



HERNANI MARTINS JÚNIOR

**O USO DA DISTRIBUIÇÃO BETA t
GENERALIZADA PARA OS ERROS DO
MODELO GARCH: UMA APLICAÇÃO NO
PREÇO DO CAFÉ**

LAVRAS - MG

2014

HERNANI MARTINS JÚNIOR

**O USO DA DISTRIBUIÇÃO BETA t GENERALIZADA PARA OS
ERROS DO MODELO GARCH: UMA APLICAÇÃO NO PREÇO DO
CAFÉ**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de "Doutor".

Orientadora

Profa.Dra. Thelma Sáfydi

**LAVRAS
MINAS GERAIS - BRASIL
2014**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Martins Júnior, Hernani.

O uso da distribuição beta T generalizada nos erros do modelo
Garch : uma aplicação no preço do café / Hernani Martins Júnior. –
Lavras : UFLA, 2014.

101 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2014.

Orientador: Thelma Sáfyadi.

HERNANI MARTINS JÚNIOR

**O USO DA DISTRIBUIÇÃO BETA t GENERALIZADA PARA OS
ERROS DO MODELO GARCH: UMA APLICAÇÃO NO PREÇO DO
CAFÉ**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de "Doutor".

APROVADA em 15 de Outubro de 2014.

Prof. Lucas Monteiro Chaves	UFLA
Prof. Joel Augusto Muniz	UFLA
Prof. Maria do Carmo Pacheco de Toledo Costa	UFLA
Prof. Denismar Alves Nogueira	UNIFAL-MG

**Profa. Dra. Thelma Sáfydi
UFLA
(Orientadora)**

**LAVRAS
MINAS GERAIS – BRASIL**

À minha filhinha Helena;

A Deus,

Dedico.

AGRADECIMENTOS

A Deus pelos ensinamentos celestiais e conforto nas horas bem precisas.

À minha querida professora e orientadora Thelma Sáfadi, que com carinho me acolheu.

Ao CNPq pela ajuda financeira e por acreditar no potencial do povo brasileiro.

Ao programa de Pós-Graduação em Estatística e Experimentação Agropecuária pela oportunidade e pelos ensinamentos.

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO	13
2.1	Distribuições de Probabilidade	13
2.1.1	Distribuição Beta	13
2.1.2	Distribuição Beta Generalizada	13
2.1.3	Generalização da Distribuição Beta	14
2.1.4	Distribuição Generalização da Beta (<i>Generalized Beta Generated GBG</i>)	18
2.1.4.1	Momentos da GBG	21
2.1.4.2	Processo de estimação da GBG	27
2.1.5	A Distribuição Generalizada Beta t	29
2.1.5.1	Geração de números aleatórios	32
2.1.5.2	Estimação	33
2.2	Modelos de Séries Temporais	35
2.2.1	Modelos ARIMA	36
2.2.2	Modelos ARCH	40
2.2.2.1	Determinação da Ordem	41
2.2.2.2	Estimação	41
2.2.2.3	Verificação de um modelo ARCH	44
2.2.2.4	Previsão em um modelo ARCH	45
2.2.3	Modelos GARCH	45
2.2.3.1	Curtose nos modelos GARCH	50
2.2.3.2	Estimação	51
2.3	Métodos de Otimização	54
2.4	Crerios de Avaliação de Modelos	56
3	METODOLOGIA	59

3.1	A construção de uma nova proposta de estimação para o modelo GARCH.....	59
4	APLICAÇÃO.....	64
4.1	Aplicação para uma série temporal do preço de café. Abordagem clássica.....	64
4.2	Modelo ARIMA-GARCH para uma série temporal do preço de café. Sob a pressuposição de distribuição “ t de Student” para o resíduo.....	74
4.3	Modelo ARIMA-GARCH – Aplicação para uma série temporal do preço de café. Sob a pressuposição de distribuição “Beta t de Student” para o resíduo.....	78
4.4	Modelo ARIMA-GARCH - Aplicação para uma série temporal do preço de café. Abordagem GBt.....	81
5	CONCLUSÕES	86
	BIBLIOGRAFIA.....	87
	ANEXOS	91

RESUMO

MARTINS JÚNIOR, Hernani. **O uso da distribuição beta t generalizada para os erros do modelo garch: uma aplicação no preço do café** 2014. 99 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.¹

Este trabalho analisa algumas distribuições de probabilidade flexíveis, especificamente a classe das *Generalized Beta Generated*, estuda algumas de suas propriedades e propõe o uso delas como pressuposição para o erro, componente aleatório em modelos de regressão. Relaxando as pressuposições de normalidade que geralmente o circundam. O método da máxima verossimilhança é utilizado para a estimação de parâmetros através do software R. Dentre os algoritmos de estimação existentes foi utilizado o de Nelder-Mead. Diferentes distribuições foram utilizadas como base para o modelo de regressão ARMA-GARCH aplicado a série de preços de café. Foram elas: Distribuição *t* de Student, Distribuição Beta *t*, Distribuição Generalizada Beta *t*, além da distribuição Normal. A matriz hessiana para os modelos estimados, avaliada nas estimativas paramétricas obtidas, também foi calculada numericamente utilizando o R. Os resultados mostram que um ganho pode ser obtido no que diz respeito à qualidade do ajuste, quando comparados os desempenhos destas distribuições com o desempenho de distribuições clássicas então utilizadas. O teste de razão de verossimilhanças foi feito com vias à comparação de modelos aninhados.

Palavras-chave: função objetiva, otimização, distribuição de probabilidade, estimação, modelo.

¹ **Orientadora:** Thelma Sáfiadi – UFLA

ABSTRACT

MARTINS JÚNIOR, Hernani. **The use of Beta t Generalized Distribution used as assumption to GARCH models errors: A coffee price application.** 2014. XXX 99 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.¹

This work study some flexible distribution of Generalized Beta class, actually the *Generalized Beta Generated*, study some of its properties and show how can they be used as assumption to the residuals of regression models in the way to supplant the no normality in some situations. The likelihood method is used to find the parameters estimative by R software. The Nelder-Mead algorithm was chosen to objective function optimization. Some distributions has been used as base to ARMA-GARCH regression models applied to a coffee price series. They were Student t Distribution, Beta t Distribution, Generalized Beta t distribution, and the Normal Distribution. The hessian matrix for estimated models were numerically computed, evaluated on the estimated values and again by support of R software. The results shows a improvement when compared with the classical approach. The likelihood test was done to compare nested models.

Keywords: Objective function, optimization, probability distribution, estimation, model.

¹ **Advisor:** Thelma Sáfaci – UFLA

1 INTRODUÇÃO

Atualmente existem inúmeras distribuições de probabilidade disponíveis na literatura. Profissionais de todo o mundo estudam esta classe especial de funções. Distribuições de probabilidade, há muito são estudadas: ver Fisher (1924), Jones (2001), Eugene, Lee e Famoye (2002), Jones (2004), Nadarajah e Gupta (2007), Alexander et al. (2012). O estudo das distribuições é ferramenta preponderante no estudo de variáveis aleatórias as quais representam os mais diversos problemas nas mais diferentes áreas, apenas para citar algumas: em séries temporais; na regressão clássica; em análise de sobrevivência; ou mesmo na análise de valores extremos.

A estatística, em muitas de suas áreas, está calcada na distribuição normal ou na pressuposição de normalidade. Quando essa característica não se comprova, surgem os mais diversos problemas (como falta de ajustamento, e/ou o não atendimento de pressuposições) que alimentam várias linhas de pesquisa onde se busca resolver tais especificidades, como pode ser visto em Gupta e Kundu (2001), Pascoa, et al. (2010), e Ortega, Cordeiro e Carrasco (2011)

Sendo assim, seria desejável um modelo com generalizado poder de explicação. Um modelo que possa abarcar tanto as situações mais comuns quanto as menos corriqueiras. Que seja uma ferramenta, que uma vez difundida, seja capaz de resolver uma gama maior de problemas.

O relaxamento da pressuposição de normalidade é assunto corrente na literatura Jones (2004), Akinsete, Famoye e Lee (2008), Barreto, Santos e Cordeiro (2010), Alexander et al. (2012).

Neste sentido este trabalho busca a utilização de uma distribuição de probabilidade recentemente apresentada por Alexander et al. (2012), como alternativa à modelagem de dados e, especificamente, à modelagem do erro em modelos de regressão. Esta nova distribuição, possuindo mais parâmetros (o que dificulta o processo de estimação) apresenta uma grande flexibilidade

no comportamento de suas curvas, e é exatamente isto que se espera de uma distribuição que possa modelar diferentes tipos de dados, mesmo que isto tenha um custo computacional elevado.

Desta forma o objetivo principal deste trabalho é o estudo da distribuição beta t generalizada (um caso especial da *Generalized Beta Generated*), apresentando suas propriedades e potencialidades na modelagem de dados. Uma comparação com a distribuição t de Student é intentada como forma de se verificar o desempenho dessa distribuição.

Seguindo a idéia apresentada em Jones (2004) escolheu-se, dentre as distribuições abarcadas pela GBG, apresentadas por Alexander et al. (2012), escolheu-se a distribuição Beta t Generalizada.

Esta distribuição teve suas principais propriedades estudadas, no intuito de utilizá-la para modelar conjuntos de dados que apresentem elevado grau de assimetria ou curtose. Um outro objetivo intentado é, com o uso dela minorar o erro de predição em modelos de regressão para séries temporais, ajustando o resíduo nestes modelos.

Parte do processo inferencial deve ser refeito, uma nova função de verossimilhança é escrita com base nesta nova distribuição, fato que inevitavelmente acarretará mudanças no processo decisório. A busca pelas estimativas para os parâmetros do modelo continua a ser um desafio uma vez que o incremento de novos parâmetros aumentam em muito a complexidade do espaço paramétrico, dificultando ainda mais o processo de otimização.

2 REFERENCIAL TEÓRICO

2.1 Distribuições de Probabilidade

2.1.1 Distribuição Beta

A distribuição Beta é uma das mais importantes distribuições na estatística e na teoria da probabilidade. Seu grande uso se dá em parte pelo fato de sua função poder ser monotonicamente crescente ou decrescente, em forma de U ou unimodal. É uma função contínua definida entre 0 e 1 com dois parâmetros de curva estritamente positivos, a e b . Sua densidade é dada por

$$f_{beta}(x; a, b) = B(a, b)^{-1} x^{a-1} [1-x]^{b-1}, \quad 0 < x < 1. \quad (1)$$

Nesta expressão $B(a, b)$, representa uma constante normalizadora, é a função Beta, e é dada por $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, sendo $\Gamma(\cdot)$ a função

Gama, dada por $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$, Mood, Graybill e Boes (1974).

2.1.2 Distribuição Beta Generalizada

A distribuição Beta Generalizada foi introduzida por McDonald (1984) sendo sua utilização apenas possível quando todos os parâmetros (a, b, c) forem maiores que zero. Considerando $U \sim Beta(a, b)$ e considerando a transformação $U^{1/c}$,

$$P(U^{1/c} \leq x) = P(U \leq x^c) = F_{Beta}(x^c) = \int_0^{x^c} \frac{1}{B(a,b)} w^{a-1} (1-w)^{b-1} dw. \quad (2)$$

Derivando $F_{Beta}(x^c)$ tem-se a densidade da chamada distribuição Beta Generalizada possuindo 3 parâmetros, a , b e c , cuja densidade é dada por:

$$f_{gb}(x; a, b, c) = cB(a, b)^{-1} x^{ac-1} [1 - x^c]^{b-1}, \quad 0 < x < 1. \quad (3)$$

A Equação 3 é facilmente obtida conforme os passos abaixo:

$$\begin{aligned} \frac{d}{dx} F(x^c) &= f_{beta}(x^c) cx^{c-1} \\ &= \frac{1}{B(a,b)} (x^c)^{a-1} (1-x^c)^{b-1} cx^{c-1} \\ &= \frac{c}{B(a,b)} x^{ca-c} (1-x^c)^{b-1} x^{c-1} \\ &= \frac{c}{B(a,b)} x^{ca-1} (1-x^c)^{b-1}. \end{aligned}$$

Outras importantes distribuições podem advir desta distribuição, como é o caso da distribuição Kumaraswamy, Kwmaraswamy (1980), que vem à tona com $a=1$. Com $c=1$ tem-se a própria distribuição $Beta(a,b)$.

2.1.3 Generalização da Distribuição Beta

Tomando $X = F^{-1}(U)$ com $U \sim Beta(a,b)$ sendo $Beta(a,b)$ a distribuição beta e F representando uma distribuição de probabilidade acumulada qualquer. Sendo assim:

$$P[X \leq x] = P[F^{-1}(U) \leq x] = P[U \leq F(x)] = \int_0^{F(x)} \frac{1}{B(a,b)} u^{a-1} (1-u)^{b-1} du \quad (4)$$

logo, a variável aleatória X tem distribuição equivalente a uma Generalização da Beta (o termo em inglês é *Generalized Beta Generated*, corrente na literatura), cuja densidade é dada por.

$$f_{bg}(x) = B(a, b)^{-1} f(x) F(x)^{a-1} [1 - F(x)]^{b-1}, \quad x \in \mathbb{R}. \quad (5)$$

Uma característica interessante da densidade da Generalização da Beta é que permanece simétrica se $a = b$ fato que nos levará a seguinte densidade (6) para a igualdade entre $a = b$:

$$f_{bg}(x) = B(a, a)^{-1} f(x) F(x)^{a-1} [1 - F(x)]^{a-1}, \quad x \in \mathbb{R}. \quad (6)$$

Observe que $f_{bg}(x) = f_{bg}(-x)$ como condição para simetria, equivale a fazer:

$$\begin{aligned} f_{bg}(x) &= B(a, a)^{-1} f(x) F(x)^{a-1} [1 - F(x)]^{a-1} \\ &= F(x)^{a-1} [1 - F(x)]^{a-1} \\ &= (1 - \alpha)\alpha. \\ f_{bg}(-x) &= B(a, a)^{-1} f(x) F(x)^{a-1} [1 - F(x)]^{a-1} \\ &= F(-x)^{a-1} [1 - F(-x)]^{a-1} \\ &= \alpha(1 - \alpha). \end{aligned}$$

Assim, $F(-x) = 1 - F(x)$.

Desta última igualdade está provada a simetria, considerando α um número qualquer entre 0 e 1, um valor de probabilidade.

A primeira distribuição da classe Gerada Beta a ser estudada foi a distribuição Beta Normal em Eugene et al. (2002) onde tomando $f(x) = \phi(x)$ e $F(x) = \Phi(x)$ como sendo a densidade e a acumulada da normal padrão e $X = \Phi^{-1}(U)$, $U \sim Beta(a, b)$, então X tem distribuição Beta Normal Padrão denotada por $bn(a, b, 0, 1)$, cuja densidade é dada por:

$$f_{bn}(x; a, b, 0, 1) = B(a, b)^{-1} \phi(x) \Phi(x)^{a-1} [1 - \Phi(x)]^{b-1}, \quad x \in \mathbb{R}. \quad (7)$$

A distribuição Beta Normal Padrão admite diretamente a transformação de locação-escala, uma vez que se:

$$P[\sigma X + \mu \leq x] = P[X \leq (x - \mu) / \sigma]$$

$$P[\Phi^{-1}(U) \leq (x - \mu) / \sigma] = P[U \leq \Phi_{\mu, \sigma^2}(x)] = P[\Phi^{-1}(U)],$$

o que equivalentemente seria o mesmo que tomando $X \sim f_{bn}(x; a, b, 0, 1)$ então $\sigma X + \mu = Y \sim f_{bn}(y; a, b, \mu, \sigma)$, que é chamada de Beta Normal.

Os parâmetros a e b controlam a assimetria e a curtose. Sendo a função de densidade da distribuição Beta Normal for unimodal e sendo $a=b$ a densidade da beta normal é simétrica, se $a < b$ a densidade da beta normal é assimétrica positivamente e se $a > b$ a densidade da beta normal é assimétrica negativamente. Quando $a=b > 1$ a densidade da beta normal possui curtose positiva, quando $a=b < 1$ possui curtose negativa. Todavia a variação da assimetria e da curtose é bem limitada para as diferentes parametrizações da Beta Normal, fato que pode ser visto em Eugene, Lee e Famoye (2002) onde os autores mostram que a assimetria sempre se encontra no intervalo $[-1, 1]$ e a máxima curtose encontrada é 4,18 para valores de a e b variando de 0,01 a 100. Este fato impossibilita um ajuste com qualidade a dados cujas medidas de curtose e assimetria extrapolam estes limites.

A distribuição Beta Normal é bimodal para certos valores de a e b e a solução analítica para a e b onde a distribuição é bimodal não pode ser obtida algebricamente para tanto utiliza-se solução numérica para um número de raízes das derivadas da Beta Normal (Famoye, Lee e Eugene 2004). No caso de bimodalidade a distribuição é assimétrica a direita quando $a < b$ e assimétrica a esquerda quando $a > b$. Assim esta distribuição é muito versátil, permitindo descrever dados simétricos, assimétricos, unimodais e bimodais.

Eugene, Lee e Famoye (2002) descrevem a região de bimodalidade utilizando função de regressão a partir de gráfico indicativo de unimodalidade-bimodalidade construído para uma sequência de valores de

a e b . Esta função é construída a partir da $bn(a, b, \mu = 0, \sigma = 1)$ e o resultado é válido para os demais valores de μ e σ por que a propriedade de bimodalidade de BN é independente dos parâmetros μ e σ . Para uma prova completa, ver Eugene, Lee e Famoye (2002).

São diversas as distribuições oriundas a partir da Gerada Beta dependendo de qual classe de função é adotada para $F(x)$. Caso $F(x)$ seja a função de distribuição acumulada da distribuição Gumbel tem-se a distribuição Beta Gumbel introduzida por Nadarajah e Kotz (2004). Caso $F(x)$ seja a função de distribuição acumulada da distribuição Fréchet tem-se a distribuição Beta Fréchet introduzida por Nadarajah e Gupta (2004). Em 2006 propuseram a Beta Exponencial tomando $F(x)$ como sendo a função de distribuição acumulada da distribuição Exponencial, Nadarajah e Kotz (2006), neste trabalho os autores apresentam expressões para a função geradora de momentos e função característica, apresentam também a distribuição para a soma e para a razão da distribuição Beta Exponencial, além das entropias de Rényi e Shannon. Muitas outras distribuições podem ser deduzidas nesta mesma sistemática. Em Alexander et al. (2012) é apresentado as funções de densidades de diversas distribuições oriundas deste processo, são elas: Beta Normal, Beta Log-Normal, Beta t de Student, Beta Laplace, Beta Exponencial, Beta Weibull, Beta Gumbel, Beta Birnbaum-Sauders, Beta Gama, Beta Pareto, Beta Logística.

Distribuições da classe da Generalização da Beta têm sido usadas com sucesso em diversas situações. Pode-se ver a Beta Normal modelando problemas do tipo dose resposta, Razzaghi (2009), ou sendo utilizada para modelar dados de enchentes, Akinsete, Famoye e Lee (2008) onde lançam mão da Beta Pareto, já Jones e Larsen (2004) ajustaram as distribuições t assimétrica e log F para dados de temperatura via distribuição Gerada Beta.

A própria distribuição Beta surge como um caso particular da GBG se para a distribuição paterna, $F(x)$, é tomada a distribuição uniforme

contínua. Por fim duas outras importantes distribuições são amplamente discutidas: a distribuição log F Fisher (1924) e a distribuição t assimétrica proposta por Jones (2004) e melhor estudada por Jones e Faddy (2003). Em casos de $a=b$ que como já foi dito resulta em uma distribuição simétrica, surge como caso particular a distribuição t de Student. Estas duas distribuições possuem diferentes conformações caudais, o que em certas situações proporcionam melhorias no ajuste.

2.1.4 Distribuição Generalização da Beta (*Generalized Beta Generated - GBG*)

A distribuição utilizada como base para este trabalho é uma Generalização da distribuição (GBG sigla do Inglês de *Beta Generalized Generated*) que por abreviação será denotada por GBG. Considerando que a distribuição Generalizada Beta tem em seu núcleo uma variável aleatória u . Fazendo $u = F(x; \tau)$, sendo $F(x; \tau)$ uma função de distribuição de probabilidade acumulada com vetor de parâmetros τ , a nova distribuição que surge desta transformação é a GBG. A diferença desta nova distribuição em relação à classe das distribuições Gerada Beta, é que $F(x; \tau)$ é exponenciada pelo parâmetro c .

A densidade da GBG é, então, dada por:

$$f_{gbg}(x; \tau, a, b, c) = cB(a, b)^{-1} f(x; \tau) F(x; \tau)^{ac-1} [1 - F(x; \tau)^c]^{b-1}, \quad x \in \square. \quad (8)$$

Esta distribuição pode ser obtida com a seguinte transformação:

$$\begin{aligned}
F_{gbg}(x) &= \int_0^{F^c(x)} \frac{1}{B(a,b)} u^{a-1} (1-u)^{b-1} du \\
\frac{dF(x)}{dx} &= \frac{1}{B(a,b)} f(x) (F^c(x))^{a-1} (1-F^c(x))^{b-1} c F^{c-1} \\
&= \frac{1}{B(a,b)} (F(x)^{ac-1}) (1-F(x)^c)^{b-1} f(x),
\end{aligned}$$

e foi introduzida por Alexander et al. (2012), e apresenta outras distribuições como casos particulares.

Uma grande variedade de distribuições pode ser utilizada como sendo $F(x; \tau)$ e muitas das densidades oriundas desta transformação podem ser vistas em Alexander et al. (2012).

A distribuição beta generalizada, por exemplo, que fora introduzida por McDonald (1984), vem a tona, a partir da GBG, se para o núcleo desta for tomada uma distribuição uniforme e se, além disto, a for igual a 1 tem-se a Kumaraswamy Generalizada, se, ainda, c for igual a 1 tem-se a distribuição Beta.

Peça importante no processo de estimação dos parâmetros, usando o pacote *fitdistrplus* do Software R, é a função distribuição acumulada representada por sua forma estocástica, que necessariamente deve ser especificada no processo de estimação. Supondo que $X \sim F^{-1}(U^{1/c})$, $U \sim B(a, b)$, a função de distribuição acumulada pode ser escrita como:

$$F(x; \tau, a, b, c) = B(a, b)^{-1} \int_0^{F(x; \tau)^c} w^{a-1} (1-w)^{b-1} dw, \quad (9)$$

função de distribuição esta, que pôde ser facilmente interpretada pelo R.

Alguns trabalhos na última década se dedicam ao estudo desta distribuição Jones (2004): se dedica ao estudo de algumas características da distribuição Gerada Beta mas focando especialmente em casos de distribuições paternas simétricas em torno de zero como a distribuição normal e a distribuição t de Student, objetivando com isto o estudo das

características desta distribuição com nenhum outro parâmetro que não sejam os parâmetros de locação e escala, além de pretender uma distribuição gerada com suporte em toda a reta Real.

Neste caso (com $F(x; \tau)$ simétrica) a e b governam a assimetria e o peso das caudas. Com $a=b$ a distribuição permanece simétrica; tornando as caudas mais pesadas à medida que a é aumentado; e tornado-as mais leves à medida em que a é diminuído. Na medida em que a se torna diferente de b a distribuição se torna assimétrica. E o sinal desta diferença é que dirá se a assimetria é positiva ou negativa. Sendo assim, o fato de a distribuição paterna ser simétrica facilita em muito o entendimento do comportamento da distribuição à medida que se variam os demais parâmetros.

A distribuição GBG pode ser escrita como uma combinação linear de distribuições F-exponenciadas cujas propriedades foram estudadas por diversos autores. Nadarajah e Gupta (2007) e Muldholkar et al. (1995) sendo que neste são apresentadas os k-ésimos momentos da função de distribuição Weibull Exponenciada. A densidade da GBG pode ser alternativamente escrita como:

$$f_{gbg}(x, a, b, c) = \sum_{i=1}^{\infty} w_i g_{a(i+c)}(x), \quad (10)$$

e sua distribuição pode ser escrita como:

$$F_{gbg}(x, a, b, c) = \sum_{i=1}^{\infty} w_i G_{a(i+c)}(x), \quad (11)$$

em que $g_{a(i+c)}(x)$ é a densidade da $Exp^{a(i+c)}$, $G_{a(i+c)}(x)$ é a distribuição da $Exp^{a(i+c)}$

$$w_i = \frac{(-1)^i \binom{b}{i}}{(a+i)B(a, b+1)}.$$

Outra forma conveniente de definir a distribuição GBG é através da função confluyente hipergeométrica

$$F_{gbg}(x, a, b, c) = \frac{cF(x)^a}{aB(a, b)} \frac{\Gamma(c)}{\Gamma(a)\Gamma(1-b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(1-b+j)}{\Gamma(c+j)} \frac{x^j}{j!} \quad (12)$$

com $c = a + 1$. Esta função é relacionada com as equações diferenciais cuja aplicação é extensa em diversas áreas da ciência.

Desta forma propriedades das GBG podem ser obtidas diretamente a partir das propriedades da função hipergeométrica Gradshteyn e Ryzhik (2000).

2.1.4.1 Momentos da GBG

Com objetivo de se descrever as propriedades das distribuições de probabilidades, comumente se usa os momentos para descrever a esperança, a variância, a curtose e a assimetria.

A esperança de uma variável aleatória com distribuição GBG é dada por:

$$E[X] = \int_{-\infty}^{\infty} xf_{gbg}(x)dx, \quad (19)$$

$f_{gbg}(x; \tau, a, b, c) = cB(a, b)^{-1} f(x; \tau) F(x; \tau)^{ac-1} [1 - F(x; \tau)^c]^{b-1}$, e a variância é dada por:

$$E[X^2] - E^2[X] = \int_{-\infty}^{\infty} x^2 f_{gbg}(x)dx - \left(\int_{-\infty}^{\infty} xf_{gbg}(x)dx \right)^2, \quad (20)$$

(19) e (20) representam o primeiro momento (média) e o segundo momento centrado no primeiro (variância). Já o terceiro e quarto momentos (assimetria e curtose), como são utilizados classicamente, não têm sua aplicação indicada para a distribuição em questão.

Os momentos da GBG são dados por:

$$\mu'_s = E\left[\left(F^{-1}(U^{1/c})\right)^s\right], \quad U \sim B(a, b)$$

$$\mu'_s = \int_0^1 \frac{1}{B(a, b)} \left(F^{-1}(u^{1/c})\right)^s u^{a-1} (1-u)^{b-1} du$$

fazendo $u^{1/c} = x$

$$\mu'_s = \int_0^1 \frac{1}{B(a, b)} \left(F^{-1}(x)\right)^s x^{ac-c} (1-x^c)^{b-1} cx^{-x} x^{-1} dx$$

$$\mu'_s = \int_0^1 \frac{c}{B(a, b)} \left(F^{-1}(x)\right)^s x^{ac-1} (1-x^c)^{b-1} dx$$

$$\mu'_s \approx cB(a, b) \int_0^1 \left(F^{-1}(x)\right)^s x^{ac-1} (1-x^c)^{b-1} dx \quad (21)$$

Fazendo uso dos momentos de probabilidade ponderados

$t(s, r) = E\left[Y^s F(Y)^r\right]$, em que Y segue a distribuição núcleo, o que pode ser obtido para a maioria das distribuições, a expressão para os momentos pode ser reescrita de outras formas como:

Para a e c inteiros,

$$\mu'_s \approx cB(a, b) \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} t(s, a(i+c)-1),$$

Por outro lado, se a ou c forem reais e não inteiros a expressão para os momentos é dada por,

$$\mu'_s = cB(a, b) \sum_{i=0}^{\infty} \sum_{j=k}^{\infty} (-1)^{k+i+j} \binom{a(i+c)-1}{j} \binom{j}{k} \binom{b}{i} t(s, a(i+c)-1). \quad (22)$$

Segundo Alexander et al. (2012) distribuições com caudas pesadas, como a distribuição em questão, nem sempre possuem terceiro e quarto momentos bem definidos.

Comumente para demonstrar como os parâmetros a , b e c estão associados a assimetria e curtose, utilizam-se medidas baseadas em quantis. Quando se consideram distribuições com caudas pesadas (o que é buscado

no escopo deste trabalho) é comum que a curtose clássica seja infinita, fato que torna este momento não informativo. Sendo assim usa-se a assimetria de Bowley (Kenney e Keeping 1962), cuja função é dada por:

$$B = \frac{Q(3/4) - 2Q(1/2) + Q(1/4)}{Q(3/4) - Q(1/4)},$$

em que $Q(\cdot)$ são quartis. E usa-se a curtose de Moors (1998), cuja função é dada por:

$$M = \frac{Q(7/8) - Q(5/8) - Q(3/8) + Q(1/8)}{Q(6/8) - Q(2/8)},$$

em que $Q(\cdot)$ são octis.

Esta concepção percentflica de momentos é muito usada na literatura, recentemente muitos trabalhos como, por exemplo, Alexander et al. (2012) e Leão, et al. (2013) lançam mão desta técnica.

A distribuição GBG, assim como todas as distribuições oriundas desta classe (GBN, GBW, GBt, entre outras) apresentam como inovação o incremento do parâmetro c . Alexander et al. (2012) estudam tanto a assimetria de Bowley quanto a curtose de Moors a luz deste novo parâmetro. Para o caso da GBt a assimetria de Bowley tem a seguinte conformação:

$$B = 0 \text{ se } c = 1$$

$$B > 0 \text{ se } c > 1$$

$$B < 0 \text{ se } c < 1.$$

Quando $c \neq 1$ o grau de assimetria aumenta com a curtose da distribuição paterna e diminui com os valores de a e b .

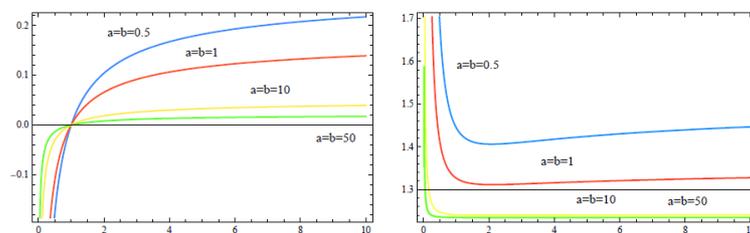
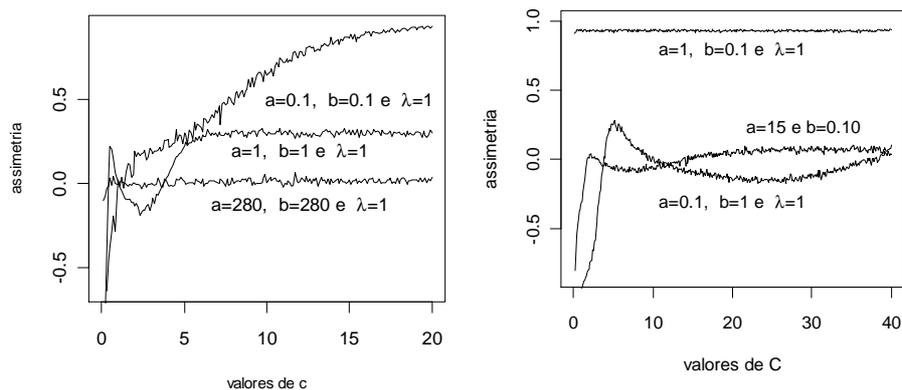


Figura 1 Assimetria de Bowley (esquerda) e curtose de Moors (direita) para a densidade da GBt com 5 graus de liberdade, em função de c (variando de 0 a 10)

para $a=b=0,5$, para $a=b=1$, para $a=b=10$ e para $a=b=50$. (Fonte Alexander et al. (2012))

Pode-se ver que para algumas parametrizações da GBt a curtose de Moors pode ser extremamente grande e positiva, fato que ocorre somente para valores pequenos de c , mas partindo-se de zero, onde a curtose parece ser máxima, a curtose decresce rapidamente, atingindo o mínimo em $c=2$, vide Figura 1. Seguindo a metodologia proposta por Alexander et al. (2012), gerou-se números aleatórios da distribuição GBt com diferentes parametrizações, considerando-se a assimetria conforme mostrado na Figura 2 e considerando a curtose conforme mostrado na Figura 3. Em ambos os casos cada ponto representa a média dos coeficientes calculados de cada uma das 100 amostras aleatórias geradas, cada amostra composta de 160 observações.



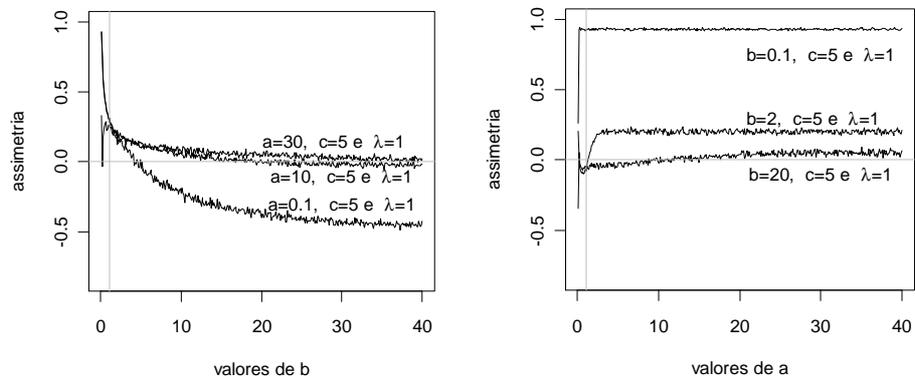
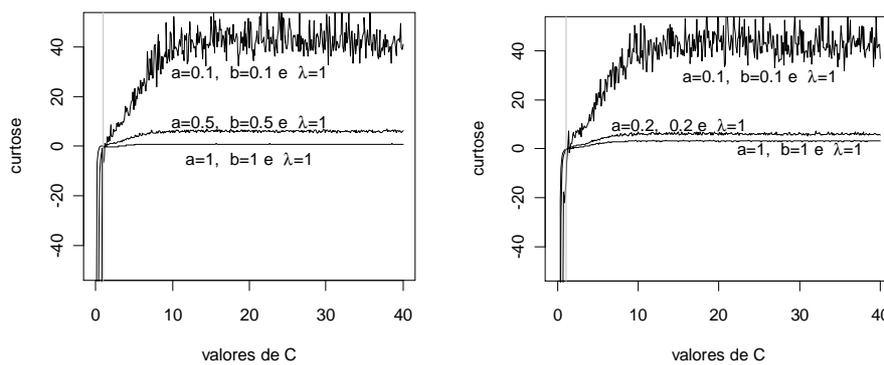


Figura 2 Coeficiente de assimetria de Bowley para dados simulados da GBt em diferentes parametrizações (o traço vertical representa uma reta passando em 1 nas abscissas).

Outro aspecto relevante é que para $c \neq 1$ a curtose da GBt é maior que a curtose da distribuição t de Student. Para pequenos valores de c a curtose de Moors se torna excessivamente grande ao mesmo tempo que a assimetria de Bowley, se torna negativa e com elevado valor em módulo. Sendo assim pode-se concluir que o parâmetro c é um parâmetro de assimetria.



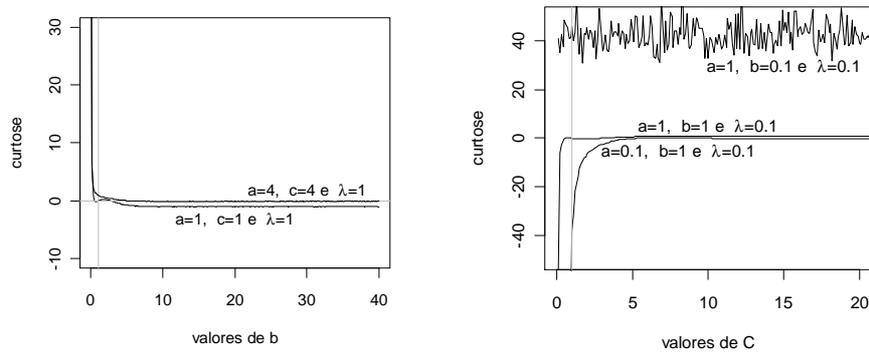


Figura 3 Coeficiente de curtose de Moors para dados simulados da GBT em diferentes parametrizações (o traço vertical representa uma reta passando em 1 nas abscissas).

O parâmetro λ não interfere na assimetria, é um parâmetro de escala como vai ser mostrado mais adiante. Todavia esta afirmação a respeito da não interferência na curtose pode ser vista na Figura 4.

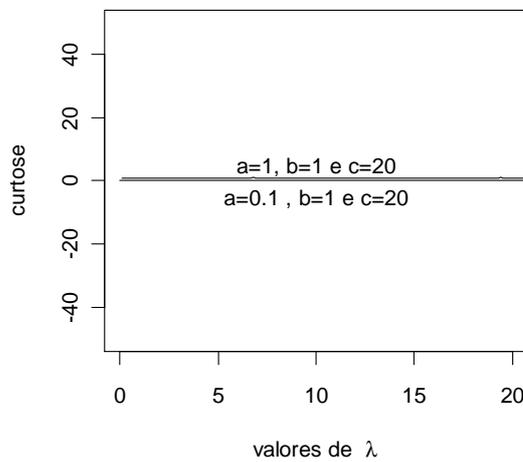


Figura 4 coeficiente de curtose calculado em função de diferentes valores do parâmetro λ .

2.1.4.2 Processo de estimação da GBG

Seja $X = (x_1, \dots, x_n)$ uma variável aleatória tal que $X \sim F_{gbg}(F(x, \tau); a, b, c)$, em que τ representa o vetor de parâmetros de $F(x)$. A função de log verossimilhança, é obtida do log do produtório da densidade da distribuição, e definida por:

$$l(\tau, a, b, c; x) = n \log(c) - n \log[B(a, b)] + \sum_{i=1}^n \log f(x_i; \tau) + (ac - 1) \sum_{i=1}^n \log F(x_i; \tau) + (b - 1) \sum_{i=1}^n \log(1 - F^c(x_i; \tau)). \quad (23)$$

Esta função deve ser maximizada em relação a cada um dos parâmetros, derivando-se em relação a cada parâmetro e igualando-se a zero, fato que gerará as funções escores, suporte para a estimação propriamente dita. As funções escores obtidas em Alexander et al. (2012) são apresentadas abaixo:

$$U_a(\theta) = -n[\psi(a) - \psi(a + b)] + c \sum_{i=1}^n \log[F(x_i; \tau)],$$

$$U_b(\theta) = -n[\psi(a) - \psi(a + b)] + \sum_{i=1}^n \log[1 - F^c(x_i; \tau)],$$

$$U_c(\theta) = \frac{n}{c} + a \sum_{i=1}^n \log[f(x_i; \tau)] - (b - 1) \sum_{i=1}^n \frac{\log[F^c(x_i; \tau)] \log[F(x_i; \tau)]}{1 - \log[F^c(x_i; \tau)]}$$

$$U_\tau(\theta) = \sum_{i=1}^n \frac{f'(x_i; \tau)}{f(x_i; \tau)} + (ac - 1) \sum_{i=1}^n \frac{F'(x_i; \tau)}{F(x_i; \tau)} + c(b - 1) \sum_{i=1}^n \frac{F^{c-1}(x_i; \tau) F'(x_i; \tau)}{1 - F^c(x_i; \tau)}. \quad (24)$$

θ representa o vetor com todos os parâmetros aqui denotados por a, b, c (parâmetros incorporados no processo de geração da nova distribuição) e por

τ , um vetor de parâmetros representando os parâmetros da distribuição núcleo.

Estas funções escores têm natureza fortemente não linear, fato que pode ser visto nas próprias funções escores, que são derivadas em relação a um parâmetro e se apresenta em função de outros.

Este sistema não possui solução analítica, comumente se recorre a métodos numéricos implementados através de diversas rotinas computacionais disponíveis em diversos softwares.

A segunda derivada em relação aos parâmetros resulta na matriz da informação esperada comumente chamada de matriz hessiana. As observações da diagonal principal da matriz hessiana inversa estão associadas a dispersão em torno da estimativa pontual, enquanto que as demais observações representam a correlação entre os parâmetros componentes do modelo.

Para a construção de intervalos de confiança e testes de hipóteses para os parâmetros é necessário a matriz da informação observada $J(\theta)$, que sob certas condições do espaço paramétrico e no contexto de grandes amostras, limite quando n tende ao infinito, possui distribuição normal multivariada $N_{P+3}(0, I(\theta)^{-1})$ em que P é o número de parâmetros de $F(x; \theta)$, e $I(\theta)$ é a matriz da informação esperada. Vale ressaltar que a matriz da informação esperada é difícil de ser encontrada, então se usa como aproximação $J(\hat{\theta})$ que é a matriz de informação observada avaliada em $\hat{\theta}$. A partir daí pode-se construir intervalos de confiança e testes de hipóteses para os parâmetros individualmente, o que certamente contribui para o conveniente entendimento dos parâmetros.

2.1.5 A Distribuição Generalizada Beta t

A distribuição Generalizada Beta t, por abreviação GBt, foi apresentada primeiramente na literatura em Alexander et al. (2012) onde foi apresentado apenas sua densidade. A construção da distribuição GBt se dá de maneira simples. Basta incorporar a função acumulada da distribuição t de Student no núcleo da função acumulada da GBG

Utilizando a parametrização utilizada por Alexander et al. (2012), tomando $F(x; \lambda)$ como sendo a acumulada da distribuição t de Student, dada por:

$$F(x; \lambda) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{\lambda + x^2}} \right), \quad (25)$$

e sua densidade densidade é dada por:

$$f(x; \lambda) = \frac{\lambda}{(\lambda + x^2)^{3/2}}, \quad (26)$$

então a função densidade da GBt fazendo 25 e 26 em 8 é:

$$f(x; \lambda, a, b, c) = \frac{\lambda c}{2^{ac} B(a, b)} \frac{1}{(\lambda + x^2)^{3/2}} \left(1 + \frac{x}{\sqrt{\lambda + x^2}} \right)^{ac-1} \left[1 - \frac{1}{2^c} \left(1 + \frac{x}{\sqrt{\lambda + x^2}} \right)^c \right]^{b-1}, \quad (27)$$

a função (27) é apresentada em Alexander et al. (2012) como fruto da incorporação de (25) e (26) em (8).

Onde λ representa o parâmetro da função núcleo, $f(x;\lambda)$ e $F(x;\lambda)$ representam a função densidade e função de densidade acumulada da função núcleo, respectivamente.

A função (26) é chamada de Gerada Beta t , uma vez que é afirmado que $f(x;\tau)$ e $F(x;\tau)$ representariam a densidade e a acumulada da distribuição t de Student. Entretanto como visto que as funções (25) e (26) não apresentam a parametrização da distribuição t de Student tal qual é classicamente apresentada na literatura. A função densidade t de Student é dada por $f(x;\lambda) = \Gamma\left(\frac{\lambda+1}{2}\right) \left(\Gamma(\lambda/2) \sqrt{\lambda\pi}\right)^{-1} \left(1+x^2/\lambda\right)^{-(\lambda+1)/2}$, como pode ser visto em Mood, Graybill e Boes (1974), Tsay (2002), Morettin e Toloí 2004 e Bueno (2008).

Sendo assim a função densidade (25) e a função acumulada (26) parece ter propriedades interessantes e pode ser uma via bem mais simples de se obter (27) uma vez que a tradicional t de Student não possui forma fechada para sua função acumulada.

Todavia, mesmo que a função (27) não possua em seu núcleo uma t de Student tal qual é mais conhecida, ela apresentou facilidade de convergência quando utilizada como base de um modelo de regressão para séries temporais. Também apresenta suporte em toda a reta Real, e escrita em função da função Beta ao invés da função Gama, fato pode parecer inócuo, mas conforme frisa Sucarrat (2013), a função Gama representa um dos desafios em processos de otimização uma vez que existe uma barreira computacional (atualmente) para a função gama em 173, ou seja, gama de 173 é o máximo que pode ser computado, uma vez que a função gama generaliza o fatorial para números inteiros. Já quando se utiliza a função Beta esta barreira passa a ser mais distante, podendo chegar a 286.

É claro que uma das maneiras de estender os limites para a variação paramétrica seria reescrever a função gama em função de uma função beta. Isto é de fato um complicador a mais, que é desnecessário quando se usa da função (27), configurando uma vantagem a mais.

A função de distribuição acumulada da GBG, apresentada em sua forma estocástica, é dada por:

$$F(x; \lambda, a, b, c) = B(a, b)^{-1} \int_0^{F(x; \lambda)^c} w^{a-1} (1-w)^{b-1} dw, \quad (28)$$

que é facilmente implementada utilizando o software R, uma maneira simples para obter quantis da GBt é fazer um *grid* com integrações desta função acumulada. Sendo assim tomando uma sequência (tão grande e tão ampla quanto desejada) no suporte da função de distribuição, pode-se fazer uma operação de integração para cada valor da sequência criada, tal procedimento seria o equivalente ao da construção de uma tabela para a distribuição. O gráfico da função quantílica advém a partir dos pares ordenados (valor da sequência, resultado da integral), dispostos graficamente.

De posse da função de distribuição acumulada, a função quantílica viria de maneira trivial através da inversa de $F(x; \lambda, a, b, c)$. Entretanto esta operação de inversão não é tão simples quanto parece. Diversos *softwares* foram utilizados com o intuito de se conseguir a função inversa $F^{-1}(x; \lambda, a, b, c)$, mas sem sucesso, mesmo utilizando alguns dos mais poderosos *softwares* existentes como o *Maple* ou o *Wolfram/Alpha*. Segundo a literatura consultada Alexander et al. (2012) a função quantílica é dada por:

$$Q_{gbg}(u; \lambda, a, b, c) = F^{-1}\left(Q_{B(a,b)}(u)^{1/c}\right), \quad (29)$$

em que $Q_{B(a,b)}(u)$ é a função quantílica de $B(a,b)$, mas esta forma que também não é simples de ser trabalhada, requerendo a inversão da função Beta.

2.1.5.1 Geração de números aleatórios

Tomando $X = F^{-1}(U)$ com $U \sim Beta(a,b)$ sendo $Beta(a,b)$ a distribuição Beta e F representando uma distribuição de probabilidade acumulada qualquer, números aleatórios com distribuição GBG podem ser gerados facilmente. Supondo uma função núcleo, cuja função de distribuição acumulada seja dada por (25) fazendo a inversão de (25), considerando $U \sim Beta(a,b)$ e $w = U^{1/c}$, então números aleatórios de uma variável aleatória $X \sim GBG(a,b,c,\lambda)$ serão obtidos pela função (30),

$$X = \frac{\sqrt{\lambda}(2w-1)}{\sqrt{1-(2w-1)^2}}. \quad (30)$$

Como se pode ver pela equação (30), o parâmetro λ funciona como um parâmetro de escala, aumentando ou diminuindo a amplitude da variável aleatória. Assim quanto maior for λ maior será a amplitude da variável aleatória, independentemente dos demais parâmetros. Os parâmetros a , b e c controlam em conjunto a assimetria e a curtose. Diferentes combinações de a , b e c possibilitam o ajuste de um mesmo valor de curtose, isto equivale a dizer que um sistema possui diferentes soluções. Vale lembrar que a curtose representa o peso relativo das caudas em relação ao todo da função de densidade, sendo assim estes três parâmetros também determinam a dispersão uma vez que caudas mais pesadas representam dados mais dispersos. Desta forma a , b e c cumprem o mesmo papel de λ .

A geração de números aleatórios permite o estudo de algumas propriedades da função. Pode-se analisar como se comportam os momentos da distribuição quando da variação dos parâmetros. Este tipo de análise é preponderante quando do estudo de novas distribuições. A seguir algumas propriedades da distribuição Beta t Generalizada são demonstradas a partir da geração de números aleatórios.

As rotinas para geração de números aleatórios da distribuição GBt bem como da construção de gráficos de assimetria e curtose estão no Anexo A.

2.1.5.2 Estimação

Supondo que se deseje ajustar um conjunto de dados X a uma distribuição GBt. Este processo de estimação pode ser feito pelo método da máxima verossimilhança. Assim a partir da função de densidade de probabilidade da GBt cuja densidade é dada por (27), a função de verossimilhança é

$$L(a, b, c, \lambda; X) = \prod_{i=1}^l \left[\frac{\lambda c}{2acB(a, b)} \frac{1}{(\lambda + x_i^2)^{3/2}} \left(1 + \frac{x_i}{\sqrt{\lambda + x_i^2}} \right)^{ac-1} \left[1 - \frac{1}{2^c} \left(1 + \frac{x_i}{\sqrt{\lambda + x_i^2}} \right)^c \right]^{b-1} \right], \quad (31)$$

e a função de log verossimilhança será dada por:

$$l = \log(L(a, b, c, \lambda; X)). \quad (32)$$

As estimativas dos parâmetros são obtidas maximizando a função de log verossimilhança em relação a cada um dos parâmetros. Estas funções oriundas da derivação são as chamadas funções escores e dão suporte ao

processo de estimação, que é feito através de métodos iterativos de solução de equações não lineares. Para o caso em questão foram obtidas 4 funções:

$$\frac{\partial l}{\partial a}, \quad \frac{\partial l}{\partial b}, \quad \frac{\partial l}{\partial c}, \quad \frac{\partial l}{\partial \lambda}.$$

Tanto as funções escores como a função de log verossimilhança por serem muito grandes foram apresentadas no Anexo B.

2.2 Modelos de Séries Temporais

Atualmente existem inúmeras distribuições de probabilidade disponíveis na literatura e profissionais de todo o mundo estudam estas distribuições que possuem papel preponderante na construção e no entendimento de modelos de regressão.

Os modelos para séries temporais são modelos bastante peculiares, dada a ocorrência de correlação das observações ao longo do tempo e muitas vezes há uma heterocedasticidade condicional. Estas características são típicas de diversas séries financeiras.

As séries temporais envolvendo preços do café comumente representam um desafio à modelagem estatística. Problemas envolvendo as pressuposições a respeito da natureza do erro comumente conduzem a grosseiras aproximações.

A economia brasileira tem o seu suporte na produção do café, correspondendo a 15% do valor total da safra agrícola do País. Este “ouro verde” chegou ao Brasil por volta de 1730 e, desde então, ele se destaca na receita cambial, favorecendo outras atividades agrícolas não só na formação de capital, mas também na geração de empregos, CEPEA (2012).

Diante da importância desta cultura e da dificuldade de se fazer uma modelagem apropriada deste tipo de série é que se propõe a construção de um modelo de regressão baseado em uma distribuição mais flexível. As distribuições comumente utilizadas, Normal e t de Student, nem sempre são capazes de modelar o elevado grau de curtose existente nesta série. Espera-se, portanto, suplantar esta dificuldade, implementando uma função de probabilidade mais flexível, na verossimilhança para a variância condicional.

2.2.1 Modelos ARIMA

A pressuposição de independência é importantíssima em diversas situações. É grande simplificadora no contexto de proposição de modelos, estimação de parâmetros e processo inferencial.

Mas a pressuposição de independência introduz sérias limitações ao estudo de séries temporais onde a correlação é fato corrente, onde os erros são auto-correlacionados e influenciam a evolução do processo, Morettin e Tolo (2004).

Desta forma os modelos ARIMA (*Auto-regressive Integrated Moving Average* – Autorregressivo Integrado Médias Móveis) são uma ferramenta útil para a modelagem de dados correlacionados. Estes modelos são utilizados para descrever processos lineares estacionários, que podem ser representados por:

$$Z_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \sum_{k=0}^{\infty} \psi_k a_{t-k}, \psi_0 = 1,$$

em que a_t é um ruído branco, $\mu = E(Z_t)$ e $\psi_1, \psi_2 \dots$ é uma sequência de parâmetros tal que o somatório de suas estimativas ao quadrado seja finito.

Uma das pressuposições mais importantes envolvendo os modelos ARIMA, é a estacionaridade. Para uma revisão detalhada a respeito da estacionaridade, ver Bueno (2008).

Para se obter a estacionaridade, uma das técnicas mais difundidas é tomar a diferença entre as observações, Morettin e Tolo (2004); Tsay (2002).

As diferenças na série ao mesmo tempo em que permitem a obtenção de uma série estacionária, em se tratando de séries financeiras, tem uma

interpretação direta que muito interessa ao investidor ou ao profissional que lida diretamente neste campo, representa o retorno da série. O retorno da série representa quanto a série variou no período compreendido pela diferença, podendo representar uma alta no movimento da série se o retorno for positivo, ou pode representar uma baixa no movimento da série se o retorno for negativo.

Considere-se uma série de diferenças (retornos) $\{a_t, t = 1, \dots, n\}$, observadas em intervalos de tempo eqüidistantes. Esta série pode ser considerada como parte de uma realização de um processo estocástico $\{a_t, t \in Z\}$, cuja especificação fica completa a partir do conhecimento das distribuições finito-dimensionais

$$F(a_1, \dots, a_n; t_1, \dots, t_n) = P(a(t_1) \leq x_1, \dots, a(t_n) \leq x_n)$$

para qualquer tempo. Esta expressão pode ser reescrita em termos das distribuições condicionais, e será dada por:

$$F(a_1, \dots, a_n) = F_1(a_1) F_2(a_2 | a_1) \dots F_n(a_n | a_1, \dots, a_{n-1})$$

muitas das vezes o interesse reside em saber como as distribuições condicionais evoluem ao longo do tempo. É importante supor que a distribuição dos retornos seja não degenerada, ou seja, converge em distribuição para uma distribuição conhecida quando submetida a um processo de “normalização”.

A definição desta distribuição é peça fundamental no processo de estimação. Se X é uma distribuição estável então pode ser representada como uma soma de variáveis aleatórias estáveis. Suponha que X seja uma variável aleatória, e que existam constantes normalizadoras a_n, b_n para cada n , tais que $a_n X + b_n$ tem a mesma distribuição que a estatística de ordem (n)

$X_1 + \dots + X_n$) da variável aleatória X . Então diz-se que X é uma variável aleatória estável. Exemplo destas são a distribuição Normal e Cauchy com média e variância conhecidas e desconhecidas respectivamente Morettin e Toloi (2004).

Sob a condição de estabilidade $a_n = n^{1/\alpha}$ com $0 < \alpha \leq 2$. O coeficiente α é chamado índice e determina o peso das caudas. Se $\alpha = 2$ então se tem a distribuição Normal, caso contrário tem-se uma distribuição com caudas mais pesadas que a da Normal. Se $1 < \alpha < 2$, X têm média finita, e se $0 < \alpha \leq 1$ X teria média infinita como o caso da distribuição Cauchy.

Se a distribuição X possuir caudas mais pesadas, geralmente seus momentos de segunda ordem serão infinitos, todavia os retornos geralmente possuem momentos de segunda ordem finitos, o que limitaria o uso destas distribuições estáveis, exceto, é claro, no caso da distribuição Normal, que para o caso da modelagem de retornos, não é recomendada segundo a literatura.

Como pode ser visto, as distribuições de probabilidade desempenham papel crucial em todas as fases da modelagem deste tipo de dado. As características das distribuições têm como objetivo primordial a modelagem de uma realidade observada e a evolução do uso delas se dá no sentido de sempre melhorar o ajuste do modelo. Aumentando o poder preditivo dos modelos.

De qualquer forma, independentemente da distribuição em questão, desde que seja contínua, a obtenção de estimativas para seus parâmetros se dá via o método da máxima verossimilhança. Supondo que as distribuições condicionais $f_i(a_i | a_1, \dots, a_{i-1})$ com função de densidade dada por $f_i(a_i; \theta)$, então a função de verossimilhança será dada por:

$$f(a_1 \dots a_t; \theta) = f_1(a_1; \theta) \prod_{t=2}^n f_t(a_t; \theta).$$

O estimador de máxima verossimilhança de θ é obtido maximizando a função de verossimilhança ou o logaritmo dela.

Os modelos ARIMA nem sempre são capazes de modelar sozinhos uma série de dados econométricos. Muitas das vezes por mais parâmetros que se incorpore no modelo, não é possível obter níveis de eficiência satisfatórios e além do mais muitas das pressuposições a respeito do resíduo acabam por não se verificar. Uma das causas comumente encontrada é o fato da variância possuir um comportamento de agrupamentos. O que poderia ser descrito como uma espécie de correlação no comportamento da variância. Ou seja, variações grandes nos retornos tendem a serem seguidas por variações igualmente grandes, assim como, variações pequenas nos retornos tendem a serem seguidas por variações pequenas. Forma-se então ao longo da série alguns agrupamentos de volatilidade, seja ela alta ou baixa.

Neste caso é necessário um modelo que capte esta variação sistemática da volatilidade, ou volatilidade estocástica. Foi com foco na resolução deste problema que Engle (1982) estudando séries inflacionárias no Reino Unido, propôs os modelos ARCH.

2.2.2 Modelos ARCH

Quando se fala em séries financeiras, é comum ver uma estrutura de volatilidade condicional, e sendo assim, os modelos lineares do tipo ARIMA não são adequados para descrever o comportamento destas séries. Para este tipo de série geralmente se usa os modelos da classe ARCH (*autoregressive conditional heteroscedasticity*) que foi introduzida por Engle (1982).

Nos modelos da classe ARCH os retornos são serialmente não correlacionados, mas a variância condicional (volatilidade) é correlacionada aos retornos por meio de uma função quadrática. É definido por:

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2,$$

sendo ε_t é uma sequência de variáveis aleatórias i.i.d com média zero e variância 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i > 0$, m é o número de parâmetros do modelo ARCH. Usualmente ε_t tem distribuição t de student com ν graus de liberdade, ou distribuição normal com média 0 e variância 1. A dedução de algumas das propriedades deste modelo pode ser vista em Morettin e Tolo (2004). Considerando como exemplo um modelo ARCH(1):

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$$

$$E(a_t) = 0; \quad V(a_t) = \frac{\alpha_0}{1 - \alpha_1}; \quad Cov(a_{t+i}, a_t) = E(a_{t+i} a_t) = 0,$$

a curtose é dada por $K = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3$.

Ao se admitir que os retornos ao quadrado sigam um modelo ARCH um dos aspectos que é peculiar a estes modelos é que ele possui curtose elevada. Este fato é considerado uma vantagem dos modelos ARCH,

entretanto os retornos positivos continuam a ser tratados da mesma maneira que os retornos negativos, fato que não se verifica na prática, configurando uma desvantagem do modelo ARCH, uma vez que este pressupõe assimetria aos retornos.

Algumas restrições paramétricas devem ser levadas em consideração garantindo que a variância da série seja finita. Na prática, este componente aleatório é assumido ter distribuição normal ou t de Student.

Na abordagem ARCH grandes diferenças ao quadrado implica em grande variância condicional σ_t^2 no retorno médio de a_t . Sendo assim este retorno tende a ser grande. Esta tendência implica maior probabilidade de ocorrência, o que gera os agrupamentos ao longo da série. Este efeito ARCH é comum para várias séries econômicas.

2.2.2.1 Determinação da Ordem

A heterocedasticidade é detectada pelo teste F e a ordem do modelo pela função de autocorrelação parcial de a_t^2 . Dentro de uma abordagem ARCH, para uma dada amostra, a_t^2 é um estimador não viesado para σ_t^2 . Todavia espera-se que a_t^2 seja linearmente relacionada da forma $a_{t-1}^2, \dots, a_{t-m}^2$, em uma estrutura similar a de um modelo autorregressivo de ordem m , ARCH (m).

2.2.2.2 Estimação

Para a estimação dos parâmetros comumente se usa duas funções num contexto de modelos ARCH. Sob a pressuposição de normalidade a função de verossimilhança para um modelo ARCH(m) é dada por:

$$L(\boldsymbol{\alpha} | A_T) = f(a_T | F_{T-1})f(a_{T-1} | F_{T-2}) \dots f(a_{m+1} | F_m)f(a_1, \dots, a_m | \boldsymbol{\alpha}) \\ = \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{a_t^2}{2\sigma_t^2}\right\} f(a_1, \dots, a_m | \boldsymbol{\alpha}),$$

sendo $A_t = (a_1, a_2, \dots, a_m)$ é a variável aleatória em questão indexada pelo tempo, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_m)'$ é um vetor de parâmetros de um modelo ARCH(m) com $m < T$ e $f(a_1, \dots, a_m | \boldsymbol{\alpha})$ é a função de densidade de probabilidade conjunta que é desconhecida e por representar um produto em T produtos pode ser retirada, transformando a verossimilhança exata em uma verossimilhança condicionada às primeiras observações, que é uma aproximação da verossimilhança exata. A log-verossimilhança é matematicamente mais simples de ser trabalhada, transformando o produtório em somatório, sendo que este pode ser facilmente maximizado. Tal função de log-verossimilhança é dada por:

$$l(\boldsymbol{\alpha} | A_T) = \sum_{t=m+1}^T \left[-\frac{1}{2} \ln(\sigma_t^2) + \frac{1}{2} \frac{a_t^2}{\sigma_t^2} \right],$$

sendo $\boldsymbol{\alpha}$ representando o vetor de parâmetros que para o caso de um ARCH(m) modelando a variância $\sigma_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2$. Em algumas situações é mais razoável assumir que ε_t possui distribuição t de Student, justamente pelo fato de ser necessário uma distribuição de caudas mais pesadas. Tomando x_ν como sendo uma variável aleatória com distribuição t de Student com ν graus de liberdade, então

$Var(x_\nu) = \nu(\nu - 2)$ para $\nu > 2$ e usando $\varepsilon_t = x_\nu \sqrt{\nu/(\nu - 2)}$ Assim a densidade de ε_t é dada por,

$$f(\varepsilon_t | \nu) = \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)\sqrt{(\nu - 2)\pi}} \left(1 + \frac{\varepsilon_t^2}{(\nu - 2)}\right)^{-\frac{(\nu + 1)}{2}},$$

usando $a_t = \sigma_t \varepsilon_t$, $F(\varepsilon_t) = F\left(\frac{a_t}{\sigma_t}\right) \Rightarrow f(\varepsilon_t) = f\left(\frac{a_t}{\sigma_t}\right) \cdot \frac{1}{\sigma_t}$, neste caso a

função de verossimilhança condicional para a_t é dada por:

$$L(\nu, \sigma_t^2 | A_T) = \prod_{t=m+1}^T \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)\sqrt{(\nu - 2)\pi}} \frac{1}{\sigma_t} \left[1 + \frac{a_t^2}{(\nu - 2)\sigma_t^2}\right]^{-(\nu + 1)/2},$$

onde $\nu > 2$ e $A_T = (a_1, a_2, \dots, a_m)$. As estimativas obtidas por este procedimento são chamadas estimativas de máxima verossimilhança sob a pressuposição de distribuição t de Student. Aplicando o logaritmo tem-se

$$l(\nu, \sigma_t^2 | a_{m+1}, \dots, a_T) = - \sum_{t=m+1}^T \left[\frac{\nu + 1}{2} \ln \left(1 + \frac{a_t^2}{(\nu - 2)\sigma_t^2}\right) + \frac{1}{2} \ln \sigma_t^2 \right].$$

Considerando que o grau de liberdade seja pré-definido, o valor usualmente fica entre 3 e 6. Caso o grau de liberdade não seja pré-definido, então o logaritmo da verossimilhança fica:

$$l(\nu, \sigma_t^2 | a_{m+1}, \dots, a_T) = - \sum_{t=m+1}^T \left[\frac{\nu + 1}{2} \ln \left(1 + \frac{a_t^2}{(\nu - 2)\sigma_t^2}\right) + \frac{1}{2} \ln \sigma_t^2 \right] + (T - m)[\ln(\Gamma((\nu + 1)/2)) - \ln(\Gamma(\nu/2)) - 0,5 \ln((\nu - 2)\pi)].$$

O grau de liberdade nem sempre é estimado. Embora teoricamente para que possa se calcular a variância basta que o grau de liberdade seja maior que 2, comumente se utiliza o grau de liberdade como sendo igual a 5, que é o

menor grau de liberdade possível para que a distribuição t de Student possua forma fechada para os 4 primeiros momentos. Uma completa explanação a respeito deste processo de estimação pode ser vista em Tsay (2002).

Um aspecto negativo do uso da distribuição t de Student no processo de estimação de modelos econométricos é que apesar de esta distribuição possuir caudas mais pesadas que as da normal esta característica é mais acentuada num contexto de graus de liberdade menores. Assintoticamente as duas distribuições são equivalentes. E vale ressaltar que quando se fala em séries temporais comumente lida-se com grandes amostras, ou seja, grandes graus de liberdade. Tsay (2002) quando apresenta o que ele chama de inovação t , sugere a não estimação do parâmetro Grau de Liberdade, outrossim, fixa-o em $GL=5$, no intuito de assegurar a característica de caudas pesadas para distribuição de base.

Há suposição de uma distribuição t de Student tenta suprir a falta de ajustamento nas regiões caudais. Esta preocupação com a qualidade do ajuste nas regiões caudais se torna ainda mais relevante quando se trata de dados financeiros, onde sabidamente é comum um excesso de curtose. Os modelos sempre evoluíram no sentido de captar cada vez mais a natureza e o mecanismo de variação de determinado conjunto de dados. Entretanto, no que diz respeito à natureza do resíduo, comumente se opta pela aproximação normal, ou quando muito por uma aproximação via t de Student. De qualquer forma continua a se fazer aproximações. Neste sentido, este trabalho propõe a construção de um modelo baseado na distribuição Beta t Generalizada. Esta distribuição, por apresentar variadas combinações de assimetria e curtose, pode ser promissora neste intento.

2.2.2.3 Verificação de um modelo ARCH

Supondo o resíduo de um modelo ARIMA, ao qual esteja sendo ajustado um modelo ARCH(m), segundo Morettin e Toloi (2004), as observações padronizadas ou resíduos do modelo,

$$\tilde{a}_t = \frac{a_t}{\sqrt{\sigma_{t-1}^2}},$$

são variáveis aleatórias i.i.d com distribuição normal padrão ou t de Student. Para se verificar a adequação do modelo calcula-se a estatística Q de Ljung-Box para a sequência \tilde{a}_t , além disso o cálculo dos coeficientes de assimetria e curtose além de uma análise gráfica a partir de um QQ-Plot podem ser utilizadas para aferir o ajustamento.

2.2.2.4 Previsão em um modelo ARCH

As previsões de volatilidade um modelo ARCH(m), para um horizonte $t+l$ são obtidas por

$$\hat{\sigma}_t^2(1) = \alpha_0 + \alpha_1 a_t^2 + \dots + \alpha_m a_{t-m}^2,$$

as previsões l passos a frente a partir da origem fixada em t são dadas por

$$\hat{\sigma}_t^2(l) = \alpha_0 + \sum_{i=1}^T \alpha_i \hat{\sigma}_t^2(l-i), \quad \hat{\sigma}_t^2(l-i) = a_{t+l-i}^2, \quad \text{se } l-i \leq 0.$$

2.2.3 Modelos GARCH

Bollerslev (1986), sugeriu uma generalização para o modelo ARCH, chamado de *Generalized ARCH* ou GARCH cuja evolução em relação ao seu precursor é o fato de poder descrever a volatilidade condicional com um menor número de parâmetros. Seja uma série qualquer e assumindo que a

média do processo pode ser descrita por um modelo ARMA, então a_t segue um modelo Generalizado Autorregressivo Condicionalmente Heteroscedástico, GARCH(p,q), se

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i a_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

$\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j > 0$, $\sum_{j,i=1}^{\max(p,q)} (\alpha_i + \beta_j) < 1$, então para o próximo período de previsão da variância tem-se uma mistura da previsão do último período e do último retorno ao quadrado.

Para entender bem as propriedades de um modelo GARCH, é importante lançar mão de uma pequena transformação. Fazendo $\eta_t = a_t^2 - \sigma_t^2$ então $\sigma_t^2 = a_t^2 - \eta_t$, então se pode reescrever a equação do modelo GARCH da seguinte forma:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i \beta_i) a_{t-i}^2 + \eta_t + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

é visível que se s for igual a zero retorna-se ao modelo ARCH(m). Já η_t não é uma sequência i.i.d., é uma diferença martingal, possuindo $E[\eta_t] = 0$ e $\text{cov}[\eta_t, \eta_{t-j}] = 0$, $j \geq 0$. Este resultado pode ser diretamente obtido de Tsay (2002) ou Morettin e Tolo (2004).

Um modelo GARCH pode ser visto como uma aplicação da idéia de um ARMA à série ao quadrado a_t^2 . Usando esta idéia tem-se que

$$E(a_t^2) = \frac{\alpha_0}{1 - \sum_{j,i=1}^{\max(p,q)} (\alpha_i + \beta_j)}.$$

Pode-se ver que como o modelo ARCH(1) é um modelo AR(1) para o quadrado do resíduo, um modelo GARCH(1,1) é um modelo ARMA(1,1) para o quadrado do resíduo. Fazendo a substituição $v_t = a_t^2 - \sigma_t^2$, tem-se

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ a_t^2 - v_t &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 (a_{t-1}^2 - v_{t-1}) \\ a_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + v_{t-1} - \beta_1 v_{t-1}.\end{aligned}$$

que é um ARMA(1,1) para o quadrado do resíduo.

A variância incondicional do quadrado do resíduo é

$$\begin{aligned}\text{Var}(a_t^2) &= E[a_t^2] - (E[a_t])^2 \\ &= E[a_t^2] = E[\sigma_t^2 \varepsilon_t^2] = E[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 E[a_{t-1}^2] + \beta_1 \sigma_{t-1}^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) E[a_{t-1}^2].\end{aligned}$$

Como a_t é um processo estacionário, $\text{Var}(a_t) = \alpha_0 / (1 - \alpha_1 - \beta_1)$ e uma vez que $a_t = \sigma_t \varepsilon_t$, a variância incondicional dos retornos é também $\alpha_0 / (1 - \alpha_1 - \beta_1)$, $E[\sigma_t^2] = E[a_t^2]$.

Da mesma forma que uma modelo ARMA(1,1) pode ser escrito como um modelo AR(∞), um modelo GARCH(1,1) pode ser escrito como um ARCH(∞),

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 a_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2 (\alpha_0 + \alpha_1 a_{t-3}^2 + \beta_1 \sigma_{t-3}^2) \\
&\quad \vdots \\
&= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=1}^{\infty} a_{t-1-i}^2 \beta_1^i.
\end{aligned}$$

Pode-se ver então que a variância condicional no tempo t é uma soma ponderada do resíduo ao quadrado do tempo anterior, entretanto o peso destes resíduos na ponderação cai na medida em que se avança no passado.

Dado que a variância não condicional dos retornos é $E[\sigma_t^2] = \alpha_0 / (1 - \alpha_1 - \beta_1)$, a equação do modelo GARCH(1,1) pode ser reescrita por

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= (1 - \alpha_1 - \beta_1) E[\sigma_t^2] + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.
\end{aligned}$$

Desta forma pode-se ver que a variância não condicional do próximo período é uma ponderação entre as variâncias não condicionais dos retornos $E[\sigma^2]$, entre os retornos ao quadrado do último período a_{t-1}^2 , e a variância não condicional do retorno no tempo anterior σ_{t-1}^2 , com pesos que somam 1, $(1 - \alpha_1 - \beta_1)$, α_1 , β_1 .

Isto é importante para se fazer previsões para o as variâncias tanto do próximo retorno quanto também para l passos adiante. A partir da equação do modelo GARCH(1,1) pode-se escrever a equação de previsão do próximo período para a variância, $\hat{\sigma}_{t+l}^2 = \sigma^2 + (\alpha_1 + \beta_1)^l (\sigma_t^2 - \sigma^2)$. Tsay (2002) é uma fonte para estas demonstrações. A variância para o próximo passo é deduzida a seguir:

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\
\hat{\sigma}_{t+1}^2 &= \alpha_0 + \alpha_1 E[a_t^2 | I_{t-1}] + \beta_1 \sigma_t^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 + \beta_1 \sigma_t^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1) (\sigma_t^2 - \sigma^2), \\
\hat{\sigma}_{t+2}^2 &= \alpha_0 + \alpha_1 E[a_{t+1}^2 | I_{t-1}] + \beta_1 E[\sigma_{t+1}^2 | I_{t-1}] \\
&= \alpha_0 + (\alpha_1 + \beta_1) \hat{\sigma}_t^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1) (\hat{\sigma}_{t+1}^2 - \sigma^2) \\
&= \sigma^2 + (\alpha_1 + \beta_1)^2 (\sigma_t^2 - \sigma^2), \\
&\vdots \\
\hat{\sigma}_{t+l}^2 &= \alpha_0 + (\alpha_1 + \beta_1) \hat{\sigma}_{t+l-1}^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1)^l (\hat{\sigma}_{t+l-1}^2 - \sigma^2) \\
&= \sigma^2 + (\alpha_1 + \beta_1)^l (\sigma_t^2 - \sigma^2).
\end{aligned}$$

Em que I_{t-l} é a informação no tempo $t-l$.

Pontos positivos e negativos dos modelos GARCH são os mesmos do modelo ARCH e assim como os modelos ARIMA, são uma generalização dos modelos autorregressivos, da mesma forma os modelos do tipo GARCH propiciam, via de regra, melhor ajuste que um modelo ARCH, e com a vantagem de considerar um menor número de parâmetros.

Dentre os modelos do tipo GARCH mais utilizados está o GARCH(1,1), segundo Morettin e Toloi (2004). Partindo-se do modelo, cuja volatilidade é expressa por

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad 0 \leq \alpha, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1.$$

Primeiramente pode-se ver que se os componentes $\alpha_1 a_{t-1}^2$ ou $\beta_1 \sigma_{t-1}^2$ forem grandes, σ_t^2 também o será, daí o comportamento de agrupamento de volatilidade, ou seja, se houver uma grande variação tende a haver outra

igualmente grande, assim como se houver uma pequena variação, tende a haver outra igualmente pequena.

2.2.3.1 Curtose nos modelos GARCH

A volatilidade é sempre um importante aspecto a ser levado em conta no processo de estimação, entretanto é uma dificuldade que pode ser transposta. Para se ter uma idéia correta da variabilidade da volatilidade estocástica, deve-se sempre considerar a curtose dos modelos com volatilidade. Considerando o modelo GARCH(1,1), é mostrado como a curtose está sistematizada. Tal modelo pode ser escrito por

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

com $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 > 0$, $(\alpha_1 + \beta_1) < 1$, e ε_t é uma sequência i.i.d. onde

$$E(\varepsilon_t) = 0, \quad \text{var}(\varepsilon_t) = 1, \quad E(\varepsilon_t^4) = K_\varepsilon + 3.$$

Em que K_ε é o excesso de curtose para a distribuição assumida para ε_t .

Tendo que $V(a_t) = E(\sigma_t^2) = \alpha/[1 - (\alpha_1 + \beta)]$, e que $E(\sigma_t^4) = (K_\varepsilon + 3)E(\sigma_t^2)$, se σ_t^4 existir.

Tomando o quadrado do modelo de volatilidade tem-se

$$\sigma_t^4 = \alpha_0^2 + \alpha_1^2 a_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2\alpha_0 \alpha_1 a_{t-1}^2 + 2\alpha_0 \beta_1 \sigma_{t-1}^2 + 2\alpha_1 \beta_1 \sigma_{t-1}^2 a_{t-1}^2.$$

Tomando a esperança de σ_t^4 e utilizando algumas propriedades, tem-se

$$E(\sigma_t^4) = \frac{\alpha_0^2(1 + \alpha_1 + \beta_1)}{[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(K_\varepsilon + 2) - (\alpha_1 + \beta_1)^2]}.$$

Assegurando-se que sendo $2\alpha_1 - (\alpha_1 + \beta_1)^2 > 0$ tem-se que

$$K_a = \frac{E(a_t^4)}{[E(a_t^2)]^2} - 3 = \frac{(K_\varepsilon + 3)[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2 - K_\varepsilon\alpha_1^2} - 3.$$

Sob a pressuposição de normalidade $K_\varepsilon = 0$ então K_a reduz-se a

$$K_a^N = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} - 3,$$

O sobrescrito N indica que a curtose é calculada sob a pressuposição de normalidade. E após algumas manipulações algébricas pode ser reescrita por

$$K_a^N = \frac{6\alpha_1^2}{1 + 2\alpha_1^2 + (\alpha_1 + \beta_1)^2}.$$

Este resultado tem duas importantes implicações. A primeira é que a curtose só existirá se o denominador for maior que zero. A segunda é que se $\alpha_1 = 0$ então a curtose será zero. Ou seja, o excesso de curtose do modelo GARCH(1,1) considerando $\alpha_1 = 0$ não existe, o mesmo que dizer que o modelo não possui caudas pesadas. Afora esta segunda condição, mesmo sob a pressuposição de normalidade, o modelo GARCH(1,1) possui caudas pesadas.

2.2.3.2 Estimação

A estimação pelo método da máxima verossimilhança faz uso da função de distribuição conjunta. A função de verossimilhança é pensada

como uma função dos parâmetros dado um conjunto de dados,
 $L(\Theta|x_1, \dots, x_n)$.

Supõe-se que o resíduo é independente. Esta suposição permite que a função de distribuição conjunta seja dada pelo produtório das funções de densidade de probabilidade marginais. No contexto GARCH a independência é pouco factível, entretanto ainda é possível escrever a função de probabilidade conjunta como um produto de densidades marginais.

$$\begin{aligned} f(a_1, \dots, a_T) &= f(a_T|a_1, \dots, a_{T-1})f(a_1, \dots, a_{T-1}) \\ &= f(a_T|a_1, \dots, a_{T-1})f(a_{T-1}|a_1, \dots, a_{T-2})f(a_1, \dots, a_{T-2}) \\ &\quad \vdots \\ &= f(a_T|a_1, \dots, a_{T-1})f(a_{T-1}|a_1, \dots, a_{T-2}) \cdots f(a_1). \end{aligned}$$

Conquanto que a amostra seja suficientemente grande o último termo da direita $f(a_1)$, pode ser desconsiderado Morettin e Toloi (2004).

Considerando o modelo GARCH(1,1),

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

sob a pressuposição de normalidade a função de verossimilhança e dada por

$$L(\alpha_0, \alpha_1, \beta_1, \mu|a_1, \dots, a_T) = \prod_{t=2}^T (2\pi\sigma_{t-1}^2)^{-0.5} \exp(-(a_{t-1} - \mu)^2 / 2\sigma_{t-1}^2).$$

E a função de log verossimilhança é dada por

$$l(\alpha_0, \alpha_1, \beta_1, \mu|a_1, \dots, a_T) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=2}^T \log(\sigma_t^2) - \frac{1}{2} \sum_{t=2}^T \left(\frac{(a_t - \mu)^2}{2\sigma_t^2} \right),$$

com $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

O relaxamento da pressuposição de normalidade para o resíduo é fato corrente na literatura a respeito de séries temporais. Comumente se utiliza a distribuição t de Student como distribuição base para o resíduo, neste caso a função de verossimilhança é dada por:

$$L(\alpha_0, \alpha_1, \beta_1, \nu | a_1, \dots, a_T) = \prod_{t=1}^T \left[\Gamma\left(\frac{\nu+1}{2}\right) \left(\Gamma(\nu/2) \sqrt{(\nu-2)\pi\sigma_t^2}\right)^{-1} \exp\left(-\frac{\varepsilon_t^2}{(\nu-2)\sigma_t^2}\right)^{\frac{\nu+1}{2}} \right],$$

com $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

No software R, há implementado, rotinas de estimação do modelo GARCH através da função `nlminb()` e da função `optim` (com o método “L-BFGS-B”), entretanto em todos os casos o nó górdio é o calculo da função de log verossimilhança que necessita de valores para a variância condicional, cujo cálculo, segundo Wurtz, Chalabi e Luksan (2008), requer um duplo *loop* para a sua implementação mais simples.

```
# Código para computação da Variância Condicional
N = 10000; eps = round(rnorm(N), digits = 2) / 10
omega = 0.1; alpha = c(0.1, 0.05); gamma = c(0, 0); beta = c(0.4, 0.3);
delta = 2
u = length(alpha); v = length(beta); uv = max(u,v); h = rep(0.1, uv)
# Case I: Conditional Variances as Double for-Loop:
for (i in(uv+1):N ) {
  ed = 0
  for (j in 1:u) {
    ed = ed+alpha[j]*(abs(eps[i-j])-gamma[j]*eps[i-j])^delta
  }
  h[i] = omega + ed + sum(beta*h[i-(1:v)])
}
```

Entretanto o tempo gasto pelo duplo loop pode ser resumido usando a função `filter()` que possibilita rapidez no processo de iteração. O uso da função `filter()` é demonstrado nas linhas abaixo.

```

# Usando a função filter() ao invés de loops para a variância condicional:
edelta = (abs(eps)-gamma*eps)^delta
edeltat = filter(edelta, filter = c(0, alpha), sides = 1)
c = omega/(1-sum(beta))
h = c( h[1:uv], c + filter(edeltat[-(1:uv)], filter = beta, method =
"recursive", init = h[uv:1]- c))

```

O cálculo da variância condicional pelo método do filtro, aumenta a velocidade da computação da função objetiva enormemente, dependendo do tamanho da série. Para a obtenção do erro padrão das estimativas é necessário a matriz hessiana e/ou a matriz gradiente que podem ser calculadas numericamente.

2.3 Métodos de Otimização

Não há como falar em modelos não lineares, como é o caso dos modelos GARCH sem falar de métodos de estimação. Para a otimização de parâmetros em modelos GARCH há diversas funções implementadas e diversos algoritmos são utilizados para esta proposta. No caminhar por entre o estado da arte, no que diz respeito a estimação em modelos GARCH, é impossível não se deparar com siglas tais como: BFGS, SANN, CG, L-BFGS-B, Nelder-Mead, entre outros algoritmos de otimização, que são compilados dentro de funções diversas para otimização tais como “lme”, “optim”, “nlminb”.

Dentre os métodos de estimação o proposto por Nelder e Mead (1965) é um método robusto, relativamente lento uma vez que usa os valores absolutos da função objetiva, e é o mais utilizado para otimização de funções não diferenciáveis, é o método padrão da função *optim()*. Esta função pode ser usada tanto para um parâmetro quanto para muitos parâmetros. Também opera quando não há nenhum parâmetro, caso em que apenas computa o valor da função.

Assim como o método de Nelder-Mead, também se encontra implementado o método BFGS que é um método do tipo Quase-Newton. Esse método foi proposto por Broyden, Fletcher, Goldfarb e Shanno. Extensão deste é o método L-BFGS-B que permite a constrição do espaço paramétrico ao qual deve conter os valores iniciais do processo de iteração. Este método foi proposto por Byrd et al. (1995), e além desta nova funcionalidade em relação ao seu predecessor, apresenta ainda a vantagem de ser mais rápido.

Fletcher e Reeves (1964) propuseram o método CG (abreviação de *Conjugate Gradients*), método que é menos eficiente que o BFGS, porém é melhor em solução de sistemas mais complexos.

Outro método que assim como o de Nelder-Mead, utiliza apenas valores da função objetiva (não sendo portanto auxiliado pela função gradiente) é o método do recozimento, conhecido na literatura por SANN (sigla de *Simulated Annealing*) Belisle (1992). Pelo fato de utilizar apenas valores da função objetiva, é um método relativamente lento e depende massivamente dos parâmetros de controle. Não é um método muito utilizado na estatística mas figura entre um dos melhores na tarefa de minimização de superfícies de respostas complexas.

A existência de grande número de algoritmos bem como de funções de otimização se dá pelo fato de ser complexa a tarefa de se encontrar um valor paramétrico que minimize a função objetiva para estes modelos.

Parte da solução deste desafio vem em muitos casos da utilização de uma abordagem de “dois estágios”, segundo Wurtz, Chalabi e Luksan (2008). O primeiro estágio seria encontrar uma solução (que seja alguma perto da solução ótima do sistema de equações) utilizando algum algoritmo que possa limitar a variação da estimativa dentro de limites de constrição previamente definidos, o que permite a obtenção de estimativas dentro do

espaço paramétrico. Este primeiro estágio pode ser alcançado via métodos como L-BFGS-B e CG, que permitem a constrição do espaço paramétrico. O segundo estágio consiste em utilizar as estimativas obtidas no primeiro estágio como valores iniciais para o segundo estágio no qual é utilizado um método de estimação que não permite a constrição do espaço paramétrico, neste estágio pode se usar métodos de otimização como Nelder-Mead ou SANN (método do recozimento). Uma revisão a respeito destes métodos pode ser vista em Nocedal e Wright (1999).

Esta abordagem integrada em muitos casos permite a solução do problema, e está implementada em diversos pacotes estatísticos como Matlab, Gauss e Ox. No R ela está implementada pelo método BFGS-NM.

O algoritmo padrão para a estimação de modelos GARCH é chamado de sqp (*Sequential quadratic programming*) constitui parte de diversos softwares estatísticos como Ox e Matlab. Algoritmos baseados em sqp são extremamente eficientes como métodos não lineares. Wurtz, Chalabi e Luksan (2008) escreveu em R o programa SQP Fortran, e consiste num método de programação recursiva utilizado para otimização de modelos GARCH.

2.4 Critérios de Avaliação de Modelos

Em se tratando de séries temporais é comum diferentes pesquisadores obterem diferentes modelos considerando um mesmo conjunto de dados então para minorar este fato e tentar obter resultados consistentes uma importante ferramenta para medir a qualidade do ajuste é o teste não paramétrico da estatística Q.

Dentre os testes não paramétricos talvez o mais utilizado seja o teste de Ljung-Box cuja estatística Q aplicada ao resíduo ao quadrado,

comumente aplicado ao resíduo de um modelo ARMA(p,q), comumente ponto de partida para o ajuste de um modelo GARCH(p,q). A estatística deste teste é dada por:

$$Q(m) = T((T-2)) \sum_{i=1}^m \frac{\hat{\rho}_i a_i^2}{T-i},$$

em que T é o comprimento da série, a_t^2 representa a série de resíduos ao quadrado, m é um número escolhido de correlações a serem computadas e $\hat{\rho}_i$ é a autocorrelação estimada do i -ésimo lag da série a_t^2 . A estatística do teste $Q(m)$ possui, assintoticamente, distribuição qui-quadrado com $m - p - q$ graus de liberdade. Esta estatística é importante na detecção de heteroscedasticidade condicional de a_t e é assintoticamente equivalente ao teste dos multiplicadores de Lagrange.

Este teste equivalente ao teste F para $\alpha_i = 0$ em uma regressão linear do tipo

$$\alpha_t^2 = \alpha_0 + \alpha_1 a_{t-1} + \dots + \alpha_m a_{t-m} + e_t, \quad t = m+1, \dots, T,$$

m é um número inteiro qualquer, T é o comprimento da série e_t é um componente aleatório, per figurando o erro do modelo. Então a estatística F será dada por:

$$F = \frac{\left(\sum_{t=m+1}^T (a_t^2 - \bar{w}) - \sum_{t=m+1}^T (\hat{e}_t^2) \right)}{m} \frac{T-2m-1}{\sum_{t=m+1}^T (\hat{e}_t^2)} \sim \chi^2(m) \text{ sob } H_0,$$

em que \bar{w} representa a média de a_t^2 e \hat{e}_t^2 representa os resíduos da regressão linear.

O princípio da parcimônia, sempre presente na construção de modelos, tenta coadunar o número de parâmetros à efetividade do modelo.

Quanto mais parâmetros, melhor é o ajuste, conclusão um tanto quanto óbvia. Então geralmente o número de parâmetros deve constituir uma função penalizadora que deve compor alguma função que subsidie a seleção de um modelo dentre outros. A maioria dos programas já tem em seus pacotes estes critérios de seleção compilados. De tal forma que após o ajuste de um modelo o cálculo destas quantidades se dá de forma automática. Dentre os mais utilizados estão o critério de Akaike (AIC, sigla proveniente do inglês *Akaike Information Criterion*), o critério de Informação Bayesiano (BIC, sigla proveniente do inglês *Bayesian Information Criterion*), o critério de Informação de Schwarz (SIC) e o critério de Hannan e Quinn (HQC, sigla do inglês de *Hannan Quinn Criterion*). Estes critérios são amplamente utilizados em qualquer livro que trate de análise de séries temporais como Tsay (2002), Morettin e Tolo (2004) e Bueno (2008). As estatísticas dos referidos testes são dadas por:

$$AIC_{\alpha}(k) = -2l(\hat{\theta}) + 2n/T,$$

$$BIC(k) = -2l(\hat{\theta}) + n \frac{\log(T)}{T},$$

$$SIC(k) = \frac{-2l(\hat{\theta})}{T} + \log\left(\frac{n + 2 * T}{T}\right),$$

$$HQC(k) = -2l(\hat{\theta}) + 2n \frac{\log(\log(T))}{T}.$$

Em que $l(\hat{\theta})$ é a estimativa de máxima verossimilhança de $l(\theta)$, n é o número de parâmetros e T é o comprimento da série.

3 METODOLOGIA

3.1 A construção de uma nova proposta de estimação para o modelo GARCH

A chamada inovação t apresenta um progresso no processo de estimação, entretanto ajustes caudais ainda são demandados. É neste sentido que a proposta deste trabalho se dirige, uma vez que sob certas condições o processo da volatilidade é estritamente estacionário, mas não fracamente estacionário, uma vez que não possui os dois primeiros momentos.

A utilização de uma distribuição mais razoável vem no sentido de no momento de se assumir uma distribuição para o erro de um modelo GARCH (ε_t), possa-se adotar uma distribuição que não cause distorções, que realmente seja capaz de captar o real comportamento do resíduo, que tenha caudas pesadas, assim como sugere o processo da volatilidade.

Considerando um modelo GARCH(1,1), cuja construção é dada por:

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

então supondo que ε_t tenha distribuição GBt cuja função de densidade é dada por:

$$f(\varepsilon_t | a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1) = \frac{\lambda c}{2^{ac} B(a, b)} \frac{1}{(\lambda + \varepsilon_t^2)^{3/2}} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^{ac-1} \left[1 - \frac{1}{2^c} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^c \right]^{-b-1},$$

Com $\varepsilon_t = a_t / \sigma_t$, e $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ fato que demanda uma transformação de variável do tipo

$$a_t = \sigma_t \varepsilon_t, \quad F(\varepsilon_t) = F\left(\frac{a_t}{\sigma_t}\right) \Rightarrow f(\varepsilon_t) = f\left(\frac{a_t}{\sigma_t}\right) \cdot \frac{1}{\sigma_t},$$

a função de verossimilhança para a_t será dada por:

$$L(a_t | a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1) = \prod_{t=2}^T \left[\frac{\lambda c}{2^{ac} B(a, b) (\lambda + (a_t / \sigma_t)^2)^{3/2}} \left(1 + \frac{(a_t / \sigma_t)}{\sqrt{\lambda + (a_t / \sigma_t)^2}} \right)^{ac-1} \right]^{b-1} \frac{1}{\sigma_t}$$

$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ e a função de log verossimilhança será dada por:

$$l(a_t | a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1) = \log \left(L(a_t | a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1) \right),$$

assumindo que o resíduo do modelo ARIMA a_t tem distribuição GBt e tomando $\theta = (a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1)^T$ como sendo o verdadeiro vetor de parâmetros. As estimativas dos parâmetros são obtidas maximizando a função de log verossimilhança em relação a cada um dos parâmetros. Estas funções oriundas da derivação da função l são as chamadas funções escores e dão suporte ao processo de estimação, que é feito através de métodos iterativos de solução de equações não lineares. Para o caso em questão foram obtidas o vetor de funções escores é dado por

$$U = U(\theta) = \left(\frac{\partial l}{\partial a}, \frac{\partial l}{\partial b}, \frac{\partial l}{\partial c}, \frac{\partial l}{\partial \lambda}, \frac{\partial l}{\partial \alpha_0}, \frac{\partial l}{\partial \alpha_1}, \frac{\partial l}{\partial \beta} \right)^t$$

e por serem um pouco complexas, são apresentadas somente em torno desta notação, mas podem ser obtidas via integração analítica utilizando softwares como o Maple.

Para intervalo de confiança e teste de hipóteses a respeito dos parâmetros do modelo, é necessário a obtenção da matriz da informação que para o caso será uma matriz $I(\theta)_{7 \times 7}$, dada por:

$$I(\theta) = \begin{pmatrix} \frac{\partial^2 l}{\partial a \partial a} & \frac{\partial^2 l}{\partial a \partial b} & \frac{\partial^2 l}{\partial a \partial c} & \frac{\partial^2 l}{\partial a \partial \lambda} & \frac{\partial^2 l}{\partial a \partial \alpha_0} & \frac{\partial^2 l}{\partial a \partial \alpha_1} & \frac{\partial^2 l}{\partial a \partial \beta} \\ \frac{\partial^2 l}{\partial b \partial a} & \frac{\partial^2 l}{\partial b \partial b} & \frac{\partial^2 l}{\partial b \partial c} & \frac{\partial^2 l}{\partial b \partial \lambda} & \frac{\partial^2 l}{\partial b \partial \alpha_0} & \frac{\partial^2 l}{\partial b \partial \alpha_1} & \frac{\partial^2 l}{\partial b \partial \beta} \\ \frac{\partial^2 l}{\partial c \partial a} & \frac{\partial^2 l}{\partial c \partial b} & \frac{\partial^2 l}{\partial c \partial c} & \frac{\partial^2 l}{\partial c \partial \lambda} & \frac{\partial^2 l}{\partial c \partial \alpha_0} & \frac{\partial^2 l}{\partial c \partial \alpha_1} & \frac{\partial^2 l}{\partial c \partial \beta} \\ \frac{\partial^2 l}{\partial \lambda \partial a} & \frac{\partial^2 l}{\partial \lambda \partial b} & \frac{\partial^2 l}{\partial \lambda \partial c} & \frac{\partial^2 l}{\partial \lambda \partial \lambda} & \frac{\partial^2 l}{\partial \lambda \partial \alpha_0} & \frac{\partial^2 l}{\partial \lambda \partial \alpha_1} & \frac{\partial^2 l}{\partial \lambda \partial \beta} \\ \frac{\partial^2 l}{\partial \alpha_0 \partial a} & \frac{\partial^2 l}{\partial \alpha_0 \partial b} & \frac{\partial^2 l}{\partial \alpha_0 \partial c} & \frac{\partial^2 l}{\partial \alpha_0 \partial \lambda} & \frac{\partial^2 l}{\partial \alpha_0 \partial \alpha_0} & \frac{\partial^2 l}{\partial \alpha_0 \partial \alpha_1} & \frac{\partial^2 l}{\partial \alpha_0 \partial \beta} \\ \frac{\partial^2 l}{\partial \alpha_1 \partial a} & \frac{\partial^2 l}{\partial \alpha_1 \partial b} & \frac{\partial^2 l}{\partial \alpha_1 \partial c} & \frac{\partial^2 l}{\partial \alpha_1 \partial \lambda} & \frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_0} & \frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_1} & \frac{\partial^2 l}{\partial \alpha_1 \partial \beta} \\ \frac{\partial^2 l}{\partial \beta \partial a} & \frac{\partial^2 l}{\partial \beta \partial b} & \frac{\partial^2 l}{\partial \beta \partial c} & \frac{\partial^2 l}{\partial \beta \partial \lambda} & \frac{\partial^2 l}{\partial \beta \partial \alpha_0} & \frac{\partial^2 l}{\partial \beta \partial \alpha_1} & \frac{\partial^2 l}{\partial \beta \partial \beta} \end{pmatrix}.$$

$I(\theta)$ é chamada de matriz hessiana ou matriz de derivadas segundas (Mood, Graybill e Boes 1974). A variância mínima de um estimador é dada pelo Limite Inferior de Crámer-Rao, que é $LI = 1/(-E[I(\theta)])$, entretanto $E[I(\theta)]$ é desconhecida, podendo ser estimada por $E[I(\hat{\theta})]$ com $\hat{\theta}$ estimador de θ .

Sendo assim, se tomarmos o inverso da matriz hessiana ($I(\theta)^{-1}$) temos uma matriz de variâncias e covariâncias, onde os elemento da diagonal principal representam as variâncias enquanto que os demais

elementos representam as covariâncias. Por se tratar de uma matriz cuja diagonal principal é composta por estimativas de variâncias, que são quantidades estritamente positivas, estamos falando de uma matriz positiva definida, cuja estimativa pode ser dada por:

$$I(\theta)^{* -} \cong I(\theta)^{-} G(\theta) G(\theta)' I(\theta)^{-},$$

em que $G(\theta)$ é a matriz gradiente, ou seja, uma matriz de derivadas parciais. Esta operação assemelha-se a uma restrição na solução, garantido que $I(\theta)^{* -}$ seja uma estimativa de $I(\theta)^{-}$ e que seja positiva definida. Esta aproximação é via de regra utilizada em métodos numéricos, parte da premissa de que $G(\theta)G(\theta)'$ é aproximadamente igual a $I(\theta)$, ou seja, o quadrado da matriz de derivadas primeiras como aproximadamente igual à matriz de derivadas segundas Huber (1967).

Sendo assim os elementos da diagonal principal de $I(\theta)^{* -}$ representam estimativas de variância dos estimadores. O estimador de máxima verossimilhança $\hat{\theta}$ de θ é calculado numericamente a partir da solução do sistema de equações escores $U(\theta_i) = 0$. Sob certas condições de regularidade, não estando θ no limite do espaço paramétrico, $\sqrt{n}(\hat{\theta} - \theta)$ possui distribuição assintótica normal multivariada $N_7(0, I(\theta)^{-1})$.

Intervalos de confiança para os parâmetros do modelo podem facilmente ser obtidos:

$$IC_{(1-\alpha)}(\theta_i): \hat{\theta}_i \pm Z_{\alpha/2} \sqrt{I(\theta)^{* -}_{i,i}},$$

e $I(\theta)^{* -}_{i,i}$ denota o i -ésimo elemento da diagonal principal de $I(\theta)^{* -}$, com $i = 1, \dots, 7$ e $Z_{\alpha/2}$ é um quantil da distribuição normal padrão.

A normalidade assintótica também pode ser útil para verificar a qualidade do ajuste da distribuição, comparando a distribuição GBt com um de seus modelos aninhados como por exemplo as distribuições: t de Student, Beta e Kumaraswamy. Esta comparação pode ser feita via o teste da razão de verossimilhança cuja estatística é dada por LR:

$$LR = 2(l(\hat{\theta}) - l(\tilde{\theta})),$$

sendo $l(\tilde{\theta})$ a log verossimilhança do modelo aninhado ao modelo cuja log verossimilhança é expressa por $l(\hat{\theta})$. LR tem distribuição assintótica χ_k^2 em que k representa a diferença entre as dimensões de $\hat{\theta}$ e $\tilde{\theta}$.

Outros critérios de seleção de modelos são comumente utilizados, como AIC, BIC, HQC, entre outros. Estes critérios são baseados na estimativa da função de verossimilhança calculada com base nas estimativas dos parâmetros estimados, e tem esta grandeza penalizada pela quantidade de parâmetros. Quanto maior o número de parâmetros maior será a função penalizadora.

4 APLICAÇÃO

4.1 Aplicação para uma série temporal do preço de café. Abordagem clássica

Dada a instabilidade da economia cafeeira ao longo dos anos, foi analisado as variações ocorridas no preço do café a partir do Modelo Autorregressivo-Médias Móveis, ARMA, de Box-Jenkins para Séries Estacionárias, utilizando uma série que abrange as cotações diárias dos preços do café desde o dia 02/09/1996 até 15/01/2013, correspondendo a um total de 4074 observações. Figura 1. O preço mínimo de cotação observado foi de R\$101,48, em dezembro do ano de 2001, e o máximo, em maio de 2011, R\$555,19.

Observando o gráfico da série original, Figura 5, verifica-se a existência de períodos de pouca oscilação e períodos de grande oscilação dos preços, o que caracteriza a volatilidade da série. Os períodos de pouca oscilação são indícios de mercado estável, assim como os de grande oscilação caracterizam períodos de instabilidade de preços. Os grupos de comportamento, mais instáveis ou menos instáveis, por si só já são indícios de que a série é correlacionada.

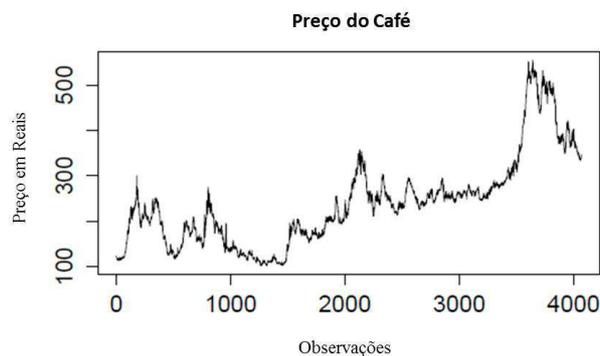


Figura 5 Preço do café 02/09/1996 até 15/01/2013

Na análise de séries temporais a estacionaridade é uma das pressuposições mais importantes. Tradicionalmente atinge-se a estacionaridade fazendo o uso das variações diárias da série, chamadas de retornos. Os retornos representam o quanto foi a variação no fechamento do dia, e é dado por: $R_t = \log x_t - \log x_{t-1}$. Podemos analisar claramente as variações dia após dia a partir do gráfico para a série de retornos na Figura 6:

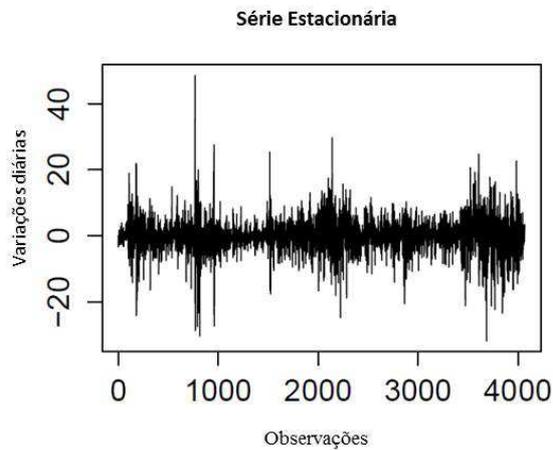


Figura 6 Série estacionária. Retornos do preço do café, variações diárias

As variações de preço positivas podem ocorrer devido à perspectivas de queda de safra, eventos climáticos que prejudiquem a lavoura, entre outros. De outro modo, as variações de preço negativas podem ser ligadas a incertezas do mercado, como queda no consumo ou aumento dos estoques de café.

O intervalo correspondente ao período de 1996 a 2000 (1ª a 1000ª observação) é caracterizado por grandes oscilações, ocorridas

significativamente nos anos de 1997, 1999 e 2000. A alta volatilidade observada ao longo do ano de 1997 pode ser explicada pelo baixo estoque aliado às estimativas de safra para a época. Nos anos de 1999 e 2000 fatores climáticos afetaram a produção, geadas ocorridas em regiões produtoras caracterizaram o período de grandes variações.

O período correspondente ao início de 2003 até primeiro semestre de 2005 (aproximadamente 1500^a a 2100^a observação) é caracterizado por maior estabilidade, possivelmente ocasionado por um equilíbrio entre consumo e produção, além da boa posição do Brasil entre os países produtores e exportadores de café na época.

Como podemos observar através do gráfico, há períodos de baixa volatilidade e períodos de alta instabilidade dos preços. As variações negativas podem ser explicadas pelas altas safras, assim como as variações positivas podem ser explicadas pela recuperação de preços no mercado internacional.

Como pode ser visto na Figura 7, a variância da série logarítmica de retornos foi controlada. Este fato é consagrado pela literatura, já que o *log* transforma o efeito multiplicativo da variância em efeitos aditivos.

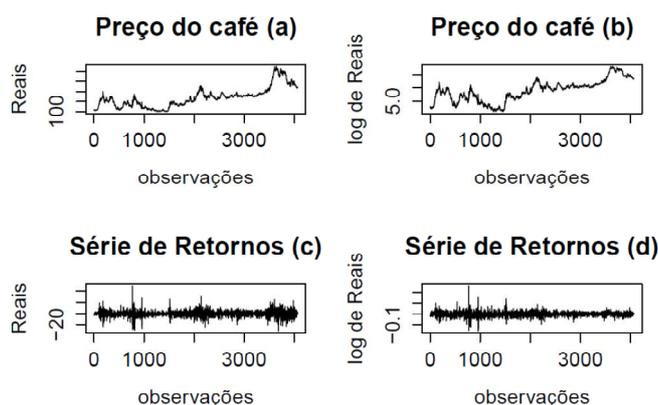


Figura 7 (a) Série Original de preços do café em reais; (b) Série de Log do preço; (c) Série de Retornos, diferença entre a observação no tempo t e no tempo $t-1$; (d): Série de Retornos na escala logarítmica.

De posse de uma série estacionária e transformada como mostrado na Figura.7 (d) ajustou-se o modelo ARMA. Para esses procedimentos foi utilizado o programa R com o pacote *tseries*.

Dentre os modelos ARMA ajustados, o que apresentou menor AIC (-19902,75) foi o ARMA (2,2), cujas estimativas são apresentadas na Tabela 1. Modelos com mais parâmetros foram estimados, com o objetivo de se obter um resíduo independente. Mas nestes casos o incremento de mais parâmetros provocou um incremento da função penalizadora que compõe o Critério de Akaike.

Tabela 1 Estimativas do modelo ARMA(2,2) para a série de retornos do preço do café.

Parâmetro	Estimativas	Erro-Padrão
AR1	-0,8818	0,1099
AR2	-0,6784	0,1348
MA1	0,8502	0,1182
MA2	0,6120	0,1457

O intercepto do modelo não apresentou significância, portanto, foi retirado do modelo. As estimativas acima deverão compor o modelo ARMA cuja forma geral é:

$$\hat{Z}_t = AR_1 \hat{Z}_{t-1} + AR_2 \hat{Z}_{t-2} + a_t - MA_1 a_{t-1} - MA_2 a_{t-2}.$$

Assim, o modelo estimado é:

$$\hat{Z}_t = -0,8818 \hat{Z}_{t-1} - 0,6784 \hat{Z}_{t-2} + a_t - 0,8502 a_{t-1} - 0,6120 a_{t-2}.$$

Os resíduos do ARMA (2,2) foram considerados não correlacionados, fato que pode ser visto na função de autocorrelação mostrada na Figura 8:

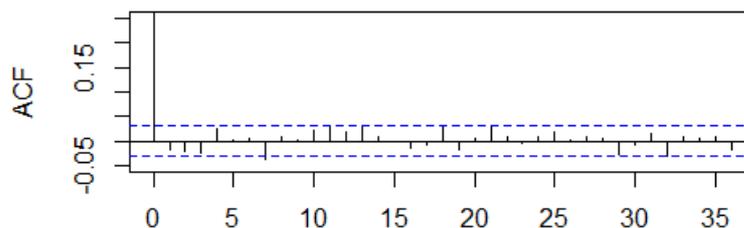


Figura 8 Função de autocorrelação para o resíduo do modelo ARMA(2,2)

Segundo Morettin e Tolo (2004), em Séries de Preços, caso o quadrado dos resíduos seja correlacionado, deve-se ajustar um modelo da ARCH para os resíduos dos modelos ARMA.

Os modelos GARCH(p,q) devem ser ajustados em três passos:

1. Estimação de um bom modelo ARMA

$$\hat{Z}_t = AR_1 \hat{Z}_{t-1} + AR_2 \hat{Z}_{t-2} + a_t - MA_1 a_{t-1} - MA_2 a_{t-2}.$$

2. Cálculo da correlação dos resíduos ao quadrado;

$$\hat{\rho} = \frac{\sum_{t=1}^T (\hat{\epsilon}_t^2 - \hat{\sigma}_t^2)(\hat{\epsilon}_{t-1}^2 - \hat{\sigma}_{t-1}^2)}{\sum_{t=1}^T (\hat{\epsilon}_t^2 - \hat{\sigma}_t^2)^2}.$$

3. Os resíduos ao quadrado devem apresentar normalidade condicional e serem independentes. Estas pressuposições são verificadas pelo teste de Ljung-Box e pelo teste de Jarque-Bera respectivamente.

Tal procedimento foi realizado, assumindo distribuição t de Student para os resíduos, ajustando-se um ARMA(2,2) (cujas estimativas são apresentadas na Tabela 1) e para o seu resíduo de um modelo Garch(1,1), cujas estimativas são apresentadas, Tabela 2:

Tabela 2 Estimativas para os parâmetros do GARCH(p,q)

Parâmetro	Estimativa	Erro-Padrão	T	Probabilidade ($> t $)
α_0	$1,23 \cdot 10^{-5}$	$1,5 \cdot 10^{-6}$	8,22	$2,22 \cdot 10^{-16***}$
α_1	$1,08 \cdot 10^{-1}$	$4,52 \cdot 10^{-3}$	23,94	$< 2 \cdot 10^{-16***}$
β_1	$8,66 \cdot 10^{-1}$	$7,28 \cdot 10^{-3}$	118,9	$< 2 \cdot 10^{-16***}$

Assim o modelo GARCH(1,1) para o resíduo do ARMA(2,2) fica:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

$$\sigma_t^2 = 1,23 \cdot 10^{-5} + 1,08 \cdot 10^{-1} a_{t-1}^2 + 8,66 \cdot 10^{-1} \sigma_{t-1}^2.$$

O valor da função de verossimilhança para este modelo foi de 5275.879, foram necessárias 23 iterações e a convergência foi apenas relativa. Os critérios de informação foram os seguintes: AIC=-5.1372, BIC=-5.1325, SIC=-5.1372, HQC=5.1335. Na análise de resíduo do modelo GARCH, obteve-se pelo teste de Jarque-Bera, que os resíduos satisfizeram a condição de normalidade condicional. E pelo teste de Ljung -Box, que os resíduos ao quadrado são não correlacionados. Desta forma, completa-se a análise da série temporal dos preços do café. Este algoritmo de estimação é consagrado pela literatura, Tsay (2002). E aparentemente não fere nenhum grande pressuposto, naturalmente esta análise seria corroborada pelos valores preditos para os retornos. Entretanto na busca constante por melhorias, não deixa de inquietar uma peculiaridade do resíduo em questão, fato adiante discutido.

Na Figura 9 são mostrados os resíduos do primeiro modelo estimado ARMA(2,2) e os resíduos do segundo modelo ajustado ARMA(2,2)-GARCH(1,1) com visível melhoria no ajuste. Exceto pelos extremos quantis, a modelagem da heterocedasticidade condicional (Modelo GARCH), incrementa em muito a qualidade do ajuste. Todavia elementos caudais permanecem fora do ajuste. Estes valores extremos comumente são ignorados dada a sua natureza de pouca frequência, entretanto em se tratando de séries financeiras, estes valores devem ter especial atenção, por que estão relacionados com grandes perdas ou grandes ganhos. Agrave-se a isto o fato de serem bem mais frequentes do que seria em dados biológicos ou agrônômicos, configurando o que a literatura chama de caudas pesadas, característica comum a dados financeiros. Destarte uma metodologia que bem modele estas caudas pesadas é bem vinda no trato de dados desta natureza.

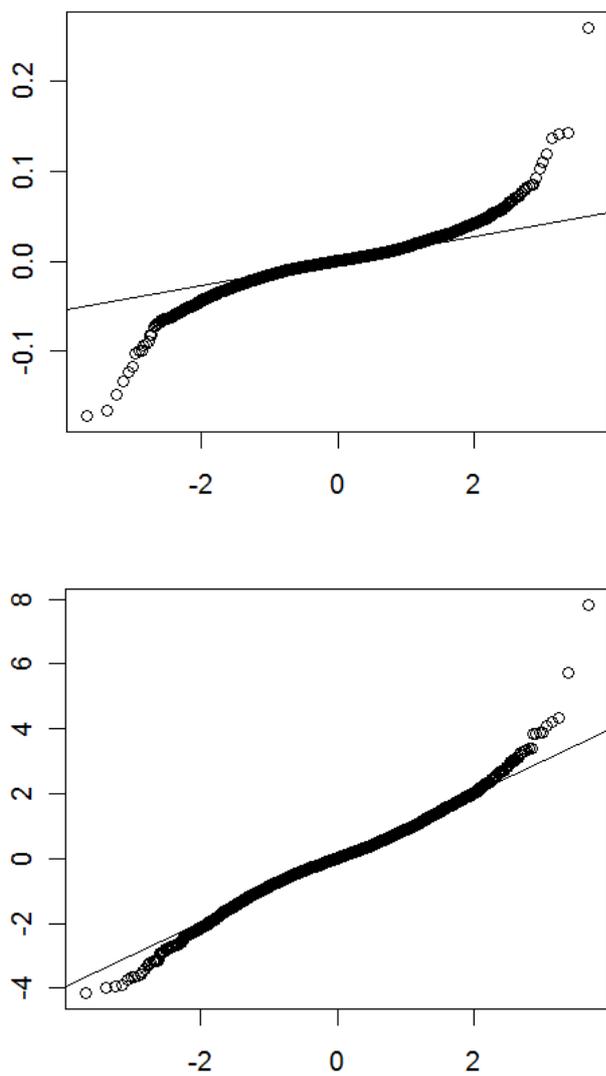


Figura 9 (Acima) – Resíduo para o modelo ARMA(2,2); (Abaixo) – Resíduo para o modelo ARMA(2,2)-GARCH(1,1)

Embora graficamente os resíduos tenham uma densidade aparentemente normal, Figura 10, o teste de Shapiro-Wilks para a normalidade, refuta esta hipótese a estatística do teste foi $W = 0,9686$, valor de probabilidade $< 2,2e-16$. O teste de Kolmogorov-Smirnov usado para medir a aderência dos resíduos (que são apresentados na Figura 9) para com uma distribuição normal simulada, com os parâmetros média zero e a variância igual à

variância dos resíduos (Figura 10). Este teste também sugere a não aderência, uma vez que a estatística foi $D = 0,4846$, valor de probabilidade $< 2,2e-16$. Sendo assim dois importantes testes estão em consonância em negar a normalidade deste resíduo. Lembrando que a normalidade é pressuposição para o modelo em questão.

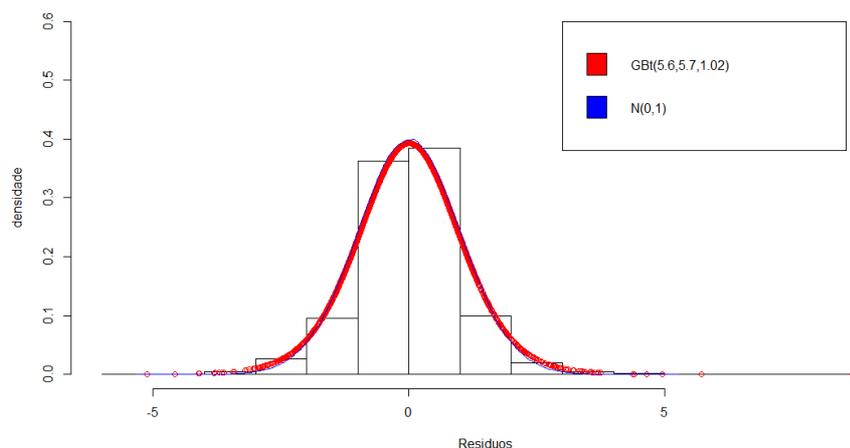


Figura 10 Histograma dos resíduos Densidade da normal e densidade da GBt

Apenas por curiosidade ajustou-se a distribuição GBt aos resíduos, utilizando o método da máxima verossimilhança e o Software R. Os parâmetros foram: $a=5,799$ (0,850); $b=5,867$ (0,0704); e $c=1,01$ (0,038). Os valores entre parênteses representam o erro padrão de cada estimativa. Todos os parâmetros podem ser considerados diferentes de zero uma vez que possuem erro padrão pequeno em relação à magnitude da respectiva estimativa. A função de log verossimilhança avaliada nas estimativas apresentou valor igual a $-5687,204$; AIC igual a $11380,40$ e BIC igual a $11399,34$. A matriz de correlação para os parâmetros é apresenta na Tabela 3.

Tabela 3 Matriz de correlação dos parâmetros ajustados da GBt

	a	b	c
a	1,0000000	0,9467583	-0,6642245
b	0,9467583	1,0000000	-0,3948495
c	-0,6642245	-0,3948495	1,0000000

A assimetria de Bowley (B) e a curtose de Moore (M) foi calculada para o resíduo em questão, obtendo-se $B=0,0000$ e $M=-0,0054$. Indicando uma assimetria igual a zero, característica idêntica a da distribuição normal e uma curtose menor que a normal, indicando um comportamento levemente platicúrtico em relação à distribuição normal.

Desta forma temos uma aparente incongruência deste resíduo em relação à pressuposição de normalidade. Um novo modelo foi ajustado objetivando uma melhoria na qualidade do ajuste bem como a captação deste comportamento da falta de ajuste sistemática ao longo das caudas das distribuições.

4.2 Modelo ARIMA-GARCH para uma série temporal do preço de café. Sob a pressuposição de distribuição “*t* de Student” para o resíduo.

O primeiro passo para a estimação do modelo sob a pressuposição “*t* de Student” foi a construção da função da função de verossimilhança, tendo a distribuição “*t* de Student” como base. Rotinas foram feitas utilizando o Software R.

```
# escrevendo a função de verossimilhança
llht=function(par){
  v <- par[1]; omega <- par[2]; alpha <- par[3]; beta <- par[4]
  z <- aii^2 ; Mean = mean(z^2)
  e <- alpha0 + alpha1*c(Mean, z[-length(aii)]^2)
  h <- filter(e, beta,"recursive", init = Mean)
  ei <- z/h
  sum(log(0.5)+log(v)-1.5*log(v+ei^2)-0.5*log(h),2,length(z))
}
```

Vale lembrar que a distribuição apresentada aqui não configura a distribuição *t* de Student tal qual conhecemos, razão pela qual foi escrita entre aspas, outrossim, a função base aqui apresentada é a que Alexander et al. (2012) chama de *t* de Student. A variância foi calculada recursivamente utilizando a função *filter()* e está denotada por *h*.

Considerando um modelo GARCH(1,1):

$$a_t = \sigma_t \varepsilon_t \text{ e } \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Pressupondo uma distribuição *t* de Student para o resíduo. A verossimilhança é expressa por:

$$l(a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1; \varepsilon_t) = \sum_{t=2}^T \left(\log \left(\frac{\lambda c}{2^{ac} B(a, b)} \frac{1}{(\lambda + \varepsilon_t^2)^{3/2}} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^{ac-1} \left[1 - \frac{1}{2^c} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^c \right]^{b-1} \right) \right) - \sum_{t=2}^T (0.5 \log(\sigma_t^2)),$$

com $\varepsilon_t = \sigma_t / a_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ e com $a = b = c = 1$.

Vale ressaltar que com $c=1$ nenhuma assimetria extra é incorporada e com $a=b=1$ a distribuição Beta (a,b) reduz-se a uma distribuição uniforme contínua, não imprimindo nenhum tipo de ponderação na distribuição núcleo. Desta forma a função de verossimilhança se resume a verossimilhança da distribuição t de Student:

$$l(1,1,1,\lambda,\alpha_0,\alpha_1,\beta_1;\varepsilon_t) = \sum_{t=2}^T \left(\log \left(\frac{\lambda}{(\lambda + \varepsilon_t^2)^{3/2}} \right) - 0.5 \log(\sigma_t^2) \right),$$

com $\varepsilon_t = \sigma_t / a_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

Para otimização desta função fez-se uso do algoritmo de otimização Nelder-Mead que no R pode ser implementado pela função *optim()*. Entretanto este mesmo algoritmo está implementado em diversos softwares estatísticos. Neste passo é necessário o arbitramento de chutes iniciais para início do processo de estimação. Também são calculadas numericamente a matriz hessiana e a matriz gradiente, com o objetivo de se obter uma estimativa da matriz de covariâncias. A raiz quadrada da diagonal da matriz de covariâncias é tomada como estimativa do erro padrão. Um teste t é realizado para a estimativa de cada parâmetro. Estes cálculos foram computados utilizando o Software R e os resultados são sumarizados na Tabela 2.4

```
# ajustando os parâmetros
fit = optim(start = c(v=1, omega = var(aii), alpha = 0.108, beta =
0.8662), objective = llht, control = list(omega=5))
# calculando a variância
hes = hessian(llht,c(fit$par[1],fit$par[2],fit$par[3],fit$par[4]))
g = grad(llht,c(fit$par[1],fit$par[2],fit$par[3],fit$par[4]))
rob = ginv(hes)%*%g%*%t(g)%*%ginv(hes)
R = sqrt(diag(rob))
# realizando um teste t
Estimativas = fit$par; t.cal=fit$par/R;Prob=pt(t.cal,4060, lower.tail=F)
```

data.frame(estimativas,t.cal,Prob)

Também foram calculados os critérios para seleção de modelos: foram calculados o Critério da Informação de Akaike (AIC), o Critério da Informação Bayesiano (BIC), o Critério da Informação de Schwarz e o Critério de Hannan Quinn (HQC) que também são apresentados na Tabela 4:

Tabela 4 Estimativas, erro padrão, valor de t e valor de probabilidade para modelo GARCH(1,1) sob a pressuposição de Distribuição “t de Student” para o resíduo

Parâmetro	Estimativas (Erro Padrão)	t calculado	Valor de Prob.
λ	1,306 (0,022)	45,1047	< 0,0000
α_0	0,1994 (0,004)	40,9029	< 0,0000
α_1	0,0062 (4,670)	0,0013	0,4995
β_1	0,7205 (0,280)	2,5741	0,0050
AIC = -0,9209 BIC = -0,9147 SIC = -0,9209 HQIC = -0,9187			

O processo de otimização deste tipo de função se mostrou desafiador. Inúmeros obstáculos tiveram de ser superados, dentre eles a escolha do método de otimização. Atualmente há na literatura diversos métodos disponíveis e diversos algoritmos implementados, cada um trazendo consigo suas especificidades, que podem ou não contribuir na resolução do problema em questão.

Algoritmos que possibilitam contração do espaço paramétrico, a princípio, figuram como os mais promissores, mas com algum tempo passa-se a perceber que as estimativas obtidas quase sempre esbarram nos limites impostos, levando a crer numa possível irreabilidade das estimativas ou mesmo numa possível ineficácia do método. Dentre os métodos que possibilitam a contração do espaço paramétrico o L-BFGS-B ou o BFGS figuram entre os mais utilizados.

Os métodos que não possibilitam a contração do espaço paramétrico costumam apresentar estimativas que não tem interpretação física, ou sem

sentido prático, ou seja, fora do espaço paramétrico. Já dentre os métodos que não possibilitam a constrição do espaço paramétrico o método SANN ou o de Nelder-Mead são os mais utilizados, este último é talvez o mais tradicional de todos, sendo utilizado como padrão para diversas funções de otimização como a função *optim()* por exemplo, embora não se possa restringir o espaço paramétrico, pode-se utilizar parâmetros de controle que controlam o ritmo de caminhar do método através do espaço paramétrico. Ficam imunes ao viés, as vezes impresso pelo pesquisador, ao impor uma restrição paramétrica, apresentam boa taxa de convergência e são consistentes no valor das estimativas, mesmo com chutes iniciais distintos, fato que indica consistência do método.

Os limites computacionais de computação também representam um desafio. A literatura já fala, e este fato é facilmente comprovado na prática, um limite computacional de $a=173$ para função Gama(a) e um limite computacional de $a=b=286$ para a função Beta(a,b). Esta facilidade analítica de cálculo da função Beta em relação a função Gama, é muitas vezes utilizada em favor dos processos de otimização. A distribuição t de Student como é conhecida classicamente, tem sua constante de normalização representada por uma função Gama que entretanto pode ser reescrita em função de uma função Beta.

De qualquer maneira, quando se ultrapassa estes limites, o processo é automaticamente interrompido, e uma mensagem de erro é apresentada. Volta-se a estaca zero, analisa-se possíveis falhas na construção do modelo, possíveis erros na escrita da função, e por fim questiona-se a adequabilidade do método escolhido.

Diversos métodos de otimização utilizam a matriz gradiente como ferramenta para o processo de otimização, uma vez que a matriz gradiente aponta para a direção de maior crescimento da função (princípio utilizado no método de Newton-Raphson). Estes métodos comumente extrapolam os valores de algumas estimativas impossibilitando a computação da função

objetiva, seja por incapacidade de cálculos, seja pela explosão ao infinito da função objetiva.

São muitos os fatores de que depende o sucesso do processo de estimação e muitas as nuances envolvidas em cada um dos métodos e algoritmos implementados cujo domínio da linguagem bem como das funcionalidades, por si, já é um grande desafio. Para o problema em questão a função *optim()* cujo método padrão de otimização é o Nelder-Mead, por fim, se mostrou a mais adequada e consistente, sendo então a escolhida para a aplicação proposta.

A computação numérica da função hessiana para modelos com muitos parâmetros também representou outro desafio. É extremamente comum em problemas desta natureza “o surgimento” de matrizes hessianas, cujas inversas, apresentam valores negativos na diagonal principal, o que é um contra senso uma vez que estes valores seriam estimativas de variâncias. A solução também veio da literatura, que sugere o uso de uma aproximação robusta para a computação da matriz hessiana, este estimador sanduíche para a variância é utilizado em todo o processo de estimação dos novos modelos.

4.3 Modelo ARIMA-GARCH – Aplicação para uma série temporal do preço de café. Sob a pressuposição de distribuição “Beta *t* de Student” para o resíduo.

O próximo passo foi a estimação de um segundo modelo, que incorpora 2 novos parâmetros, *a* e *b* correspondentes aos parâmetros da distribuição Beta.

Considerando um modelo GARCH(1,1):

$$a_t = \sigma_t \varepsilon_t \text{ e } \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Pressupondo uma distribuição Beta *t* Generalizada para o resíduo. A verossimilhança é expressa por:

$$l(a, b, c = 1, \lambda, \alpha_0, \alpha_1, \beta_1; \varepsilon_t) =$$

$$\sum_{t=2}^T \left(\log \left(\frac{\lambda}{2^a B(a, b)} \frac{1}{(\lambda + \varepsilon_t^2)^{3/2}} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^{a-1} \left[1 - \frac{1}{2} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right) \right]^{b-1} \right) \right)$$

$$- \sum_{t=2}^T (0.5 \log(\sigma_t^2)),$$

com $\varepsilon_t = \sigma_t / a_t$ e $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

De forma análoga ao procedimento do passo anterior a função de verossimilhança foi otimizada. Utilizando o mesmo algoritmo, através do software R para a computação, e passando pelas mesmas dificuldades já relatadas, embora agora avultadas devido a presença de um maior número de parâmetros, e devido à maior complexidade do espaço paramétrico, utilizando a seguinte rotina:

```
#escrevendo a função de verossimilhança
llhBETA=function(par){
  a <- par[1]; b <- par[2]; v <- par[3]; omega <- par[4]; alpha <- par[5];
  beta <- par[6]
  z <- aii^2 ; Mean = mean(z^2)
  e <- omega + alpha*c(Mean, z[-length(aii)]^2)
  h <- filter(e, beta,"recursive", init = Mean)
  ei <- z/h
  sum(log((v/((2^(a))*beta(a,b)*(v+ei^2)^1.5))*((1+ei/(sqrt(v+ei^2))))^(
    (a-1))*(1-(1/2)*((1+ei/sqrt(v+ei^2))))^(b-1))-
    0.5*log(h),2,length(z))
}
#estimando parâmetros
fit2= optim(c(a=.9,b=.9,v=1, omega = 0.1*var(aii), alpha = 0.13, beta =
0.86), llhBETA, control=list(maxit=1000,beta=1, alpha=0.5))
#calculando a variância
hes2 =
hessian(llhBETA,c(fit2$par[1],fit2$par[2],fit2$par[3],fit2$par[4],fit2$pa
r[5],fit2$par[6]))
g =
grad(llhBETA,c(fit$par[1],fit$par[2],fit$par[3],fit$par[4],fit$par[5],fit$
par[6]))
robusta = ginv(hes2)%*%g%*%t(g)%*%ginv(hes2)
R = sqrt(diag(robusta))
#fazendo o teste t
```

```

estimativas = fit$par; t.cal=fit$par/R; Prob=pt(t.cal,4060, lower.tail=F)
data.frame(estimativas,t.cal,Prob)
# Calculando os critérios de informação:
N = length(aii); NPAR = length(fit$par); fit$ics = c(
  AIC = (-2*fit$value)/N + 2 * NPAR/N,
  BIC = (-2*fit$value)/N + NPAR * log(N)/N,
  SIC = (-2*fit$value)/N + log((N+2*NPAR)/N),
  HQIC = (-2*fit$value)/N + (2*NPAR*log(log(N)))/N ); fit$ics

```

as estimativas para o modelo “Beta *t*” são apresentadas na Tabela 5 e como se pode ver, os critérios de informação (AIC, BIC, SIC, HQIC) foram bem menores que no modelo *t* cujas estimativas de parâmetros e demais estatísticas estão dispostas na Tabela 4.

Tabela 5 Estimativas, erro padrão, valor de *t* e valor de probabilidade para modelo GARCH(1,1) sob a pressuposição de Distribuição “Beta *t* de Student” para o resíduo

Parâmetro	Estimativas (Erro Padrão)	<i>t</i> calculado	Valor de Prob.
<i>a</i>	0,8930 (0,7644)	1,1681	0,1241
<i>b</i>	0,8850 (0,7616)	1,1620	0,1263
λ	1,0052 (1,0052)	0,9999	0, 5869
α_0	0,1179 (0,1179)	0,9999	0,5871
α_1	0.0210 (0,0031)	6,7006	< 0.0000
β_1	0.6880 (0,3123)	2,2027	< 0.0000
AIC = -1,43 BIC = -1,42 SIC = -1,43 HQIC = -1,43			

O valor da verossimilhança foi 844.0789, num processo com 261 iterações até atingir a convergência. Mais uma vez o grande desafio é os chutes iniciais. A busca por valores iniciais adequados não é tarefa simples em uma estimação simultânea de número tão grande parâmetros. Este fato torna o processo exaustivamente trabalhoso e demorado.

Cada processo de estimação também gasta um tempo considerável quando comparado com as funções já existentes em pacotes para análise de séries temporais do R. Este fato nos faz pensar que não estamos fazendo da melhor forma o processo de estimação, entretanto o que ocorre é outra coisa.

Em pacotes já compilados, como por exemplo fGarch, existem funções que permitem ajustar modelos GARCH para dados com diversos níveis de assimetria e curtose, utilizando diversas distribuições como por exemplo a distribuição de Laplace, Cauchy, t-Assimétrica entre outras. Estas distribuições têm parte de seus parâmetros estimados diretamente a partir dos dados. Uma vez calculadas estas estatísticas são lançadas na função de verossimilhança e a partir de então é que começa o processo de estimação propriamente dito.

Desta forma o processo nestes pacotes se dá em dois passos. Primeiro busca-se encontrar uma distribuição mais adequada para os dados, cuja função é parcialmente escrita em função de estimativas já calculadas e parcialmente escrita em função de parâmetros, estes sim a serem estimados.

4.4 Modelo ARIMA-GARCH - Aplicação para uma série temporal do preço de café. Abordagem GBt

O último modelo estimado foi segundo o modelo completo incorporando além dos dois parâmetros anteriores (a e b) um terceiro parâmetro c correspondente ao parâmetro que deu gênese à distribuição Generalizada Beta t a partir da Beta t. Este novo como visto é um parâmetro de curva responsável por imprimir assimetria. Então considerando um modelo GARCH(1,1): $a_t = \sigma_t \varepsilon_t$ e $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, pressupondo uma distribuição GBt para o resíduo, a verossimilhança é expressa por:

$$l(a, b, c, \lambda, \alpha_0, \alpha_1, \beta_1; \varepsilon_t) = \sum_{t=2}^T \left(\log \left(\frac{\lambda c}{2^{ca} B(a, b)} \frac{1}{(\lambda + \varepsilon_t^2)^{3/2}} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^{ac-1} \left[1 - \frac{1}{2} \left(1 + \frac{\varepsilon_t}{\sqrt{\lambda + \varepsilon_t^2}} \right)^c \right]^{b-1} \right) \right) - (T-2)0.5 \log(\sigma_t^2),$$

com $\varepsilon_t = a_t / \sigma_t$ e $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

Esta função de verossimilhança reescrita no código abaixo, seguida da rotina de otimização do modelo:

```
# escrevendo a função de verossimilhança
llhBETA=function(par){
  a <- par[1]; b <- par[2]; c <- par[3]; v <- par[4]; omega <- par[5]; alpha
<- par[6]; beta <- par[7]
  z <- aii^2 ; Mean = mean(z^2)
  e <- omega + alpha*c(Mean, z[-length(aii)]^2)
  h <- filter(e, beta,"recursive", init = Mean)
  ei <- z/h

  sum(log((v*c/((2^(a*c))*beta(a,b)*(v+ei^2)^1.5))*((1+ei/(sqrt(v+ei^2)))^
(a*c-1))*(1-((1/2)^c)*((1+ei/sqrt(v+ei^2)))^c)^(b-1))-
log(h), 2,length(z))
}

# estimando parâmetros
fit= optim(c(a=.9,b=.9,c=1,v=1, omega = 0.1*var(aii), alpha = 0.13, beta
= 0.86), llhBETA, control=list(maxit=1000,beta=1, alpha=0.5))

#calculando a variância
hes =
hessian(llhBETA,c(fit$par[1],fit2$par[2],fit2$par[3],fit2$par[4],fit2$pa
r[5],fit2$par[6],fit2$par[7]))
g =
grad(llhBETA,c(fit$par[1],fit$par[2],fit$par[3],fit$par[4],fit$par[5],fit$
par[6],fit2$par[7])
robusta = ginv(hes2)%*%g%*%t(g)%*%ginv(hes2)
R = sqrt(diag(robusta))
#fazendo o teste t
estimativas = fit$par; t.cal=fit$par/R; Prob=pt(t.cal,4060, lower.tail=F)
data.frame(estimativas,t.cal,Prob)

# Calculando os critérios de informação:
N = length(aii); NPAR = length(fit$par); fit$ics = c(
  AIC = (-2*fit2$value)/N + 2 * NPAR/N,
  BIC = (-2*fit2$value)/N + NPAR * log(N)/N,
  SIC = (-2*fit2$value)/N + log((N+2*NPAR)/N),
  HQIC = (-2*fit2$value)/N + (2*NPAR*log(log(N)))/N ); fit$ics
```

Os resultados do processo de estimação deste novo modelo são apresentados na Tabela 6:

Tabela 6 Estimativas, erro padrão, valor de t e valor de probabilidade para modelo GARCH(1,1) sob a pressuposição de Distribuição “Generalizada Beta t de Student” para o resíduo.

Parâmetro	Estimativas (Erro Padrão) Prob.	t calculado	V. de Prob.
a	0,8841 (0,5046) 0,0356	1,8037	
b	0,8866 (0,5046) 0,1946	0,8608	
c	1,0045 (0,7654) 0,0947	1,3123	
λ	1,0036 (0,5046) 0,1587	0,9998	
α_0	0,1137 (0,1138) 0,1587	0,9998	
α_1	0,0975 (0,0530) 0,0329	1,8391	
β_1	0,6641 (0,3363) 0,0241	1,9745	
AIC=-1,5424		BIC=-1,5316	SIC=-1,5424 HQIC=-1,5386

O valor da verossimilhança foi 1795.75, num processo com 274 iterações até atingir a convergência. O resultado desta análise se mostrou surpreendente, redundando em valores ainda menores de todos os critérios de informação calculados. A matriz de hessiana para o modelo proposto pode ser obtida numericamente a partir da rotina acima descrita.

A diminuição constante dos critérios de informação ao longo dos 3 modelos, diminuição esta que se deu na medida em que se aumentou o número de parâmetros, demonstrou a consistência da idéia proposta, este fato pode ser visto na Tabela 7, na qual são mostrados os resultados dos testes da Razão de Verossimilhanças para os modelos aninhados.

Tabela 7 Teste de razão de Razão de Verossimilhanças para modelos GBt, Bt e t

Teste	Estatística	Valor de Probabilidade
GBt x Bt	443,3574	< 0,0001
GBt x t	2534,9366	< 0,0001

Bt x t 2091,5791 < 0,0001

Como pode ser visto pela Tabela 7, todos os modelos são significativamente diferentes entre si. Houve uma continuada diminuição do valor do critério de informação na medida em que foram sendo acrescentados novos parâmetros. Os modelos propostos parecem incrementar a qualidade do ajuste.

Na Tabela 8 são apresentados diferentes modelos e seus respectivos critérios de informação. São apresentados 3 modelos estimados sob a pressuposição de normalidade e os três modelos aqui propostos.

Tabela 8 Comparação de diferentes modelos baseando-se no critério da informação de Akaike

Modelo	Estimativa da Verossimilhança	Valor do AIC
Normal (time series)	-14190,88	6,97
Normal (fGarch)	-10454,77	5,14
Normal	-40432,29	5,13
<i>t</i>	1877,71	-0,91
Beta <i>t</i>	2923,50	-1,43
GB <i>t</i>	3145,18	-1,54

Três dos modelos apresentados foram estimados sob a pressuposição de normalidade, cada um seguindo uma metodologia de estimação. O primeiro (primeira linha da tabela) foi estimado utilizando o pacote timeSeries do software R, o segundo foi estimado utilizando o pacote fGarch do software R, o terceiro e último modelo estimado sob a pressuposição de normalidade foi estimado programando-se a própria função de verossimilhança, utilizando o mesmo procedimento utilizado para se estimar os modelos propostos *t*, Beta *t* e GB*t*. Como pode ser visto todos os modelos propostos foram melhores, consoante o critério de Akaike, que os classicamente utilizados sob a pressuposição de normalidade.

5 CONCLUSÕES

Os modelos propostos apresentaram uma melhoria na qualidade do ajuste. Os modelos propostos, que são eles modelo t de Student, modelo Beta t de Student, e modelo Generalizado Beta t de Student, propiciaram uma melhoria na qualidade do ajuste na medida em que se aumentou o número de parâmetros.

O modelo t de Student apresentou AIC igual a -0,91 ao passo que o modelo Beta t de Studente apresentou AIC igual a -1,43, o incremento de dois novos parâmetros referentes a distribuição Beta propiciou um incremento na melhoria do ajuste, fato que foi atestado pelo teste de Razão de Verossimilhanças, cujo valor de probabilidade foi menor que 0,00001

O modelo Generalizado Beta t (GBt), apresentou AIC igual a -1,54, valor inferior ao obtido no modelo t de Student e no modelo Beta t de Student. Esta diferença mais uma vez foi atestada pelo teste de razão de verossimilhanças, que apontou superioridade do modelo GBt em relação aos outros dois submodelos, e cujos valores de probabilidade para ambas as comparação foi menor que 0,00001.

Considerando como critério de comparação, apenas o valor do AIC, todos os modelo propostos apresentaram valores maiores (1877,71 considerando a distribuição t; 2923,50 considerando a distribuição Beta t; 3145,18 considerando a distribuição GBt) que o valor apresentado pelo modelo baseado na distribuição Normal (-40432,29). Neste caso não é possível construir um teste de razão de verossimilhanças uma vez que o modelo Normal não é aninhado a nenhum dos demais.

BIBLIOGRAFIA

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, *Notre Dame*, v. 19, n.6, p. 716-723, dec. 1974. Disponível em: <http://www.unt.edu/rss/class/Jon/MiscDocs/Akaike_1974.pdf>. Acesso em: 10 jan. 2013.
- AKINSETE, A.; FAMOYE, F.; LEE, C. The beta pareto distribution. **Statistics**, Huntington, v. 42, n. 6, p. 547-563, oct. 2008.
- ALEXANDER, C. et. al. Generalized beta-generated distributions. **Computational statistic & data analysis**, United Kingdom, v. 56, n. 6, p. 1880-1897, sep. 2012. Disponível em: <<http://dx.doi.org/10.1016/j.csda.2011.11.015>>. Acesso em: 10 de jan. de 2013.
- BARRETO, W. SOUZA; SANTOS, A. H. S.; CORDEIRO, G. M. The beta generalized exponential distribution. **Journal of Statistical Comutations and Simulation**, Recife, v. 80, p. 159-172, 10 sep. 2008. Disponível em: <<http://arxiv.org/pdf/0809.1889.pdf>>. Acesso em: 22 abr. 2011.
- BELISLE, C. J. P. Convergence theorems for a class of simulated annealing Algorithms. **J. Applied Probability**, Michigam, v. 29, n. 4, p. 885-895, dec. 1992. Disponível em: <<http://www.jstor.org/stable/3214721>>. Acesso em: 20 fev. 2014.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. **Journal of Econometrics**, San Diego, v. 31, p. 307-327, feb. 1986.
- BOYCE, W. **Equações diferenciais elementares e problemas de valores de contorno**. 9. ed. Rio de Janeiro: LTC, 2010.
- BUENO, R. L. S. **Econometria de Séries Temporais**. São Paulo: Cengage Learning, 2008.
- BYRD, R. H. et al. A limited memory algorithm for bound constrained optimization. **SIAMJ Scientific Computing**, Illinois, p. 1190-1208, may 1994.
- ENGLE, R. F. Autorregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, England, v. 50, n. 4, p. 987-1007, jul. 1982.

EUGENE, N.; LEEA, C.; FAMOYE, F. Beta Normal Distribution and its applications. **Communication in statistics Theory and Methods**, Hamilton, v. 31, n. 4, p. 497-512, sep. 2002.

FAMOYE, F.; LEEA, C.; EUGENE, N. Beta-Normal Distribution: Bimodality Properties and Application. **Journal of Modern Applied Statistical Methods**, Hamilton, v. 3, n. 1, p. 84-103, 2004.

FISHER, R. A. On a Distribution yielding the error functions of several well know statistics. **Proceedings of the International Mathematical Congress**, Toronto, p. 805-813, 1924.

FLETCHER, R.; REEVES, C. M. Function Minimization by Conjugate Gradients. **Computer Journal**, Oxford, v. 7, n. 2, p. 148-154, 1964.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of Integrals, series and products**. 7. ed. San Diego: Academic Press, 2007.

GUPTA, R. D.; KUNDU, D. Exponentiated exponential family: an alternative to gamma and Weibull. **Biometrical Journal**, New Jersey, n. 43, p. 117-130, 2001.

HUBER, P. J. The Behavior of Maximum Likelihood under Nonstandard Conditions. **Proceedings of the Fifth Berkeley Symposium on Mathematical and Probability**, Berkeley, v. 1, p. 805-813, 1967.

JONES, M. C. A Skew t distribution. In: CHARALAMBIDES, M. V.; KOUTRAS, M. V.; BALAKRISHNAN, N. (ed.) **Probability and Statistical Models with Applications**. London: Chapman and Hall, CRC, 2001. p. 269-277.

_____. Families of distributions arising from distributions of order statistics. **Test**, London, v. 13, n. 1, p. 1-43, jun. 2004.

JONES, M. C.; FADDY, M. J. A skew extension of t distribution, with applications. **Journal of the Royal Statistical Society**, New Jersey, v. 65, n. 1, p. 159-174, 2003.

JONES, M. C.; LARSEN, P. V. Larsen. Multivariate distributions with support above the diagonal. **Biometrika**, Oxford, v. 91, n. 4, p. 159-174, 2004.

KENNEY, J. F.; KEEPING, E. S. **Mathematics of Statistics**. New Jersey: Princenton, 1962.

- KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, New Orleans, v. 46, n. 1-2, p. 79-88, 1980.
- LEÃO, J. et al. On some properties of beta Inverse Rayleigh distribution. **Chilean Journal of Statistics**, Chile, v. 4, n. 2, p. 111-131, sept. 2013.
- MCDONALD, J. B. Some generalized functions for the size distribution of income. **Econometrica**, Cleveland, n. 52, p. 647-663, 1984.
- MOOD, A. M.; GRAYBILL, F. A; BOES, D. C. **Introduction to the Theory of Statistics**. 3. ed. Mc Graw Hill, 1974.
- MOORS, J. J. A. A quantile alternative for Kurtosis. **Journal of the Royal Statistical Society D, The Statistician**, Tilburg, n. 37, p. 25-32, 1998.
- MORETTIN, P. A; TOLOI, C. **Análise de Séries Temporais**. São Paulo: Edgar Bluncher, 2004.
- NADARAJAH, S.; GUPTA, A. K. The exponentiated gamma distribution with application to drought data. **Calcutta Statistical Association Bulletin**, Mumfordganj Allahabad, n. 59, p. 29-54, 2007.
- _____. The beta Fréchet distribution. **Far East Journal of Theoretical Statistics**, Allahabad, v.14, n. 1 p. 15-24, 2004.
- NADARAJAH, S.; KOTZ, S. Kotz. The Beta Gumbel Distribution. **Mathematical Problems in Engineering**, London, n. 4, p. 323-332, 2004.
- _____. The Beta Exponential Distribution. *Reliability. Engineering and System Safety*, New York, p. 689-697, 2006.
- NELDER, J. A; MEAD, R. A Simplex Algorithm for Function Minimization. **Computer Journal**, Oxford, v.7, n. 4, p. 308-313, 2009.
Disponível em:
<<http://www.ime.unicamp.br/~sandra/MT853/handouts/Ref3%28NelderMead1965%29.pdf>>. Acesso em: 22 dez. 2013.
- NOCEDAL, J., e S. J. Wright. **Numerical Optimization**. New York, Springer, 1999.
- ORTEGA, E.; CORDEIRO, G.; CARRASCO, J. The log-generalized modified Weibull regression model. **Brazilian Journal of Probability and Statistics**, São Paulo, v. 25, n.1, p. 1-29, 2011.

PASCOA, M. et al. The Kumaraswamy-Generalized Gamma Distributions with Applications in Survival Analysis. **Statistical Methodology**, New York, v. 8, n. 5, p. 411-433, 2010.

RAZZAGHI, M. Beta-normal distribution in dose-response modeling and risk assessment for quantitative responses. **Environmental and Ecological Science**, Edimburgo, v. 16, n. 1, p. 21-36, 2009.

SUCARRAT, G. Betategarch: simulation, estimation and forecasting of first-order beta-skew-t-EGARCH models. **The R Journal**, Yongstown, v.5, n. 2, p. 137-147, 2013. Disponível em: <<http://journal.r-project.org/archive/2013-2/sucarrat.pdf>>. Acesso em: 07 jan. 2014.

TSAY, R. S. **Analysis of financial time series**. New York: John Wiley & Sons, 2002.

WURTZ, D.; CHALABI, Y.; LUKSAN, L. Parameter Estimation of ARMA Models with GARCH/APARCH errors An R and SPlus software implementation. **Journal of Statistical Software**, Los Angeles, v. vv, n. 2, p. 1-41, 2008. Disponível em: <<http://faculty.washington.edu/ezivot/econ589/WurtzEtAlGarch.pdf>>. Acesso em 07 jan 2014.

ANEXOS

Anexo A

Rotinas em R para geração de números aleatórios da distribuição GBG, para cálculo dos coeficientes de assimetria e de curtose, e construção das Figuras 2, 3 e 4.

- a) Simulação de dados aleatórios e cálculo do coeficiente de curtose, com o parâmetro c variando e demais fixos.

```
#####  
c <- seq(0,40,0.1)  
medias1 <- as.vector( matrix(0,1,401))  
for(j in 1:401){  
  a=0.10; b=0.1; v=1;  
  bi=matrix(0,401,160)  
  for(i in 1:100){  
    bi[i,] = (sqrt(v)*(2*((rbeta(160,a,b))^(1/c[j]))-1)) /((1-  
(2*((rbeta(160,a,b))^(1/c[j]))-1)^2)^0.5)  
    bi[i,] = sort(bi[i,])  
  }  
  R <- bi  
  curtose=matrix(0,100,7)  
  for(i in 1:100) {  
    curtose[i,7]=(R[i,140]-R[i,100]-R[i,60]+R[i,20])/(R[i,120]-  
R[i,40])  
    curtose[i,6]=R[i,140]  
    curtose[i,5]=R[i,120]  
    curtose[i,4]=R[i,100]  
    curtose[i,3]=R[i,60]  
    curtose[i,2]=R[i,40]  
    curtose[i,1]=R[i,20]  
  }  
  medias1[j] <- mean(curtose[,7])  
}  
plot(medias1~c, type="l", ylab="curtose",ylim=c(-50,50), xlab="valores  
de C");  
text(25,30,expression (paste ("a=0.1, ", "b=0.1 e ", lambda,"=1")))  
#####
```

```

medias2 <-as.vector( matrix(0,1,401))
for(j in 1:401){
  a=1; b=1; v=1;
  bi=matrix(0,401,160)
  for(i in 1:100){
    bi[i,] = (sqrt(v)*(2*((rbeta(160,a,b))^(1/c[j]))-1)) /((1-
(2*((rbeta(160,a,b))^(1/c[j]))-1)^2)^0.5)
    bi[i,] = sort(bi[i,])
  }
  R <- bi
  curtose=matrix(0,100,7)
  for(i in 1:100) {
    curtose[i,7]=(R[i,140]-R[i,100]-R[i,60]+R[i,20])/(R[i,120]-
R[i,40])
    curtose[i,6]=R[i,140]
    curtose[i,5]=R[i,120]
    curtose[i,4]=R[i,100]
    curtose[i,3]=R[i,60]
    curtose[i,2]=R[i,40]
    curtose[i,1]=R[i,20]
  }
  medias2[j] <- mean(curtose[,7])
}

lines(medias2~c, type="l", ylab="curtose")
text(30,0,expression (paste ("a=1, ", "b=1 e ", lambda,"=1")))
#####
medias3 <-as.vector( matrix(0,1,401))
for(j in 1:401){
  a=0.2; b=0.2; v=1;
  bi=matrix(0,401,160)
  for(i in 1:100){
    bi[i,] = (sqrt(v)*(2*((rbeta(160,a,b))^(1/c[j]))-1)) /((1-
(2*((rbeta(160,a,b))^(1/c[j]))-1)^2)^0.5)
    bi[i,] = sort(bi[i,])
  }
  R <- bi
  curtose=matrix(0,100,7)
  for(i in 1:100) {

```

```

curtose[i,7]=(R[i,140]-R[i,100]-R[i,60]+R[i,20])/(R[i,120]-
R[i,40])
curtose[i,6]=R[i,140]
curtose[i,5]=R[i,120]
curtose[i,4]=R[i,100]
curtose[i,3]=R[i,60]
curtose[i,2]=R[i,40]
curtose[i,1]=R[i,20]
}
medias3[j] <- mean(curtose[,7])
}
lines(medias3~c, type="l", ylab="curtose")
text(15,10,expression (paste ("a=0.2, ", "0.2 e ", lambda,"=1")))
abline(v=1, col= "gray")

```

- b) Simulação de dados aleatórios e cálculo do coeficiente de assimetria, com o parâmetro c variando e demais fixos.

```

#####
c1 <- seq(0,20,0.1)
medias1 <-as.vector( matrix(0,1,201))
for(j in 1:201){
  a=0.10; b=0.10; v=1;
  bi=matrix(0,100,100)
  for(i in 1:100){
    bi[i,] = (sqrt(v)*(2*((rbeta(100,a,b))^(1/c[j]))-1)) /((1-
(2*((rbeta(100,a,b))^(1/c[j]))-1)^2)^0.5)
    bi[i,] = sort(bi[i,])
  }
  R <- bi
  assimetria=matrix(0,100,4)
  for(i in 1:100) {
    assimetria[i,4]=(R[i,75]-2*R[i,50]+R[i,25])/(R[i,75]-
R[i,25])
    assimetria[i,3]=R[i,75]
    assimetria[i,2]=R[i,50]
    assimetria[i,1]=R[i,25]
  }
  medias1[j] <- mean(assimetria[,4])
}

```

```

}
plot(medias1~c1, type="l", ylab="assimetria");
text(15,0.5,expression (paste ("a=0.1, ", "b=0.1 e ", lambda,"=1")))
#####
for(j in 1:201){
  a=1; b=1; v=1;
  bi=matrix(0,100,100)
  for(i in 1:100){
    bi[i,] = (sqrt(v)*(2*((rbeta(100,a,b))^(1/c[j]))-1)) /((1-
(2*((rbeta(100,a,b))^(1/c[j]))-1)^2)^0.5)
    bi[i,] = sort(bi[i,])
  }
  R <- bi
  assimetria=matrix(0,100,4)
  for(i in 1:100) {
    assimetria[i,4]=(R[i,75]-2*R[i,50]+R[i,25])/(R[i,75]-
R[i,25])
    assimetria[i,3]=R[i,75]
    assimetria[i,2]=R[i,50]
    assimetria[i,1]=R[i,25]
  }
  medias[j] <- mean(assimetria[,4])
}
lines(medias~c, type="l", ylab="assimetria")
text(12,0.2,expression (paste ("a=1, ", "b=1 e ", lambda,"=1")))
#####3
c <- seq(0,20,0.1)
medias2 <-as.vector( matrix(0,1,201))
for(j in 1:201){
  a=280; b=280; v=1;
  bi=matrix(0,100,100)
  for(i in 1:100){
    bi[i,] = (sqrt(v)*(2*((rbeta(100,a,b))^(1/c[j]))-1)) /((1-
(2*((rbeta(100,a,b))^(1/c[j]))-1)^2)^0.5)
    bi[i,] = sort(bi[i,])
  }
  R <- bi
  assimetria=matrix(0,100,4)
  for(i in 1:100) {

```

```

        assimetria[i,4]=(R[i,75]-2*R[i,50]+R[i,25])/(R[i,75]-
R[i,25])
        assimetria[i,3]=R[i,75]
        assimetria[i,2]=R[i,50]
        assimetria[i,1]=R[i,25]
    }
    medias2[j] <- mean(assimetria[,4])
}
lines(medias2~c, type="l", ylab="assimetria");
text(12,-0.12,expression (paste ("a=280, ", "b=280 e ", lambda,"=1")))
#####
plot(medias1~c1, type="l", ylab="assimetria", xlab="valores de c",
cex.lab=0.8);
text(15,0.5,expression (paste ("a=0.1, ", "b=0.1 e ", lambda,"=1")))
lines(medias~c, type="l", ylab="assimetria")
text(12,0.2,expression (paste ("a=1, ", "b=1 e ", lambda,"=1")))
lines(medias2~c, type="l", ylab="assimetria");
text(12,-0.12,expression (paste ("a=280, ", "b=280 e ", lambda,"=1")))

```

Anexo B

Função de verossimilhança GBt

$$l(a, b, c, \lambda; X) = n \ln \left(\frac{1}{2 \cdot a^c \beta(a, b) (x_i^2 + \lambda)^{3/2}} \left(\lambda^c \left(1 + \frac{x_i}{\sqrt{x_i^2 + \lambda}} \right)^{ac-1} \left(1 - 1.2 \cdot^{-1.c} \left(1 + \frac{x_i}{\sqrt{x_i^2 + \lambda}} \right)^c \right)^b \right) \right)$$

Funções escores para estimação de parâmetros da distribuição GBt

$$\frac{d}{d a} l = -0.6931471806 n c - \frac{1. n \left(\frac{\partial}{\partial a} \beta(a, b) \right)}{\beta(a, b)} + \sum_{i=1}^n c \ln \left(1 + \frac{x_i}{\sqrt{x_i^2 + \lambda}} \right)$$

$$\frac{d}{d b} l = \frac{1}{\beta(a, b)} \left(n \left(\beta(a, b) \ln \left(1 + \frac{x_i}{\sqrt{x_i^2 + \lambda}} \right) - e^{-0.6931471806 c} \left(\frac{\sqrt{x_i^2 + \lambda} + x_i}{\sqrt{x_i^2 + \lambda}} \right)^c \right) - \left(\frac{\partial}{\partial b} \beta(a, b) \right) \right)$$

