{\theta}}^{(k)}), \tag{3.10}$$

em que

$$\begin{aligned}Q_{1i}(\alpha, \beta, \phi|\hat{\boldsymbol{\theta}}^{(k)}) =& -\log(\phi) - \frac{1}{2\phi}\hat{q}_i^{(k)}(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\hat{x_i}^{(k)})^\top(\mathbf{z}_i - \mathbf{a} - \mathbf{b}\hat{x_i}^{(k)}) \\ & - \frac{1}{2\phi}\hat{\Lambda}_x^{(k)}\mathbf{b}^\top\mathbf{b}, \\ Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}) =& -\log(\phi_x) - \frac{1}{2\phi_x}(\hat{\Lambda}_x^{(k)} + \hat{q}_i^{(k)}(\hat{x_i}^{(k)} - \hat{\mu_x}^{(k)})^2), \\ Q_{3i}(\omega|\hat{\boldsymbol{\theta}}^{(k)}) =& -\frac{\omega}{2}(\hat{v}_i^{(k)} + \hat{q}_i^{(k)}) + \frac{1}{2}\log(\omega) + \omega,\end{aligned}$$

com $\hat{\Lambda}_x^{(k)}$, $\hat{x}_i^{(k)}$, $\hat{x}_i^{2(k)}$ e $c^{(k)}$ já definidos em (3.8).

De (2.24), temos que $U_i|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{Z}_i = \mathbf{z}_i \sim \text{GIG}(\omega, \omega + d(\mathbf{z}_i), \lambda = -3/2)$, com $d(\mathbf{z}_i) = (\mathbf{z}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$. Assim, temos que

$$\begin{aligned}\hat{q}_i^{(k)} &= \text{E}[U_i^{-1}|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{Z}_i = \mathbf{z}_i] \\ &= \frac{3}{\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)}} + \sqrt{\frac{\hat{\omega}^{(k)}}{\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)}}}\frac{K_{1/2}(\sqrt{\hat{\omega}^{(k)}(\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)})})}{K_{3/2}(\sqrt{\hat{\omega}^{(k)}(\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)})})} \\ \hat{v}_i^{(k)} &= \text{E}[U_i|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{Z}_i = \mathbf{z}_i] \\ &= \sqrt{\frac{\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)}}{\hat{\omega}^{(k)}}}\frac{K_{1/2}(\sqrt{\hat{\omega}^{(k)}(\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)})})}{K_{3/2}(\sqrt{\hat{\omega}^{(k)}(\hat{\omega}^{(k)} + \widehat{d(\mathbf{z}_i)}^{(k)})})}.\end{aligned}$$

Desta maneira, temos o seguinte algoritmo EM:

Passo E: Dado $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, calcule $\hat{q}_i^{(k)}$, $\hat{x}_i^{(k)}$, $\hat{x}_i^{2(k)}$ e $\hat{v}_i^{(k)}$ para $i = 1, ..., n$ a partir de $d(\hat{\mathbf{z}_i})^{(k)}$.

Passo M: Atualize $\hat{\boldsymbol{\theta}}^{(k+1)}$ maximizando $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$ sobre $\boldsymbol{\theta}$, o que leva às seguintes expressões fechadas:

$$
\begin{aligned}
\hat{\alpha}^{(k+1)} &= \frac{\sum_{i=1}^n \hat{q}_i^{(k)}(Y_i - \hat{\beta}^{(k)}\hat{x_i}^{(k)})}{\sum_{i=1}^n \hat{q}_i^{(k)}}, \\
\hat{\beta}^{(k+1)} &= \frac{\sum_{i=1}^n (Y_i - \hat{\alpha}^{(k)})\hat{q}_i^{(k)}\hat{x_i}^{(k)}}{\sum_{i=1}^n (\hat{\Lambda}_x^{(k)} + \hat{q}_i^{(k)}\hat{x}_i^{2(k)})}, \\
\hat{\mu_x}^{(k+1)} &= \frac{\sum_{i=1}^n \hat{q}_i^{(k)}\hat{x_i}^{(k)}}{\sum_{i=1}^n \hat{q}_i^{(k)}}, \\
\hat{\phi_x}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n \hat{\Lambda}_x^{(k)} + \hat{q}_i^{(k)}(\hat{x}_i^{2(k)} - 2\hat{\mu_x}^{(k)}\hat{x_i}^{(k)} + (\hat{\mu_x}^{(k)})^2), \\
\hat{\phi}^{(k+1)} &= \frac{1}{2n}\sum_{i=1}^n \hat{q}_i^{(k)}[X_i^2 - 2X_i\hat{x_i}^{(k)} + \hat{x}_i^{2(k)} + Y_i^2 - \hat{\alpha}^{(k)}Y_i - 2\hat{\beta}^{(k)}Y_i\hat{x_i}^{(k)} \\
&\quad + 2\hat{\alpha}^{(k)}\hat{\beta}^{(k)}\hat{x_i}^{(k)} + (\hat{\alpha}^{(k)})^2 + (\hat{\beta}^{(k)})^2\hat{x}_i^{2(k)}], \\
\hat{\omega}^{(k+1)} &= \frac{n}{\sum_{i=1}^n \hat{q}_i^{(k)} + \sum_{i=1}^n \hat{v}_i^{(k)} - 2n}.
\end{aligned}
$$

Da mesma maneira que ocorre com o algoritmo anterior, os passos são repetidos até a convergência do algoritmo EM, sendo a regra de convergência utilizada $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|$ suficientemente pequeno.

Valor Inicial: Os valores iniciais, como no caso para ω fixo, são obtidos através do método dos momentos sob distribuição normal. Logo, serão dados por:

$$
\begin{aligned}
\hat{\mu_x}^{(0)} &= \overline{X}, \\
\hat{\beta}^{(0)} &= \frac{S_{YY} - S_{XX} + \sqrt{(S_{YY} - S_{XX})^2 + 4S_{XY}^2}}{2S_{XY}}, \\
\hat{\alpha}^{(0)} &= \overline{Y} - \hat{\beta}^{(0)}\overline{X}, \\
\hat{\phi_x}^{(0)} &= \frac{S_{YY}}{\hat{\beta}^{(0)}}, \\
\hat{\phi}^{(0)} &= S_{XX} - \hat{\phi_x}^{(0)},
\end{aligned}
$$

com S_{XY}, S_{XX} e S_{YY} dados por

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}),$$
$$S_{XX} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$
$$S_{YY} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Já para obter o valor inicial do parâmetro ω, serão utilizados o valores iniciais $\boldsymbol{\theta}^{*(0)} = (\alpha^0, \beta^0, \mu_x^0, \phi_x^0, \phi^0)$ já encontrados anteriormente e assim através função de log-verossimilhança condicional

$$\mathrm{L}(\omega) = \sum_{i=1}^{n}\ln f(z_i|\boldsymbol{\theta}^* = \boldsymbol{\theta}^{*(0)}),$$

maximizado em função de ω, podemos encontrar o valor inicial de ω,

$$\hat{\omega}^0 = \arg\max_{\omega} \mathrm{L}(\omega).$$

3.2.3 Teste de Hipóteses

Discutiremos agora o teste da razão de verossimilhança (TRV) para algumas hipóteses de interesse do modelo de regressão NIG com erro nas variáveis. Serão considerados os seguintes problemas específicos que são de interesse prático:

- $H_{01} : \alpha = 0$ vs $H_{11} : \alpha \neq 0$,
- $H_{02} : \beta = 1$ vs $H_{12} : \beta \neq 1$,
- $H_{03} : \alpha = 0$ e $\beta = 1$ vs $H_{13} : \alpha \neq 0$ e $\beta \neq 1$.

A estatística do TRV é dada por

$$Q_{LR} = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\},$$

em que $\hat{\boldsymbol{\theta}}$ é o EMV de $\boldsymbol{\theta}$ irrestrito e $\tilde{\boldsymbol{\theta}}$ é o EMV de $\boldsymbol{\theta}$ sob H_0, em que sob a hipótese nula, a estatística do teste da razão de verossimilhança segue uma distribuição χ^2 com graus de liberdade igual ao número de restrições da hipótese nula (p). Desta maneira, temos que

$$Q_{LR} \sim \chi^2_{(p)}.$$

Assim, a seguir serão apresentados os algoritmos EM para obter as EMV sob H_0. Para os cálculos será considerado o modelo NIG com ω parâmetro desconhecido, assim temos:

- EMV sob $H_0 : \alpha = 0$ (Intercepto nulo), com p=1

  Quando consideramos o intercepto nulo do modelo, temos que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(0, \beta, \phi|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{3i}(\omega|\hat{\boldsymbol{\theta}}^{(k)}), \qquad (3.11)$$

  assim este sofre alterações no Passo M do algoritmo EM. Neste caso temos então um novo vetor de parâmetros $\boldsymbol{\theta} = (\beta, \mu_x, \phi_x, \phi, \omega)^\top$ e desta maneira maximizando condicionalmente $Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}^{(k)})$ obtemos as novas estimativas dos parâmetros no passo $(k+1)$, através das fórmulas fechadas abaixo:

$$\begin{aligned}
\tilde{\beta}^{(k+1)} &= \frac{\sum_{i=1}^{n} Y_i\ \tilde{q}_i^{(k)} \tilde{x_i}^{(k)}}{\sum_{i=1}^{n} (\tilde{\Lambda}_x^{(k)} + \tilde{q}_i^{(k)} \tilde{x}_i^{2(k)})}, \\
\tilde{\mu_x}^{(k+1)} &= \frac{\sum_{i=1}^{n} \tilde{q}_i^{(k)} \tilde{x_i}^{(k)}}{\sum_{i=1}^{n} \tilde{q}_i^{(k)}}, \\
\tilde{\phi_x}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \tilde{\Lambda}_x^{(k)} + \tilde{q}_i^{(k)} (\tilde{x}_i^{2(k)} - 2\tilde{\mu_x}^{(k)} \tilde{x_i}^{(k)} + (\tilde{\mu_x}^{(k)})^2), \\
\tilde{\phi}^{(k+1)} &= \frac{1}{2n} \sum_{i=1}^{n} \tilde{q}_i^{(k)} [X_i^2 - 2X_i \tilde{x_i}^{(k)} + \tilde{x}_i^{2(k)} + Y_i^2 - 2\tilde{\beta}^{(k)} Y_i \tilde{x_i}^{(k)} + (\tilde{\beta}^{(k)})^2 \hat{x}_i^{2(k)}], \\
\tilde{\omega}^{(k+1)} &= \frac{n}{\sum_{i=1}^{n} \tilde{q}_i^{(k)} + \sum_{i=1}^{n} \hat{v}_i^{(k)} - 2n},
\end{aligned}$$

  com $\tilde{\Lambda}_x^{(k)}$, $\tilde{x}_i^{(k)}$ e $\tilde{x}_i^{2(k)}$ já definidos em (3.8).

  Os valores iniciais também sofrem alteração quando consideramos o intercepto nulo do modelo. Também são obtidos através do método dos momentos considerando uma distribuição normal e serão dados por:

$$\begin{aligned}
\tilde{\mu_x}^{(0)} &= \overline{X}, \\
\tilde{\beta}^{(0)} &= \frac{\overline{Y}}{\overline{X}}, \\
\tilde{\phi_x}^{(0)} &= \frac{S_{XY}\overline{X}}{\overline{Y}}, \\
\tilde{\phi}^{(0)} &= S_{XX} - \frac{S_{XY}\overline{X}}{\overline{Y}}, \\
\tilde{\omega}^{0} &= \arg\max_{\omega} \sum_{i=1}^{n} \ln f(z_i|\boldsymbol{\theta}^* = \boldsymbol{\theta}^{*(0)}),
\end{aligned}$$

  onde $\boldsymbol{\theta}^{*(0)} = (\beta^0, \mu_x^0, \phi_x^0, \phi^0)$.

- EMV sob $H_0 : \beta = 1$, com p=1

  Quando consideramos o parâmetro $\beta = 1$, temos que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(\alpha, 1, \phi|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{3i}(\omega|\hat{\boldsymbol{\theta}}^{(k)}), \qquad (3.12)$$

assim este sofre alterações no Passo M do algoritmo EM. Temos então um novo vetor de parâmetros $\boldsymbol{\theta} = (\alpha, \mu_x, \phi_x, \phi, \omega)^\top$. Assim, maximizando condicionalmente $Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}^{(k)})$ obtemos as novas estimativas dos parâmetros no passo $(k+1)$, através das fórmulas fechadas abaixo:

$$
\begin{aligned}
\tilde{\alpha}^{(k+1)} &= \frac{\sum_{i=1}^{n} \tilde{q}_i^{(k)}(Y_i - \tilde{x_i}^{(k)})}{\sum_{i=1}^{n} \tilde{q}_i^{(k)}}, \\
\tilde{\mu_x}^{(k+1)} &= \frac{\sum_{i=1}^{n} \tilde{q}_i^{(k)}\tilde{x_i}^{(k)}}{\sum_{i=1}^{n} \tilde{q}_i^{(k)}}, \\
\tilde{\phi_x}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n} \tilde{\Lambda}_x^{(k)} + \tilde{q}_i^{(k)}(\tilde{x}_i^{2(k)} - 2\tilde{\mu_x}^{(k)}\tilde{x_i}^{(k)} + (\tilde{\mu_x}^{(k)})^2), \\
\tilde{\phi}^{(k+1)} &= \frac{1}{2n}\sum_{i=1}^{n} \tilde{q}_i^{(k)}[X_i^2 - 2X_i\tilde{x_i}^{(k)} + \tilde{x}_i^{2(k)} + Y_i^2 - \tilde{\alpha}^{(k)}Y_i - 2Y_i\tilde{x_i}^{(k)} \\
&\quad + 2\tilde{\alpha}^{(k)}\tilde{x_i}^{(k)} + (\tilde{\alpha}^{(k)})^2 + \hat{x}_i^{2(k)}], \\
\tilde{\omega}^{(k+1)} &= \frac{n}{(\sum_{i=1}^{n} \tilde{q}_i^{(k)} + \sum_{i=1}^{n} \tilde{v}_i^{(k)} - 2n)}.
\end{aligned}
$$

Os valores iniciais também sofrem alteração quando consideramos $\beta = 1$ no modelo. Também são obtidos através do método dos momentos considerando uma distribuição normal e serão dados por:

$$
\begin{aligned}
\tilde{\mu_x}^{(0)} &= \overline{X}, \\
\tilde{\alpha}^{(0)} &= \overline{Y} - \overline{X}, \\
\tilde{\phi_x}^{(0)} &= S_{XY}, \\
\tilde{\phi}^{(0)} &= S_{XX} - S_{XY}, \\
\tilde{\omega}^{0} &= \arg\max_{\omega} \sum_{i=1}^{n} \ln f(z_i|\boldsymbol{\theta}^* = \boldsymbol{\theta}^{*(0)}),
\end{aligned}
$$

onde $\boldsymbol{\theta}^{*(0)} = (\alpha^0, \mu_x^0, \phi_x^0, \phi^0)$.

- EMV sob $H_0 : \alpha = 0$ e $\beta = 1$, com p=2

  Quando consideramos $\alpha = 0$ e $\beta = 1$ no modelo, temos que

$$
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_{1i}(0, 1, \phi|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\mu_x, \phi_x|\hat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{3i}(\omega|\hat{\boldsymbol{\theta}}^{(k)}), \qquad (3.13)
$$

  assim este sofre alterações no Passo M do algoritmo EM. Temos então um novo vetor de parâmetros desconhecidos $\boldsymbol{\theta} = (\mu_x, \phi_x, \phi, \omega)^\top$, e desta maneira, maximizando condicionalmente $Q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}^{(k)})$ obtemos as novas estimativas dos parâmetros no passo

$(k+1)$, através das fórmulas abaixo:

$$\begin{aligned}
\tilde{\mu_x}^{(k+1)} &= \frac{\sum_{i=1}^{n} \tilde{q}_i^{(k)} \hat{x_i}^{(k)}}{\sum_{i=1}^{n} \tilde{q}_i^{(k)}}, \\
\tilde{\phi_x}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \tilde{\Lambda}_x^{(k)} + \tilde{q}_i^{(k)} (\tilde{x}_i^{2(k)} - 2\tilde{\mu_x}^{(k)} \tilde{x_i}^{(k)} + (\tilde{\mu_x}^{(k)})^2), \\
\tilde{\phi}^{(k+1)} &= \frac{1}{2n} \sum_{i=1}^{n} \tilde{q}_i^{(k)} [X_i^2 - 2X_i \tilde{x_i}^{(k)} + \tilde{x}_i^{2(k)} + Y_i^2 - 2Y_i \tilde{x_i}^{(k)} + 2\tilde{x_i}^{(k)} + \hat{x}_i^{2(k)}], \\
\tilde{\omega}^{(k+1)} &= \frac{n}{(\sum_{i=1}^{n} \tilde{q}_i^{(k)} + \sum_{i=1}^{n} \tilde{v}_i^{(k)} - 2n)}.
\end{aligned}$$

3.3 Estudos de Simulação

Para analisar a performance dos algoritmos EM propostos na Seção 3.2, foram realizados estudos de simulação para investigar o desempenho do método de estimação dos parâmetros do modelo estudado. Foram considerados diferentes tamanhos amostrais, $n = 50, 100, 200, 500$ e 1000, tanto para o modelo que considera o parâmetro ω pré-fixado, quanto para quando consideramos ω parâmetro desconhecido do modelo, em que são observados 3 casos diferentes, $\omega = 0.5, 1$ e 5, sendo portanto, um caso em que $\omega < 1$, outro em que $\omega = 1$ e outro $\omega > 1$. Para cada caso foram utilizadas 1000 replicas, que foram obtidos através da representação estocástica dada em (2.26).

Para este estudo, serão analisados dois cenários distintos:

- **Cenário 1**: $\alpha = 0$, $\beta = 1$, $\mu_x = 3$, $\phi_x = 8$, $\phi = 2$ e $\omega = 0.5, 1$ e 5,
- **Cenário 2**: $\alpha = 2$, $\beta = 5$, $\mu_x = 3$, $\phi_x = 8$, $\phi = 2$ e $\omega = 0.5, 1$ e 5,

em que serão analisados, para cada tamanho amostral, a estimativa $\hat{\theta}_k$, a média dada por $E[\hat{\theta}_k]$, o viés amostral dado por $\text{Viés}(\hat{\theta}_k) = \dfrac{1}{N}\sum_{j}^{N}(\hat{\theta}_{kj} - \theta_k)$ e o erro quadrático médio amostral (EQM) definido como $\text{EQM}_k = (E[\hat{\theta}_k - \theta_k]^2)$, para $k = 1, ..., 6$. Os valores dos parâmetros foram escolhidos ao acaso.

3.3.1 Cenário 1

Primeiramente, através do Cenário 1 será estudado por meio de simulações a eficácia do modelo para o caso, em que $\alpha = 0$ e $\beta = 1$, para os casos em que ω é pré-fixado e também quando é parâmetro do modelo. Os valores dos outros parâmetros serão dados por $\mu_x = 3$, $\phi_x = 8$, $\phi = 2$ e $\omega = 0.5, 1$ e 5.

Cenário 1 considerando ω pré-fixado

Tabela 1 – Medidas resumo para as estimativas dos parâmetros α, β e μ_x considerando ω pré-fixado para o Cenário 1.

ω	n	estimativa de α			estimativa de β			estimativa de μ_x		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
0.5	50	-0,0630	0,0630	0,3787	1,0213	0,0213	0,0245	3,0098	0,0098	0,1199
	100	-0,0190	0,0190	0,1698	1,0071	0,0071	0,0101	3,0066	0,0066	0,0524
	200	-0,0102	0,0102	0,0751	1,0065	0,0065	0,0063	3,0058	0,0058	0,0218
	500	-0,0086	0,0086	0,0268	1,0026	0,0026	0,0021	3,0025	0,0025	0,0102
	1000	0,0060	0,0060	0,0102	0,9990	0,0010	0,0010	3,0009	0,0009	0,0053
1	50	-0,0466	0,0466	0,2531	1,0154	0,0154	0,0204	3,0105	0,0105	0,1401
	100	-0,0164	0,0164	0,1085	1,0066	0,0066	0,0093	3,0070	0,0070	0,0669
	200	-0,0046	0,0046	0,0458	1,0076	0,0076	0,0043	2,9970	0,0030	0,0273
	500	-0,0034	0,0034	0,0174	0,9993	0,0007	0,0011	2,9974	0,0026	0,0100
	1000	-0,0030	0,0030	0,0103	1,0006	0,0006	0,0008	3,0014	0,0014	0,0067
5	50	-0,0247	0,0247	0,2052	1,0106	0,0106	0,0130	3,0121	0,0121	0,1859
	100	-0,0156	0,0156	0,0887	1,0069	0,0069	0,0062	2,9886	0,0114	0,0834
	200	-0,0098	0,0098	0,0390	1,0022	0,0022	0,0024	3,0032	0,0032	0,0372
	500	-0,0023	0,0023	0,0105	1,0028	0,0028	0,0009	3,0006	0,0006	0,0134
	1000	-0,0009	0,0009	0,0094	1,0002	0,0002	0,0005	2,9998	0,0002	0,0083

Aqui através das Tabelas 1 e 2 temos os resultados das simulações quando consideramos o modelo com ω pré-fixado e podemos perceber que com o aumento do parâmetro ω temos uma diminuição dos valores do EQM, com exceção do parâmetro μ_x que aumenta seu EQM com o aumento do parâmetro. Já foi mostrado anteriormente na Figura 18 que quanto maior for o valor de ω, mais a distribuição NIG converge para uma distribuição normal, assim temos que o EQM diminui quanto mais próxima a distribuição for da normal, mas quanto mais leptocúrtica for a distribuição menor o EQM para μ_x, o que era esperado já que os dados estão mais concentrados em torno da média. Podemos facilmente observar que com o aumento do número amostral o valor do EQM também diminui, também podemos notar que quando n cresce, os EQMs diminuem, como já era esperado.

Nas Figuras 19-21 podemos ver a convergência dos estimadores dos parâmetros modelo proposto com o aumento da amostra e também perceber que o mesmo tem uma boa convergência dos estimadores para valores amostrais altos, mas quando utilizados valores amostrais muito pequenos ele já não é tão eficaz para estimar os parâmetros do modelo.

Um fato importante que ainda pode ser notado através das estimativas e dos gráficos é que o estimador do parâmetro ϕ_x tem uma convergência mais lenta que os demais parâmetros do modelo, sendo então necessário grandes valores amostrais ou então valores de ω maiores para sua melhor convergência do estimador, em quanto estimadores dos outros parâmetros aceitam tamanhos amostrais menores.

Cenário 1 considerando ω parâmetro desconhecido

Tabela 2 – Medidas resumo para as estimativas dos parâmetros ϕ_x e ϕ considerando ω pré-fixado para o Cenário 1.

ω	n	estimativa de ϕ_x			estimativa de ϕ		
		Média	Viés	EQM	Média	Viés	EQM
0.5	50	7,9192	0,0808	9,6342	1,8783	0,1217	0,3390
	100	7,9279	0,0721	5,0031	1,9502	0,0498	0,1396
	200	7,9502	0,0498	2,5304	1,9751	0,0249	0,1094
	500	7,9515	0,0485	0,9343	1,9905	0,0095	0,0386
	1000	8,0034	0,0034	0,4355	1,9953	0,0047	0,0136
1	50	7,9524	0,0476	7,2374	1,9228	0,0771	0,2559
	100	7,9734	0,0266	3,1616	1,9533	0,0467	0,1495
	200	7,9927	0,0073	1,6153	1,9699	0,0301	0,0671
	500	8,0052	0,0052	0,6283	1,9950	0,0050	0,0224
	1000	7,9957	0,0042	0,2566	2,0029	0,0029	0,0117
5	50	7,9529	0,0470	4,8439	1,9330	0,0670	0,1031
	100	7,9549	0,0450	2,3326	1,9551	0,0449	0,0878
	200	8,0233	0,0233	1,1141	1,9773	0,0227	0,0395
	500	7,9989	0,0011	0,3874	1,9935	0,0065	0,0101
	1000	8,0009	0,0009	0,1564	1,9989	0,0010	0,0089

Aqui através das Tabelas 3 e 4, temos os resultados das simulações quando consideramos o modelo com ω parâmetro desconhecido e podemos perceber da mesma maneira do caso anterior, em que ω é pré-fixado, que com o aumento do parâmetro ω temos uma diminuição dos valores do EQM, mas neste caso com exceção dos parâmetros μ_x e ω que aumentam seu EQM com o aumento do parâmetro. Também podemos facilmente observar que com o aumento do número amostral o valor do EQM também diminui.

Nas Figuras 22-24 podemos ver a convergência dos estimadores dos parâmetros no modelo proposto com o aumento da amostra e também perceber que temos uma melhor estimação dos parâmetros para valores amostrais altos, como ocorre no caso anterior utilizar valores amostrais pequenos torna o modelo menos eficaz para estimar os seus parâmetros.

Também podemos ver novamente que o estimador do parâmetro ϕ_x tem uma convergência mais lenta que os demais estimadores dos parâmetros do modelo, sendo então necessário grandes valores amostrais ou então valores de ω maiores para sua melhor convergência.

3.3.2 Cenário 2

Já através do Cenário 2 será estudado por meio de simulações a eficácia do método de estimação dos parâmetros do modelo para o caso, em que $\alpha \neq 0$ e $\beta \neq 1$, sendo escolhidos os valores de $\alpha = 2$ e $\beta = 5$, para os casos em que ω é pré-fixado e também quando é parâmetro desconhecido do modelo. Os outros parâmetros foram escolhidos como $\mu_x = 3$, $\phi_x = 8$, $\phi = 2$ e $\omega = 0.5, 1$ e 5, da mesma maneira que ocorreu no Cenário 1.

Tabela 3 – Medidas resumo para as estimativas dos parâmetros α, β e μ_x considerando ω parâmetro desconhecido do modelo para o Cenário 1.

ω	n	estimativa de α			estimativa de β			estimativa de μ_x		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
0.5	50	-0,0607	0,0607	0,3435	1,0155	0,0155	0,0329	2,9894	0,0105	0,1177
	100	-0,0186	0,0186	0,1709	1,0070	0,0070	0,0148	2,9918	0,0082	0,0617
	200	-0,0099	0,0099	0,0814	1,0060	0,0060	0,0073	2,9983	0,0017	0,0276
	500	-0,0079	0,0079	0,0327	1,0023	0,0023	0,0029	3,0012	0,0012	0,0120
	1000	-0,0059	0,0059	0,0176	1,0009	0,0009	0,0015	2,9987	0,0012	0,0060
1	50	-0,0455	0,0455	0,2688	1,0154	0,0154	0,0209	3,0116	0,0116	0,1496
	100	-0,0092	0,0092	0,1271	1,0065	0,0065	0,0101	3,0016	0,0016	0,0754
	200	-0,0091	0,0091	0,0695	1,0011	0,0011	0,0051	3,0013	0,0013	0,0371
	500	-0,0051	0,0051	0,0266	1,0006	0,0006	0,0019	3,0009	0,0009	0,0150
	1000	0,0027	0,0027	0,0132	1,0002	0,0002	0,0010	3,0008	0,0008	0,0076
5	50	-0,0228	0,0228	0,2092	1,0071	0,0071	0,0139	3,0119	0,0119	0,1916
	100	-0,0151	0,0151	0,0972	1,0063	0,0063	0,0068	3,0097	0,0097	0,0934
	200	-0,0066	0,0066	0,0502	1,0021	0,0021	0,0034	2,9974	0,0026	0,0459
	500	-0,0018	0,0018	0,0200	1,0002	0,0002	0,0013	3,0022	0,0022	0,0190
	1000	-0,0001	0,0001	0,0095	1,0001	0,0001	0,0006	2,9998	0,0002	0,0091

Tabela 4 – Medidas resumo para as estimativas dos parâmetros ϕ_x, ϕ e ω considerando ω parâmetro desconhecido do modelo para o Cenário 1.

ω	n	estimativa de ϕ_x			estimativa de ϕ			estimativa de ω		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
0.5	50	8,0177	0,0177	9,1118	1,8995	0,1004	0,4299	0,5418	0,0418	0,0056
	100	8,0185	0,0185	5,0009	1,9510	0,0489	0,2280	0,5198	0,0198	0,0021
	200	7,9863	0,0136	2,4284	1,9768	0,0231	0,1160	0,5093	0,0093	0,0009
	500	7,9972	0,0027	1,0240	1,9872	0,0127	0,0484	0,5043	0,0043	0,0003
	1000	7,9940	0,0059	0,5013	1,9961	0,0039	0,0233	0,5023	0,0023	0,0001
1	50	8,0102	0,0102	7,2984	1,9275	0,0975	0,2942	1,0455	0,0455	0,0094
	100	7,9801	0,0199	3,1800	1,9557	0,0442	0,1581	1,0244	0,0244	0,0041
	200	8,0060	0,0060	1,6348	1,9794	0,0205	0,0733	1,0126	0,0126	0,0019
	500	7,9879	0,0120	0,7101	1,9951	0,0048	0,0322	1,0052	0,0052	0,0007
	1000	8,0052	0,0052	0,3305	1,9983	0,0016	0,0165	1,0018	0,0018	0,0003
5	50	7,9957	0,0042	4,8727	1,9314	0,0685	0,1895	5,0634	0,0634	0,0277
	100	7,9864	0,0136	2,3164	1,9583	0,0416	0,0967	5,0311	0,0311	0,0135
	200	7,9866	0,0133	1,1983	1,9810	0,0189	0,0470	5,0141	0,0141	0,0067
	500	7,9990	0,0009	0,4963	1,9888	0,0112	0,0198	5,0071	0,0071	0,0025
	1000	8,0009	0,0009	0,2600	2,0009	0,0009	0,0098	5,0021	0,0021	0,0012

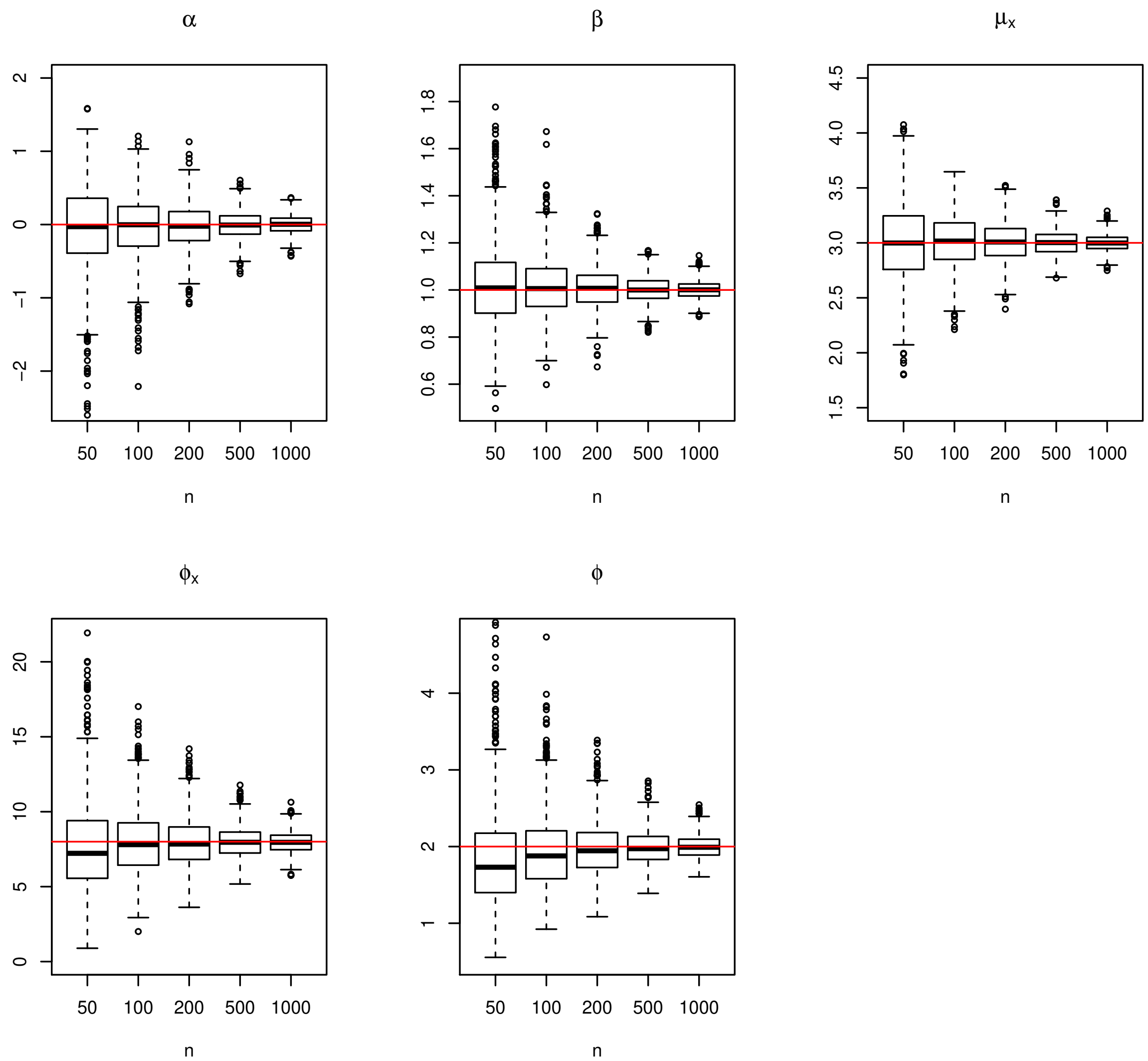


Figura 19 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x e ϕ quando consideramos $\omega = 0.5$ pré-fixado no modelo de regressão NIG com erro nas variáveis para o Cenário 1.

Cenário 2 considerando ω pré-fixado

Aqui através das Tabelas 5 e 6 temos os resultados das simulações quando consideramos o modelo com ω pré-fixado. Podemos perceber como também ocorreu no Cenário 1, que com o aumento do valor parâmetro ω temos uma diminuição dos valores do EQM, com exceção novamente do parâmetro μ_x que aumenta seu EQM com o aumento do parâmetro.

Nas Figuras 25-27 podemos ver a convergência dos EMV quando aumentamos o tamanho da amostra e também perceber que o mesmooferece boas estimativas para valores amostrais altos, mas novamente quando utilizados tamanhos amostrais pequenos ele já não é tão eficaz para estimar os parâmetros do modelo.

De modo semelhante que ocorreu no Cenário 1, pode ser notado através das

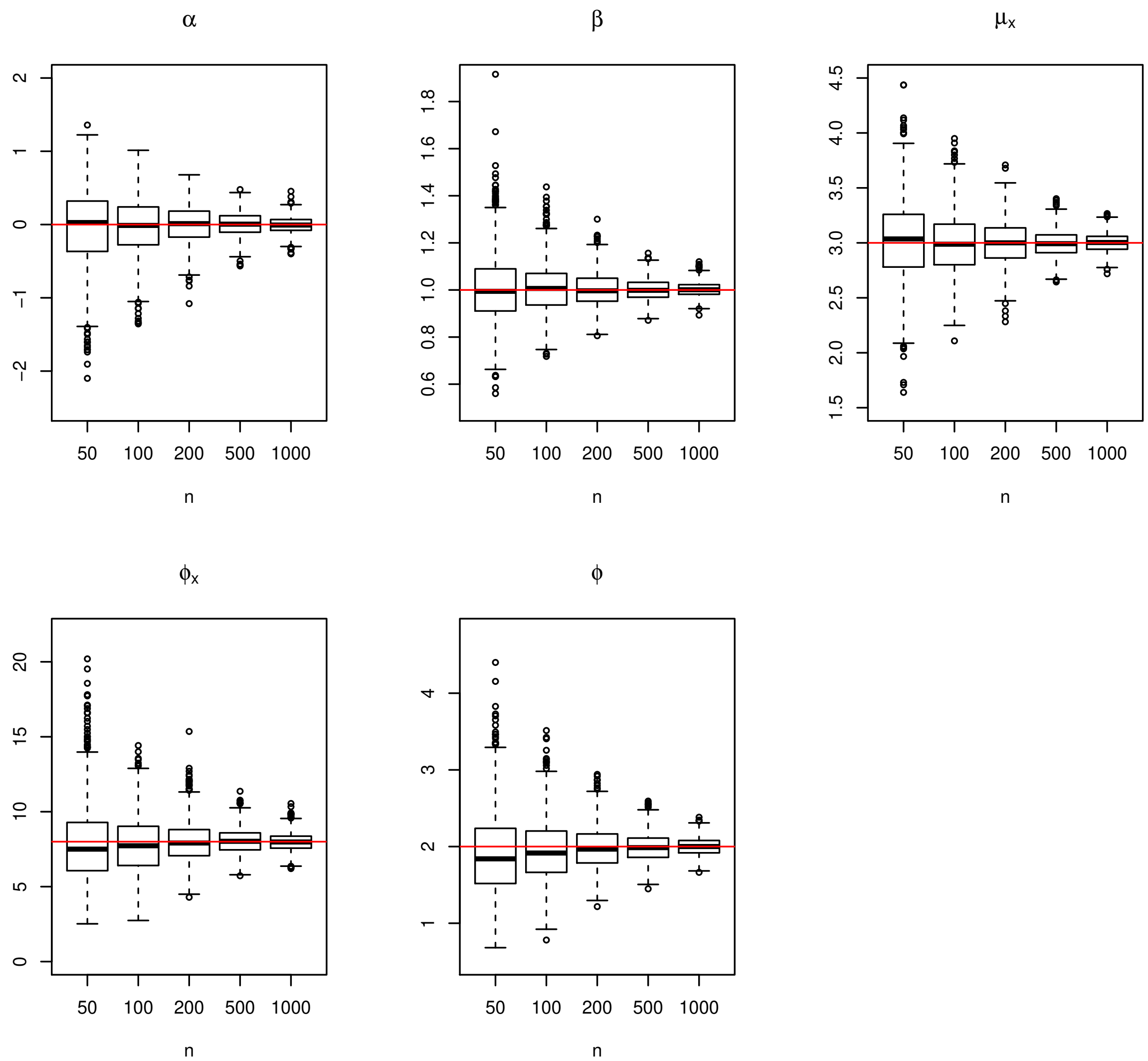


Figura 20 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x e ϕ quando consideramos $\omega = 1$ pré-fixado no modelo de regressão NIG com erro nas variáveis para o Cenário 1.

estimativas e dos gráficos que o estimador do parâmetro ϕ_x tem uma convergência mais lenta que os demais estimadores do modelo, sendo portanto necessário também a utilização de grandes amostras ou então valores de ω maiores para sua melhor convergência.

Quando comparado com o Cenário 1, temos que de maneira geral que o Cenário 2 apresenta na maioria dos casos valores maiores de EQM. Existindo então uma melhor convergência dos estimadores quando temos $\alpha = 0$ e $\beta = 1$.

Cenário 2 considerando ω parâmetro desconhecido

Aqui através das Tabelas 7 e 8 temos os resultados das simulações quando consideramos o modelo com ω parâmetro desconhecido. Podemos perceber da mesma maneira do caso do Cenário 1 temos que com o aumento do parâmetro ω, temos uma diminuição dos valores do EQM, mas com exceção dos parâmetros μ_x e ω que aumentam

Tabela 5 – Medidas resumo para as estimativas dos parâmetros α, β e μ_x considerando ω pré-fixado para o Cenário 2.

ω	n	estimativa de α			estimativa de β			estimativa de μ_x		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
0.5	50	1,7533	0,2467	4,7890	5,0801	0,0801	0,4185	3,0057	0,0057	0,1205
	100	1,8528	0,1472	2,2288	5,0438	0,0438	0,2000	3,0052	0,0052	0,0561
	200	1,9023	0,0977	1,1001	4,9800	0,0200	0,0836	3,0040	0,0040	0,0240
	500	1,9732	0,0268	0,3586	5,0091	0,0091	0,0293	2,9961	0,0038	0,0104
	1000	2,0111	0,0111	0,2102	4,9975	0,0025	0,0159	2,9975	0,0025	0,0056
1	50	1,8862	0,1138	3,3894	5,0386	0,0386	0,2664	3,0052	0,0052	0,1438
	100	1,9480	0,0520	1,6217	5,0270	0,0270	0,1290	3,0058	0,0058	0,0781
	200	1,9491	0,0509	0,7903	5,0110	0,0110	0,0581	3,0064	0,0064	0,0304
	500	1,9838	0,0162	0,2384	5,0076	0,0076	0,0262	2,9951	0,0048	0,0144
	1000	2,0058	0,0058	0,1120	5,0017	0,0017	0,0122	2,9956	0,0044	0,0068
5	50	1,9319	0,0681	2,3864	5,0316	0,0316	0,1473	2,9912	0,0088	0,1907
	100	1,9458	0,0542	1,2201	5,0205	0,0205	0,0756	2,9964	0,0036	0,0856
	200	1,9763	0,0237	0,5065	5,0103	0,0103	0,0295	2,9965	0,0035	0,0390
	500	1,9894	0,0106	0,1402	5,0044	0,0044	0,0118	2,9966	0,0034	0,0160
	1000	2,0055	0,0055	0,1047	4,9988	0,0012	0,0068	2,9986	0,0014	0,0085

Tabela 6 – Medidas resumo para as estimativas dos parâmetros ϕ_x e ϕ considerando ω pré-fixado para o Cenário 2.

ω	n	estimativa de ϕ_x			estimativa de ϕ		
		Média	Viés	EQM	Média	Viés	EQM
0.5	50	7,9089	0,0911	9,1826	1,8841	0,1159	0,4313
	100	7,9471	0,0529	5,0313	1,9335	0,0665	0,2211
	200	8,0489	0,0489	2,5476	1,9648	0,0352	0,1146
	500	8,0186	0,0186	1,0755	1,9889	0,0111	0,0474
	1000	7,9944	0,0055	0,5282	1,9922	0,0078	0,0240
1	50	8,1031	0,1031	7,03235	1,9053	0,0947	0,3212
	100	7,9424	0,0576	3,7248	1,9459	0,0541	0,1675
	200	7,9809	0,0190	1,8031	1,9719	0,0281	0,0879
	500	8,0286	0,0286	0,7536	1,9917	0,0083	0,0333
	1000	8,0080	0,0080	0,3738	1,9917	0,0083	0,0152
5	50	7,9008	0,0992	4,6263	1,9224	0,0776	0,1974
	100	7,9558	0,0442	2,2282	1,9665	0,0335	0,1000
	200	7,9862	0,0138	1,1212	1,9757	0,0243	0,0402
	500	8,0179	0,0179	0,4700	1,9891	0,0109	0,0186
	1000	8,0072	0,0072	0,2293	1,9945	0,0055	0,0100

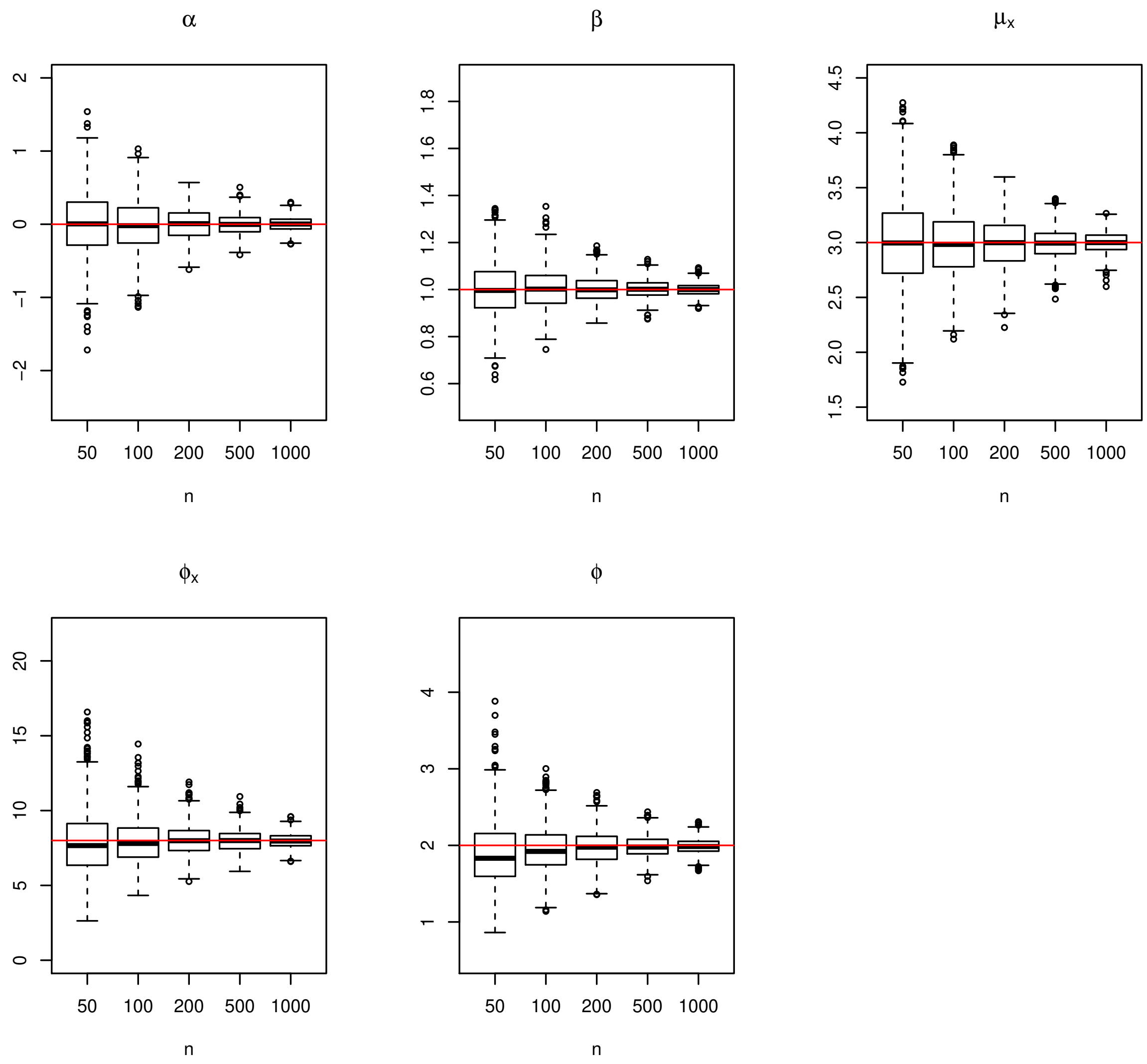


Figura 21 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x e ϕ quando consideramos $\omega = 5$ pré-fixado no modelo de regressão NIG com erro nas variáveis para o Cenário 1.

seu EQM com o aumento do valor do parâmetro.

Pelas Figuras 28-30, podemos ver a convergência dos estimadores e com o aumento da amostra e também perceber que o mesmo oferece boas estimativas para valores amostrais altos, como ocorre nos casos anteriores utilizar valores pequenos de n torna o modelo menos eficaz para estimar os parâmetros do modelo.

Temos neste caso que os estimadores dos parâmetros α e ϕ_x tem uma convergência mais lenta que os demais parâmetros do modelo, sendo então necessário maiores valores de amostra para uma estimação eficiente.

Novamente quando comparamos o Cenário 1 com o Cenário 2, temos que de maneira geral o Cenário 1 parece apresentar na maioria dos casos uma melhor convergência na estimação dos parâmetros, pois apresenta na maior parte dos casos valores menores de

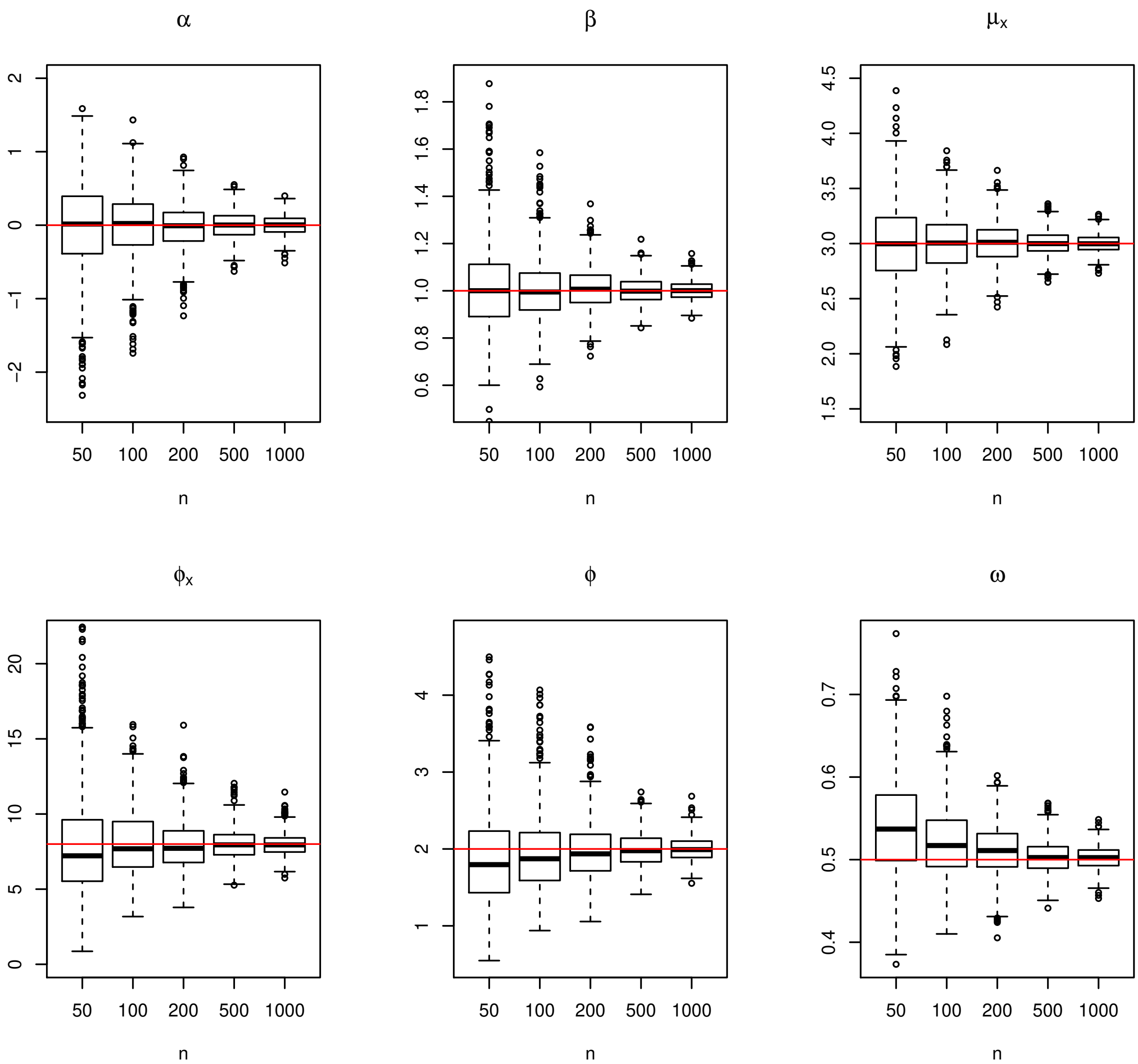


Figura 22 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x, ϕ e ω, quando consideramos $\omega = 0.5$ parâmetro desconhecido do modelo de regressão NIG com erro nas variáveis para o Cenário 1.

EQM.

Tabela 7 – Medidas resumo para as estimativas dos parâmetros α, β e μ_x considerando ω parâmetro desconhecido do modelo para o Cenário 2.

ω	n	estimativa de α			estimativa de β			estimativa de μ_x		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
0.5	50	1,8095	0,1905	4,7469	5,0661	0,0661	0,4118	3,0051	0,0051	0,1303
	100	1,9060	0,0940	2,0963	5,0287	0,0287	0,1834	3,0048	0,0048	0,0653
	200	1,9515	0,0485	1,0090	5,0177	0,0177	0,0919	3,0028	0,0028	0,0315
	500	1,9769	0,0231	0,4308	5,0082	0,0082	0,0372	2,9984	0,0015	0,0122
	1000	1,9836	0,0163	0,2144	5,0017	0,0017	0,0163	3,0013	0,0013	0,0063
1	50	1,9121	0,0878	3,4059	5,0398	0,0398	0,2663	2,9952	0,0048	0,1519
	100	1,9741	0,0258	1,7209	5,0136	0,0136	0,1316	2,9950	0,0040	0,0777
	200	1,9833	0,0167	0,8674	5,0104	0,0104	0,0671	2,9951	0,0048	0,0306
	500	1,9922	0,0078	0,3288	5,0045	0,0045	0,0246	2,9980	0,0020	0,0149
	1000	2,0097	0,0097	0,1716	5,0018	0,0018	0,0125	2,9985	0,0015	0,0060
5	50	1,9190	0,0810	2,4240	5,0316	0,0316	0,1567	2,9973	0,0027	0,1893
	100	1,9520	0,0479	1,2614	5,0150	0,0150	0,0849	3,0026	0,0026	0,0954
	200	1,9696	0,0304	0,6042	5,0100	0,0100	0,0382	2,9964	0,0036	0,0484
	500	1,9892	0,0108	0,2329	5,0049	0,0049	0,0145	3,0032	0,0032	0,0190
	1000	2,0033	0,0033	0,1187	4,9994	0,0006	0,0074	2,9990	0,0010	0,0095

Tabela 8 – Medidas resumo para as estimativas dos parâmetros ϕ_x, ϕ e ω considerando ω parâmetro desconhecido do modelo para o Cenário 2.

ω	n	estimativa de ϕ_x			estimativa de ϕ			estimativa de ω		
		Média	Viés	EQM	Média	Viés	EQM	Média	Viés	EQM
0.5	50	7,9033	0,0967	9,1804	1,8971	0,1029	0,4301	0,5406	0,0406	0,0054
	100	7,9072	0,0928	5,0317	1,9357	0,0642	0,2211	0,5216	0,0216	0,0023
	200	8,0393	0,0393	2,3428	1,9765	0,0235	0,1139	0,5105	0,0105	0,0010
	500	8,0165	0,0165	0,9978	1,9894	0,0106	0,0433	0,5043	0,0043	0,0004
	1000	8,0057	0,0057	0,5358	1,9930	0,0069	0,0240	0,5018	0,0018	0,0001
1	50	8,0163	0,0163	7,0057	1,8819	0,1181	0,3172	1,0456	0,0456	0,0092
	100	7,9744	0,0256	3,4459	1,9445	0,0554	0,1614	1,0250	0,0250	0,0043
	200	8,0075	0,0075	1,7559	1,9778	0,0221	0,0803	1,0116	0,0116	0,0018
	500	8,0060	0,0060	0,7536	1,9994	0,0006	0,0333	1,0045	0,0045	0,0007
	1000	7,9974	0,0026	0,3506	1,9994	0,0006	0,0147	1,0023	0,0023	0,0003
5	50	7,9055	0,0945	4,2677	1,9123	0,0877	0,1961	5,0650	0,0650	0,0273
	100	7,9524	0,0476	2,3229	1,9635	0,0365	0,0939	5,0304	0,0304	0,0130
	200	7,9581	0,0419	1,1233	1,9768	0,0232	0,0501	5,0195	0,0195	0,0071
	500	8,0143	0,0143	0,4470	1,9947	0,0053	0,0195	5,0074	0,0074	0,0026
	1000	7,9912	0,0087	0,2332	1,9951	0,0049	0,0096	5,0039	0,0039	0,0013

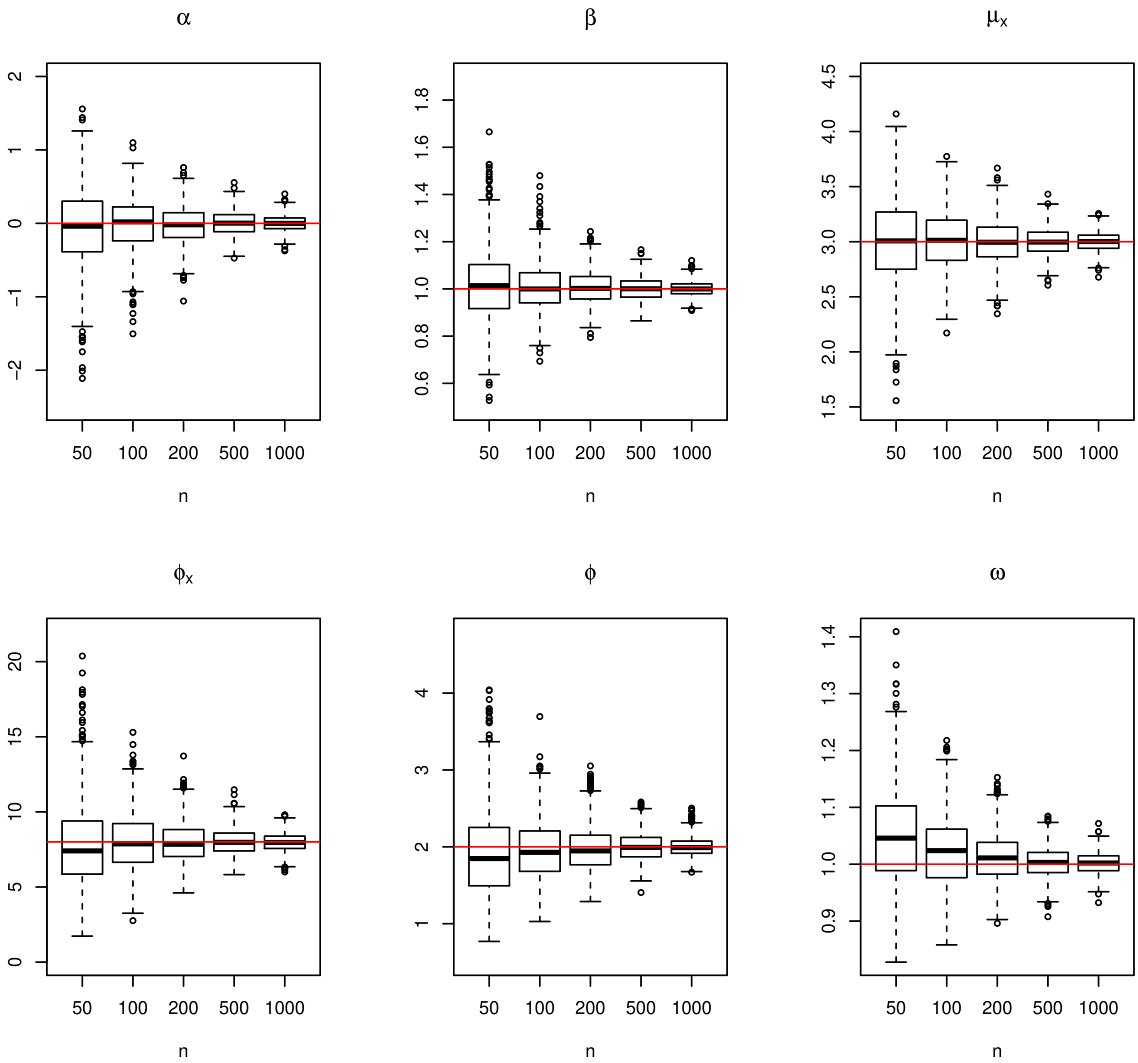


Figura 23 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x, ϕ e ω, quando consideramos $\omega = 1$ parâmetro desconhecido do modelo de regressão NIG com erro nas variáveis para o Cenário 1.

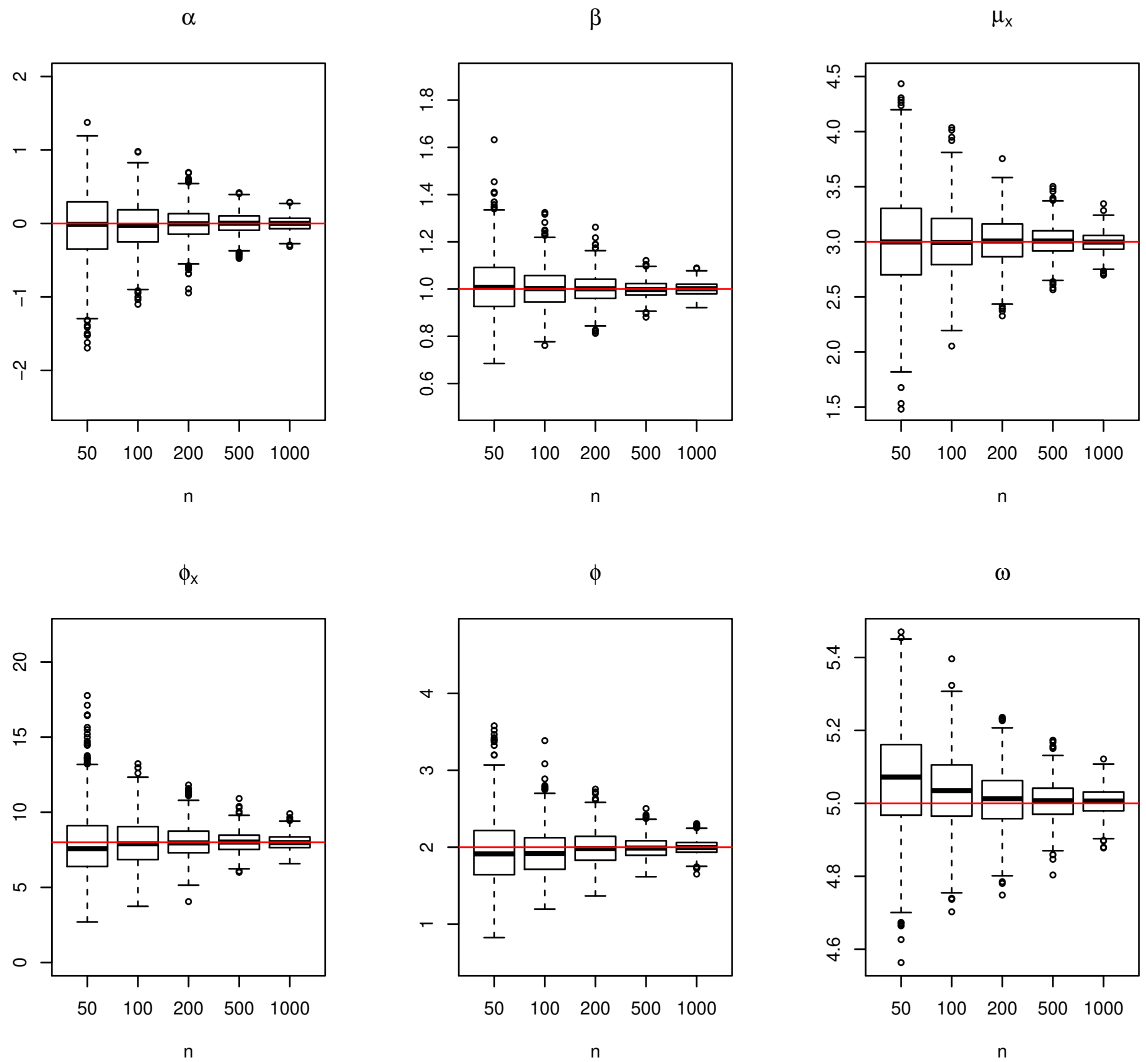


Figura 24 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x, ϕ e ω, quando consideramos $\omega = 5$ parâmetro desconhecido do modelo de regressão NIG com erro nas variáveis para o Cenário 1.

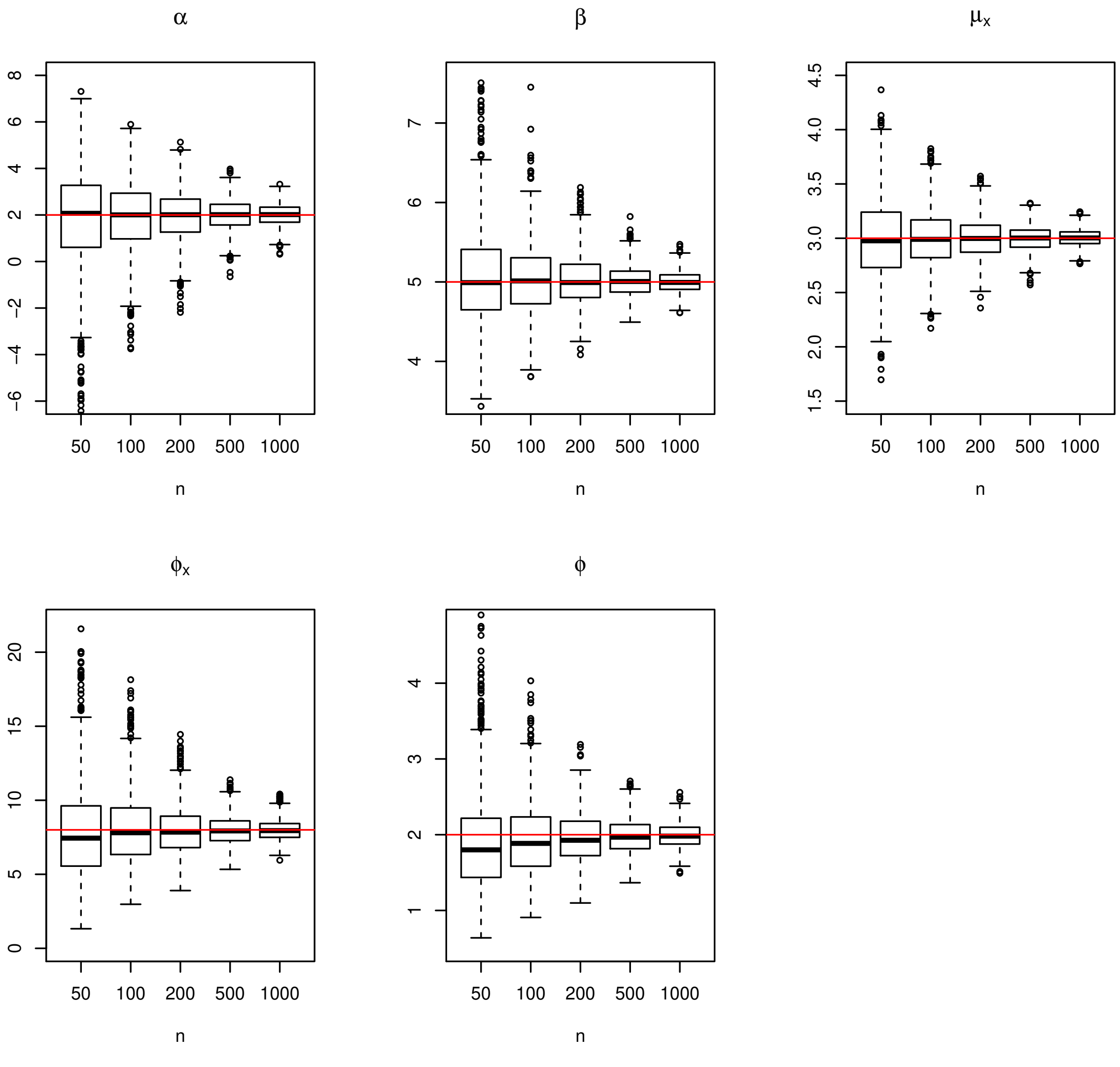


Figura 25 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x e ϕ quando consideramos $\omega = 0.5$ pré-fixado no modelo de regressão NIG com erro nas variáveis para o Cenário 2.

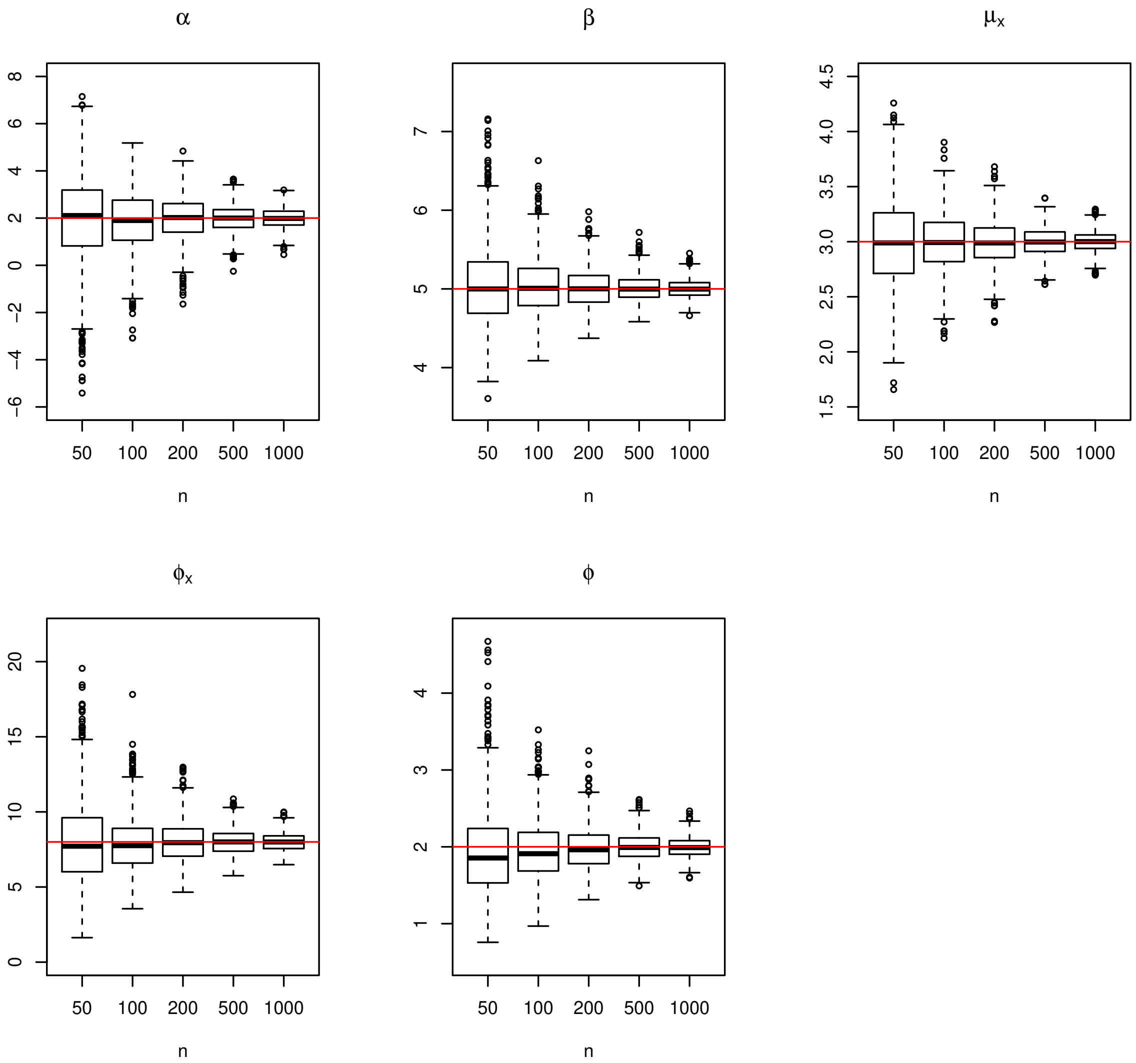


Figura 26 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x e ϕ quando consideramos $\omega = 1$ pré-fixado no modelo de regressão NIG com erro nas variáveis para o Cenário 2.

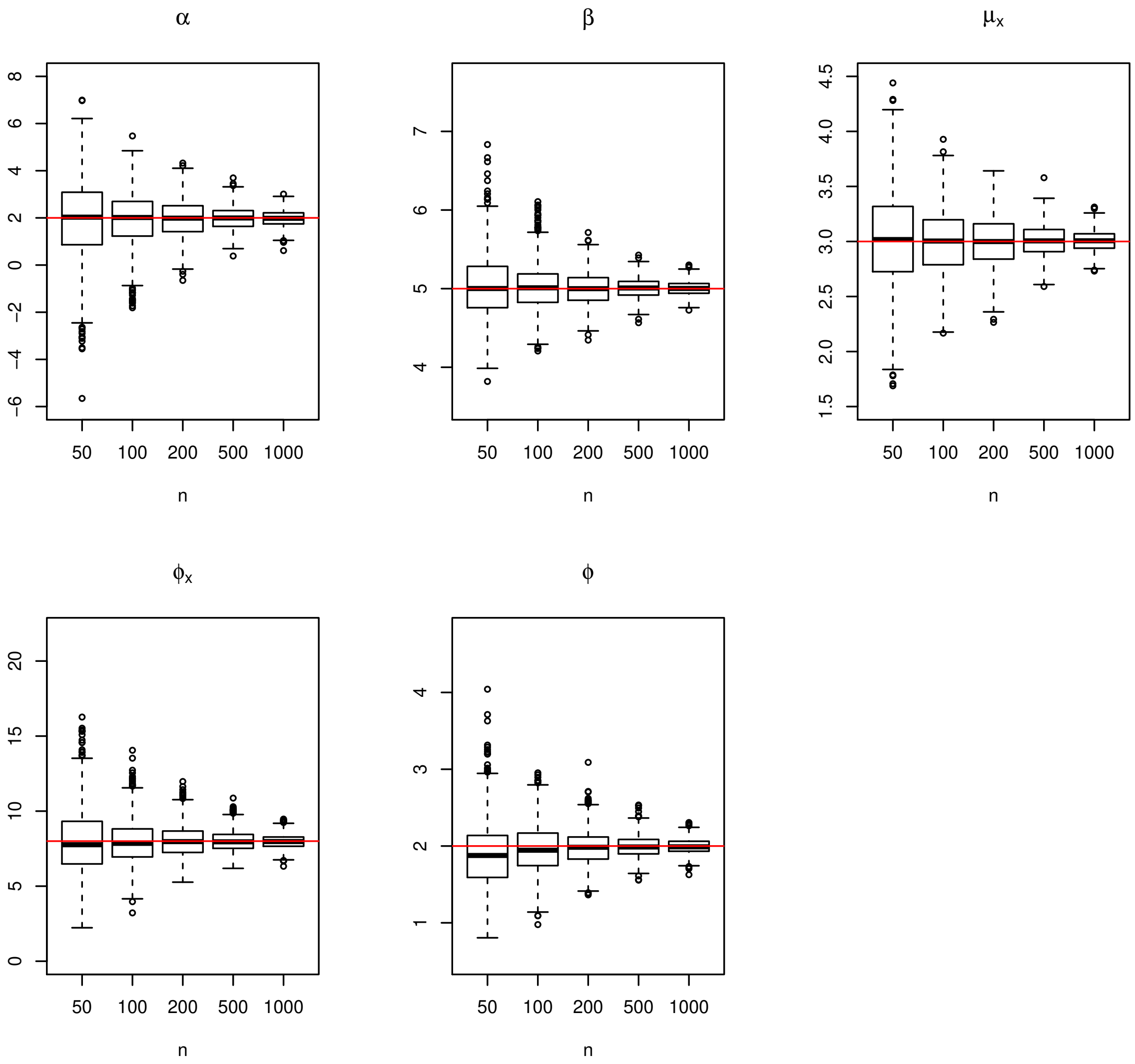


Figura 27 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x e ϕ quando consideramos $\omega = 5$ pré-fixado no modelo de regressão NIG com erro nas variáveis para o Cenário 2.

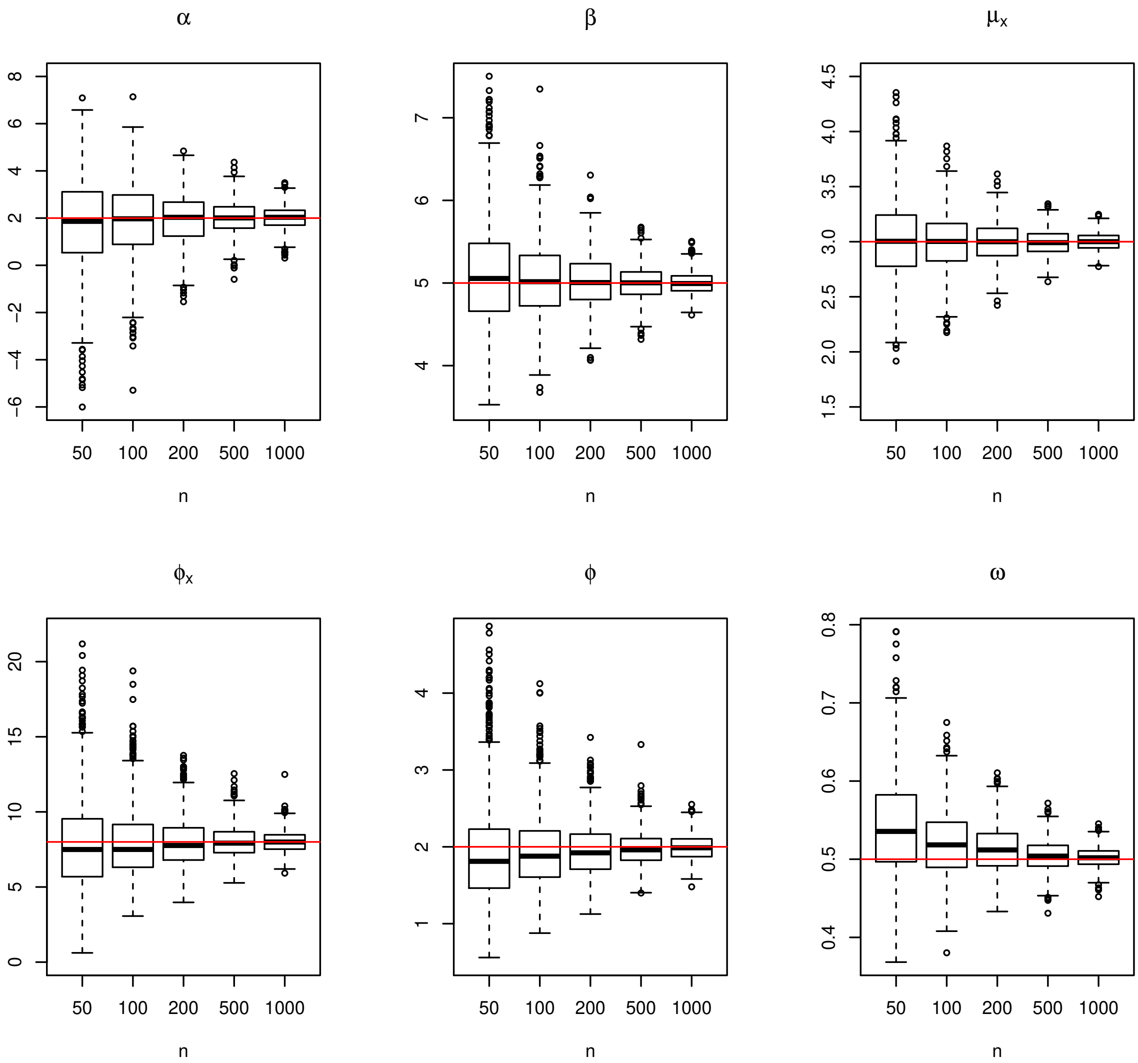


Figura 28 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x, ϕ e ω quando consideramos $\omega = 0.5$ parâmetro desconhecido do modelo de regressão NIG com erro nas variáveis para o Cenário 2.

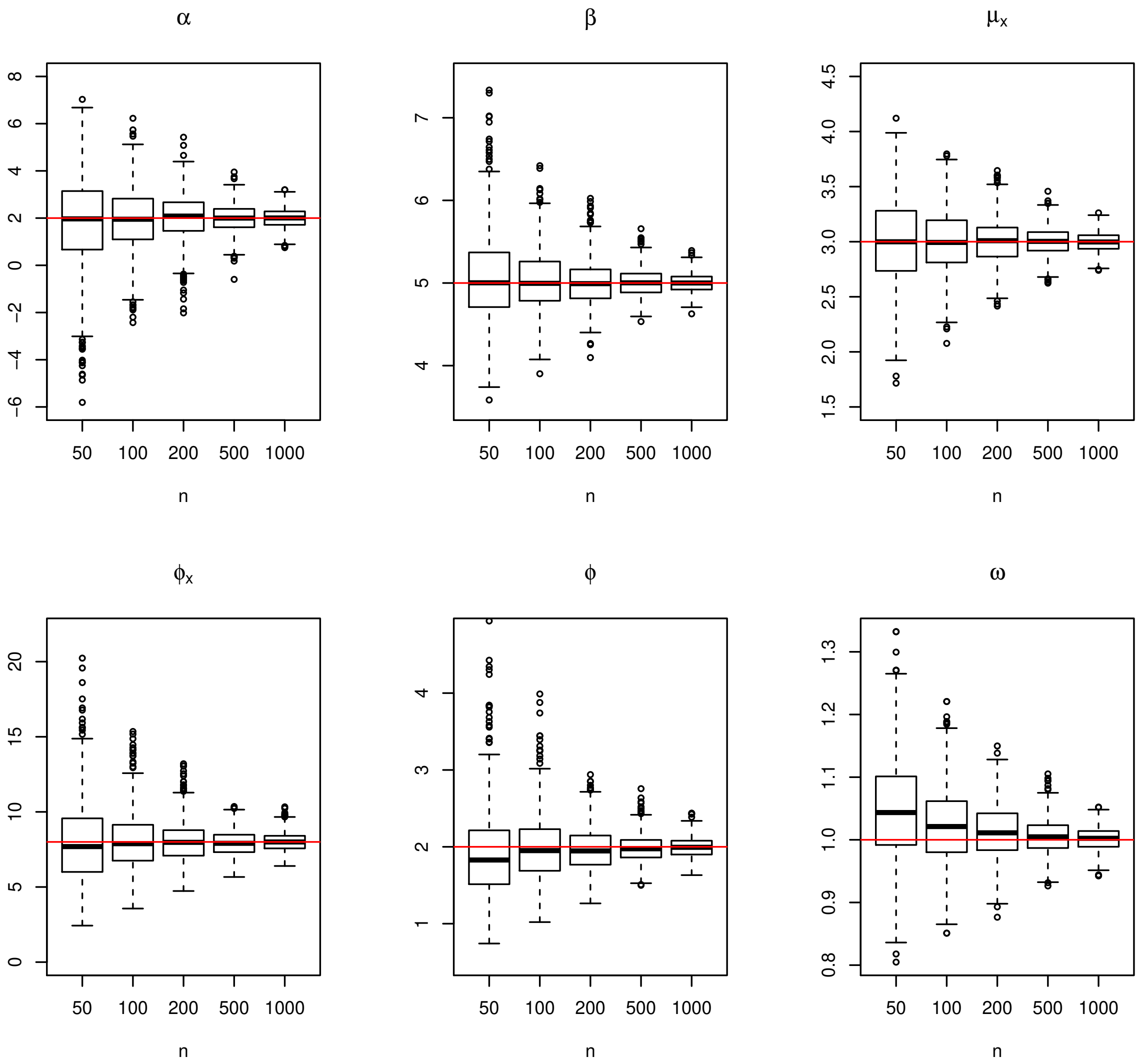


Figura 29 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x, ϕ e ω quando consideramos $\omega = 1$ parâmetro desconhecido do modelo de regressão NIG com erro nas variáveis para o Cenário 2.

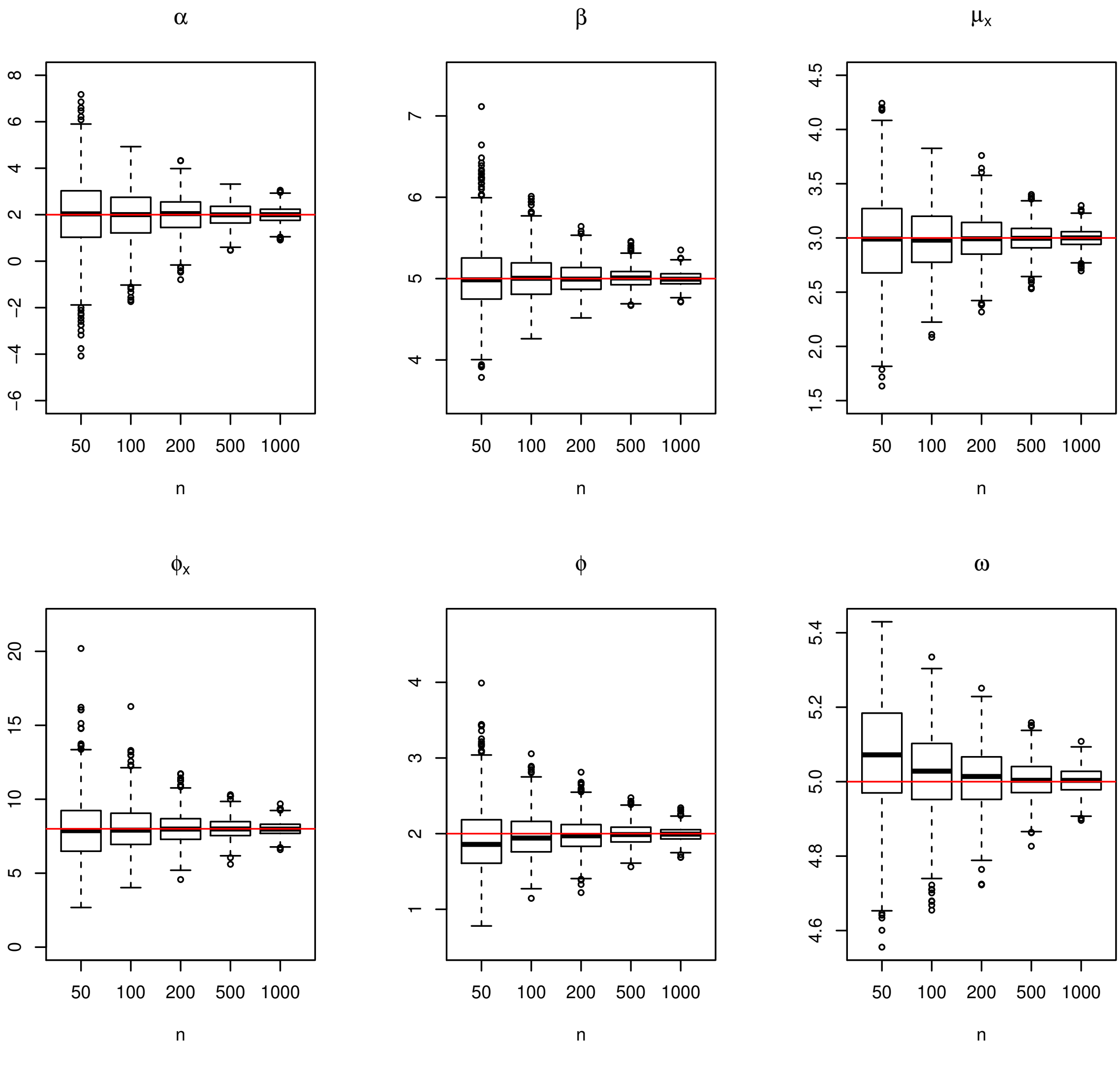


Figura 30 – Box-plot para as 1000 estimativas dos parâmetros α, β, μ_x, ϕ_x, ϕ e ω quando consideramos $\omega = 5$ parâmetro desconhecido do modelo de regressão NIG com erro nas variáveis para o Cenário 2.

4 Análise de Diagnóstico

Uma grande quantidade técnicas de diagnósticos de modelo podem ser aplicadas para realizar a seleção do modelo e analisar a adequação do seu ajuste. Nesta seção serão analisadas algumas maneiras de como este diagnóstico pode ser feito, já que a escolha do modelo apropriado, do ponto de vista estatístico, é um tópico extremamente importante na análise de dados (Bozdongan, 1987). Entre as inúmeras metodologias utilizadas para selecionar o melhor modelo ajustado, neste trabalho serão utilizados os critérios de informação de Akaike (AIC) e Bayesiano (BIC), como também o teste de bondade de ajuste Kolmogorov–Smirnov multivariado.

E ao analisar um conjunto de dados, é essencial que as estimativas do modelo proposto sejam resistentes às pequenas perturbações do modelo, pois caso o modelo ajustado não apresente uma boa descrição dos dados, pode conduzir a inferências errôneas. Assim, é necessário realizar um estudo sobre a robustez dos resultados obtidos, como por exemplo, utilizando a distância de Cook.

4.1 Critério de Informação de Akaike (AIC)

O Critério de Informação Akaike (AIC) é uma maneira de selecionar um modelo a partir de um conjunto finito de modelos. O modelo escolhido é aquele que minimiza a distância de Kullback-Leibler entre o modelo e o modelo "real". Este critério busca um modelo que ajuste bem aos dados, mas que contenha poucos parâmetros. Assim, seja k o número de parâmetros a serem estimados no modelo, Akaike (1974) definiu seu critério de informação como:

$$\text{AIC} = -2\log\, L(\hat{\boldsymbol{\theta}}) + 2k,$$

em que $L(\hat{\boldsymbol{\theta}})$ é a função de log-verossimilhança maximizada em $\hat{\boldsymbol{\theta}}$.

4.2 Critério de Informação Bayesiano (BIC)

Desenvolvido por Schwarz (1978), o Critério de Informação Bayesiano (BIC), também conhecido como Critério de Informação Schwarz (SIC) é um critério de seleção de modelo entre um conjunto finito de modelos, e está particularmente relacionado com o critério AIC. Assim, seja k o número de parâmetros a serem estimados no modelo, o critério BIC será definido por

$$\text{BIC} = -2\log\, L(\hat{\boldsymbol{\theta}}) + k\,\log(n).$$

No ajuste de modelos é possível aumentar a verossimilhança com a adição de parâmetros, mas fazendo isso pode-se gerar modelos excessivamente complexos, assim, o critério BIC resolve este problema penalizando modelos com grande números de parâmetros, sendo esta penalidade maior do que no critério AIC.

Como podemos ver comparando os critérios AIC e BIC, estes critérios diferem apenas pelo coeficiente que multiplica o número de parâmetros, ou seja, o termo que penaliza modelos com um grande número de parâmetros. Desta maneira, os modelos escolhidos através do critério BIC tendem a ser menos complexos do que quando comparados com os modelos obtidos através do critério AIC. O modelo com menores valores de AIC e BIC é considerado o de melhor ajuste.

4.3 Teste de bondade de ajuste Kolmogorov–Smirnov multivariado

Os testes de bondade de ajuste foram desenvolvidos na maior parte para distribuições univariadas, assim poucas referências podem ser encontradas sobre testes de bondade de ajuste multivariados, que podem ser encontradas em (Krishnaiah, 1980; Kotz e Johnson, 1985; e D'Agostino e Stephens, 1986). O teste de bondade de ajuste Kolmogorov–Smirnov (KS) multivariado foi desenvolvido, para avaliar o ajuste de um modelo para quando temos dados multivariados.

A estatística de teste univariada padrão de Kolmogorov-Smirnov toma a forma

$$D = \sup_x |F_n(x) - \hat{F}(x)|,$$

em que $F_n(x)$ é a função de distribuição acumulada (fda) e $\hat{F}(x)$ é a fda do modelo ajustado (Massey, 1951). O cálculo da estatística de teste é direto no caso univariado, uma vez que as observações podem ser ordenadas diretamente do menor para o maior por magnitude. Já a fda empírica para dados multivariados pode ser expressa como

$$F_n(x_1, x_2, ..., x_p) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, ..., X_p \leqslant x_p),$$

que é a proporção de observações que satisfazem todas desigualdades, não sendo, portanto, difícil encontrar a fda empírica multivariada, mas sendo difícil encontrar a diferença máxima entre as fda para calcular a estatística multivariada KS.

Um alternativa para este problema é utilizar os valores da log-densidade para medir o ajuste do modelo. Este fornecerá um bom ajuste quando a log-densidade estiver concentrada nas mesmas áreas que os dados subjacentes. Evitando a utilização de integração numérica ou integração de Monte Carlo, pois a dimensão é reduzida em valores univariados da densidade. Desta maneira, o procedimento proposto por O'Hagan et al. (2016) é dado por

1. Selecione o modelo ótimo $\hat{F}$,

2. Ordene as observações em $\mathbf{X}$ de acordo com o seu valor da log-densidade sob $\hat{F}$,

3. Trace a fda empírica dos valores da log-densidade dos dados originais $\mathbf{X}$ sob $\hat{F}$,

4. Simule novos dados $\mathbf{Y}$ a partir de $\hat{F}$,

5. Ordene as observações em $\mathbf{Y}$ de acordo com o seu valor da log densidade sob $\hat{F}$,

6. Trace a fda empírica dos valores da log densidade dos dados simulados $\mathbf{Y}$ sob $\hat{F}$ nos mesmos eixos do Passo 3,

7. Repita os passos 4, 5 e 6 para um número grande de simulações.

A intuição deste procedimento é que, se o modelo realmente fornece um bom ajuste aos dados, então a fda empírica dos valores simulados será similar a fda empiríca derivada dos dados originais. Desta maneira, as fdas das densidades dos dados originais e dos dados simulados devem ser semelhantes entre si. Isso proporciona um diagnóstico visual útil do ajuste do modelo.

4.4 Distância de Cook

O procedimento de deleção de casos é uma das técnicas mais utilizadas em análise de diagnóstico (Cook e Weisberg, 1982), para investigar o efeito da i-ésima observação de um conjunto de dados na estimação dos parâmetros do modelo. Seja $\boldsymbol{\theta}$ o vetor de parâmetros de interesse do modelo e $\hat{\boldsymbol{\theta}}$ a estimativa de $\boldsymbol{\theta}$, quando a i-ésima observação Z_i é deletada do conjunto observado completo, $\mathbf{Y} = \{Y_1, Y_2, ..., Y_n\}$, um ajuste do modelo utilizando o mesmo método de estimação produz a estimativa $\hat{\boldsymbol{\theta}}_{(i)}$, com o índice "$(i)$" representando a estimativa de $\boldsymbol{\theta}$ com o i-ésimo caso deletado. Medidas apropriadas de diferença entre $\hat{\theta}_{(i)}$ e $\hat{\theta}$ podem ser utilizadas para quantificar a influência da i-ésima observação na estimativa dos parâmetro $\hat{\theta}$. Uma das medidas mais utilizadas é a distância de Cook, que é definida por

$$C_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top \mathbf{M}(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}), \tag{4.1}$$

em que $\mathbf{M}$ é uma matriz positiva definida apropriadamente escolhida, sendo normalmente a inversa da matriz de covariância estimada de $\hat{\boldsymbol{\theta}}$ ou $\hat{\boldsymbol{\theta}}_{(i)}$ (Cook and Weisberg, 1982).

Como é necessário encontrar $\hat{\boldsymbol{\theta}}_{(i)}$ para cada caso, quando o tamanho amostral for muito grande a carga computacional pode ser bastante pesada, desta maneira, a fim de facilitar o cálculo de $\boldsymbol{\theta}_{(i)}$ Cook e Weisberg (1982) apresentam a aproximação

$$\hat{\boldsymbol{\theta}}^1_{(i)} = \hat{\boldsymbol{\theta}} + \left\{-\ddot{\ell}(\hat{\boldsymbol{\theta}})\right\}^{-1} \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}), \tag{4.2}$$

em que $\dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}) = \partial \ell_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Cook e Weisberg (1982) também consideraram algumas escolhas de $\mathbf{M}$, sendo entre elas a mais comum, a utilização da matriz de informação observada $\mathbf{M} = -\ddot{\ell}(\hat{\boldsymbol{\theta}})$. Assim, substituindo (4.2) em (4.1), obtemos a seguinte aproximação para a distância generalizada de Cook (Xie & Wei, 2007)

$$GD_i^1 = \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}})^\top \left\{-\ddot{\ell}(\hat{\boldsymbol{\theta}})\right\}^{-1} \dot{\ell}_{(i)}(\hat{\boldsymbol{\theta}}). \tag{4.3}$$

Porém, dado que a função de log-verossimilhança do modelo de erro nas variáveis sob a distribuição NIG é dado por $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$, em que

$$\ell_i(\boldsymbol{\theta}) = -\frac{3}{4}[\log(\omega + d(\mathbf{z}_i)) - \log\, \omega] + log\ K_1(\sqrt{\omega(\omega + d(\mathbf{z}_i))}\ ) - \log\, 2\pi - \frac{1}{2}\log\, |\boldsymbol{\Sigma}|,$$

com $d(\mathbf{z}_i) = (\mathbf{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}_i - \boldsymbol{\mu})$, visto que as derivadas envolvem expressões complicadas, temos que é difícil trabalhar diretamente com a função $\ell(\boldsymbol{\theta})$. Desta maneira, vamos implementar o método de Cook através da função-Q, que corresponde à metodologia de Zhu et al.(2001), onde utilizamos a esperança condicional da função de log-verossimilhança completa, que é dada por

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{z}_c)|\hat{\boldsymbol{\theta}}^{(k)}, \mathbf{z}],$$

e pode ser encontrada em (3.10).

Quando a i-ésima observação é deletada, a função-Q, $Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, pode ser formada da mesma maneira que em (3.10), apenas retirando a i-ésima observação. Pode ser mostrado que sua primeira derivada $\dot{Q}_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ com respeito a $\boldsymbol{\theta}$, avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{(i)}$ deve ser da forma

$$\dot{Q}_{(i)}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \frac{\partial}{\partial \boldsymbol{\theta}} \left[Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{4.4}$$

sendo que as primeiras derivadas são dadas por

$$\begin{aligned}
\dot{Q}_{\alpha(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{1}{\phi} \sum_{j \in J_i} \hat{q}_j (Y_j - \alpha - \beta \hat{x}_j), \\
\dot{Q}_{\beta(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{1}{\phi} \sum_{j \in J_i} (\hat{q}_j (Y_j - \alpha - \beta \hat{x}_j)\hat{x}_j - \beta \hat{\Lambda}_x), \\
\dot{Q}_{\mu_x(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{1}{\phi_x} \sum_{j \in J_i} \hat{q}_j (\hat{x}_j - \mu_x), \\
\dot{Q}_{\phi_x(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\frac{n-1}{2\phi_x} + \frac{1}{2\phi_x^2} \sum_{j \in J_i} \left(\hat{\Lambda}_x + \hat{q}_j (\hat{x}_j - \mu_x)^2\right), \\
\dot{Q}_{\phi(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= -\frac{n-1}{2\phi} + \frac{1}{2\phi^2} \sum_{j \in J_i} \left[\hat{q}_j \left((X_j - \hat{x}_j)^2 + (Y_j - \alpha - \beta \hat{x}_j)^2\right) + \hat{\Lambda}_x (1 + \beta^2)\right], \\
\dot{Q}_{\omega(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \frac{n-1}{2\omega} + n - 1 - \frac{1}{2} \sum_{j \in J_i} (\hat{q}_j + \hat{v}_j),
\end{aligned}$$

em que $J_i = \{j \in \mathbb{N} \mid 1 \leqslant j \leqslant n, j \neq i \}$.

A fim de obter as medidas de diagnóstico para a influência local utilizando a função-Q, também é necessário computar a matriz Hessiana a partir da segunda derivada de $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, sendo esta dada por

$$\ddot{Q}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_1(\alpha,\beta,\phi) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_2(\mu_x,\phi_x) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddot{Q}_3(\omega) \end{pmatrix},$$

em que as matrizes são dadas por $\ddot{Q}_1(\alpha,\beta,\phi) = [\ddot{Q}_{1,\tau\pi}]$, $\ddot{Q}_{1,\tau\pi} = \dfrac{\partial^2 Q_1(\alpha,\beta,\phi|\hat{\boldsymbol{\theta}})}{\partial\tau\partial\pi}$, com $\tau,\pi = \alpha,\beta$ ou ϕ, $\ddot{Q}_2(\mu_x,\phi_x) = [\ddot{Q}_{2,\tau\pi}]$, $\ddot{Q}_{2,\tau\pi} = \dfrac{\partial^2 Q_2(\mu_x,\phi_x|\hat{\boldsymbol{\theta}})}{\partial\tau\partial\pi}$,com $\tau,\pi = \mu_x$ ou ϕ_x, e $\ddot{Q}_3(\omega) = \ddot{Q}_{3,\omega\omega} = \dfrac{\partial^2 Q_3(\omega|\hat{\boldsymbol{\theta}})}{\partial\omega\partial\omega}$, assim a matriz Hessiana tem elementos dados por

$$\begin{aligned}
\ddot{Q}_{1,\alpha\alpha} &= -\frac{1}{\phi}\sum_{i=1}^{n}\hat{q}_i,\\
\ddot{Q}_{1,\alpha\beta} &= -\frac{1}{\phi}\sum_{i=1}^{n}\hat{q}_i\hat{x}_i,\\
\ddot{Q}_{1,\alpha\phi} &= -\frac{1}{\phi^2}\sum_{i=1}^{n}\hat{q}_i(Y_i-\alpha-\beta\hat{x}_i),\\
\ddot{Q}_{1,\beta\beta} &= -\frac{1}{\phi}\sum_{i=1}^{n}(\hat{\Lambda}_x+\hat{x}_i^2\hat{q}_i),\\
\ddot{Q}_{1,\beta\phi} &= -\frac{1}{\phi^2}\sum_{i=1}^{n}((Y_i-\alpha-\beta\hat{x}_i)\hat{x}_i\hat{q}_i-\beta\hat{\Lambda}_x),\\
\ddot{Q}_{1,\phi\phi} &= \frac{n}{\phi^2}\left(1-\frac{\hat{\Lambda}_x}{\phi}(1+\beta^2)\right)-\frac{1}{\phi^3}\sum_{i=1}^{n}\hat{q}_i\left[(X_i-\hat{x}_i)^2-(Y_i-\alpha-\beta\hat{x}_i)^2\right],\\
\ddot{Q}_{2,\mu_x\mu_x} &= -\frac{1}{\phi_x}\sum_{i=1}^{n}\hat{q}_i,\\
\ddot{Q}_{2,\mu_x\phi_x} &= -\frac{1}{\phi_x^2}\sum_{i=1}^{n}\hat{q}_i(\hat{x}_i-\mu_x),\\
\ddot{Q}_{2,\phi_x\phi_x} &= \frac{n}{2\phi_x^2}-\frac{1}{\phi_x^3}\sum_{i=1}^{n}\left(\hat{q}_i(\hat{x}_i-\mu_x)^2+\hat{\Lambda}_x\right),\\
\ddot{Q}_{3,\omega\omega} &= -\frac{n}{\omega^2}.
\end{aligned}$$

É necessário encontrar $\hat{\boldsymbol{\theta}}_{(i)}$ para cada caso, em que a carga computacional total envolvida pode ser bastante pesada, especialmente quando o tamanho amostral for grande. Assim, uma nova aproximação $\hat{\boldsymbol{\theta}}^1_{(i)}$ utilizando a função-Q é usada para reduzir a carga computacional (Cook e Weisberg, 1982; Zhu et al., 2001),

$$\hat{\boldsymbol{\theta}}^1_{(i)} = \hat{\boldsymbol{\theta}} + \left\{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\right\}^{-1}\dot{Q}_{(i)}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}). \tag{4.5}$$

4.5 Distância de Cook Generalizada

Será utilizado o método proposto por Xie et al.(2007) baseado no algoritmo EM, assim a distância de Cook generalizada será dada por

$$GD_i = \left(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}\right)^\top \left\{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\right\} \left(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}\right), \tag{4.6}$$

que é uma combinação ponderada dos elementos da diferença $\left(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}\right)$ com a norma $-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$. Substituindo (4.5) em (4.6), obtemos a seguinte aproximação GD_i^1 de GD_i

$$GD_i^1 = \dot{Q}_{(i)}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^\top \left\{-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})\right\}^{-1} \dot{Q}_{(i)}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}). \tag{4.7}$$

É consideravelmente fácil analisar a influência do i-ésimo caso para subconjunto de parâmetros. Considere o conjunto de parâmetros $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top\right)^\top$ em que $\boldsymbol{\theta}_1^\top$ é o subconjunto de parâmetros de interesse com dimensão q_1, e $\boldsymbol{\theta}_2^\top$ é o subconjunto de parâmetros que não temos interesse imediato de analisar, com dimensão q_2, sendo então $q_1 + q_2 = q$. Assim, das equações (4.6) e (4.7), as distâncias de Cook generalizadas para os subconjuntos podem ser definidas como

$$GD_i^1(\boldsymbol{\theta}_1) = \dot{Q}_{(i)\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^\top \left\{(\mathbf{I}_{q_1}, \mathbf{0})[-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^{-1}(\mathbf{I}_{q_1}, \mathbf{0})^\top\right\} \dot{Q}_{(i)\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^\top, \tag{4.8}$$

$$GD_i^1(\boldsymbol{\theta}_2) = \dot{Q}_{(i)\boldsymbol{\theta}_2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^\top \left\{(\mathbf{0}, \mathbf{I}_{q_2})[-\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})]^{-1}(\mathbf{0}, \mathbf{I}_{q_2})^\top\right\} \dot{Q}_{(i)\boldsymbol{\theta}_2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})^\top, \tag{4.9}$$

em que $\dot{Q}_{(i)\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}_1|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ e $\dot{Q}_{(i)\boldsymbol{\theta}_2}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) = \partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}_2|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

Os valores de $GD_i^1(\omega_1)$ e $GD_i^1(\omega_2)$ revelam o impacto da i-ésima observação na estimção de ω_1 e ω_2, respectivamente. Abaixo serão listadas algumas situações de análise de influência avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, que será omitido para simplificação.

- **Caso 1:** Temos interesse em testar cada parâmetro separadamente, seja $\pi = \alpha, \beta, \mu_x, \phi_x, \phi$ ou ω, temos que a distância de Cook generalizada será dada por

$$GD_i^1(\pi) = -\left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi}\right]^\top \left(\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi^2}\right)^{-1} \left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi}\right].$$

- **Caso 2:** Temos interesse em testar dois parâmetros π e τ simultaneamente, assim, seja $\boldsymbol{\eta} = (\pi, \tau)$ com $\pi, \tau = \alpha, \beta, \mu_x, \phi_x, \phi$ ou ω, temos que a distância de Cook generalizada será dada por

$$GD_i^1(\pi, \tau) = -\left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}}\right]^\top \begin{pmatrix} \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi^2} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi \partial \tau} \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \tau \partial \pi} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \tau^2} \end{pmatrix}^{-1} \left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}}\right].$$

- **Caso 3:** Temos interesse em testar três parâmetros π, τ e ν simultaneamente, assim, seja $\boldsymbol{\eta} = (\pi, \tau, \nu)$ com $\pi, \tau, \nu = \alpha, \beta, \mu_x, \phi_x, \phi$ ou ω, temos que a distância de Cook generalizada será dada por

$$GD_i^1(\boldsymbol{\eta}) = -\left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}}\right]^\top \begin{pmatrix} \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi^2} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi \partial \tau} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \pi \partial \nu} \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \tau \partial \pi} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \tau^2} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \tau \partial \nu} \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \nu \partial \pi} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \nu \partial \tau} & \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \nu^2} \end{pmatrix}^{-1} \left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\eta}}\right].$$

- **Caso 4:** Temos interesse em testar todos os parâmetro simultaneamente,assim, temos que a distância de Cook generalizada será dada por

$$GD_i^1(\boldsymbol{\theta}) = -\left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right]^\top \left(\ddot{Q}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right)^{-1} \left[\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right].$$

4.6 Distância-Q

Outra medida de diferença entre $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{(i)}$ é a Distância-Q, que é bastante similar à distância de Cook (Zhu et al., 2001)

$$QD_i(\boldsymbol{\theta}) = 2\left\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_{(i)}|\hat{\boldsymbol{\theta}})\right\}, \tag{4.10}$$

substituindo (4.5) em (4.10), obtemos a seguinte aproximação QD_i^1 de QD_i

$$QD_i^1(\boldsymbol{\theta}) = 2\left\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}^1_{(i)}|\hat{\boldsymbol{\theta}})\right\}. \tag{4.11}$$

4.7 Resumo

A estratégia para identificação de casos influentes via deleção de caso pode ser resumida como,

- **Passo 1:** Calcule os EMV via algoritmo EM;
- **Passo 2:** Calcule $\ddot{Q}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$, $\dot{Q}_{(i)}(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}})$ e então a aproximação $\hat{\boldsymbol{\theta}}^1_{(i)}$ para $i = 1, \ldots, n$.
- **Passo 3:** Calcule GD_i^1 e QD_i^1 para $i = 1, \ldots, n$.
- **Passo 4:** Identifique as observações influentes ou conjunto de observações através de valores muito altos, discrepantes dos demais, de GD_i^1 e QD_i^1 para $i = 1, \ldots, n$.

5 Aplicação a Dados Reais

Neste capítulo, apresentamos a análise de dados reais, a fim de ilustrar a abordagem proposta. Serão consideradas duas aplicações, sendo a primeira utilizando dados de um estudo com atletas do Instituto Australiano do Esporte, já estudado por Telford e Cunningham (1991) e Lachos et al. (20009), e a segunda aplicação utilizando dados de um estudo do mercado de ações chileno, estudado por Galea et al. (2010).

5.1 Conjunto de dados do Instituto australiano do esporte (AIS)

Nesta seção serão analisados os dados de medição de atletas de alto desempenho do Instituto Australiano do Esporte (AIS), para 202 atletas (102 homens e 100 mulheres) em 13 variáveis. Estes dados já foram analisados por Telford e Cunningham (1991), que fornecem mais informações sobre como os dados foram coletados, e também por Lachos et al. (2011). Nesta aplicação será analisada apenas a relação entre o índice de massa corpórea (IMC) e a gordura corporal (GC) utilizando um modelo com erro nas variáveis. Assim, definimos o modelo como

$$IMC_i = \alpha + \beta \; gc_i + e_i, \; i = 1, \ldots, 202,$$

com gc_i sendo a porcentagem de gordura do indivíduo i e IMC_i o valor do IMC do indivíduo i. Porém, para calcular a gordura corporal podem ser utilizados inúmeros métodos diferentes que na maioria das vezes não condizem com o valor real, mas sim uma aproximação. Deste modo, consideramos que a gordura corporal é medida com erro de acordo com a equação

$$GC_i = gc_i + u_i,$$

em que GC_i é um estimador não viesado do verdadeiro valor de gc_i. Desta maneira, temos que os dados serão considerados com a seguinte distribuição

$$\mathbf{Z}_i = \begin{pmatrix} IMC_i \\ GC_i \end{pmatrix} \sim NIGS_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \omega), \; i = 1, \ldots, 202.$$

Além da distribuição NIG, os dados foram analisados utilizando as distribuições normal, t-Student, slash e normal contaminada. Porém, como a distribuição t-Student mostrou se ajustar melhor aos dados do que as demais distribuições, serão consideradas apenas a distribuição normal e t-Student como forma de comparação.

Tabela 9 – Análise descritiva dos dados do AIS.

	Média	**DP**	**Assimetria**	**Curtose**	**Min**	**Max**
IMC	22,9558	2,8639	0,9394	5,1322	16,7500	34,4200
Gordura corporal	13,5074	6,1898	0,7539	2,7987	5,6300	35,5200

5.1.1 Análise exploratória dos dados

Na Tabela 9 são apresentadas estatísticas descritivas dos dados, como a média, desvio padrão (DP), medida de assimetria e curtose. Segundo a tabela os dados parecem ter uma leve assimetria positiva. A curtose univariada é uma medida que caracteriza o achatamento da curva da fdp, podendo se definida como a razão entre o quarto momento central (μ_4) e o quadrado do segundo momento central (μ_2), sendo portanto, definida como

$$\frac{\mu_4}{\mu_2^2} = \frac{E(X - E(X))^4}{[E(X - E(X))^2]^2},$$

segundo esta definição a curtose da distribuição normal é igual a 3, e temos uma distribuição leptocúrtica, quando o valor da curtose for maior que 3. No nosso caso, temos que o IMC possui um valor de curtose consideravelmente maior que 3, indicando uma fdp leptocúrtica, já a gordura corporal possui uma valor de curtose próximo de 3, mas como nosso modelo segue uma distribuição bivariada, é plausível considerar uma distribuição de caudas pesadas para os dados. Como já utilizado em outros trabalhos, temos que a suposição de que as variáveis são independentes e identicamente distribuídos (iid) neste banco de dados é plausível.

5.1.2 Estimação e verificação de suposições

Baseado na análise exploratória realizada anteriormente, serão comparados os modelos NIG, normal e t-Student, considerando os valores de $\nu = 9$ para a distribuição t-Student bivariada e de $\omega = 10$ para a distribuição NIG bivariada quando consideramos o parâmetro ω pré-fixado no modelo, estes valores foram escolhidos para melhor acomodar as observações atípicas que aparecem sob erros normais. Comparando um conjunto de valores possíveis, os graus de liberdade, ν e ω, foram escolhidos de acordo com os valores que maximizam a função de máxima verossimilhança. Todos os resultados foram obtidos utilizando o software R.

A Tabela 10 contém as EMV para os parâmetros dos quatro modelos,com seus respectivos erros padrão calculados via bootstrap. Sendo que os parâmetros ϕ_x e ω possuem a maior variação. O modelo NIG possui a maior verossimilhança do que comparado com a distribuição normal, mas possui resultados um pouco inferiores quando comparado com a distribuição t-student, porém quando comparamos os critérios AIC e BIC temos uma penalização do modelo NIG com parâmetro desconhecido, por possuir uma parâmetro a mais em seu modelo, o que já poderia ser esperado devido o critério de escolha de ν e ω

Tabela 10 – Resultados dos ajustes do modelo utilizando as distribuições normal, t-Student e NIG considerando ω fixo e parâmetro para as dados do Instituto australiano do esporte.

Parâmetros	Normal		t-Student ($\nu = 9$)		NIG ($\omega = 10$)		NIG	
	Estimativa	DP	Estimativa	DP	Estimativa	DP	Estimativa	DP
α	21,4965	0,5053	21,7177	0,4675	21,4039	0,4600	21,4082	0,5016
β	0,1081	0,0342	0,0794	0,0344	0,1026	0,0344	0,1030	0,0365
μ_x	13,5073	0,4328	13,0521	0,4917	13,3827	0,4211	13,3527	0,4678
ϕ_x	30,3308	3,4932	27,6980	3,3309	37,2197	3,5718	37,2197	3,9276
ϕ	7,8005	1,1312	5,9405	0,7637	7,7115	0,9933	7,7101	1,0245
ω	-	-	-	-	-	-	11,0419	5,5880
$\ell(\hat{\boldsymbol{\theta}})$	-1149,403		-1148,960		-1149,074		-1149,070	
AIC	2308,806		2307,919		2308,149		2310,140	
BIC	2325,347		2324,461		2324,690		2329,990	

ter sido feito visando maximizar o valor da verossimilhança e obter o modelo que melhor se ajusta aos dados. Sendo assim, neste caso, o modelo NIG com $\omega = 10$ pré-fixado, parece ajustar melhor do que o modelo norma, porém levemente pior do que o modelo t-Student.

A Figura 31 apresenta o gráfico de dispersão dos dados com a reta de regressão para os quatro modelos analisados, sendo que o modelo sob ditribuição t-Student possui o menor coeficiente angular, enquanto o modelo sob distribuição normal possui o maior coeficiente angular, parecendo ser mais influenciável por valores atípicos.

Na Figura 32 podemos ver os Q-Q plots e seus envelopes que são baseados na distribuição da distância de Mahalanobis de $\mathbf{Z}_i$, para $i = 1, ..., n$, que para a distribuição normal segue uma distribuição $\chi^2_{(2)}$ e para a distribuição t-Student segue uma distribuição $2F(2, \nu)$. Para a distribuição NIG o gráfico quantil-quantil foi obtido via simulação. As linhas nas figuras representam o 5º percentil, a média e 95º percentil de 100 pontos simulados para cada observação. Os gráficos mostram que a distribuição normal é a que pior se ajusta aos dados e a distribuição NIG parece ajustar tão bem, os valores atípicos, quanto a distribuição t-Student.

Na Figura 33 podemos ver os gráficos de Komogorov-Smirnoff para os quatros modelos, em que a linha preta representa a fda empírica para os dados originais e as linhas cinzas representam a fda empírica para cada um dos 500 conjuntos de dados simulados do modelo ótimo. As figuras ilustram que os modelos fornecem um ajuste razoavelmente bom aos dados, com uma leve tendência a colocar muita probabilidade nas regiões das caudas da log-densidade e pouca na região central.

A Figura 34 exibe as distâncias de Mahalanobis para os quatro modelos ajustados, e a partir desta figura, podemos detectar como as principais possíveis observações atípicas os indivíduos 53, 56, 75, 160, 163 e 178. Do algoritmo EM, os pesos estimados $\hat{q}_i$ das observações 163 e 178 são as menores para os modelos t-Student e NIG, sendo a linha tracejada representa o peso dado pela distribuição normal, como pode ser visto na Figura

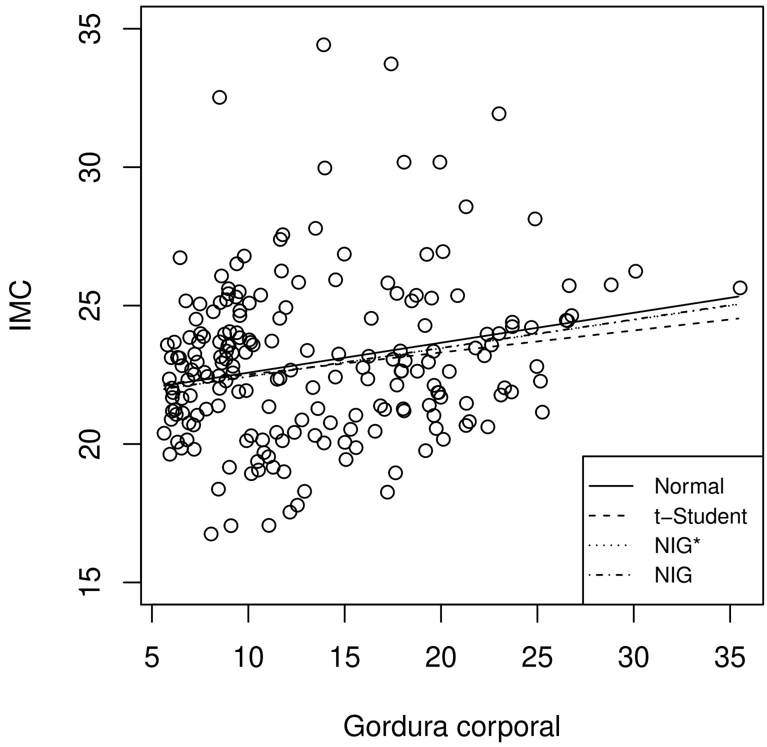


Figura 31 – Modelos de regressão lineares ajustados para a distribuição normal, t-Student, NIG com ω fixo (NIG*) e com ω parâmetro.

35, assim temos que a estimativa de $\boldsymbol{\theta}$ tende a dar pouco peso q_i para as observações com maior distância nos casos das distribuições t-Student e NIG, devido ao fato de q_i ser inversamente proporcional a distância de Mahalanobis, diferente do que ocorre na distribuição normal. Deste modo, temos que as distribuições t-Student e NIG ajustam melhor os dados atípicos do que a distribuição normal. Já no caso da distribuição NIG com ω parâmetro desconhecido do modelo, temos também o peso v_i, em que a estimativa de $\boldsymbol{\theta}$ tende a dar peso maior para as observações distantes devido ao fato de que v_i ser diretamente proporcional a distância de Mahalanobis.

Utilizando a distribuição NIG ao invés da distribuição normal tivemos resultado um pouco melhores. Porém, quando comparamos com o modelo t-Student tivemos resultados levemente inferiores, indicando talvez a distribuição NIG se ajuste melhor a dados econômicos, como será analisado adiante.

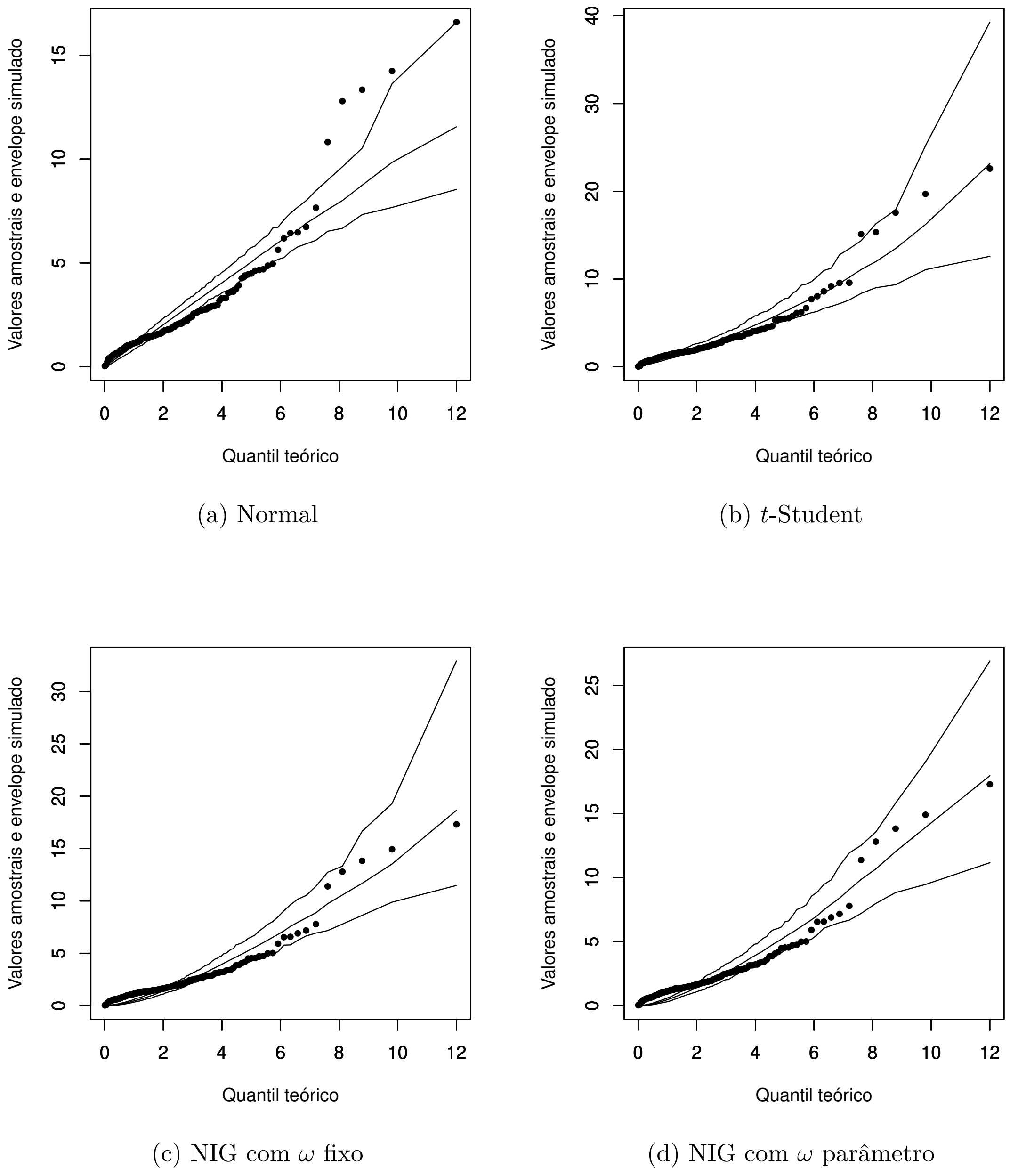
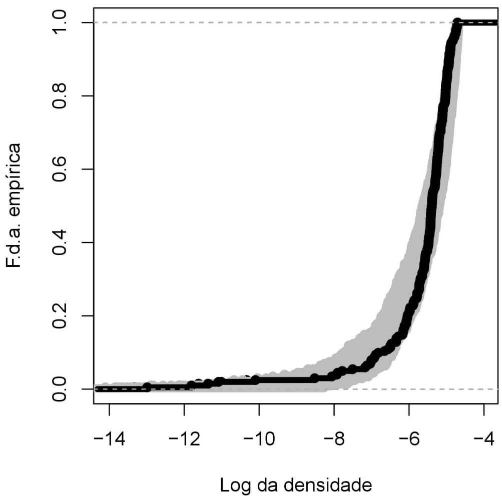
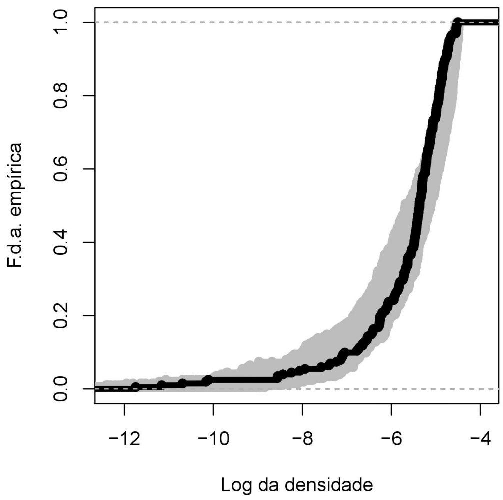


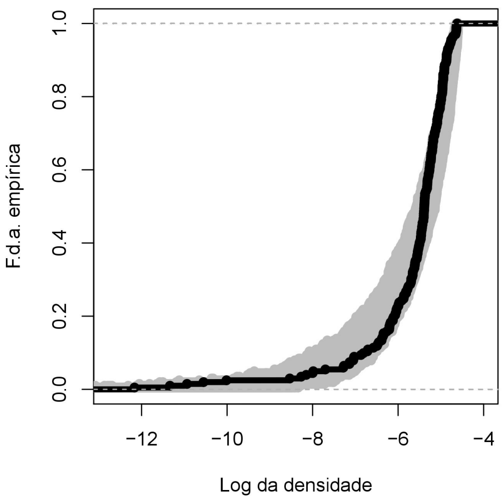
Figura 32 – Q-Q Plots e envelopes simulados para o modelo ajustado considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro para os dados do AIS.



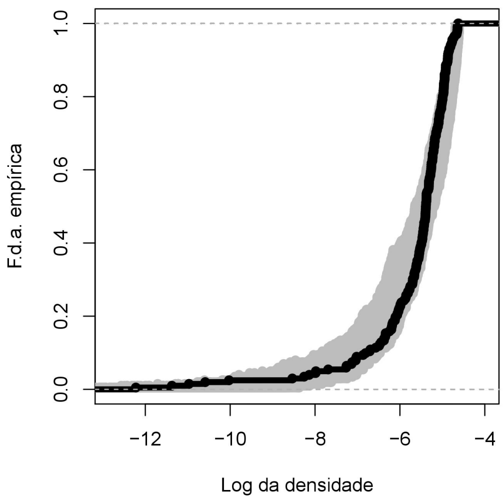
(a) Normal



(b) t-Student

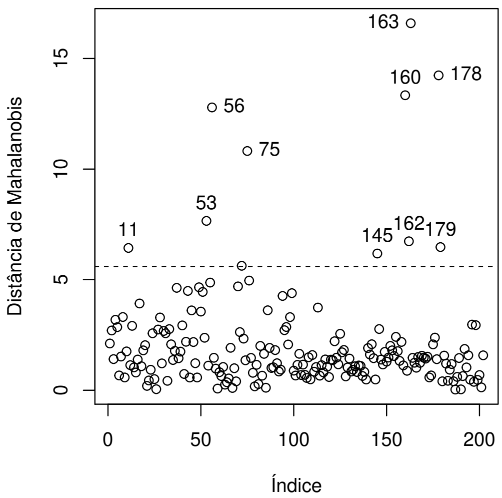


(c) NIG com ω fixo

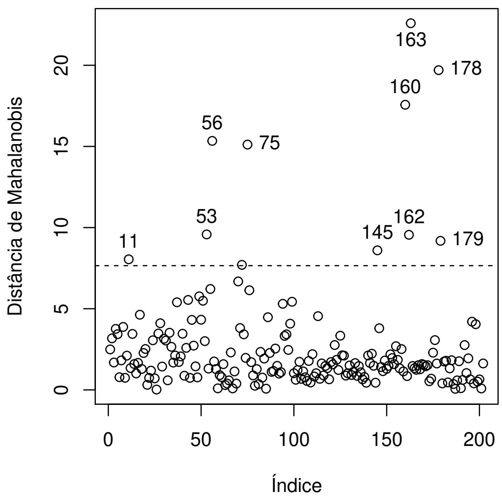


(d) NIG com ω parâmetro

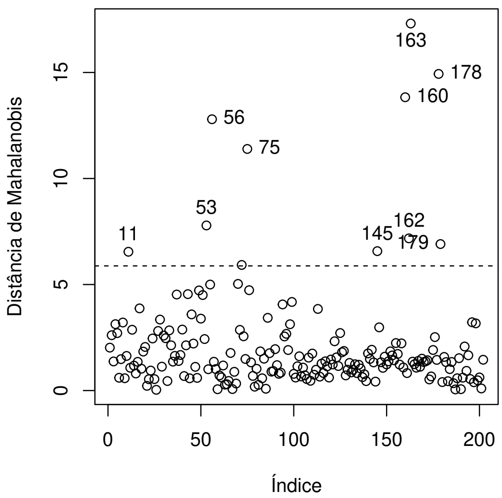
Figura 33 – Gráficos de Kolmogorov-Smirnoff para o modelo sob distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro para os dados do AIS.



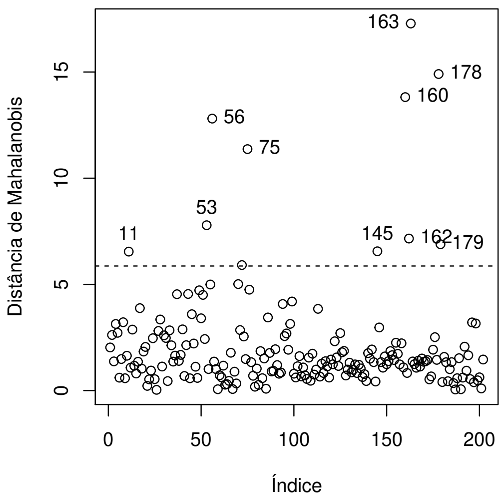
(a) Normal



(b) t-Student

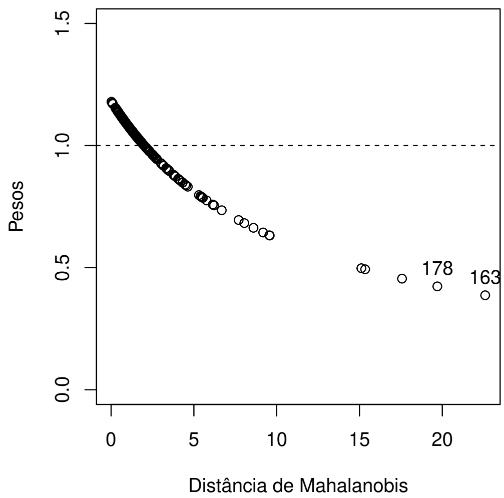


(c) NIG com ω fixo

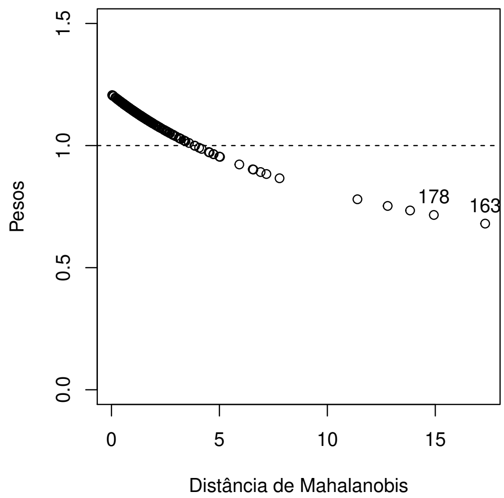


(d) NIG com ω parâmetro

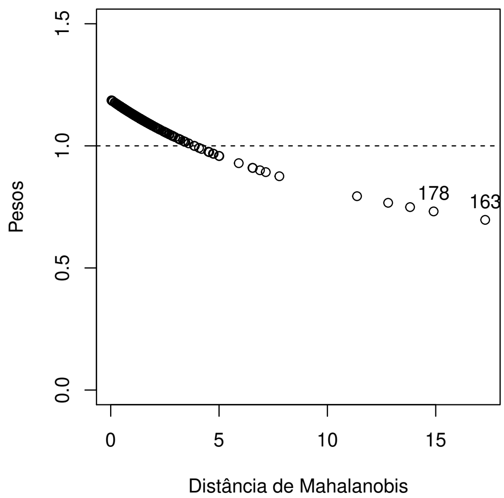
Figura 34 – Gráficos da distância de Mahalanobis considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro para os dados do AIS.



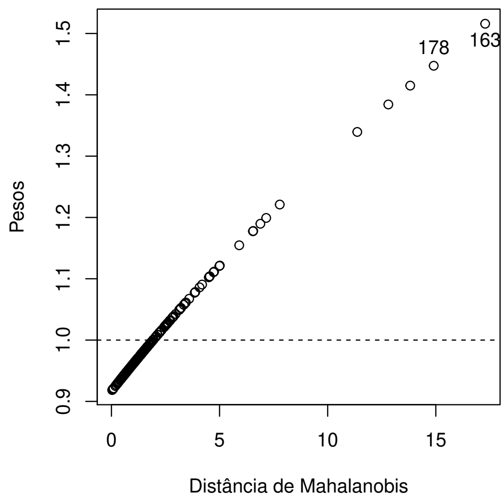
(a) t-Student (q_i)



(b) NIG com ω fixo (q_i)



(c) NIG com ω parâmetro (q_i)



(d) NIG com ω parâmetro (v_i)

Figura 35 – Estimativas de q_i para o modelo sob distribuição t-Student, NIG com ω fixo e com ω parâmetro e v_i para o modelo sob distribuição NIG com ω parâmetro para os dados do AIS.

5.2 Conjunto de dados do Mercado chileno de Ações

Os dados analisados nesta seção correspondem ao retorno mensal de ações da Bolsa de Comércio de Santiago (BCS), sendo utilizado o índice IPSA (*Indice de Precio Selectivo de Acciones*) como o retorno do mercado chileno, por ser apontado como o principal índice do mercado de ações do Chile, sendo um dos mais antigos, criado em 1977, e considerado como referência para investidores tanto do Chile como ao redor do mundo. Este índice é composto pelas 40 ações mais negociadas da BCS, sendo revisado anualmente.

Além do índice IPSA, serão analisadas 4 ações de companhias chilenas, sendo estas, Cervezas, Copec, Cuprum e Endesa. Assim, o conjunto de dados consiste dos retornos mensais de ações do mercado de ações chileno, analisados no período de janeiro de 1990 a junho de 2004. Estes dados parecem ser adequados ao trabalho, devido a distribuição NIG ser bastante utilizada para aplicações de dados financeiros.

Assim, nesta aplicação será feita uma análise dos dados utilizando o modelo com erro nas variáveis já proposto nas seções anteriores, relacionando o retorno de uma ação chilena, representada por $\mathbf{Y}$, ao retorno do mercado chileno, representado por $\mathbf{x}$. Assim, definimos nosso modelo como

$$Y_i = \alpha + \beta\, x_i + e_i,\ \ i = 1, \ldots, 174,$$

com x_i sendo o valor do retorno do mercado chileno no i-ésimo período e Y_i o valor do retorno da ação analisada no i-ésimo período. Porém, como o retorno do mercado não é diretamente observado, utilizamos o índice IPSA como uma estimativa do real valor do mercado chileno. Deste modo, consideramos que este é medido com erro de acordo com a equação

$$X_i = x_i + u_i,$$

em que X_i é o retorno do índice IPSA no i-ésimo período. Desta maneira, temos que os dados serão considerados seguindo a distribuição

$$\mathbf{Z}_i = \begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim NIGS_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \omega),\ \ i = 1, \ldots, 174.$$

Além da distribuição NIG, os dados foram analisados utilizando as distribuições normal, t-Student, slash e normal contaminada. Porém, como a distribuição t-Student se mostrou ajustar melhor aos dados do que a distribuição slash e normal contaminada, serão consideradas apenas a distribuição normal e t-Student como forma de comparação.

5.2.1 Análise exploratória dos dados

Na Tabela 11 são apresentadas estatísticas descritivas dos dados, como a média, desvio padrão (DP), medida de assimetria e curtose. Como podemos ver os retornos

Tabela 11 – Análise descritiva dos dados do mercado financeiro chileno.

Índice	**Média**	**DP**	**Assimetria**	**Curtose**	**Min**	**Max**
IPSA	0,0193	0,0697	0,0588	4,9993	-0,2985	0,2076
Cervezas	0,0236	0,1055	0,5545	4,1995	-0,2352	0,4761
Copec	0,0239	0,0902	0,5982	4,2289	-0,2000	0,3773
Cuprum	0,0488	0,1459	1,8687	8,2902	-0,3191	0,7510
Endesa	0,0223	0,0917	0,2410	4,4032	-0,3515	0,3140

possuem em sua maioria a medida de assimetria próxima de zero, indicando simetria dos dados, sendo a única exceção os retornos da Cuprum, que parecem ter uma leve simetria positiva. Em todos os casos temos valores de curtose maiores que 3, indicando que os dados analisados se ajustam bem sob distribuições com caudas pesadas. Como mostrado em (Galea et al., 2010) temos que a suposição de que os retornos são independentes e identicamente distribuídos (iid) neste banco de dados é plausível.

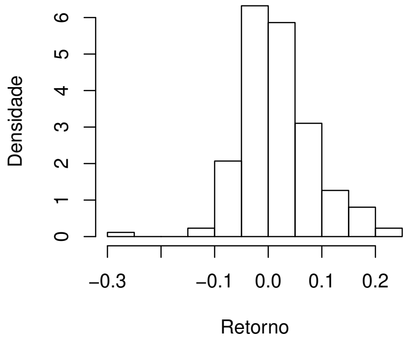
As Figuras 36 e 37 mostram através dos histogramas e Box-Plots que os retornos das variáveis do estudo parecem ter uma grande concentração em torno da média, confirmando os indícios que os dados parecem seguir uma distribuição leptocúrtica, desta maneira, a utilização de distribuições com caudas pesadas é indicada para que os valores discrepantes possam se ajustar melhor ao modelo. Assim, aparentemente, a utilização da distribuição normal não seria a mais indicada.

Como o modelo compara as ações individualmente com o índice IPSA, para uma análise mais profunda dos dados serão considerados apenas as companhias Cuprum e Endesa, devido possuírem a medida de curtose maior do que as demais, indicando seguirem distribuições mais leptocúrticas que as outras companhias, sendo nestes casos, a distribuição NIG a mais indicada. Seus gráficos de dispersão e histogramas marginais podem ser observados nas Figuras (38) e (39) respectivamente.

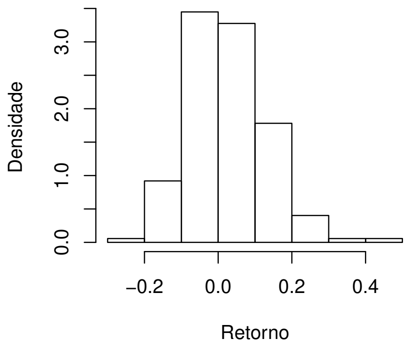
A AFP Cuprum S.A. é uma empresa de gestão de fundos de pensões privadas. Ela é encarregada de administrar uma série de fundos de pensões e investir nos mercados de capital privado e de renda fixa, foi fundada em 1981 e está situada na cidade de Santiago no Chile. Já a Endesa Chile e Empresa Nacional de Electricidad é atualmente conhecida como Enel Generación Chile S.A., e é a maior empresa de energia elétrica do Chile. Foi criada como uma subsidiária da estatal CORFO em 1 de dezembro de 1943 e foi privatizada em 1989, e também está situada na cidade de Santiago.

5.2.2 Estimação e verificação de suposições

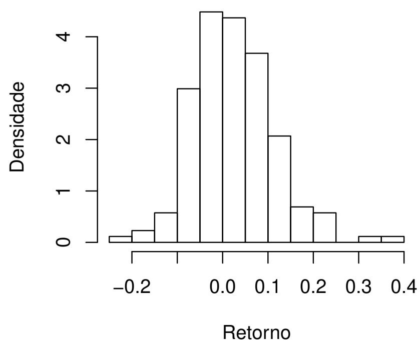
Baseado na análise exploratória realizada anteriormente, comparando os modelos NIG, normal e t-Student para analisar os retornos da companhia Cuprum foram considerados os valores de $\nu = 6$ para a distribuição t-Student e $\omega = 1$ para a distribuição



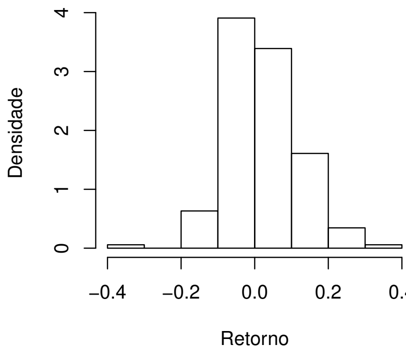
(a) IPSA



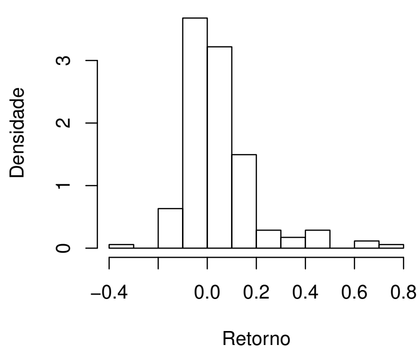
(b) Cervezas



(c) Copec

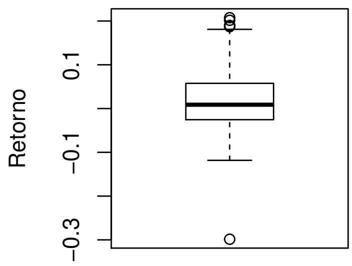


(d) Cuprum

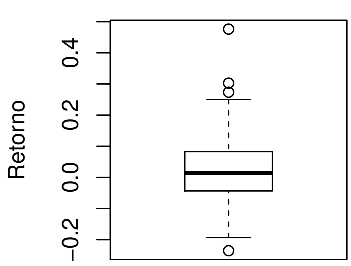


(e) Endesa

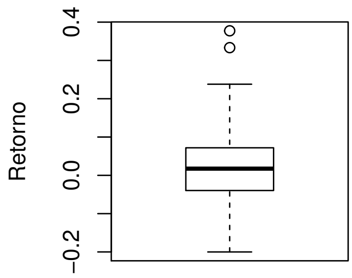
Figura 36 – Histograma dos retornos dos índices IPSA, Cervezas, Copec, Cuprum e Endesa



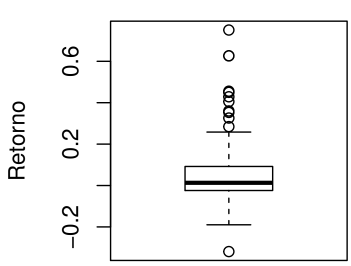
(a) IPSA



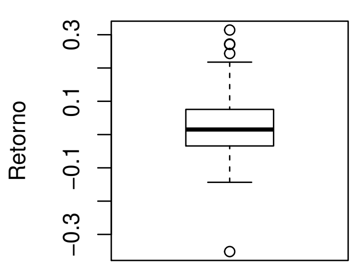
(b) Cervezas



(c) Copec



(d) Cuprum



(e) Endesa

Figura 37 – Box-plot dos retornos dos índices IPSA, Cervezas, Copec, Cuprum e Endesa

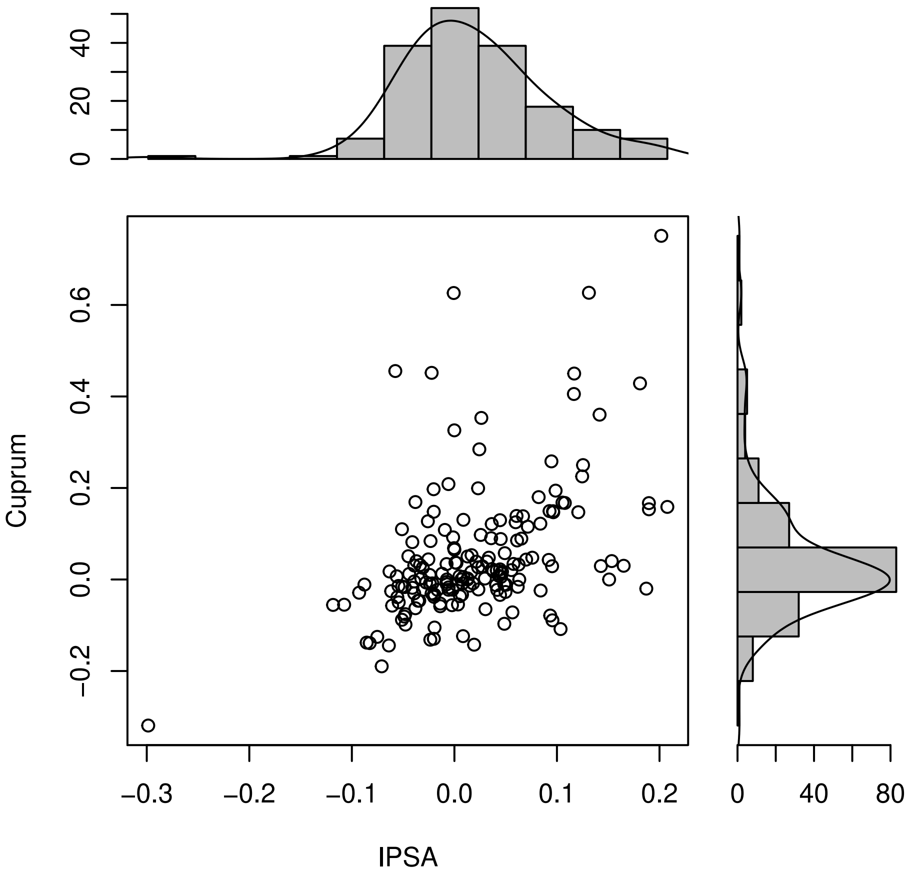


Figura 38 – Gráfico de dispersão e histogramas para IPSA e Cuprum.

NIG quando consideramos o parâmetro ω fixo no modelo. Já para analisarmos os retornos da companhia Endesa foram considerados os valores de $\nu = 8$ para a t-Student e de $\omega = 3$ para a distribuição NIG. Estes valores foram escolhidos para melhor acomodar as observações atípicas que aparecem sob erros normais. Todos os resultados foram obtidos utilizando o software R.

A Tabela 12 contém as estimativas de máxima verossimilhança dos parâmetros considerando os retornos da Cuprum e a Tabela 13 contém as estimativas de máxima verossimilhança dos parâmetros para os retornos da Endesa, considerando os quatro modelos, com seus respectivos erros padrão calculados via bootstrap. Temos que em ambos os casos os parâmetros β e ω possuem a maior variação, e temos que na aplicação da Cuprum o modelo NIG com ω parâmetro desconhecido possui a maior verossimilhança entre os quatro modelos, porém quando comparamos os critérios AIC e BIC temos uma penalização por possuir um parâmetro a mais em seu modelo, sendo assim o modelo NIG com ω pré-fixado obteve melhor ajuste quando analisamos AIC e BIC, por possuir valores menores que os demais. Já na aplicação da Endesa, tivemos que o modelo NIG obteve resultados levemente menores do que o modelo t-Student, porém significativamente

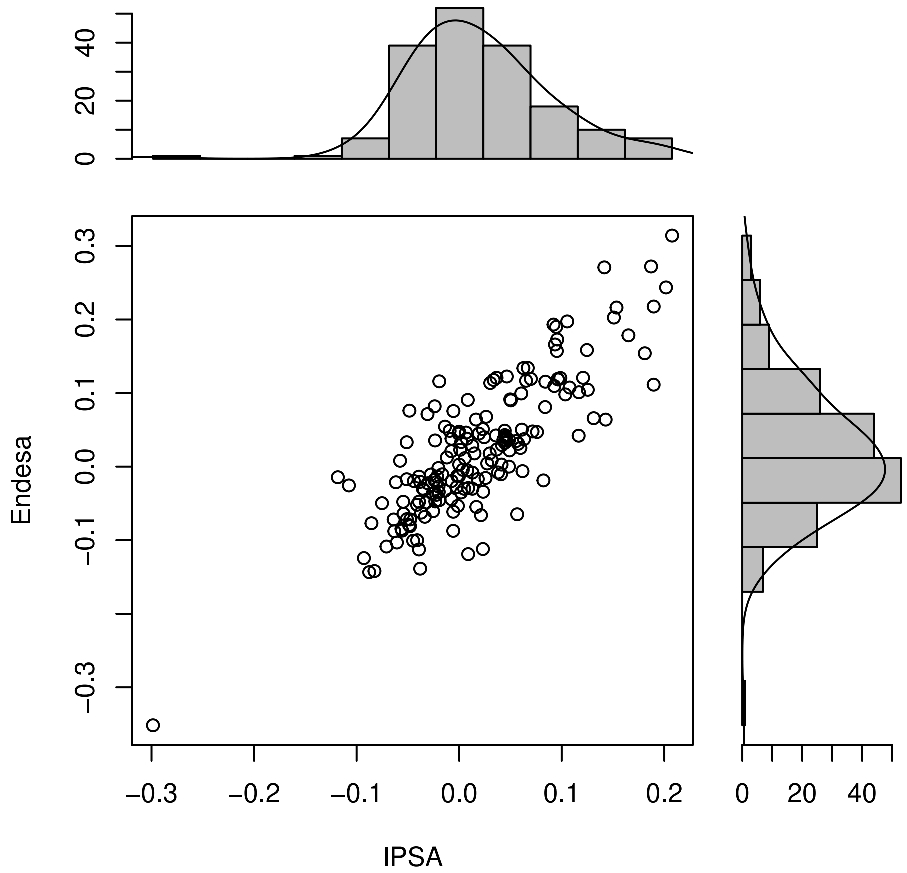


Figura 39 – Gráfico de dispersão e histogramas para IPSA e Endesa.

Tabela 12 – Cuprum: Resultados dos ajustes do modelo utilizando as distribuições normal, t-Student e NIG considerando ω pré-fixado e parâmetro para os retornos das ações.

Parâmetros	Normal		t-Student ($\nu = 6$)		NIG ($\omega = 1$)		NIG	
	Estimativa	DP	Estimativa	DP	Estimativa	DP	Estimativa	DP
α	-0,0213	0,0213	-0,0111	0,0130	-0,0257	0,0234	-0,0258	0,0238
β	3,6458	0,8290	2,7092	0,6112	3,109	0,8504	3,117	0,8575
μ_x	0,0192	0,0053	0,0134	0,0047	0,0146	0,0056	0,0145	0,0059
ϕ_x	0,0013	0,0005	0,0009	0,0003	0,0013	0,0004	0,0014	0,0004
ϕ	0,0035	0,0004	0,0021	0,0002	0,0037	0,0004	0,0037	0,0004
ω	-	-	-	-	-	-	0,8431	0,2813
$\ell(\hat{\boldsymbol{\theta}})$	327,625		361,281		364,569		365,547	
AIC	-645,250		-712,562		-719,138		-719,095	
BIC	-629,454		-696,766		-703,263		-700,141	

melhores do que o modelo normal. Apesar da penalização do parâmetro adicional do modelo NIG, quando analisamos os retornos da Cuprum, que é o caso com maior curtose, temos que comparando os valores de AIC e BIC, o modelo NIG ainda é melhor que o modelo sob a distribuição t-Student.

As Figuras 40 e 41 apresentam o gráfico de dispersão dos dados com a reta

Tabela 13 – Endesa: Resultados dos ajustes do modelo utilizando as distribuições normal, t-Student e NIG considerando ω fixo e parâmetro para os retornos das ações.

Parâmetros	Normal		t-Student ($\nu = 8$)		NIG ($\omega = 3$)		NIG	
	Estimativa	DP	Estimativa	DP	Estimativa	DP	Estimativa	DP
α	-0,0044	0,0038	-0,0005	0,0040	-0,0051	0,0041	-0,0051	0,0042
β	1,3863	0,0722	1,4171	0,0746	1,2924	0,0750	1,2925	0,07447
μ_x	0,0193	0,0049	0,0141	0,0043	0,0165	0,0044	0,0165	0,0052
ϕ_x	0,0038	0,0006	0,0160	0,0005	0,0040	0,0006	0,0040	0,0006
ϕ	0,0010	0,0001	0,0027	0,0001	0,0011	0,0001	0,0011	0,0001
ω	-	-	-	-	-	-	2,9171	3,1328
$\ell(\hat{\boldsymbol{\theta}})$	487,773		492,984		493,044		493,060	
AIC	-965,546		-975,969		-972,366		-970,431	
BIC	-949,750		-960,173		-956,570		-951,470	

de regressão para os quatro modelos analisados para as companhias Cuprum e Endesa respectivamente. Para os retornos da Cuprum o modelo sob distribuição t-Student possui um coeficiente angular menor do que os outros modelos, já para os retornos da Endesa os modelos, visualmente são muito parecidos entre si. É possível observar nestas figuras alguns indivíduos atípicos que poderiam influenciar as EMVs. Na Figura 40, os indivíduos 5, 14 e 104 foram marcados por serem possíveis valores influentes pela abordagem de influência local, já na Figura 41, temos a observação 104, que está presente em ambos os casos.

Foram testadas também algumas hipóteses de interesse do modelo, entre elas,

- $H_{01} : \alpha = 0$ vs $H_{11} : \alpha \neq 0$,
- $H_{02} : \beta = 1$ vs $H_{12} : \beta \neq 1$,
- $H_{03} : \alpha = 0$ e $\beta = 1$ vs $H_{13} : \alpha \neq 0$ e $\beta \neq 1$.

Pelas Tabelas 14 e 15 temos os resultados do testes de hipóteses sob a distribuição NIG considerando um nível de significância de 5%, e apenas quando analisamos a hipótese de intercepto nulo, $H_{01} : \alpha = 0$ contra $H_{11} : \alpha \neq 0$, para o modelo de retornos da Cuprum, que não temos evidências para rejeitar a hipótese nula, assim, temos evidências de intercepto nulo no modelo da Cuprum. Porém, quando analisamos simultaneamente $H_{03} : \alpha = 0$ e $\beta = 1$ contra $H_{13} : \alpha \neq 0$ e $\beta \neq 1$ temos evidências de rejeitar H_0, da mesma maneira que ocorre com as outras hipóteses tanto da Cuprum como da Endesa. Desta forma, temos evidências que $\alpha \neq 0$ e $\beta \neq 1$ para os modelos propostos.

Além do modelo NIG, foram realizados os testes de hipóteses para todos os modelos analisados e em todos o casos obtivemos os mesmo resultados. Assim, para a aplicação dos dados da Cuprum, temos que nosso modelo será dado por

$$
\begin{aligned}
Y_i &= \beta x_i + e_i, \\
X_i &= x_i + u_i, \qquad i = 1, \ldots, 174.
\end{aligned}
$$

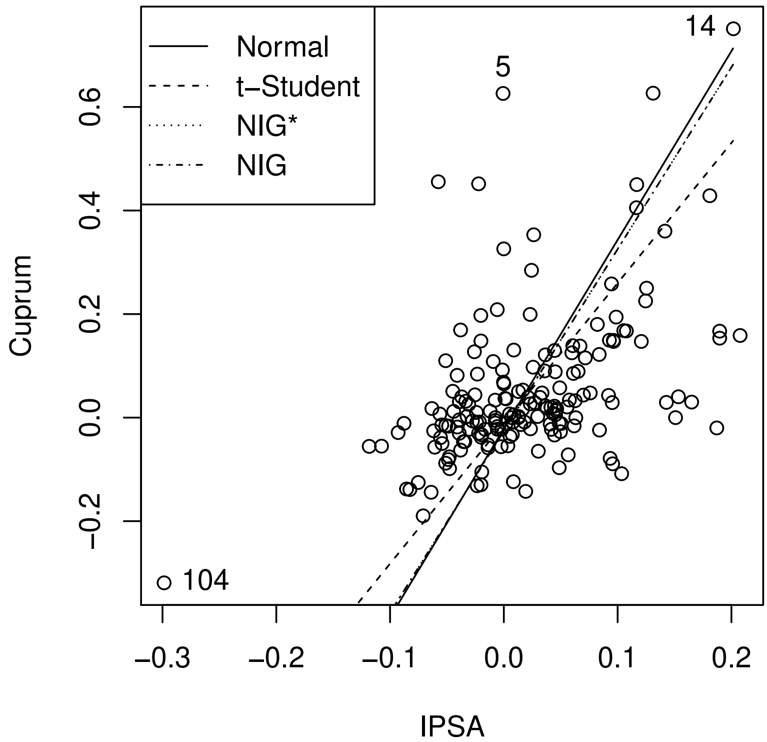


Figura 40 – Modelos de regressão lineares ajustados para a distribuição normal, t-Student, NIG com ω fixo (NIG*) e com ω parâmetro desconhecido para as variáveis IPSA e Cuprum.

Tabela 14 – Cuprum: Teste de hipóteses para o modelo.

H_0	estatística Q	p-valor	Resultado
$\alpha = 0$	1,1579	0,2818	Não rejeitamos H_0
$\beta = 1$	61,6395	≈ 0	Rejeitamos H_0
$\alpha = 0$ e $\beta = 1$	250,3158	≈ 0	Rejeitamos H_0

Tabela 15 – Endesa: Teste de hipóteses para o modelo.

H_0	estatística Q	p-valor	Resultado
$\alpha = 0$	78,8927	≈ 0	Rejeitamos H_0
$\beta = 1$	72,2183	≈ 0	Rejeitamos H_0
$\alpha = 0$ e $\beta = 1$	69,7581	≈ 0	Rejeitamos H_0

Na Figura 42 temos para a empresa Cuprum os gráficos Q-Q plots e seus envelopes que são baseados na distribuição da distância de Mahalanobis de $\mathbf{Z}_i$, para $i = 1, ..., n$, que para a distribuição normal segue uma distribuição $\chi^2_{(2)}$ e para a distribuição t-Student segue uma distribuição $2F(2, \nu)$. Para a distribuição NIG o gráfico quantil-quantil foi obtido via simulação. As linhas nas figuras representam o 5º percentil, a média e 95º

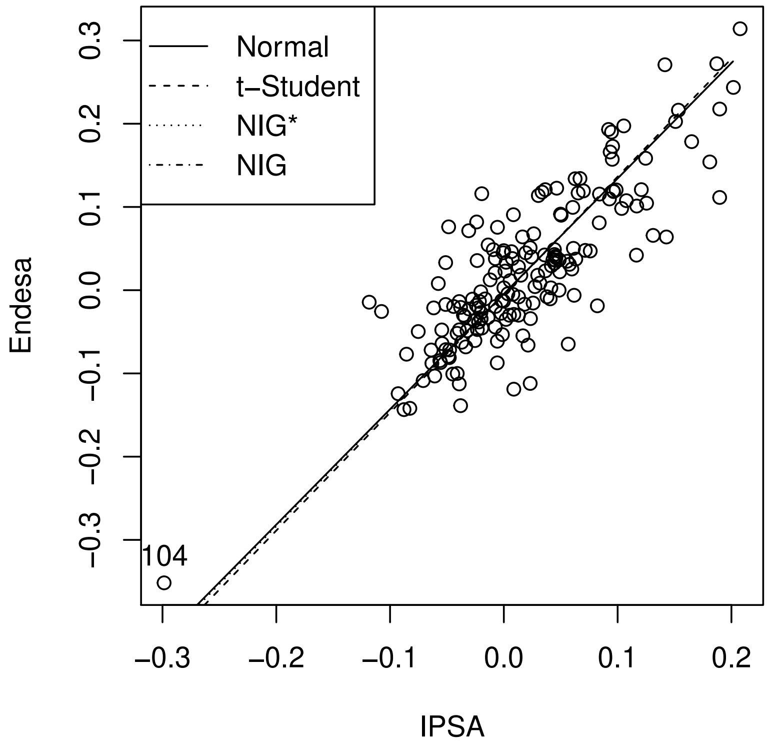
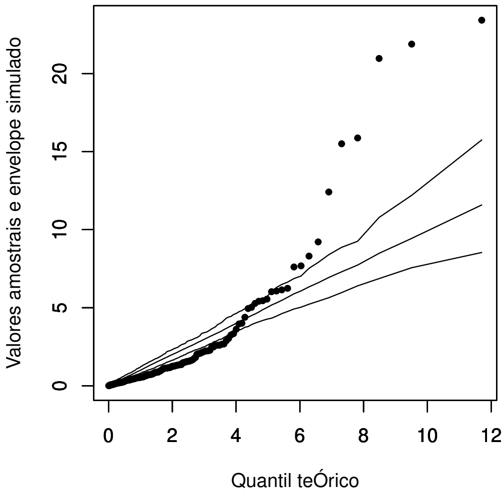


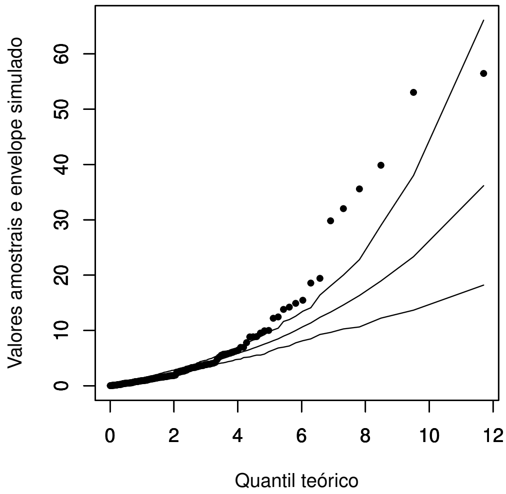
Figura 41 – Modelos de regressão lineares ajustados para a distribuição normal, t-Student, NIG com ω fixo (NIG*) e com ω parâmetro desconhecido para as variáveis IPSA e Endesa.

percentil de 100 pontos simulados para cada observação Os gráficos mostram que a distribuição normal é a que pior se ajusta aos dados e que a distribuição t-Student também não parece se ajustar muito bem, com muitas observações fora dos limites do envelope, já a distribuição NIG parece se ajustar bem aos dados. Mais uma vez, estes gráficos mostram que a distribuição NIG fornece um melhor ajuste para o conjunto de dados, até mesmo que a distribuição t-Student. Na Figura 43 para Endesa temos novamente indícios que o modelo com NIG fornece um melhor ajuste que um modelo normal, porém desta vez a diferença entre o modelo NIG e t-Student não é tão discrepante.

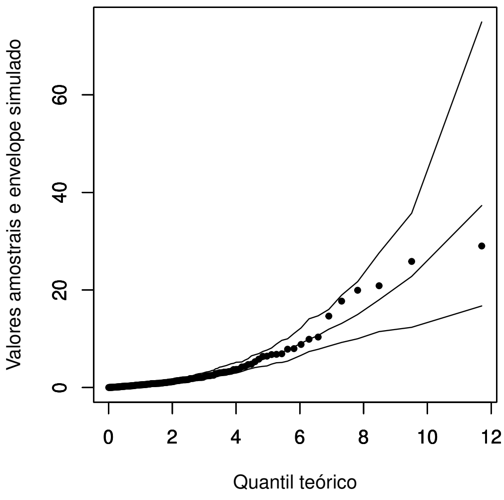
Nas Figuras 44 e 45 podemos ver os gráficos de Komogorov-Smirnoff para os quatros modelos considerando os retornos de Cuprum e Endesa respectivamente, em que a linha preta representa a fda empírica para os dados originais e as linhas cinzas representam a fda empírica para cada um dos 500 conjuntos de dados simulados do modelo ótimo. A Figura 44 ilustra que os modelos normal e t-Student fornecem um ajuste razoável aos dados, com uma tendência de superestimar a densidade na cauda esquerda e subestimar a densidade na cauda direita, já para a distribuição NIG temos que o modelo se ajusta bem aos dados, mostrando ser um ajuste melhor do que os demais modelos. A figura 45



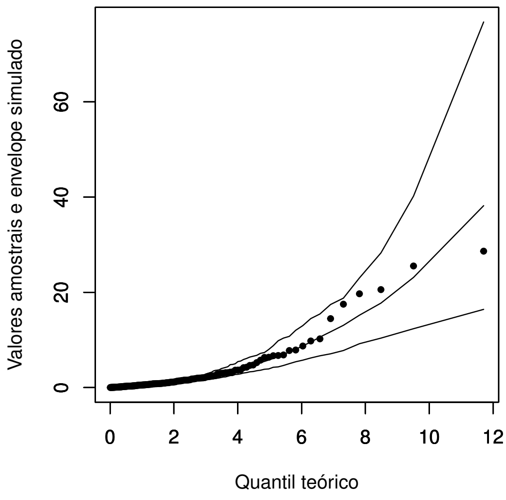
(a) Normal



(b) t-Student

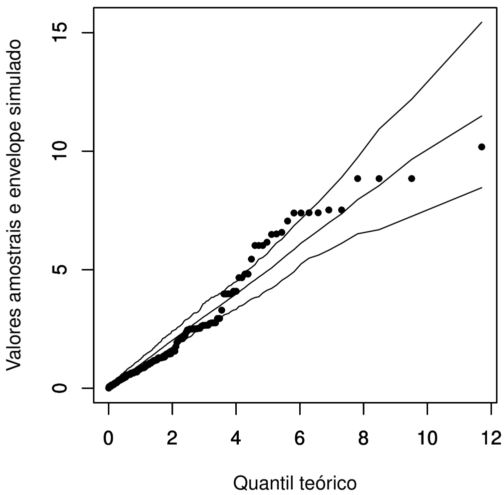


(c) Normal Inversa Guassiana com ω fixo

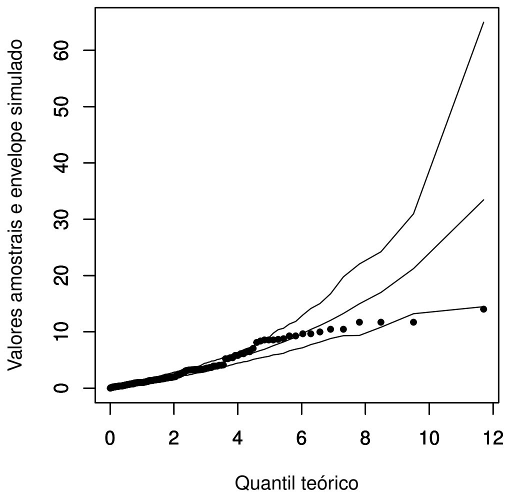


(d) Normal Inversa Guassiana com ω parâmetro

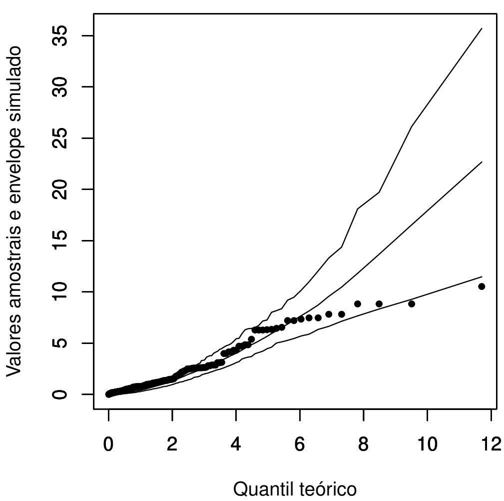
Figura 42 – Cuprum: Q-Q Plots e envelopes simulados para o modelo ajustado considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.



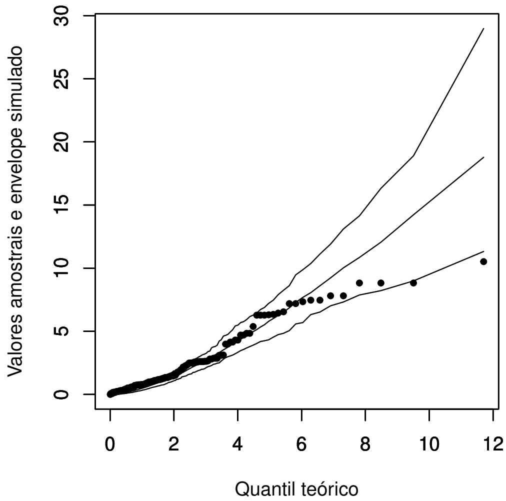
(a) Normal



(b) t-Student

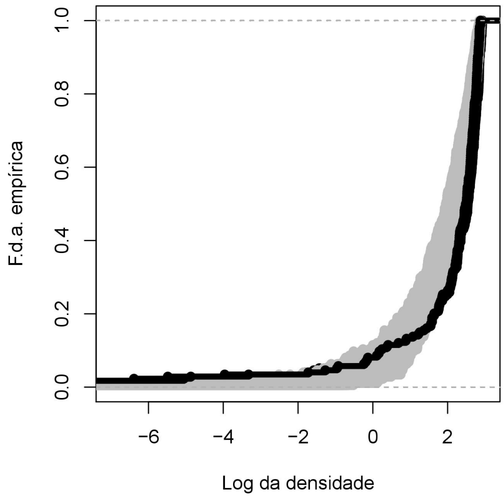


(c) Normal Inversa Guassiana com ω fixo

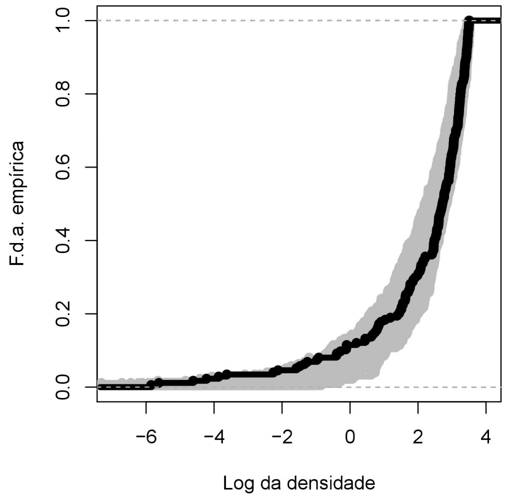


(d) Normal Inversa Guassiana com ω parâmetro

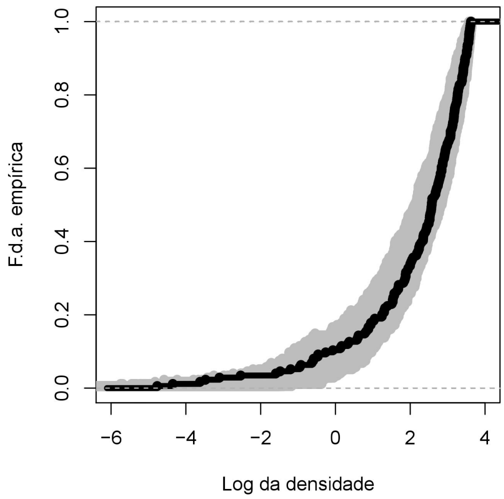
Figura 43 – Endesa: Q-Q Plots e envelopes simulados para o modelo ajustado considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.



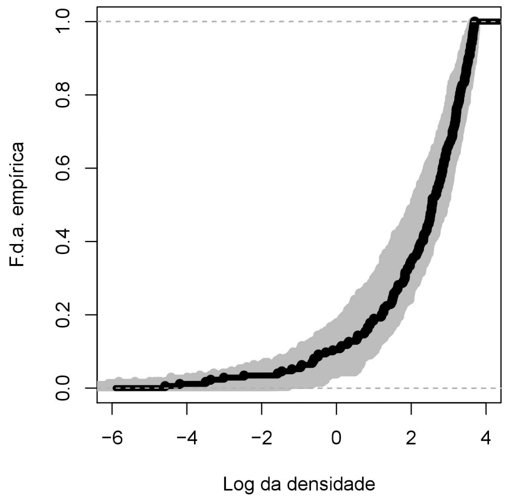
(a) Normal



(b) t-Student



(c) Normal Inversa Guassiana com ω fixo

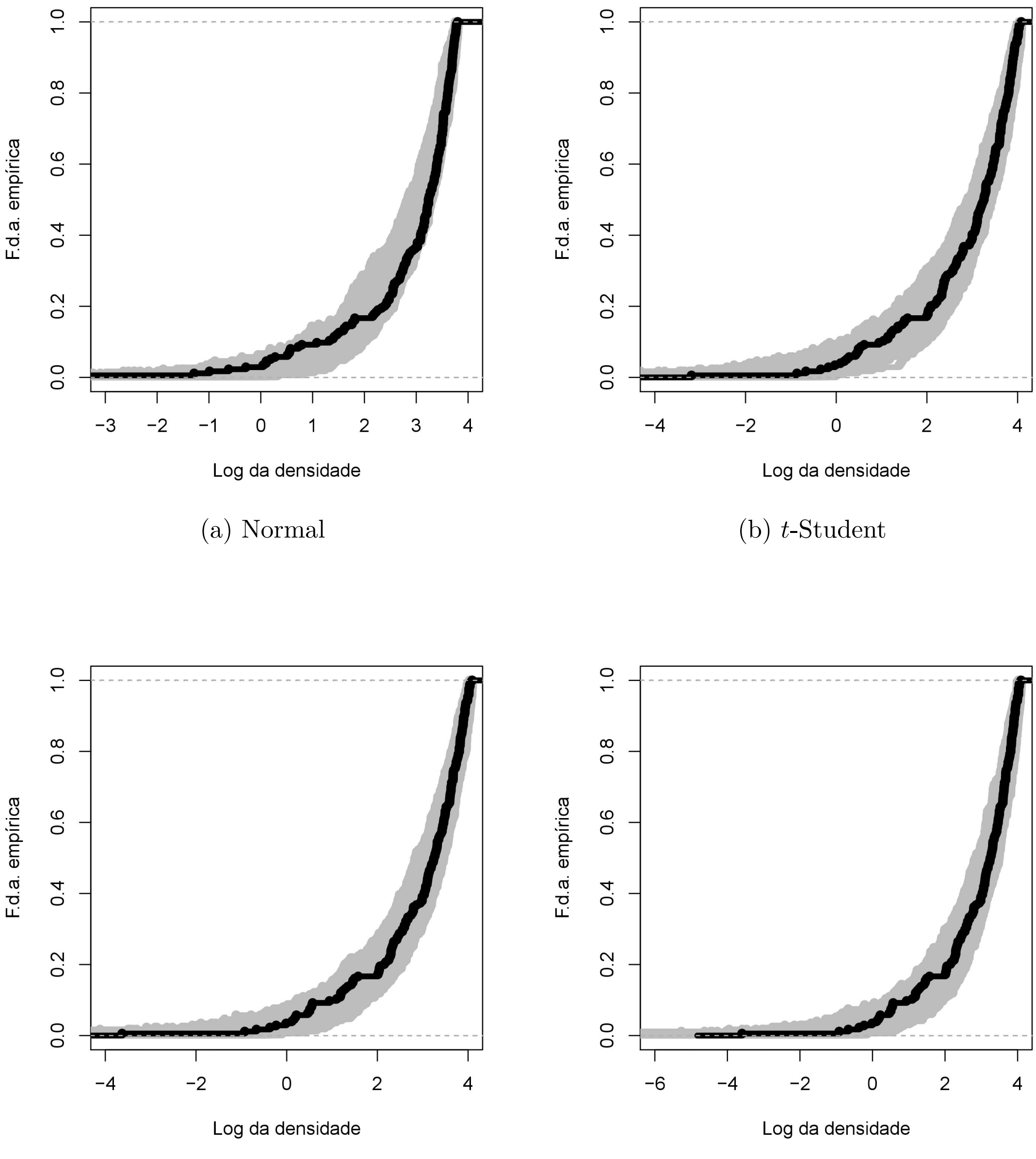


(d) Normal Inversa Guassiana com ω parâmetro

Figura 44 – Cuprum: Gráfico de Kolmogorov-Smirnoff para o modelo sob distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.

ilustra que o modelo normal fornece um ajuste razoável aos dados, com uma tendência de superestimar a densidade na cauda esquerda e subestimar a densidade na cauda direita, já para as distribuições t-Student e NIG temos que o modelo se ajusta bem aos dados, se mostrando ser um ajuste melhor que a distribuição normal.

As Figuras 46 e 47 exibem as distâncias de Mahalanobis para os quatro modelos ajustados para os retornos de Cuprum e Endesa respectivamente, e a partir destas figuras, podemos detectar como as principais possíveis observações atípicas para a companhia



(a) Normal

(b) t-Student

(c) Normal Inversa Guassiana com ω fixo

(d) Normal Inversa Guassiana com ω parâmetro

Figura 45 – Endesa: Gráfico de Kolmogorov-Smirnoff para o modelo sob distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.

Cuprum as observações 5, 6, 14, 18, 22, 37, 49, 104 e 107, e para a companhia Endesa as observações 11, 19, 48, 51, 104, 121, 124 e 172. Do algoritmo EM, os pesos estimados $\hat{q}_i$ das observações 5 e 14 para Cuprum e 104 para Endesa são as menores para os modelos t-Student e NIG, sendo a linha tracejada representa o peso dado pela distribuição normal, como pode ser visto nas Figuras 48 e 49. Podemos ver que a estimativa de $\boldsymbol{\theta}$ tende a atribuir mais peso q_i para as observações com distância de Mahalanonis pequena quando temos um modelo sob a distribuição NIG com ω parâmetro, atribuindo portanto, menos peso para as observações atípicas. O modelo sob distribuição NIG com ω parâmetro possui também o peso v_i, em que a estimativa de $\boldsymbol{\theta}$ tende a dar peso maior para as observações distantes devido ao fato de que v_i ser diretamente proporcional a distância de Mahalanobis.

De maneira geral, ambos os modelos NIG, com parâmetro pré-fixado e com ω parâmetro desconhecido, se ajustara bem aos dados, principalmente nos casos em que temos um valor elevado de curtose, os mesmo se mostraram melhor até mesmo quando comparada a distribuição t-Student. Assim, sendo extremamente indicada para dados que seguem distribuições leptocúrticas. Apesar da penalização do parâmetro adicional do modelo NIG, quando analisamos os retornos da Cuprum, que é o caso com maior curtose, temos que comparando os valores de AIC e BIC, o modelo NIG ainda é melhor que o modelo sob a distribuição t-Student. Em todos os casos os modelos NIG se motraram melhores do que o modelo normal.

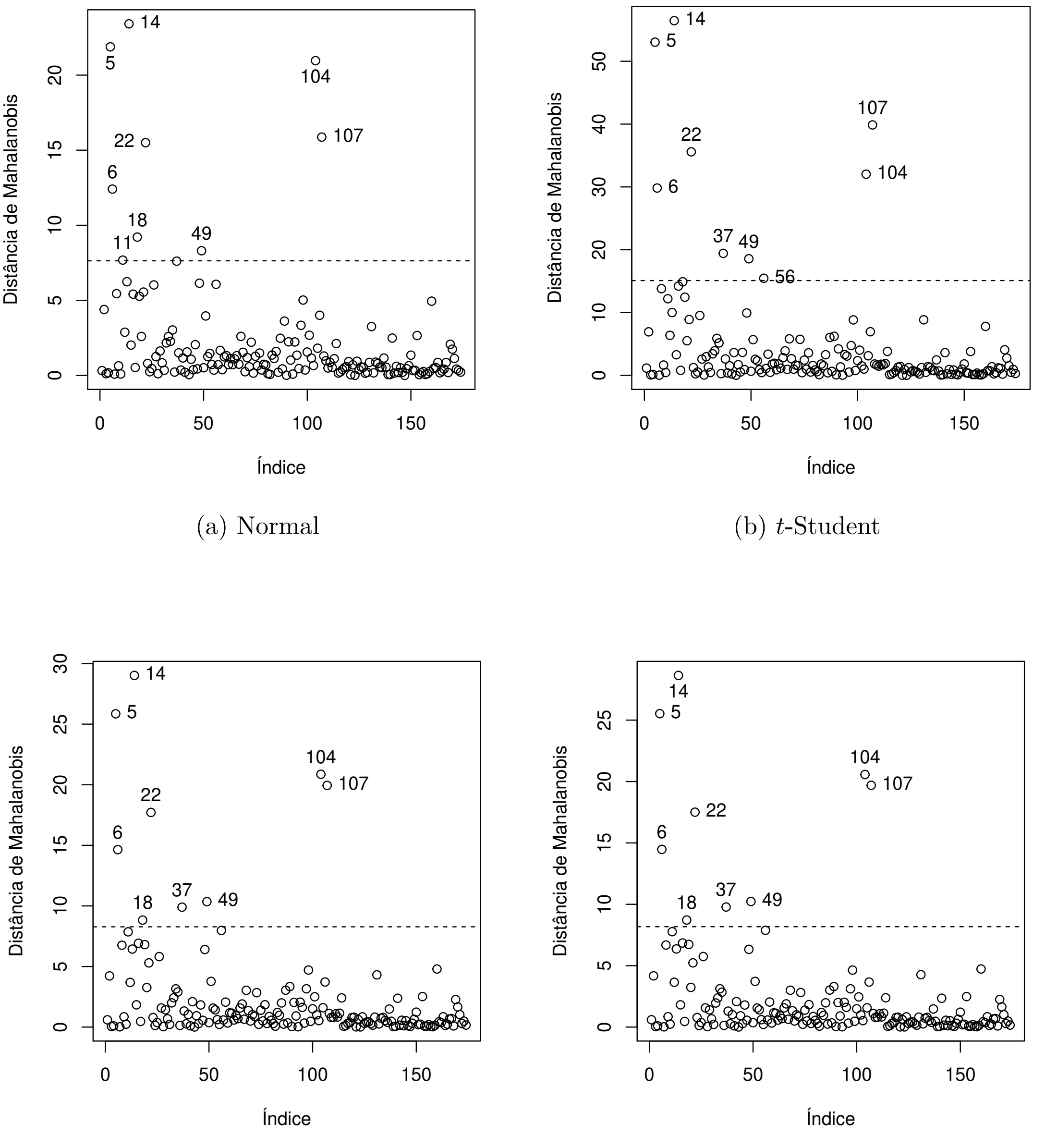
5.2.3 Análise de diagnóstico

A análise de diagnóstico dos modelos é dado pelas Figuras 50 e 51, que exibem as distâncias de Cook para os quatro modelos ajustados para Cuprum e Endesa respectivamente, e a partir destas figuras, podemos detectar novamente como as principais observações atípicas da Cuprum as observações 5, 22 e 104, e para Endesa novamente as observações 48, 51 e 104. Podemos ver que a estimativa de $\boldsymbol{\theta}$ sob o modelo normal é mais sensível a valores discrepantes, enquanto os modelos t-Student e NIG acomodam melhor esses valores do que a distribuição normal.

As observações atípicas detectados pela distância de Cook podem ser vinculadas a acontecimentos importantes para a história chilena ou então acontecimentos relevantes para as empresas analisadas, e podem ser observados abaixo,

- **Observação 5 (março 1990):** Em 11 de março de 1990, o General Augusto Pinochet entregou a presidência do Chile a Patricio Aylwin, marcando o fim de seu governo, foi de 1973 e 1990.

- **Observação 22 (outubro 1991):** Em 1991 a AFP Cuprum foi adquirida pelo Grupo Penta, que é um holding empresarial chileno, com investimentos nas áreas de previdência, seguros, finanças, saúde, imobiliária e educação.



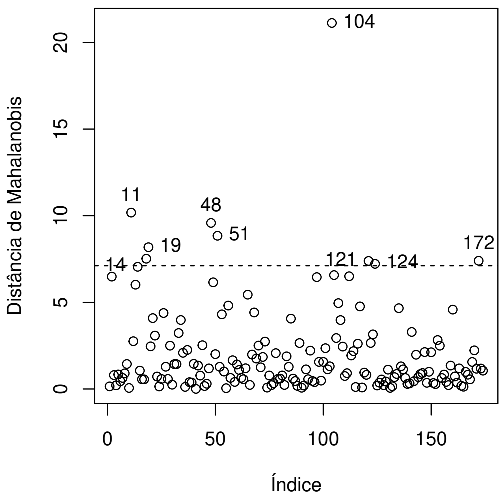
(a) Normal

(b) *t*-Student

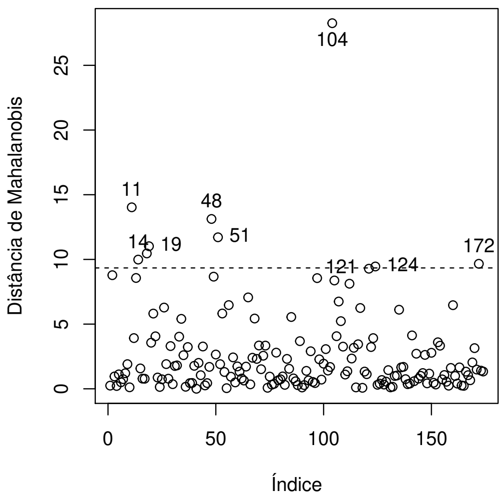
(c) Normal Inversa Guassiana com ω fixo

(d) Normal Inversa Guassiana com ω parâmetro

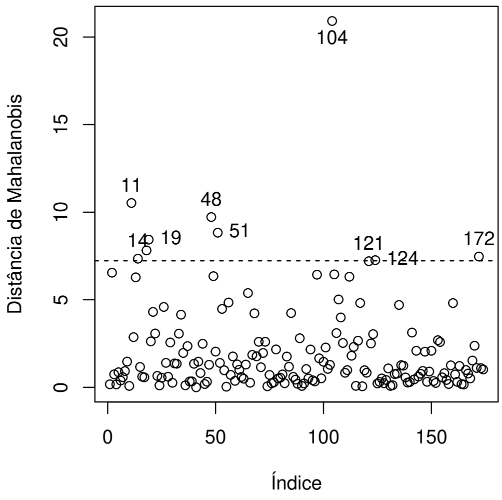
Figura 46 – Cuprum: Gráficos da distância de Mahalanobis considerando a distribuição normal, *t*-Student, NIG com ω fixo e com ω parâmetro.



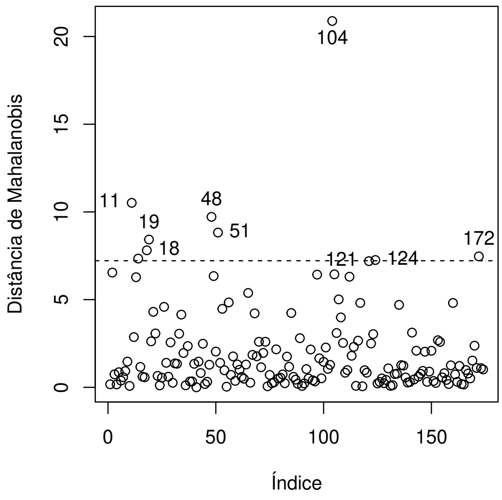
(a) Normal



(b) t-Student

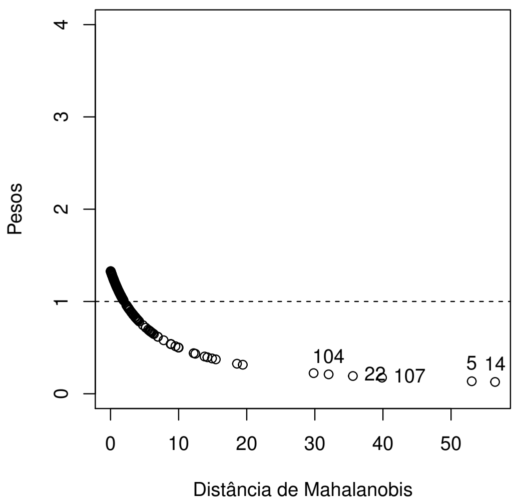


(c) Normal Inversa Guassiana com ω fixo

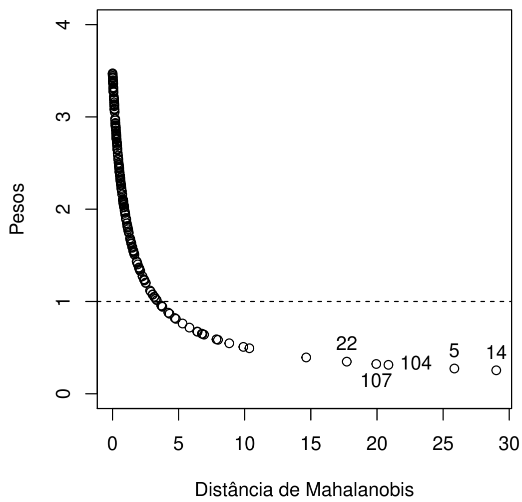


(d) Normal Inversa Guassiana com ω parâmetro

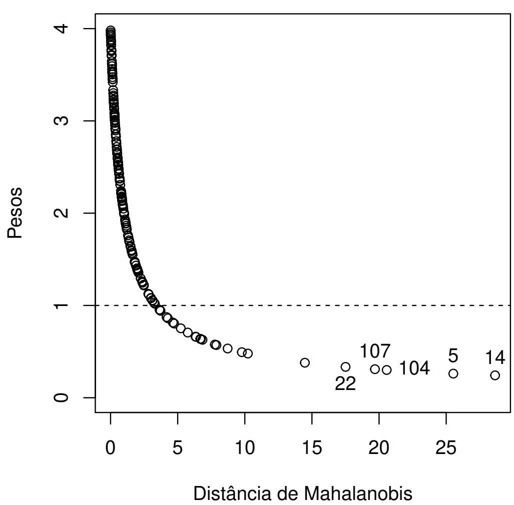
Figura 47 – Endesa: Gráficos da distância de Mahalanobis considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.



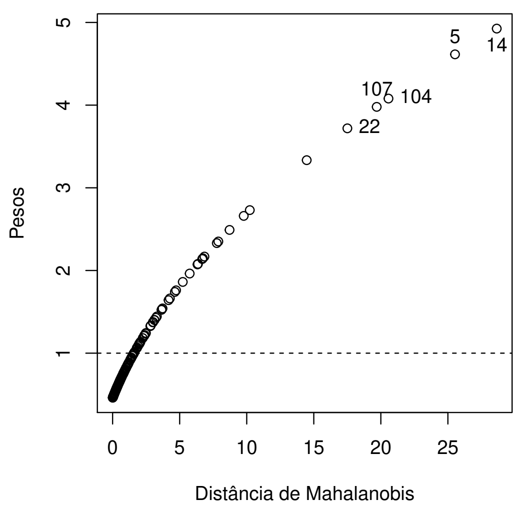
(a) t-Student (q_i)



(b) NIG com ω fixo (q_i)

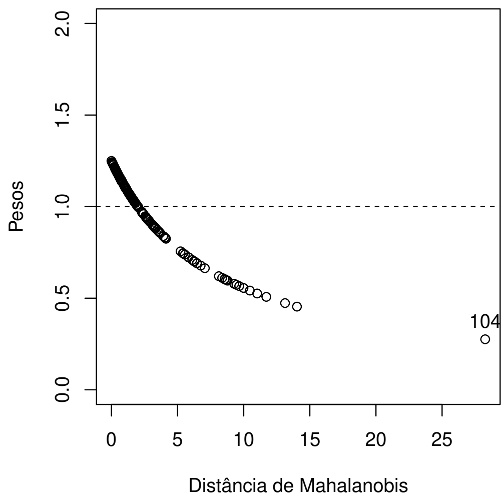


(c) NIG com ω parâmetro (q_i)

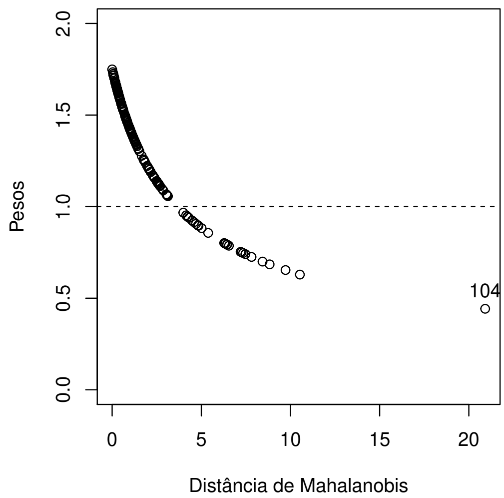


(d) NIG com ω parâmetro (v_i)

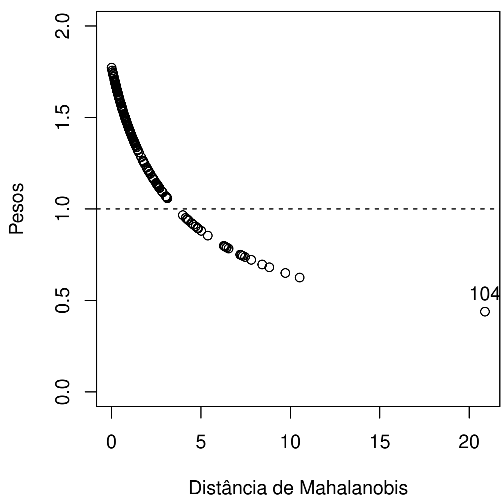
Figura 48 – Cuprum: Estimativas de q_i para o modelo sob distribuição t-Student, NIG com ω fixo e com ω parâmetro e v_i para o modelo sob distribuição NIG com ω parâmetro.



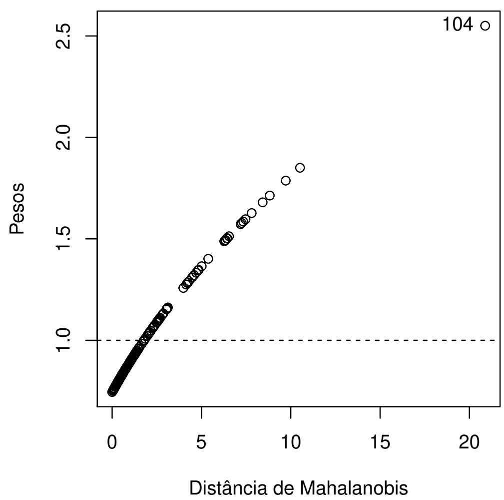
(a) t-Student (q_i)



(b) NIG com ω fixo (q_i)



(c) NIG com ω parâmetro (q_i)



(d) NIG com ω parâmetro (v_i)

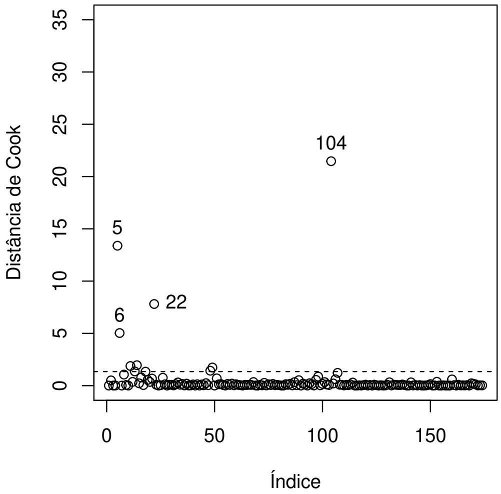
Figura 49 – Endesa: Estimativas de q_i para o modelo sob distribuição t-Student, NIG com ω fixo e com ω parâmetro e v_i para o modelo sob distribuição NIG com ω parâmetro.

- **Observação 48 (dezembro 1993):** No sábado de 11 de dezembro de 1993 ocorreram as eleições presidenciais do Chile, tendo como presidente eleito o engenheiro Eduardo Frei Ruiz-Tagle, para o período de 1994 a 2000.

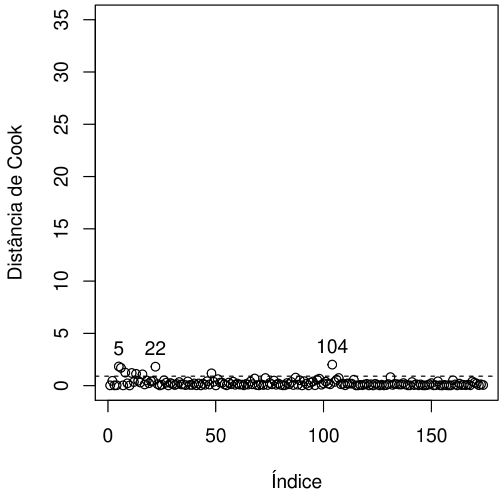
- **Observação 51 (março 1994):** Em 11 de março de 1994, Eduardo Frei Ruiz-Tagle assume a presidência da República do Chile, cargo que assume até 11 de março de 2000. Foi o segundo presidente eleito após saída de Augusto Pinochet do poder.

- **Observação 104 (agosto 1998):** Em agosto de 1998 estourou a crise financeira da Rússia, contaminada pelos efeitos do enfraquecimento das economias dos tigres asiáticos, em meio à redução do crédito no mercado internacional. Em 17 de agosto de 1998 ocorreu a moratória russa, que foi o ápice da crise, afetando não só o Chile como toda a América Latina.

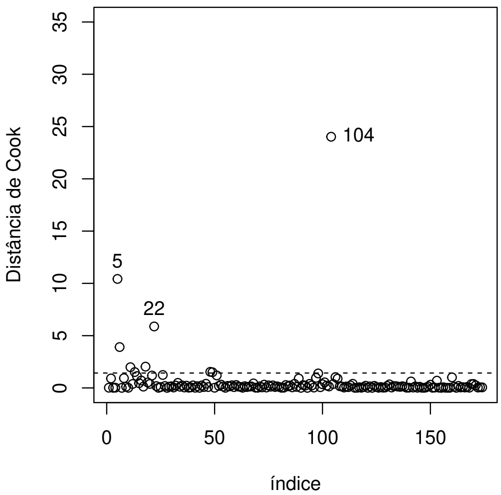
A observação 104, foi um acontecimento internacional que teve forte impacto na economia do Chile, e por isso se mostrou um valor atípico para as duas empresas analisadas. As observações 5, 48 e 51 estão vinculadas ao fim do governo Pinochet e à presidência do Chile, que interferem nos setores previdenciários e elétrico, já a observação 22 foi um fato específico que impactou apenas as ações da Cuprum. Desta maneira, temos indícios que a detecção de valores aberrantes realmente foi feita de maneira correta.



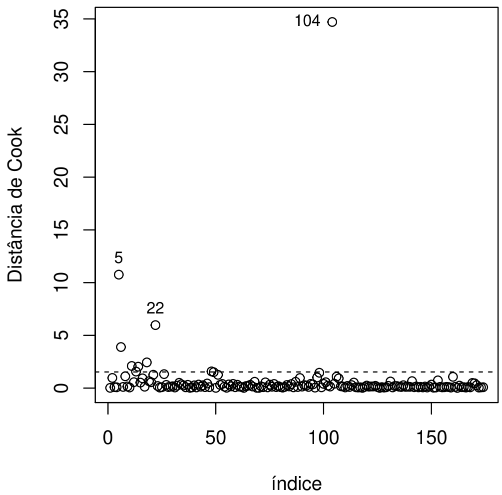
(a) Normal



(b) t-Student

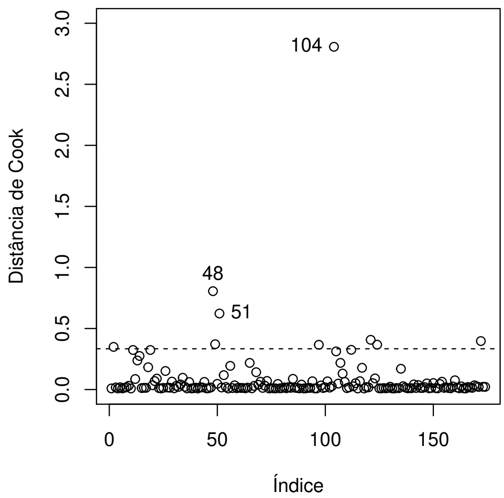


(c) Normal Inversa Guassiana com ω fixo

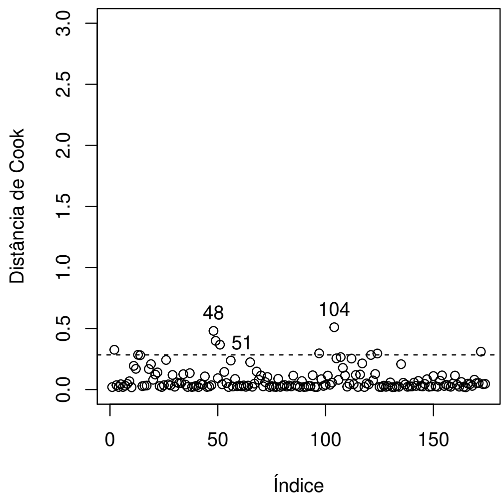


(d) Normal Inversa Guassiana com ω parâmetro

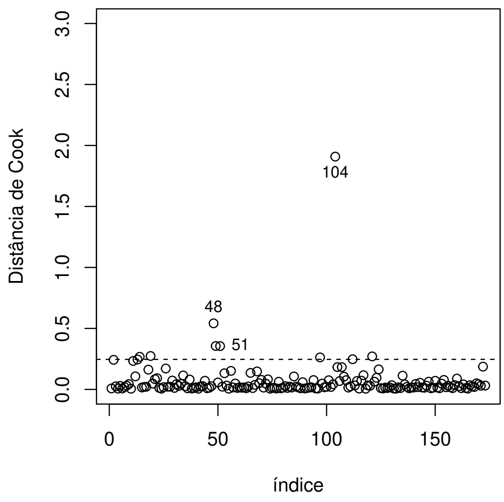
Figura 50 – Cuprum: Gráficos da distância de Cook considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.



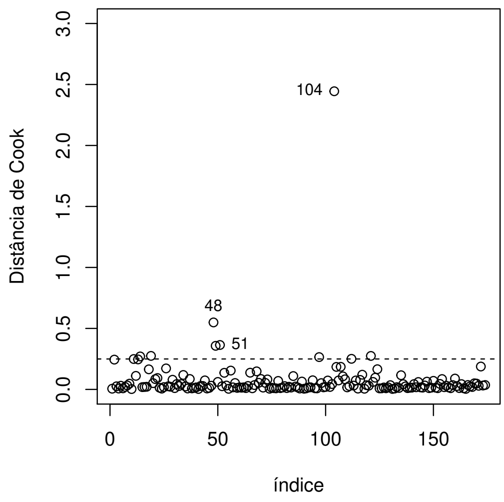
(a) Normal



(b) t-Student



(c) Normal Inversa Guassiana com ω fixo



(d) Normal Inversa Guassiana com ω parâmetro

Figura 51 – Endesa: Gráficos da distância de Cook considerando a distribuição normal, t-Student, NIG com ω fixo e com ω parâmetro.

6 Considerações Finais

Neste trabalho, desenvolvemos uma extensão do modelo com erro nas variáveis utilizando a distribuição normal inversa gaussiana simétrica, foram considerados tanto o parâmetro ω pré-fixado, como também parâmetro desconhecido do modelo. Aqui, apresentamos as distribuições Inversa Gaussiana e Hiperbólica Generalizada e apontamos algumas propriedades importantes que nos permitem definir o modelo com erro nas variáveis sob NIG simétrica. O modelo proposto foi comparado com as distribuições normal e t-Student.

Para o modelo com erro nas variáveis sob distribuição NIG simétrica, propusemos um estimador utilizando o algoritmo EM e como valores iniciais para o cálculo iterativo das estimativas de máxima verossimilhança utilizamos os estimadores obtidos pelo método dos momentos para o modelo sob distribuição normal. Esses estimadores de máxima verossimilhança foram desenvolvidos seguindo mesmo conceito de Lachos (2011).

O desempenho do modelo proposto foi avaliado através de um estudo de simulação, e verificamos a convergência do estimador de máxima verossimilhança. Os aspectos teóricos do modelo propostos neste trabalho foram ilustrados através de dois conjunto de dados reais, o primeiro analisando a relação entre IMC e gordura corporal de atletas australianos e o segundo analisando o retorno de ações do mercado chileno.

Como forma de comparação dos modelos foram utilizados os Q-Q plots, a distância de Mahalanobis e teste de bondade de ajuste de Kolmogorov-Smirnov multivariado (O'Hagan et al., 2014). Além disso, foi utilizado a distância de Cook (1986) como método de análise de diagnóstico para os modelos. O modelo NIG simétrico proposto pareceu se ajustar melhor aos dados quando temos dados bastante leptocúrticos, como é o caso dos dados econômicos.

Em termos de pesquisas futuras, podemos considerar o modelo sob distribuição NIG assimétrica e outros conjuntos de dados para complementar os resultados atingidos. Podemos considerar também o modelo de Barnett (1969), que é definido pelas equações

$$\begin{aligned} Y_{1j} &= \alpha_1 + \beta_1 x_j + e_{1j}, \\ Y_{2j} &= \alpha_2 + \beta_2 x_j + e_{2j}, \\ &\vdots \\ Y_{pj} &= \alpha_p + \beta_p x_j + e_{pj}, \end{aligned} \tag{6.1}$$

com $j = 1, \ldots, n$.

Este modelo com erros nas variáveis é conhecido como modelo de calibrações comparativas, que é usado para comparar métodos ou instrumentos de medição. Para $p = 2$, $\alpha_1 = 0$ e $\beta_1 = 1$, este modelo corresponde ao modelo de regressão simples usual com erros nas variáveis. Neste caso, dizemos que estamos comparando dois instrumentos de medições, um chamado de standard ou de precisão, que corresponde a Y_1, com um outro Y_2, que corresponde a um instrumento alternativo. Neste contexto é importante estudar se $\alpha_2 = 0$ e $\beta_1 = 1$ e neste caso os dois instrumentos são equivalentes em termo de vício aditivo (α_2) e vício multiplicativo (β_2), em que o algoritmo EM para analisar estas hipóteses já foi apresentado no Capítulo 3.

Assumindo que existe um instrumento standard de medição ($\alpha_1 = 0$ e $\beta_1 = 1$), o modelo em (6.1) pode ser escrito por

$$Y_j = \mathbf{a} + \mathbf{b}x_j + e_j,$$

onde $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top$, $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top$ e $e_i = (e_1, \ldots, e_p)^\top$, com $\boldsymbol{\alpha} = (\alpha_2, \ldots, \alpha_p)^\top$ e $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_p)^\top$. O modelo de calibração também é caracterizado em estrutural e funcional, como no caso do modelo com erro nas variáveis simples.

Para o caso estrutural, os x_i's são v.a., e portanto a distribuição do modelo é especificada ao considerar o modelo definido por

$$Y_j = \mathbf{a} + \mathbf{b}\,\mathbf{r}_j,$$

em que

$$\mathbf{r}_j = \begin{pmatrix} x_j \\ e_j \end{pmatrix} \sim NIGS_{p+1}(\boldsymbol{\eta}_x, D(\boldsymbol{\phi}_x); \omega), \; j = 1, \ldots, n,$$

com $\boldsymbol{\eta}_x = (\mu_x, \mathbf{0}^\top)^\top$ e $D(\boldsymbol{\phi}_x)$ é a matriz diagonal elementos no vetor $\boldsymbol{\phi}_x = (\phi_x, \phi_1, \ldots, \phi_p)^\top$. Pelas propriedades da distribuição NIG simétrica

$$\mathbf{Y}_i \sim NIGS_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \omega),$$

em que $\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x$ e $\boldsymbol{\Sigma} = D(\boldsymbol{\phi}) + \phi_x \mathbf{b}\mathbf{b}^\top$, onde $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$.

Assim, sejam $\mathbf{y}_1, \ldots, \mathbf{y}_n$ observações, a função de log-verossimilhança para $\mu_x, \phi_x, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}$ é dada por $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$, em que

$$\ell_i(\theta) = -\frac{p+1}{4}[\log(\omega + d(\mathbf{y}_i)) - \log \omega] + \log K_{(-p-1)/2}(\sqrt{\omega(\omega + d(\mathbf{y}_i))}\,) - \frac{p}{2}\log 2\pi - \frac{1}{2}\log |\boldsymbol{\Sigma}|,$$

com $d(\mathbf{y}_i) = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$.

A utilização do modelo de Barnett, permite uma análise mais completa, como no caso do conjunto de dados estudado no Capítulo 5, ao invés de comparar cada ação individualmente com o índice IPSA, poderia ser feita uma análise simultânea de todas as ações.

Referências

ABRAMOWITZ, M.; STEGUN, I.A. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. *Dover Publ., Inc., New York*, 1972. Citado na página 41.

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control, Boston,* v.19, n.6, p.716-723, 1974. Citado na página 77.

ARELLANO-VALLE, R.B. Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão. *Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística (IME), USP*, São Paulo, 1994. Citado na página 19.

ARELLANO-VALLE, R.B.; BOLFARINE, H. Elliptical structural models. *Communications in Statistics. Theory Methods*, v. 25, n. 10, p. 2319-2342, 1996. Citado na página 20.

ARELLANO-VALLE, R.B.; BOLFARINE, H.; LACHOS, Skew-normal linear mixed models. *JDataSci*, v. 3, p. 415– 438, 2005a. Citado na página 52.

AZZALINI, A. A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, v. 12, n. 2, p. 171-178, 1985. Citado 2 vezes nas página 17 e 18.

BARNDORFF-NIELSEN, O.E. Information and exponential families: in statistical theory. *John Wiley & Sons*, New York, 1978. Citado 3 vezes nas páginas 16, 21 e 22.

BARNDORFF-NIELSEN, O.E. Processes of normal inverse gaussian type. *Finance and stochastics*, v. 2, n. 1, p. 41–68, 1997. Citado na página 22.

BARNDORFF-NIELSEN, 0.E.; STELZER, R. Multivariate supOU process. *Annals of Applied Probability*, v. 21, n. 1, p.140-182, 2011. Citado na página 35.

BARNETT, V. Simultaneous pairwise linear structural relationships. *Biometrics*, v. 25, n. 1, p. 129-42, 1969. Citado 2 vezes nas páginas 19 e 113.

BEDRICK, E.J. An efficiente scores test for comparing several measuring devices. *Journal of Quality Technology*, v. 33, n. 1, p. 96-103, 2001. Citado na página 19.

BLAESILD, P. The Two-Dimensional Hyperbolic Distribution and Related Distributions, with an Application to Johannsen's Bean Data. *Biometrika*, v. 68, n. 1, 1981. Citado 3 vezes nas páginas 34, 35 e 41.

BOLFARINE, H; GALEA-ROJAS, M. Maximum likelihood estimation ofsimultaneous pairwise linear structural ralationships. *Biometrical Journal*, v. 37, n. 6, p. 673-689, 1995. Citado na página 19.

BOZDONGAN, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika.* v.52, n.3, p. 345-370, 1987. Citado na página 77.

BROWNE, R.P.; MCNICHOLAS, P.D. A mixture of generalized hyperbolic distributions. The Canadian Journal of Statistics, v. 43, n. 2, p. 176–198, 2015. Citado 2 vezes nas páginas 40 e 44.

CHHIKARA R.S.; FOLKS J.L. Estimation of the inverse Gaussian distribution function. *Journal of the American Statistical Association*, v. 69, n. 345, p. 2500–254, 1974. Citado na página 29.

CHRISTENSEN, R.; BLACKWOOD, L. G.Tests for precision and accuracy of multiple measuring devices. *Technometrics*, v. 35, n. 4, p. 411-420, 1993. Citado 2 vezes na páginas 19.

COOK, R. D. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society: Series B.*, v. 48, n. 2, p. 133–169, 1986. Citado na página 20.

COOK, R. D; WEISBERG, S. Residuals and Influence in Regression. *Chapman and Hall, New York*, 1982. Citado 3 vezes nas páginas 79, 80 e 81.

D'AGOSTINO, R.B.; STEPHENS, M.A. Goodness-of-Fit Techniques. *Marcel Dekker, New York*, 1986. Citado na página 78.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 39, n. 1, p. 1-38, 1977. Citado na página 49.

EBERLEIN, E.; KELLER, U. Hyperbolic Distributions in Finance. *Bernoulli.* v. 1, n. 3, p. 281-299, 1995. Citado na página 16.

FANG, K.T.; ZHANG, Y.T. Generalized multivariate analysis. *Springer-Verlag*, London, 1990. Citado na página 18.

FULLER, W.A. Measurement Error Models. *John Wiley & Sons, New York*, 1987. Citado 2 vezes nas páginas 18 e 47.

GALEA, M.; CADEMARTORI, D.; VILCA, F. The structural Sharpe model under t-distributions. Journal of Applied Statistics, v. 37, n. 12, p. 1979-1990, 2010. Citado 2 vezes nas páginas 84 e 93.

GALEA, M.; RIQUELME, M.; PAULA, G.A. Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics*, v. 14, p. 167–184, 2000. Citado na página 20.

GRUBBS, F. On estimating precision of measuring instruments and product variability. *Journal American Statistical Society*, v. 43, p. 243-264, 1948. Citado na página 19.

GRUBBS, F. Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics*, v. 15, p. 53-66, 1973. Citado na página 19.

HALPHEN, E. Sur un nouveau type de courbe de frequence. *Compte rendus de I'Academie des Sciences*, v. 213, p. 634-635, 1941 (Publicado sob o nome de D. Dugue). Citado na página 22.

HANSSEN, A.; OIGARD, T.A. The normal inverse gaussian distribution as a flexible model for heavy-tailed processes. *IEEE-EURASIP Workshop on Nonlinear Signal & Image Processing*, Baltimore, USA, v. 5, 2001. Citado na página 39.

JAECH, J.L. Statistical Analysis of Measurement Errors. *John Wiley & Sons*, New York, 1985. Citado na página 19.

JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. Continuous Univariate Distribution. *Wiley, New York*, v. 1, 1994. Citado na página 27.

JØRGENSEN, B. Lectures Notes in Statistics. Springer-Verlag, New York, Heidelberg, Berlin, v. 9, 1982. Citado na página 22.

KOTZ, S.; JOHNSON, N.L. Encyclopedia os Statistical Sciences. *Wiley, New York*, v. 6, p. 340-347. 1985. Citado na página 78.

KOTZ, S.; NADARAJAH, S. Multivariate T-Distributions and Their Applications. *Cambridge University Press*, 2004. Citado na página 22.

KRISHNAIAH, P.R. Analysis of Variance. *North-Holland Pub. Co.*, 1980. Citado na página 78.

LACHOS, V.H.; ANGOLINI, T,; ABANTO-VALLE, C.A. On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers*, v. 52, n. 3, p. 567-590, 2009. Citado 3 vezes nas páginas 51, 52 e 84.

LÉVY, P. Calcul des Probabilites. *Gauthier Villars, Paris*, 1925. Citado na página 32.

LILLESTØL, J. Risk Analysis and the NIG Distribution. *Journal of Risk*, v. 2, n. 4, 2000. Citado na página 39.

LIN, T.I.; LEE, J.C.; YEN, S.Y. Finite mixture modelling using the skew normal distribution. Statistica Sinica, v. 17, p. 909-927, 2007. Nenhuma citação no texto.

MANDELBROT, B. The Variation of Certain Speculative Prices. *The Journal of Business*, v. 36, n. 4, p. 394-419, 1963. Citado 2 vezes nas página 16 e 32.

MASSEY, F.J. The Kolmogorov–Smirnov test for goodness of fit. *Journal of the American Statistical Association*. v. 46, p. 68–78, 1951. Citado na página 78.

MCNEIL, A.; FREY, R.; EMBRECHTS, P. Quantitative Risk Management: Concepts, Techniques and Tools. *Princeton University Press, Princeton*, 2005. Citado 2 vezes nas páginas 34 e 40.

O'HAGAN, A.; MURPHY, T.B.; GORMLEYA, I.C.; MCNICHOLAS, P.D.; KARLIS, D. Clustering with the multivariate normal inverse Gaussian distribution. *Computational Statistics and Data Analysis*, v. 93, p. 18–30, 2016. Citado 2 vezes nas páginas 78 e 113.

PARETO, V. La Courbe de la Répartition de la Richesse. *Université de Lausanne, Recueil Publié par la Faculté de Droit à l'occasion de l'Exposition nationale Suisse, Genève 1896, Oevres Complètes, 8, Droz, Genève*, 1896a (1965). Citado na página 32.

PATRIOTA, A.G. Modelo heterocedástico com erros nas variáveis. *Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística (IME), USP*, São Paulo, 2010. Citado na página 19.

PATRIOTA, A.G.; BOLFARINE, H. Measurement error models with a general class of error distribution. *Statistics (Berlin)*, v. 44, p. 119-127, 2009. Citado na página 19.

PRAUSE, K. The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. *Dissertation, University of Freiburg*, 1999. Citado na página 16.

ROMEIRO, R.G.; VILCA, F.; BALAKRISHNAN, N. A robust multivariate Birnbaum-Saunders distribution: EM estimation. Statistics, A jornal of Theoretical and Applied Statistics, v.52, n. 2, p.321-344, 2017. Nenhuma citação no texto.

SCHWARZ, G. Estimating the dimensional of a model. *Annals of Statistics, Hayward,* v.6, n.2, p.461-464, 1978. Citado na página 77.

SHUSTER, J. On the inverse Gaussian distribution. *Journal of the American Statistical Association*, v. 63, p.1514–1516, 1968. Citado na página 29.

TEIXEIRA, V.M.M.; Distribuições Hiperbólicas Generalizadas: Aplicações ao Mercado Português. *Tese (Mestrado em Matemática) – Departamento de Matemática e Engenharias, Universidade da Madeira.* Funchal, 2006. Nenhuma citação no texto.

TELFORD, R.D.; CUNNINGHAM, R.B. Sex, sport and body-sixe dependency of hematology of trained athletes. *Medicine and Science in Sport and Exercise*, v. 23, p. 788-794, 1991. Citado na página 84.

WANG, C. On the Numerics of Estimating Generalized Hyperbolic Distributions. *Master's thesis, Humboldt-University of Berlin*, 2005. Citado na página 35.

XIE, F.C.; WEI, B.C. Diagnostics analysis for log-Birnbaum–Saunders regression models. *Computational Statistics & Data Analysis*, v. 51, n. 9, p. 4692-4706, 2007. Citado na página 80.

XIE, F.C.; WEI, B.C.; LIN, J.G. Case-deletion Influence Measures for the Data from Multivariate t Distributions. *Journal of Applied Statistics*, v.34, n.8, p. 907-921, 2007. Citado na página 82.

ZELLER, C.B. Distribuições Misturas de Escala Skew-Normal: Estimação e Diagnóstico em Modelos Lineares. *Tese (Doutorado em Estatística) – Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas.* Campinas, 2009. Citado na página 42.

ZHU, H,; LEE, S.Y.; WEI, B.C.; ZHOU, J. Case-deletion measures for models with incomplete data. *Biometrika*, v. 88, n. 3, p.727-737, 2001. Citado 4 vezes nas páginas 20, 80, 81 e 83.