

Presidente da República
Luiz Inácio Lula da Silva

Ministro do Planejamento, Orçamento e Gestão
Paulo Bernardo Silva

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Eduardo Pereira Nunes

Diretor-Executivo
Sérgio da Costa Côrtes

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Wasmália Socorro Barata Bivar

Diretoria de Geociências
Luiz Paulo Souto Fortes

Diretoria de Informática
Luiz Fernando Pinto Mariano

Centro de Documentação e Disseminação de Informações
David Wu Tai

Escola Nacional de Ciências Estatísticas
Sérgio da Costa Côrtes (interino)

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística - IBGE

REVISTA BRASILEIRA DE ESTATÍSTICA

volume 69 número 230 janeiro/junho 2008

ISSN 0034-7175

R. bras. Estat., Rio de Janeiro, v. 69, n. 230, p. 1-90, jan./jun. 2008

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro - RJ - Brasil

© IBGE. 2008

Revista Brasileira de Estatística, ISSN 0034-7175

Órgão oficial do IBGE e da Associação Brasileira de Estatística - ABE. Publicação semestral que se destina a promover e ampliar o uso de métodos estatísticos (quantitativos) na área das ciências econômicas e sociais, através de divulgação de artigos inéditos.

Temas abordando aspectos do desenvolvimento metodológico serão aceitos, desde que relevantes para os órgãos produtores de estatísticas.

Os originais para publicação deverão ser submetidos em três vias (que não serão devolvidas) para:

Francisco Louzada-Neto

Editor responsável – RBEs – IBGE.

Rua André Cavalcanti, 106 – Santa Teresa

20231-050 – Rio de Janeiro, RJ.

Os artigos submetidos às RBEs não devem ter sido publicados ou estar sendo considerados para publicação em outros periódicos.

A Revista não se responsabiliza pelos conceitos emitidos em matéria assinada.

Editor Responsável

Francisco Louzada-Neto (UFSCAR)

Editor de Estatísticas Oficiais

Denise Britz do Nascimento Silva (GAB/IBGE)

Editor de Metodologia

Enrico Antonio Colosimo (UFMG)

Editores Associados

Gilberto Alvarenga Paula (USP)

Dalton Francisco de Andrade (UFSC)

Ismenia Blavatsky de Magalhães (DPE/IBGE)

Helio dos Santos Migon (UFRJ)

Francisco Cribari-Neto (UFPE)

Editoração

Ismenia Blavatsky de Magalhães - Escola Nacional de Ciências Estatísticas - ENCE/IBGE

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2008.

Capa

Renato J. Aguiar – Coordenação de *Marketing*/CDDI/IBGE

Ilustração da Capa

Marcos Balster – Coordenação de *Marketing*/CDDI/IBGE

Revista brasileira de estatística/IBGE, - v.1, n.1 (jan./mar.1940), - Rio de Janeiro:IBGE, 1940-

v.

Trimestral (1940-1986), semestral (1987-).

Continuação de: Revista de economia e estatística.

Índices acumulados de autor e assunto publicados no v.43 (1940-1979) e v. 50 (1980-1989).

Co-edição com a Associação Brasileira de Estatística a partir do v.58.

ISSN 0034-7175 = Revista brasileira de estatística.

I. Estatística – Periódicos. I. IBGE. II. Associação Brasileira de Estatística.

IBGE. CDDI. Div. de Biblioteca e Acervos Especiais

CDU 31 (05)

RJ-IBGE/88-05 (rev.98)

PERIÓDICO

Impresso no Brasil/Printed in Brazil

Sumário

Nota do Editor	5
----------------------	---

Artigos

Estimação de Parâmetros de Modelos ARCH(p): Abordagem Bayes-MCMC versus Máxima Verossimilhança	7
	<i>Guilherme de Alencar Barreto Sandra Cristina de Oliveira Marinho Gomes de Andrade Filho</i>
Inferência em modelos heteroscedásticos na presença de pontos de alavanca	25
	<i>Francisco Cribari-Neto Tatiene C. Souza Klaus L.P. Vanconcellos</i>
Modelos de riscos multiplicativos e aditivos para analisar dados de sobrevivência: Uma abordagem paramétrica	51
	<i>Vera L.D. Tomazella Camila Bertini Martins</i>
Estimação Pontual em Regressão Beta: Uma Avaliação do Desempenho de Algoritmos de Otimização Não-Linear	63
	<i>Nátaly Adriana Jiménez Monroy Francisco Cribari-Neto Klaus L. P. Vasconcellos</i>
Política editorial	91

Nota do Editor

Este primeiro volume da RBEs do ano de 2008 é composto por quatro artigos. O primeiro artigo, de autoria de Guilherme de Alencar Barreto, Sandra Cristina de Oliveira e Marinho Gomes de Andrade Filho, que compara o método da máxima verossimilhança para estimação dos parâmetros de modelos ARCH com um novo procedimento Bayesianano. O segundo artigo, de autoria de Francisco Cribari-Neto, Tatiene C. Souza e Klaus L.P. Vasconcellos considera a estimação consistente da matriz de covariâncias de mínimos quadrados ordinários em um modelo de regressão linear sob heteroscedasticidade de forma desconhecida. O terceiro artigo, de autoria de Vera L.D. Tomazella e Camila Bertini Martins compara duas classes de modelos de riscos, nominalmente, multiplicativos e aditivos. O quarto artigo, de autoria de Nátaly A.J. Monroy, Francisco Cribari-Neto e Klaus L.P. Vasconcellos, avalia vários métodos de otimização não-linear da função log-verossimilhança quando o modelo de regressão beta é considerado.

Aproveito a oportunidade para agradecer a colaboração de todos os Editores Associados, revisores do periódico, autores e à equipe do IBGE. Em particular, gostaria de agradecer aos alunos de graduação, monitores da Escola Nacional de Ciências Estatísticas, Nicole P.R. Barros, Juliana F. C. Macedo, Joana A. Siqueira e Jéferson M. Ramos que colaboraram na editoração deste volume.

Uma excelente leitura.

Francisco Louzada-Neto
Editor Responsável

Estimação de Parâmetros de Modelos ARCH(p): Abordagem Bayes-MCMC *versus* Máxima Verossimilhança

Guilherme de Alencar Barreto*
Sandra Cristina de Oliveira**
Marinho Gomes de Andrade Filho***

Resumo

Neste artigo, o método da máxima verossimilhança, para estimação de parâmetros do modelo auto-regressivo com heterocedasticidade condicional (ARCH), é comparado com uma nova técnica bayesiana que utiliza simulação de Monte Carlo em Cadeias de Markov (MCMC). São simuladas séries temporais representando processos ARCH para diferentes ordens do modelo ($p = 2, 5$ e 7). Para $p = 2$, avalia-se a capacidade de ambas as técnicas de lidar tanto com modelos cuja variância incondicional é finita quanto com modelos com variância incondicional infinita. Para $p = 5$ e 7 avalia-se a influência da ordem do modelo na amplitude dos intervalos de confiança (ou credibilidade, para o caso bayesiano). A precisão das estimativas obtidas é avaliada pelo Raiz do Erro Quadrático Médio (RMSE) e o Erro Médio Percentual Absoluto (MAPE) entre as estimativas e os parâmetros conhecidos do modelo em questão, previamente usados para simular as séries. De um modo geral, a técnica proposta fornece estimativas mais precisas, mesmo para ordens elevadas, e é mais robusta que o método convencional, pois funciona independentemente do modelo ter variância finita ou não. Além disso, os intervalos de confiança obtidos pela técnica convencional violam restrições impostas aos parâmetros dos modelos ARCH, fato este não observado para a técnica bayesiana.

Endereço para correspondência: * Departamento de Engenharia de Teleinformática – DETI. Universidade Federal do Ceará – UFC, Campus do Pici. Av. Mister Hull, S/N, CP 6005, CEP 60455-760, Fortaleza-CE.

** Unidade Diferenciada de Tupã – UD-Tupã. Universidade Estadual Paulista – UNESP, Tupã, São Paulo – SP.

*** Instituto de Ciências Matemáticas e da Computação – ICMC. Departamento de Ciência da Computação e Estatística. Universidade de São Paulo – USP, Campus de São Carlos-SP. Av. Trabalhador Sancarlene, 400, CP 668, CEP 13560-970.

1. Introdução

Em séries financeiras, a variância condicional da variável observada, geralmente um retorno, varia no tempo (MORETTIN & TOLOI, 2004). Essa grandeza, também chamada de *volatilidade*, manifesta-se, na prática, de variadas formas, sendo bastante comum sua ocorrência em grupos de maior ou menor variabilidade, de modo que é possível visualizar períodos de forte oscilação (alta volatilidade) seguidos de períodos de calma relativa.

Construir modelos para volatilidade é fundamental para o mercado financeiro, influenciando na avaliação do risco de investimentos e no preço de ações. Os modelos matemáticos mais usados para quantificar a volatilidade de séries financeiras são os modelos auto-regressivos com heterocedasticidade condicional (ARCH) (ENGLE 1982), a serem estudados neste artigo, e os modelos *ARCH generalizados* (GARCH) (BOLLERSLEV 1986). Uma ampla revisão das propriedades destes modelos pode ser encontrada em Bollerslev *et al.*(1992) e Gouriéroux (1997).

No modelo ARCH, é considerado que a volatilidade em dado instante de tempo depende dos valores passados da série. Uma representação linear desta idéia foi proposta por Engle (1982), podendo o modelo ARCH de ordem p , ARCH(p), ser formulado como

$$y_t = \beta'x_t + z_t, \quad (1)$$

onde β é um vetor de parâmetros¹, x_t denota um vetor de variáveis exógenas ou elementos do conjunto condicionante Ω_{t-1} (informação até o instante $t-1$) e $z_t | \Omega_{t-1} \sim p(0, h_t)$, tal que $p(\cdot)$ é uma função de densidade de probabilidade paramétrica, usualmente Normal ou t -Student, e $h_t = E\{z_t^2 | \Omega_{t-1}\}$ é a volatilidade, dada por

$$h_t = \alpha_0 + \sum_{j=1}^p \alpha_j z_{t-j}^2, \quad (2)$$

onde $\alpha_0 > 0$, $\alpha_j \geq 0$, $j > 0$. É comum adotar $\beta = 0$, definindo z_t como

$$z_t = \sqrt{h_t} \cdot \varepsilon_t, \quad (3)$$

na qual a seqüência aleatória $\{\varepsilon_t\}$, $t \geq 0$ é assumida como ruído branco gaussiano de média zero e variância unitária. A idéia básica do modelo ARCH(p) está em assumir que a variável z_t

¹O sobrescrito (') denota o vetor transposto.

é não-correlacionada serialmente, porém a sua variância condicional (volatilidade) depende de valores passados via uma transformação quadrática.

Foi mencionado na Equação (3) que, para se ter $h_t > 0$, é necessário que $\alpha_0 > 0$ e $\alpha_j \geq 0$. Além disso, para que o processo z_t tenha variância finita, a condição necessária e suficiente é que todas as raízes do polinômio $1 - \sum_{j=1}^p \alpha_j Q^j$ estejam fora do círculo de raio unitário². Satisfeita esta condição, a variância incondicional de z_t é dada por

$$E[z_t^2] = \frac{\alpha_0}{1 - \sum_{j=1}^p \alpha_j}. \quad (4)$$

Portanto, a condição necessária para que z_t tenha variância incondicional finita é que $\alpha_0 > 0$ e $\sum_{j=1}^p \alpha_j < 1$.

O cálculo de estimativas de máxima verossimilhança (*maximum likelihood* - ML) para os parâmetros dos modelos ARCH exige a maximização de uma função não-linear (HAMILTON, 1994). Engle (1982) sugere o uso do método de Newton para estimação dos parâmetros α_j baseado na técnica de máxima verossimilhança. Este procedimento não considera as restrições impostas aos parâmetros, apresentadas no parágrafo anterior a partir da análise das Equações (3) e (4). Por outro lado, utilizar métodos de otimização não-linear que considerem tais restrições envolve uma série de dificuldades, conforme destacado por Geweke (1986). Além disso, procedimentos para identificar, ajustar e fazer diagnósticos de modelos, assim como prever valores futuros da série em análise lançam mão de propriedades da teoria assintótica, tornando a abordagem analiticamente intratável, uma vez que modelos ARCH estão muito distante da linearidade.

Como alternativa para contornar essas dificuldades, propõe-se neste artigo uma abordagem bayesiana para os modelos ARCH de séries temporais. Esta abordagem vem sendo usada com sucesso, por exemplo, em Barreto & Andrade (2004), Geweke (1992), Migon & Mazucheli (1999) e Nakatsuma (2000). Um dos primeiros autores a propor uma abordagem bayesiana para modelos ARCH foi Geweke (1989a, 1989b). Nestes trabalhos, ele considerou um tipo particular de relação entre os parâmetros do modelo que possibilita o uso de densidades *a priori* não-informativa e os parâmetros são estimados usando-se técnicas de simulação MCMC (em geral *importance sampling*).

² O operador Q simboliza o atraso unitário z^{-1} , ou seja, $Q_y(t) \equiv z^{-1}y(t) = y(t-1)$.

Uma abordagem bayesiana semiparamétrica para modelos ARCH foi proposta por Koop (1994), que usa as vantagens supracitadas da metodologia bayesiana de Geweke, mas evita a estimativa de densidades paramétricas *a posteriori*. Posteriormente, Nakatsuma (2000) propôs um método usando simulação MCMC para gerar as densidades *a posteriori* dos parâmetros de modelos de regressão linear com erros ARMA-GARCH fazendo uso de densidades *a priori* Normal.

No presente artigo, a técnica de cálculo de estimativas ML é comparada com uma técnica bayesiana que utiliza simulação MCMC. Para isto, é proposta uma reparametrização do modelo ARCH que mapeia as restrições impostas aos parâmetros em um conjunto de parâmetros irrestritos, evitando, assim, a necessidade de se considerar restrições no domínio das densidades *a priori*, permitindo inclusive a identificação de modelos com variância infinita. Neste trabalho, as condições de estacionariedade passam a ser alvo de verificação *a posteriori*.

A intenção é mostrar que a técnica bayesiana proposta fornece estimativas mais precisas, mesmo para ordens elevadas, e é mais robusta que o método convencional, pois funciona independentemente do modelo ser estacionário ou não. Além disso, os intervalos de confiança obtidos pela técnica ML violam as restrições impostas aos parâmetros de modelos ARCH, fato este não observado na técnica bayesiana.

O restante do artigo está organizado da seguinte maneira. A Seção 2 descreve brevemente a filosofia do método ML e mostra como aplicá-lo à estimação dos parâmetros de modelos ARCH. A Seção 3 faz o mesmo para a técnica bayesiana. A Seção 4 apresenta as simulações e resultados da comparação entre as duas técnicas. O artigo é concluído na Seção 5.

2. Estimação via Máxima Verossimilhança

Dada uma realização (série temporal) $Z = \{z_t\}$, $t = 1, 2, \dots, T$, do processo estocástico z_t , assume-se que a densidade de probabilidade que melhor explica este conjunto de observações tem forma conhecida (Normal, *t*-Student, etc.) e depende de um vetor de parâmetros fixos, α , porém desconhecidos.

Seja $Z_p = \{z_1, \dots, z_p\}$ o conjunto das p primeiras observações e $Z_{t-p} = \{z_{p+1}, z_{p+2}, \dots, z_t\}$ o conjunto das $t - p$ observações restantes, tal que $Z = Z_p \cup Z_{t-p}$. A função de verossimilhança de Z , condicionada nas p -primeiras observações, é então escrita como

$$L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha}) \propto \prod_{t=p+1}^T \frac{1}{\sqrt{h_t}} \exp\left\{-\frac{z_t^2}{2h_t}\right\}, \quad (5)$$

onde $\boldsymbol{\alpha} = [\alpha_0 \alpha_1 \dots \alpha_p]'$ é o vetor de parâmetros do modelo. Para simplificar a análise, determina-se a função de log-verossimilhança parcial, dada por

$$l(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha}) = \log L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha}) = \frac{1}{T} \sum_{t=p+1}^T l(z_t | \boldsymbol{\alpha}) \quad (6)$$

obtendo-se, assim, a função de log-verossimilhança da t -ésima observação

$$l(z_t | \boldsymbol{\alpha}) = -\frac{1}{2} \log(h_t) - \frac{z_t^2}{2h_t}. \quad (7)$$

Para um dado conjunto de observações \mathbf{Z} , a estimativa de máxima verossimilhança do vetor de parâmetros é definida como aquele valor de $\boldsymbol{\alpha}$ que maximiza $L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha})$ ou $l(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha})$. Como as Equações (5) e (6) são ambas não-lineares, Engle (1982) propôs o uso de técnicas de otimização numérica para calcular as estimativas ML dos parâmetros α_j , $j = 0, 1, \dots, p$, tal como o método de quasi-Newton, descrito pela seguinte equação recursiva

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + [\hat{I}^{(k)}]^{-1} \left[\frac{\partial l(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha}^{(k)})}{\partial \boldsymbol{\alpha}^{(k)}} \right], \quad (8)$$

na qual $[\hat{I}^{(k)}]$ é a matriz de informação (equivalente à matriz Hessiana) e $l(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha}^{(k)}) / \partial \boldsymbol{\alpha}^{(k)}$ é o gradiente da função de log-verossimilhança, ambos avaliados para a estimativa do vetor $\boldsymbol{\alpha}$ na iteração k , ou seja, $\boldsymbol{\alpha}^{(k)}$.

Pode-se mostrar que o gradiente da função $l(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha})$ é dado por

$$\frac{\partial l(\mathbf{Z}_{-p} | \mathbf{Z}_p, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{1}{T} \sum_{t=p+1}^T \frac{1}{2h_t} \nu_t \left(\frac{z_t^2}{h_t} - 1 \right), \quad (9)$$

em que $\nu_t = [1 \quad z_{t-1}^2 \quad \dots \quad z_{t-p}^2]'$. Já a matriz de informação $I = \{I_{ij}\}$, $i, j = 0, 1, \dots, p$, é definida como

$$I_{ij} = -E \left[\frac{1}{T} \sum_{t=p+1}^T \frac{\partial^2 l(z_t | \boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_j} \right], \quad (10)$$

tal que, usando o fato de $E[2z_t^2/h_t - 1] = 2E[z_t^2]/h_t - 1$, os elementos da matriz de informação $\{\hat{I}_{ij}\}$ podem ser consistentemente estimados por

$$\hat{I}_{ij} = \frac{1}{T} \sum_{t=p+1}^T \frac{1}{2h_t^2} \left\{ \frac{\partial h_t}{\partial \alpha_i} \frac{\partial h_t}{\partial \alpha_j} \right\}, \quad (11)$$

tal que da Equação (3) obtém-se

$$\frac{\partial h_t}{\partial \alpha_j} = \begin{cases} 1, & \text{Para } j=0 \\ z_{t-j}^2, & \text{Para } j=1, \dots, p \quad \text{e } t \geq p+1 \end{cases} \quad (12)$$

Vale enfatizar que é comum a ocorrência de problemas numéricos no método ML, causados, principalmente, pelo mau-condicionamento da matriz de informação. Esta situação causa imprecisões na inversão matricial requerida na Equação (8), particularmente para valores mais altos da ordem do modelo ARCH. Outra dificuldade do método ML é considerar as restrições impostas aos parâmetros, ou seja, $\alpha_0 > 0$ e $\sum_{j=1}^p \alpha_j < 1$. Neste artigo, estas restrições são relaxadas e somente são verificadas após os cálculos das estimativas. Estas são as principais razões que levam à proposição do método bayesiano a seguir, haja vista que não há necessidade de técnicas de otimização numérica.

3. Abordagem Bayesiana e Simulação MCMC

O método ML considera os parâmetros como constantes, já a abordagem bayesiana entende o vetor α como uma *variável aleatória* p -dimensional que obedece à determinada densidade de probabilidade conjunta. Este método requer, também, a especificação de uma *densidade de probabilidade conjunta a priori* $p(\alpha)$, que reflete o conhecimento prévio sobre a distribuição de α .

Considerando a série temporal $\mathbf{Z} = \{z_t\}$, $t = 1, 2, \dots, T$, a abordagem bayesiana para o problema de estimação dos parâmetros do modelo ARCH(p) dá-se pela combinação da função de verossimilhança $L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \alpha)$ com a densidade *a priori* $p(\alpha)$, através da *regra de Bayes*, dada por

$$p(\alpha | \mathbf{Z}) \propto p(\alpha) L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \alpha), \quad (13)$$

onde $p(\alpha | \mathbf{Z})$, chamada *densidade a posteriori* de α , caracteriza a variável aleatória α após a observação de \mathbf{Z} .

Em geral, $p(\alpha | \mathbf{Z})$ não é uma densidade de probabilidade padrão. Assim, valores médios e variâncias deverão ser obtidos por integração, o que pode se tornar uma tarefa muito difícil. Uma solução alternativa consiste em estimar $p(\alpha | \mathbf{Z})$, a partir de uma amostra aleatória gerada por simulação MCMC (GAMERMAN & LOPES, 2006; GELMAN *et al.*, 1996), cujo o objetivo é gerar uma seqüência de valores, chamada de *cadeia de Markov*, $\{\alpha^{(i)}, i = 1, \dots, m\}$ a partir das densidades condicionais $p(\alpha_j | \alpha_{-j}, \mathbf{Z})$, obtidas da densidade *a posteriori* $p(\alpha | \mathbf{Z})$, em que α_{-j} é um vetor com todos os parâmetros do modelo exceto o parâmetro α_j , ou seja,

$$\alpha_{-j} = (\alpha_0, \alpha_1, \dots, \alpha_{j-1}, \alpha_{j+1}, \dots, \alpha_p). \quad (14)$$

Para evitar estimativas viesadas, é comum descartar a parte inicial da cadeia evitando influência das condições iniciais e selecionar valores a cada K passos, para eliminar a correlação entre as amostras. A estimativa bayesiana de α , assumindo uma função de perda quadrática, é o valor esperado $E(\alpha)$ com relação à densidade *a posteriori* na Equação (13). Assim, uma estimativa do vetor α é obtida como a média amostral dos $M = m/2K$ valores restantes, ou seja,

$$\hat{\alpha} = \frac{1}{M} \sum_{j=1}^M \alpha^{(i)} \quad (15)$$

Porém, antes de se utilizar qualquer algoritmo MCMC, surge a questão sobre qual densidade *a priori* $p(\alpha)$ adotar. Para escolher uma densidade *a priori* para o vetor de parâmetros α , considera-se, primeiramente, que todos os α_j , $j = 0, 1, \dots, p$, são independentes com uma densidade *a priori* definida no intervalo $[a_j, b_j]$ com $a_j > 0$ e $b_j < 1$, tal que $a_j \leq \alpha_j \leq b_j$, visto que a restrição $\sum_{j=1}^p \alpha_j < 1$ exige que todos os $\alpha_j \in (0, 1)$. Pode-se definir também $a_0 \leq \alpha_0 \leq b_0$, com $a_0 \geq 0$ e $b_0 \leq E[z_i^2]$.

Uma vez que se deseja gerar valores para α_j próximos das médias de cada intervalo $(a_j + b_j)/2$ com mais freqüência do que os valores próximos dos limites a_j e b_j , então, propõe-se a seguinte reparametrização do problema:

$$\phi_j \log \left(\frac{\alpha_j - a_j}{b_j - \alpha_j} \right), \quad (16)$$

para $j = 0, 1, \dots, p$. Esta reparametrização conduz à escolha de uma transformação que mapeia os intervalos $(-\infty, +\infty)$ no domínio $[a_j, b_j]$ e vice-versa.

Neste artigo, adota-se a densidade Normal para $\varphi = [\phi_0 \ \phi_1 \ \dots \ \phi_p]'$ de forma que cada ϕ_j tem uma distribuição *a priori* $p(\phi_j) = N(0, \sigma_j^2)$, pois escolhendo-se um valor de σ_j , adequadamente reduz-se a taxa de rejeição do algoritmo de simulação de MCMC, acelerando também a convergência deste.

Com essa reparametrização, a função de verossimilhança é escrita da mesma forma que na Equação (5), porém com a volatilidade h_t , dada agora como função dos parâmetros ϕ_j . Assim, tem-se que

$$L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \varphi) \propto \prod_{t=p+1}^T \frac{1}{\sqrt{h_t(\varphi)}} \exp \left\{ -\frac{z_t^2}{2h_t(\varphi)} \right\}. \quad (17)$$

Em seguida, considera-se a matriz diagonal $\Sigma(\varphi)$ dada por

$$\Sigma(\varphi) = \text{diag}(h_{p+1}(\varphi), \dots, h_T(\varphi)). \quad (18)$$

Considerando a matriz diagonal $\Sigma(\varphi)$, a Equação (17) pode ser reescrita como

$$L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \varphi) \propto \frac{1}{\sqrt{|\Sigma(\varphi)|}} \exp \left\{ -\frac{1}{2} \mathbf{Z}' [\Sigma^{-1}(\varphi)] \mathbf{Z} \right\}, \quad (19)$$

tal que $|\cdot|$ denota o determinante de uma matriz. Conforme mencionado anteriormente, ao se assumir que os parâmetros ϕ_j são independentes e com densidades *a priori* $p(\phi_j) = N(0, \sigma_j^2)$, a densidade *a priori* conjunta para φ é dada por

$$p(\varphi) = \prod_{j=0}^p p(\phi_j). \quad (20)$$

De posse das Equações (19) e (20), pode-se escrever a densidade *a posteriori* para φ como

$$p(\varphi | \mathbf{Z}) \propto \left(\prod_{t=p+1}^T \frac{1}{\sqrt{h_t(\varphi)}} \exp \left\{ -\frac{z_t^2}{2h_t(\varphi)} \right\} \right) \cdot \prod_{j=0}^p p(\phi_j) \quad (21)$$

As densidades condicionais *a posteriori* para ϕ_j , obtidas da Equação (21), são então dadas por

$$p(\phi_j | \varphi_{-j}, \mathbf{Z}) \propto L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \varphi) p(\phi_j). \quad (22)$$

em que φ_{-j} é um vetor com todos os parâmetros do modelo exceto o parâmetro ϕ_j , ou seja,

$$\varphi_{-j} = (\phi_0, \phi_1, \dots, \phi_{j-1}, \phi_{j+1}, \dots, \phi_p).$$

As densidades condicionais dadas na Equação (22) são utilizadas neste artigo por um algoritmo de simulação MCMC para gerar uma amostra da densidade conjunta *a posteriori* dada na Equação (21). O algoritmo MCMC descrito a seguir consiste no uso combinado do algoritmo do amostrador de Gibbs (CASELLA & GEORGE, 1992) e do algoritmo *Metropolis-Hastings* (MH) (CHIB & GREENBERG, 1995). O uso do algoritmo MH dentro do algoritmo de Gibbs é necessário, pois as condicionais da Equação (22) não têm forma fechada.

Algoritmo MCMC: Amostrador de Gibbs com *Metropolis-Hastings*

1. Atribuir valores iniciais aos parâmetros $\varphi^{(0)} = \{\phi_j^{(0)}\}, j = 0, \dots, p$. Fazer $i = 1$.
2. Gerar um novo valor β a partir das densidades $p(\phi_j | \varphi_{-j}, \mathbf{Z})$, observando a seguinte relação:

$$\varphi^{(i)} = (\phi_0^{(i-1)}, \phi_1^{(i-1)}, \dots, \phi_{j-1}^{(i-1)}, \beta, \phi_{j+1}^{(i-1)}, \dots, \phi_p^{(i-1)})$$

3. Calcular a probabilidade de aceitação de β :

$$P(\phi_j^{(i-1)}, \beta) = \begin{cases} \min\{1, R\}, & \text{se } L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \varphi^{(i-1)}) \neq 0 \\ 1, & \text{caso contrário} \end{cases} \quad (23)$$

na qual

$$R = \frac{L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \varphi^{(i)})}{L(\mathbf{Z}_{-p} | \mathbf{Z}_p, \varphi^{(i-1)})}. \quad (24)$$

4. Gerar um número aleatório $u \sim U(0, 1)$ e executar o seguinte teste:

$$\phi_j^{(i)} = \begin{cases} \beta, & u \leq P(\phi_j^{(i-1)}, \beta) \\ \phi_j^{(i-1)}, & \text{caso contrário} \end{cases} \quad (25)$$

5. Repetir os passos 2-5 para $j = 0, \dots, p$.
6. Fazer $i = i + 1$ e retornar ao Passo 2 até as cadeias convergirem.

Após a convergência, as estimativas originais $\hat{\alpha}_j$ podem ser computadas pela transformação inversa:

$$\hat{\alpha}_j = \frac{b_j \exp\{\hat{\phi}_j\} + a_j}{1 + \exp\{\hat{\phi}_j\}}, \quad (26)$$

em que $\hat{\phi}_j$ é determinado como na Equação (15); ou seja, a partir da média das amostras geradas pelo algoritmo MCMC descrito acima.

4. Simulações e Resultados

Para comparar as abordagens ML e Bayesiana na estimação dos parâmetros α_j , foram simuladas 1 000 séries independentes³ para os modelos ARCH(2), ARCH(5) e ARCH(7). Cada série continha 1 000 amostras originalmente, porém os primeiros 500 valores foram descartados para mitigar os efeitos das condições iniciais. As 500 amostras restantes foram usadas para estimar os parâmetros do modelo ARCH. No final, cada um dos parâmetros possuía 1 000 estimativas diferentes geradas pelo método ML e 1 000 geradas pela técnica bayesiana, tal que a média aritmética delas representa o valor final do parâmetro $\hat{\alpha}_j$ (ver Equação 15).

Vale ressaltar que o estudo comparativo feito neste trabalho só pode ser aplicado em situações hipotéticas, nas quais os parâmetros verdadeiros dos modelos são conhecidos. Mesmo não sendo realista, esta situação ilustra a potencialidade de cada técnica quanto à precisão na estimação dos parâmetros na ausência do valor verdadeiro dos mesmos.

Assim, o desempenho entre as abordagens é comparado através da raiz do erro quadrático médio (RMSE) e do erro absoluto percentual médio (MAPE) entre os valores reais dos parâmetros α_j e os valores estimados $\hat{\alpha}_j$ por cada uma das técnicas:

$$RMSE = \sqrt{\frac{1}{K} \sum_{j=1}^K (\alpha_j - \hat{\alpha}_j)^2}, \quad e \quad (27)$$

$$MAPE = \frac{100}{K} \sum_{j=1}^K \left| \frac{\alpha_j - \hat{\alpha}_j}{\alpha_j} \right|. \quad (28)$$

³ Ou seja, usando seqüências diferentes do ruído branco ε_t de acordo com a Equação (3).

Outra forma aqui adotada de se avaliar a qualidade das estimativas é através dos intervalos de confiança (método ML) e intervalos de credibilidade (método bayesiano). Estes intervalos indicam o grau de certeza com que se pode encontrar o valor real do parâmetro dentro da faixa por eles especificada. Quanto menor for a largura do intervalo, mais confiável é o método de estimação. Neste artigo, usa-se um grau de certeza (nível de confiança) de 95%.

Os coeficientes dos modelos ARCH(2), ARCH(5) e ARCH(7) simulados são mostrados na Tabela 1, na qual se nota que o modelo M3 viola a restrição variância incondicional finita (i.e. $\alpha_1 + \alpha_2 < 1$). Este modelo foi adotado propositadamente para verificar a robustez das duas técnicas em situações anormais. Os parâmetros da transformação usada na técnica bayesiana (ver Equação (16)) são dados na Tabela 2.

Tabela 1 - Parâmetros dos modelos ARCH simulados

Modelo	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7
M1	1,00	0,80	0,15	–	–	–	–	–
M2	1,00	0,50	0,20	–	–	–	–	–
M3	1,00	0,90	0,20	–	–	–	–	–
$p = 5$	1,00	0,50	0,20	0,15	0,08	0,05		
$p = 7$	1,00	0,50	0,20	0,10	0,05	0,03	0,06	0,02

Os hiperparâmetros das densidades *a priori* para os cinco modelos ARCH avaliados segundo a técnica bayesiana são dados na Tabela 2.

Tabela 2 - Parâmetros da Reparametrização - Equação (16)

Modelo	M1	M2	M3	ARCH(5)	ARCH(7)
a_0, b_0	0,85; 1,15	0,85; 1,15	0,85; 1,15	0,85; 1,15	0,85; 1,15
a_1, b_1	0,65; 0,95	0,35; 0,65	0,75; 1,05	0,35; 0,65	0,35; 0,65
a_2, b_2	0,00; 0,30	0,20; 0,40	0,20; 0,40	0,20; 0,40	0,20; 0,40
a_3, b_3				0,00; 0,30	0,03; 0,17
a_4, b_4				0,00; 0,16	0,00; 0,10
a_5, b_5				0,00; 0,10	0,00; 0,06
a_6, b_6					0,00; 0,12
a_7, b_7					0,00; 0,04
σ_i^2	1,0	1,0	1,0	1,0	1,0

Na implementação do algoritmo MCMC, para cada parâmetro foi simulada uma cadeia com 5 000 iterações, descartada a metade inicial para diminuir o efeito das condições iniciais, e então tomados valores espaçados de 5 em 5, totalizando uma amostra de tamanho 500. A convergência das cadeias foi sempre observada para os hiperparâmetros escolhidos, segundo o critério de Geweke (1992).

Os gráficos da Figura 1 apresentam a distribuição empírica (histograma) das estimativas ML dos parâmetros do modelo M1. Os gráficos para os modelos M2 e M3 são semelhantes aos do modelo M1, sendo por isso omitidos. Os gráficos da Figura 2 são as densidades *a posteriori* dos parâmetros do modelo M1, omitem-se aqui, também, os gráficos para os modelos M2 e M3.

Figura 1 - Histogramas das estimativas de máxima verossimilhança

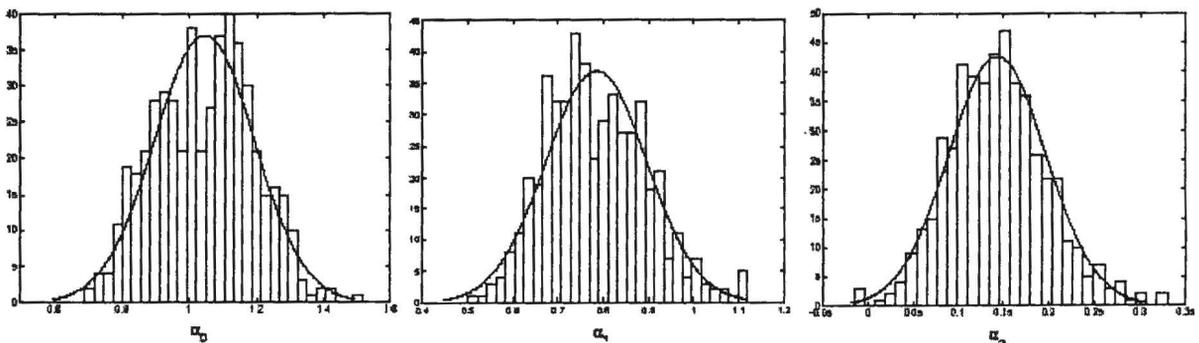
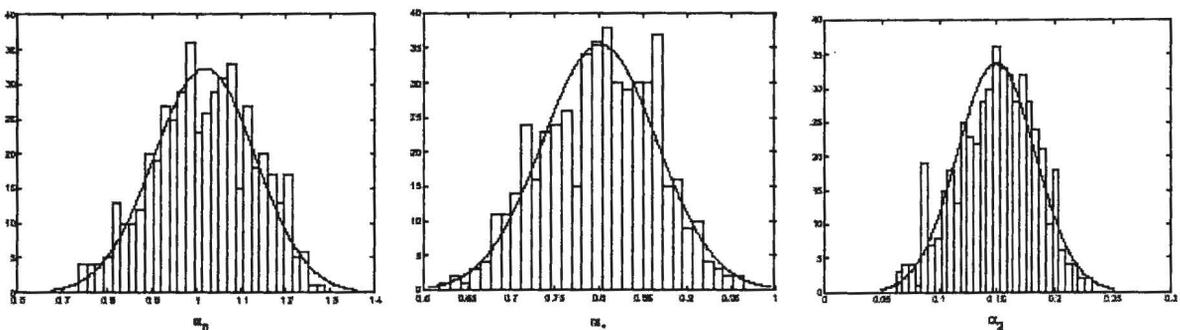


Figura 2 - Densidades *a posteriori* dos parâmetros do Modelo M1-ARCH(2)



Os erros de estimação para cada um dos modelos ARCH(2) listados na Tabela 1 são apresentados na Tabela 3. Estes resultados mostram que a abordagem bayesiana apresenta estimativas mais precisas que as estimativas ML e este comportamento não muda mesmo que os parâmetros não satisfaçam às condições de estacionariedade requeridas pelo modelo ARCH.

Tabela 3 - ML versus Bayes - modelo ARCH(2)

Modelo M1		$\hat{\alpha}$	RMSE	MAPE(%)	CI (95%)
a_0	Bayes	1,0164	0,1145	8,78	[0,8082; 1,2024]
	MLE	1,0365	0,2115	9,18	[0,7081; 1,5019]
a_1	Bayes	0,8009	0,0375	6,88	[0,6853; 0,8514]
	MLE	0,7830	0,1546	12,30	[0,4894; 1,0649]
a_2	Bayes	0,1950	0,0673	17,56	[0,0855; 0,2079]
	MLE	0,1435	0,0741	36,16	[0,0024; 0,3026]
Modelo M2		$\hat{\alpha}$	RMSE	MAPE(%)	CI (95%)
a_0	Bayes	1,0098	0,1028	8,49	[0,8099; 1,1982]
	MLE	1,0261	0,1665	13,55	[0,7418; 1,4299]
a_1	Bayes	0,5025	0,0920	14,69	[0,3232; 0,6683]
	MLE	0,4836	0,1180	18,73	[0,2615; 0,7058]
a_2	Bayes	0,2202	0,0554	22,25	[0,1171; 0,3279]
	MLE	0,1935	0,0827	33,05	[0,1851; 0,3726]
Modelo M2		$\hat{\alpha}$	RMSE	MAPE(%)	CI (95%)
a_0	Bayes	1,0041	0,1056	8,62	[0,7971; 1,1890]
	MLE	1,0315	0,2108	9,45	[0,6622; 1,5060]
a_1	Bayes	0,8981	0,0887	9,75	[0,7309; 1,1164]
	MLE	0,8750	0,1543	15,42	[0,5599; 1,1625]
a_2	Bayes	0,2017	0,0416	21,75	[0,1208; 0,3205]
	MLE	0,1949	0,0746	25,77	[0,0429; 0,3401]

Para verificar o efeito da ordem do modelo na precisão das estimativas, foram considerados dois modelos de ordem mais elevada, um modelo ARCH(5) e um modelo ARCH(7). A Tabela 4 mostra os valores obtidos para RMSE e MAPE, assim como os intervalos de confiança 95% para as estimativas ML e os intervalos de credibilidade 95% para as estimativas bayesianas para o modelo ARCH(5).

Tabela 4 - ML versus Bayes - modelo ARCH(5)

Modelo M1		$\hat{\alpha}$	RMSE	MAPE(%)	CI (95%)
a_0	Bayes	1,0076	0,0976	7,94	[0,7978; 1,1734]
	MLE	1,0959	0,3328	25,33	[0,5461; 1,8299]
a_1	Bayes	0,4981	0,0924	14,72	[0,3144; 0,6650]
	MLE	0,4905	0,1263	19,86	[0,2458; 0,7347]
a_2	Bayes	0,2208	0,0571	23,12	[0,1157; 0,3273]
	MLE	0,1928	0,0985	39,51	[0,0103; 0,3934]
a_3	Bayes	0,1496	0,3007	17,79	[0,0908; 0,2059]
	MLE	0,1483	0,0898	47,63	[-0,0263; 0,3424]
a_4	Bayes	0,0937	0,0225	22,53	[0,0599; 0,1302]
	MLE	0,0719	0,0711	70,04	[-0,0556; 0,2215]
a_5	Bayes	0,0494	0,0071	11,59	[0,0361; 0,0633]
	MLE	0,0445	0,0586	93,59	[-0,0533; 0,1764]

No cálculo das estimativas ML dos modelos ARCH(5) e ARCH(7), foram geradas 1 000 séries, mas o método iterativo de cálculo das estimativas só convergiu para 901 dessas séries. Como critério de aceitabilidade da estimativa, foi avaliado o condicionamento R da matriz $[\hat{I}^{(k)}]$, dado pela razão entre o maior e o menor de seus autovalores. O valor $B = 1/R$ é uma medida prática da qualidade da estimativa ML, pois se $B \approx 1$ a matriz $[\hat{I}^{(k)}]$ é bem-condicionada, porém se $B \approx 0$ então esta matriz é malcondicionada. Neste artigo, as estimativas ML geradas quando a matriz $[\hat{I}^{(k)}]$ tem $B < 10^{-4}$ foram descartadas.

Para a estimação bayesiana procedeu-se como para os modelos ARCH(2), ou seja, foi implementado o algoritmo MCMC e para cada parâmetro foi simulada uma cadeia com 5 000 amostras, descartando-se a metade inicial das cadeias e tomado-se valores espaçados de 5 em 5, totalizando uma amostra de tamanho 500. A convergência das cadeias também foi verificada pelo critério de Geweke (1992). Os histogramas representando as distribuições empíricas das estimativas ML e as densidades *a posteriori* dos parâmetros do modelo são apresentados nas Figuras 3 e 4, respectivamente.

Figura 3 - Distribuições empírica dos estimadores do modelo ARCH(5)

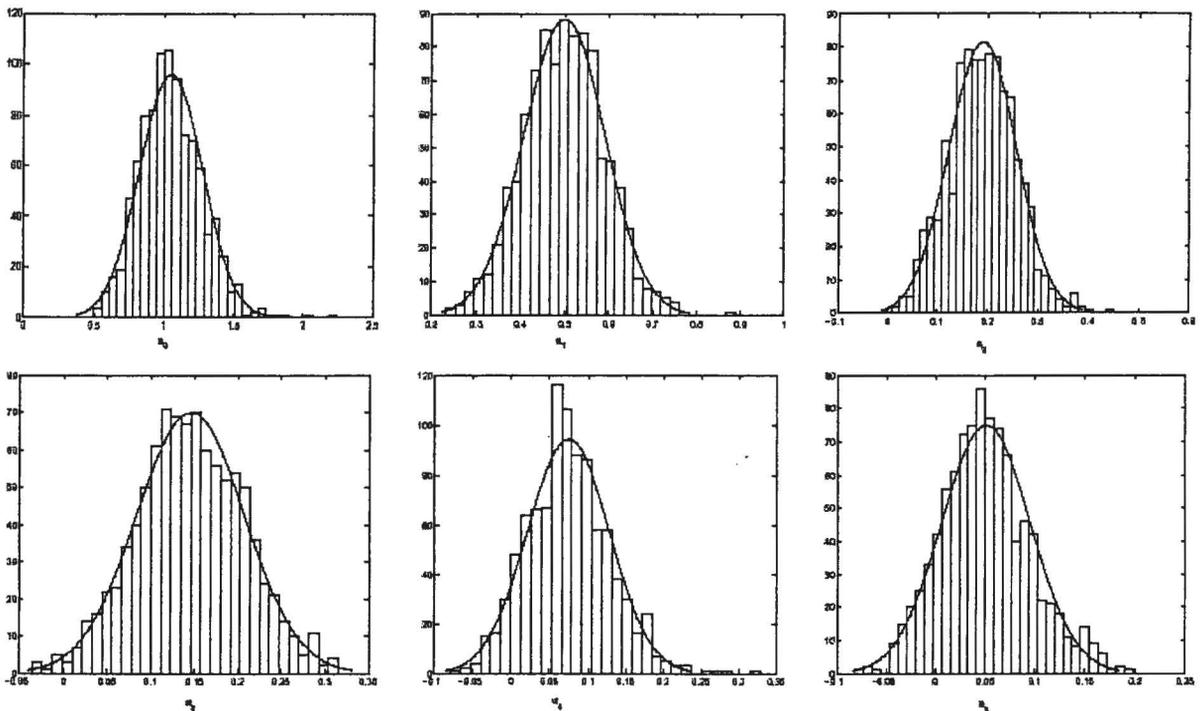
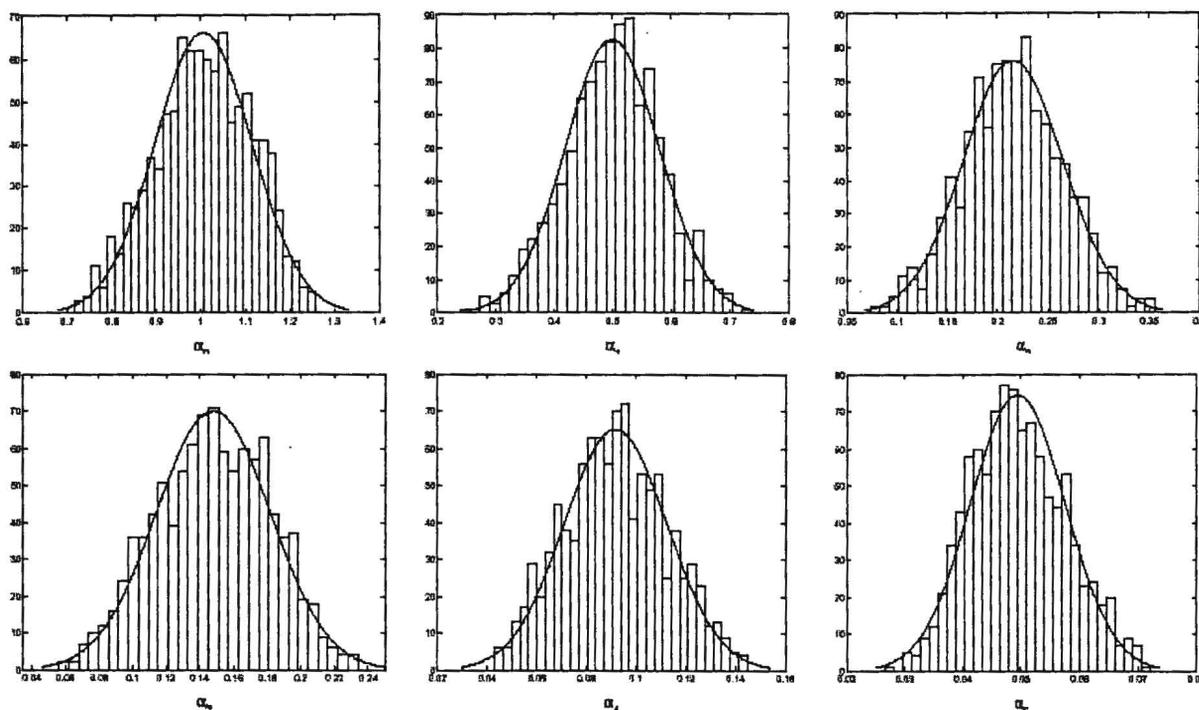


Figura 4 - Densidades *a posteriori* dos parâmetros do modelo ARCH(5)



Semelhantemente ao ocorrido para o modelo ARCH(2), os resultados permitem concluir que as estimativas bayesianas apresentam maior precisão (menores RMSE e MAPE) que as estimativas ML. Além disso, os intervalos de credibilidade bayesianos são mais estreitos do que os intervalos de confiança ML, indicando maior confiabilidade da técnica bayesiana.

É importante enfatizar que algumas das estimativas ML dos parâmetros $\alpha_j, j = 3, 4$ e 5, apresentaram valores negativos, violando, assim, as condições de positividade impostas aos parâmetros de modelos ARCH (ver Equação(3)). Os resultados da comparação entre as abordagens ML e bayesiana para o modelo ARCH(7) são apresentados na Tabela 5.

Comparando os valores de RMSE e MAPE na Tabela 5 com os da Tabela 4, conclui-se que as estimativas ML são tanto menos precisas quanto maior for ordem do modelo. Assim como observado para o modelo ARCH(5), algumas estimativas ML e seus respectivos intervalos de confiança apresentaram valores negativos.

Tabela 5 - MLE versus Bayes - modelo ARCH(7)

	Técnica	$\hat{\alpha}$	RMSE	MAPE(%)	CI (95%)
a_0	Bayes	1,0227	0,0968	7,96	[0,8298; 1,1911]
	MLE	1,0980	0,4224	21,62	[0,5911; 1,6971]
a_1	Bayes	0,4973	0,0629	10,29	[0,3712; 0,6082]
	MLE	0,4920	0,1839	19,17	[0,2335; 0,7025]
a_2	Bayes	0,2215	0,0544	21,17	[0,1165; 0,3153]
	MLE	0,1876	0,2849	39,37	[0,0334; 0,3336]
a_3	Bayes	0,0995	0,0179	14,54	[0,0650; 0,1337]
	MLE	0,0933	0,1646	73,80	[-0,1270; 0,2357]
a_4	Bayes	0,0492	0,0063	10,33	[0,0378; 0,0335]
	MLE	0,0447	0,1586	137,13	[-0,1677; 0,1827]
a_5	Bayes	0,0298	0,0028	7,46	[0,0249; 0,0335]
	MLE	0,0332	0,1505	201,60	[-0,1163; 0,1558]
a_6	Bayes	0,0596	0,0092	12,45	[0,0428; 0,0776]
	MLE	0,0715	0,2472	102,79	[-0,0466; 0,1742]
a_7	Bayes	0,0198	0,0016	6,59	[0,0170; 0,0234]
	MLE	0,0241	0,1981	317,21	[-0,1269; 0,1266]

5. Conclusão

Este trabalho comparou o método tradicional de estimação de parâmetros por máxima verossimilhança com uma abordagem bayesiana que utiliza simulação MCMC com uma nova reparametrização, mais geral que as atualmente disponíveis. Foram simuladas várias séries temporais representando processos auto-regressivos com heterocedasticidade condicional (ARCH) para diferentes ordens do modelo ($p = 2$, $p = 5$ e $p = 7$).

De um modo geral, a técnica bayesiana proposta fornece estimativas mais precisas, mesmo para ordens elevadas e é mais robusta que o método ML convencional, pois funciona independentemente do modelo ser estacionário ou não. Uma outra vantagem da abordagem bayesiana proposta está na estabilidade da convergência, diferentemente do método ML cuja solução numérica está bastante sujeita à não-convergência, principalmente para ordens elevadas do modelo ARCH.

Além disso, os intervalos de confiança obtidos pela técnica ML violam restrições de positividade impostas aos parâmetros dos modelos ARCH, fato este não observado para a técnica bayesiana.

Referências Bibliográficas

- Barreto, G. A. & Andrade, M. G. (2004). Robust bayesian approach for AR(p) models applied to streamflow forecasting, *Journal of Applied Statistical Science* 13(3): 269–292.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31: 307-327.
- Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* 52(1): 5–59.
- Casella, G. & George, E. I. (1992). Explaining the Gibbs sampler, *The American Statistician* 46(8): 167–174.
- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician* 49(4): 327–335.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, *Econometrica* 50(4): 987–1007.
- Gamerman, D. & Lopes, H. (2006). Monte Carlo Markov Chain: *Stochastic Simulation for Bayesian Inference*, 2nd ed, Chapman & Hall.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (1996). *Bayesian Data Analysis*, 2nd edn, Chapman & Hall.
- Geweke, J. (1986). Exact inference in the inequality constrained normal linear regression model, *Journal of Applied Econometrics* 1(2): 127–141.
- Geweke, J. (1989a). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* 57(6): 1317–1339.
- Geweke, J. (1989b). Exact predictive densities for linear models with ARCH disturbances, *Journal of Econometrics* 40: 63–86.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith (eds), *Bayesian Statistics*, Vol. 4, Clarendon Press, Oxford, UK.
- Gourieroux, C. (1997). *ARCH Models and Financial Applications*, Springer-Verlag.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press.
- Koop, G. (1994). Bayesian semi-nonparametric ARCH models, *The Review of Econometrics and Statistics* 76(1): 176–181.
- Migon, H. S. & Mazucheli, J. (1999). Modelos GARCH bayesianos: Métodos aproximados e aplicações, *Revista Brasileira de Econometria* 19(1): 111–138.
- Morettin, P. A. & Toloi, C. M. (2004). *Análise de Séries Temporais*, 1st ed, ABE - Projeto Fisher.
- Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: A markov chain sampling approach, *Journal of Econometrics* 95: 57–69.

Abstract

In this paper the maximum likelihood (ML) parameter estimation method is compared with a new bayesian technique that uses Markov chain Monte Carlo (MCMC) simulation. Several time series representing autoregressive conditional heteroskedastic (ARCH) processes are independently generated for different model orders ($p = 2, 5$ and 7). For $p = 2$, we evaluate the ability of both techniques to handle models with finite as well infinite conditional variance. For $p = 5$ and 7 we assess the influence of the model order on the spread of the confidence (or credible) intervals. Accuracy of the estimates is evaluated through RMSE and MAPE values computed between the estimates and the known coefficients of the simulated ARCH processes. The bayesian approach has shown to be more accurate, even for higher model orders, and is more robust than the ML method, since the former is equally able to estimate the parameters of models with finite and infinite variance. Furthermore, unlike the bayesian method, the confidence intervals computed for the ML approach violate positivity constraints imposed to the parameters of the ARCH models.

Agradecimentos

M. G. Andrade agradece à FAPESP (Processo 02/01186- 6) e G. A. Barreto agradece ao programa CAPES/PRODOC pelo apoio financeiro.

Inferência em modelos heteroscedásticos na presença de pontos de alavanca

Francisco Cribari-Neto*
Tatiene C. Souza**
Klaus L.P. Vanconcellos*

Resumo

Este artigo considera a estimação consistente da matriz de covariâncias de mínimos quadrados ordinários em um modelo de regressão linear sob heteroscedasticidade de forma desconhecida. O estimador mais usado é aquele proposto por Halbert White conhecido como HC0. Consideramos também outros estimadores consistentes: HC3, HC4, o estimador *bootstrap* ponderado proposto por Wu (1986) e o estimador *bootstrap* ponderado inversamente ajustado proposto por Cribari-Neto & Zarkos (2004). Nossas avaliações numéricas (Monte Carlo) mostram que os testes quasi-*t* baseados nos estimadores HC4 e *invwu* apresentam as menores distorções de tamanho. Uma aplicação empírica também é apresentada.

Palavras-chave: Estimação da matriz de covariâncias, heteroscedasticidade, pontos de alta alavancagem, regressão linear.

Endereço para correspondência: * Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, CEP 50740-540.

** Departamento de Estatística, Universidade Federal da Bahia, Campus Universitário de Ondina, Salvador/BA, 40170-110.

1. Introdução

A análise de regressão é um dos ramos mais utilizados da teoria estatística na pesquisa científica. O modelo clássico de regressão teve origem nos trabalhos de astronomia elaborados por Gauss no período de 1809 a 1821; trata-se de uma técnica adequada para estudar o efeito que variáveis explicativas exercem sobre uma variável resposta. Técnicas clássicas de regressão linear assumem que os erros do modelo têm variância constante, ou seja, apresentam a propriedade de *homoscedasticidade*. Porém, esta suposição é bastante forte e, em uma relevante parte dos problemas práticos, muito pouco razoável.

Sob suposições feitas para este modelo, a estimação dos parâmetros é comumente feita utilizando-se o método de mínimos quadrados ordinários (*MQO*). Este método, através de operações matriciais de fácil implementação computacional, fornece estimadores de simples interpretação e que possuem propriedades desejáveis, como não-viés, consistência, eficiência e normalidade assintótica. Quando a suposição de homoscedasticidade é violada, ou seja, quando a variância dos erros não é constante, dizemos que há *heteroscedasticidade* no modelo. A presença de heteroscedasticidade ocorre com frequência quando se trabalha com dados de corte transversal.

Sob heteroscedasticidade, o estimador de *MQO* mantém algumas propriedades desejáveis, tornando-se, contudo, ineficiente na classe dos estimadores lineares não-viesados, isto é, ele não é mais o melhor estimador linear não-viesado (*BLUE – Best Linear Unbiased Estimator*). O estimador usual de sua matriz de covariâncias torna-se viesado e não-consistente, tornando pouco confiáveis estimativas intervalares e testes de hipóteses nele baseados. Uma maneira de contornar este problema é obter estimadores alternativos ao estimador de *MQO*. Torna-se razoavelmente fácil estimar os parâmetros de um modelo de regressão quando os erros são heteroscedásticos com um padrão de heteroscedasticidade que é determinado por uma função cedástica conhecida, ou mesmo quando os parâmetros da função cedástica são desconhecidos, mas com forma de heteroscedasticidade conhecida. Contudo, quase sempre o processo gerador das diferentes variâncias é desconhecido.

Um procedimento de estimação muito usado na prática para o modelo de regressão linear na presença de heteroscedasticidade de forma desconhecida consiste em utilizar para o vetor de parâmetros β o estimador de mínimos quadrados ordinários, que permanece não-viesado e consistente, juntamente com um estimador para sua matriz de covariâncias que possua propriedades assintóticas desejáveis. Este estimador consistente da matriz de covariâncias do estimador de *MQO* possibilita a realização de inferências no modelo

heteroscedástico sem que seja necessário fazer suposições sobre a forma da heteroscedasticidade.

O objetivo do presente artigo é abordar a estimação consistente da matriz de covariâncias do estimador de mínimos quadrados ordinários sob heteroscedasticidade de forma desconhecida, avaliando, por meio de simulações de Monte Carlo, os desempenhos de estimadores alternativos propostos por diversos autores, inclusive estimadores *bootstrap*. Em particular, dois estimadores propostos recentemente na literatura serão avaliados de forma comparativa.

2. Modelo e estimadores

O modelo de interesse é o modelo de regressão linear, definido como

$$y = X\beta + e, \quad (1)$$

onde y é um vetor $n \times 1$ de observações de uma variável dependente, X é uma matriz $n \times p$ ($p < n$) de regressores fixos com $\text{posto}(X) = p$, β é um vetor $p \times 1$ de parâmetros desconhecidos e e é um vetor $n \times 1$ de erros aleatórios não-correlacionados, cada um tendo média zero e variância σ_i^2 , para $i = 1, \dots, n$ a matriz de covariâncias de e é dada por $\Psi = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$. O estimador de mínimos quadrados ordinários do vetor de parâmetros β é dado por $b = (X'X)^{-1}X'y$, possuindo média β (isto é, é não-viesado) e estrutura de variância $(X'X)^{-1}X'\Psi X(X'X)^{-1}$.

Sob homoscedasticidade, temos que $\sigma_i^2 = \sigma^2 > 0$, isto é, $\Psi = \sigma^2 I_n$, onde I_n é a matriz identidade de ordem n . Então, a matriz de covariâncias de b é dada por $\sigma^2 (X'X)^{-1}$ e pode ser facilmente estimada por $\hat{\sigma}^2 (X'X)^{-1}$, onde $\hat{\sigma}^2 = e'e/(n-p)$, $e = [I - X(X'X)^{-1}X']y$ sendo o vetor n -dimensional de resíduos de mínimos quadrados.

Na presença de heteroscedasticidade, o estimador $\hat{\sigma}^2 (X'X)^{-1}$ não é mais consistente nem não-viesado para a matriz de covariâncias de b . Como em um grande número de aplicações os erros são heteroscedásticos, é importante considerar estimadores mais confiáveis para a variância de b . Um estimador consistente da matriz de covariâncias de b muito utilizado foi proposto por Halbert White (1980), sendo aqui denotado por HC0. O estimador de White é obtido estimando-se o i -ésimo elemento diagonal de Ψ pelo i -ésimo resíduo ao quadrado, resultando em

$$HC0 = (X'X)^{-1} X' \hat{\Psi}_0 X (X'X)^{-1},$$

em que $\hat{\Psi}_0 = \text{diag}\{\hat{e}_1^2, \dots, \hat{e}_n^2\}$.

Apesar de ser um estimador consistente da matriz de covariâncias de b tanto sob heteroscedasticidade como sob homoscedasticidade, o estimador HC0 tende a subestimar, em amostras finitas, a variância do estimador de MQO, sendo tipicamente bastante viesado quando o tamanho da amostra é pequeno. Através de simulações de Monte Carlo, Long & Ervin (2000) constataram que este estimador não é confiável quando o tamanho da amostra é menor ou igual a 250. Adicionalmente, Cribari-Neto & Zarkos (1999) e MacKinnon & White (1985), em estudos de simulação, mostraram que o estimador HC0 pode ser muito viesado quando o tamanho da amostra não é grande.

Assim como o tamanho amostral, o desenho da regressão, ou seja, a estrutura da matriz de regressores X , é outro fator que também interfere no desempenho do estimador de White em amostras finitas. Quando a matriz de regressores X contém observações de alta alavancagem, isto é, observações que possuem potencial para exercer grande influência sobre o valor a ser predito, o estimador HC0 tende a ser muito viesado; ver Chesher & Jewitt (1987). Dessa maneira, o uso de estimativas obtidas a partir de HC0 pode levar a inferências errôneas.

Há na literatura várias regras para identificação de pontos de alavanca. Uma regra muito usada é considerar pontos de alavanca àquelas observações cujo respectivo h_i (isto é, i -ésimo elemento diagonal da matriz chapéu H) excede $2p/n$ ou $3p/n$, ou seja, duas ou três vezes a médias de todos os h_i 's; ver Judge *et al.* (1988, p.893).

MacKinnon & White (1985) consideraram três estimadores alternativos, dois dos quais obtidos como modificações de HC0. Um ajustamento simples, sugerido por Hinkley (1977), consiste em usar uma correção de graus de liberdade, multiplicando HC0 por $n/(n-p)$. Com esta correção obtém-se a versão conhecida como HC1, dada por

$$HC1 = \frac{n}{n-p} (X'X)^{-1} X' \hat{\Psi}_0 X (X'X)^{-1}.$$

O ajustamento pelos graus de liberdade em HC1 não é o único caminho para compensar o fato de que os resíduos de mínimos quadrados ordinários tendem a subestimar os erros verdadeiros. Se não há heteroscedasticidade, temos que

$$\begin{aligned}
E[\hat{e}\hat{e}'] &= E[(y - \hat{y})(y - \hat{y})'] \\
&= E\{(I - H)y[(I - H)y]'\} \\
&= E[(I - H)yy'(I - H)'] \\
&= ME[yy']M' \\
&= \sigma^2 M,
\end{aligned}$$

em que $H = X(X'X)^{-1}X'$ e $M = (I - H)$, usando o fato de que a matriz M é simétrica e idempotente. Assim,

$$\begin{aligned}
E[\hat{e}_i^2] &= \sigma^2 m_i \\
&= \sigma^2 (1 - h_i),
\end{aligned}$$

onde m_i e h_i são os i -ésimos elementos diagonais de M e H , respectivamente. Neste sentido, Horn, Horn & Duncan (1975) sugeriram usar

$$\hat{\sigma}_i^2 = \hat{e}_i^2 / (1 - h_i).$$

Obtém-se, assim, o estimador referido como HC2 e definido da seguinte forma:

$$HC2 = (X'X)^{-1} X' \hat{\Psi}_2 X (X'X)^{-1},$$

onde

$$\hat{\Psi}_2 = \text{diag} \left\{ \frac{\hat{e}_1^2}{1 - h_1}, \dots, \frac{\hat{e}_n^2}{1 - h_n} \right\}.$$

Uma terceira variante do estimador HC0 é baseada na técnica denominada *jackknife*¹. A idéia básica desta técnica consiste em recalcular n vezes as estimativas de mínimos quadrados ordinários para o vetor β , cada vez retirando uma das observações, e então usar a variabilidade das estimativas obtidas como estimativa da variância do estimador de MQO original. Seja $b_{(i)}$ o estimador de mínimos quadrados ordinários de β baseado em todas as observações exceto a i -ésima. É possível mostrar que

$$b_{(i)} = b - (X'X)^{-1} X_i' \hat{e}_i,$$

¹ Para detalhes sobre o método *jackknife*, ver Efron (1982).

onde X_i denota a i -ésima linha de X e $\hat{\varepsilon}_i = \hat{e}_i / (1 - h_i)$. Dessa forma, o estimador *jackknife* da matriz de covariâncias de b é dado por

$$\frac{n-1}{n} \sum_{i=1}^n \left[b_{(i)} - n^{-1} \sum_{s=1}^n b_{(s)} \right] \left[b_{(i)} - n^{-1} \sum_{s=1}^n b_{(s)} \right]'$$

Após algumas manipulações, esta quantidade pode ser escrita como

$$\frac{n-1}{n} (X'X)^{-1} [X' \Psi^* X - (n^{-1})(X'e^*e^*X)] (X'X)^{-1},$$

onde Ψ^* é uma matriz diagonal $n \times n$ cujo i -ésimo elemento diagonal é $\hat{\varepsilon}_i^2$.

Davidson & MacKinnon (1993, pp. 553-554) definem um estimador, denotado por HC3, que também é uma modificação do estimador de White, cujo comportamento é similar ao do estimador *jackknife*, sendo uma aproximação simples deste estimador no sentido que os dois estimadores apresentam comportamentos muito similares em amostras finitas. O estimador HC3 é dado por

$$HC3 = (X'X)^{-1} X' \hat{\Psi}_3 X (X'X)^{-1},$$

onde

$$\hat{\Psi}_3 = \text{diag} \left\{ \frac{\hat{\varepsilon}_1^2}{(1-h_1)^2}, \dots, \frac{\hat{\varepsilon}_n^2}{(1-h_n)^2} \right\}.$$

Cribari-Neto (2004) propõe uma modificação do estimador HC3, denominada HC4, que leva em consideração o impacto de observações com alta alavancagem em amostras finitas, incorporando fatores de descontos definidos pela razão entre os graus individuais de alavancagem e o grau médio de alavancagem. A importância de considerar o efeito da alta alavancagem de algumas observações pode ser verificada através dos resultados obtidos em Cribari-Neto & Zarkos (2001), que mostraram que a presença de pontos de alavanca na estrutura da matriz dos regressores é mais decisiva para o comportamento de estimadores consistentes para a matriz de covariâncias de b em amostras finitas, isto é, amostras pequenas ou de tamanho moderado, do que o próprio grau de heteroscedasticidade.

O estimador HC4 é definido da seguinte forma:

$$HC4 = (X'X)^{-1} X' \hat{\Psi}_4 X (X'X)^{-1},$$

onde

$$\hat{\Psi}_4 = \text{diag} \left\{ \frac{\hat{e}_1^2}{(1-h_1)^{\delta_1}}, \dots, \frac{\hat{e}_n^2}{(1-h_n)^{\delta_n}} \right\}$$

e

$$\delta_i = \min \left\{ 4, \frac{h_i}{\bar{h}} \right\} = \min \left\{ 4, \frac{nh_i}{\sum_{j=1}^n h_j} \right\},$$

com $\bar{h} = n^{-1} \sum_{j=1}^n h_j$, i.e., \bar{h} é a média dos h_i 's. Ou seja,

$$\delta_i = \min \left\{ 4, \frac{nh_i}{p} \right\}.$$

Aqui, usamos o fato de que a soma de todos os níveis individuais de alavancagem é igual a p .

O expoente controla o nível de desconto do i -ésimo resíduo quadrado e é determinado pela razão entre h_i e a média dos h_i 's, \bar{h} . Como $0 < 1-h_i < 1$ e $\delta_i > 0$, segue que $0 < (1-h_i)^{\delta_i} < 1$. O i -ésimo resíduo ao quadrado deverá ser tanto mais fortemente inflacionado quanto maior for h_i relativamente a \bar{h} . Este desconto linear é truncado em 4, que equivale a duas vezes o grau de desconto usado pelo estimador HC3. Ou seja, usa-se $\delta_i = 4$ quando $h_i > 4\bar{h} = 4p/n$. Os resultados numéricos obtidos por Cribari-Neto (2004) mostram que este novo estimador tem comportamento superior ao do estimador HC3 em amostras finitas no que diz respeito a testes quasi- t associados.

3. *Bootstrap* em modelos de regressão

Bootstrap é um método que pode ser usado para avaliar a precisão de estimativas estatísticas com base em simulações. O método *bootstrap* tipicamente produz uma aproximação para a distribuição da estatística de interesse que pode ser consideravelmente

mais precisa do que sua aproximação assintótica de primeira ordem; ver, em geral, Hall (1992). A aplicação de *bootstrap* em modelos de regressão foi estudada em detalhes por Wu (1986). Mais recentemente, vários autores investigaram o uso deste método em econometria, entre eles Horowitz (1997), Jeong e Maddala (1993), Li e Maddala (1996) e Vinod (1993). Em sua forma mais simples, o procedimento *bootstrap* para o modelo de regressão linear $y = X\beta + e$ pode ser descrito como a seguir:

1. Retire uma amostra aleatória e_1^*, \dots, e_n^* de \hat{e} com reposição;
2. Forme uma amostra *bootstrap*: $y^* = X\beta + e^*$, onde $e^* = (e_1^*, \dots, e_n^*)'$;
3. Calcule o estimador de MQO, $b^* = (X'X)^{-1} X'y^*$;
4. Repita os passos de 1 a 3 um grande número de vezes (digamos, B);
5. Calcule a variância dos B vetores das estimativas obtidas usando o esquema descrito nos passos de 1 a 4.

É comum multiplicar os resíduos no passo 1 pelo fator $\sqrt{n/(n-p)}$. O método *bootstrap* descrito acima é geralmente conhecido como *bootstrap* não-paramétrico não ponderado, *bootstrap* ingênuo não-paramétrico ou *bootstrap* naïve. O *bootstrap* paramétrico não-ponderado é semelhante ao descrito acima, mas com o passo 1 substituído por:

- 1.a Retire uma amostra aleatória e_1^*, \dots, e_n^* de uma distribuição normal com média zero e variância $\hat{\sigma}^2$.

Nenhum dos esquemas *bootstrap* descritos acima leva em consideração a possibilidade de haver heteroscedasticidade. De fato, eles não são nem consistentes nem assintoticamente não-viesados sob heteroscedasticidade; ver Shao (1988) e Wu (1986). Um estimador robusto à presença de heteroscedasticidade foi proposto por Wu (1986) e pode ser descrito como a seguir:

- 1.b Para cada i , $i=1, \dots, n$, retire t_i^* de uma população com média zero e variância unitária.
- 2.b Forme a amostra *bootstrap* (y^*, X) , onde $y_i^* = X_i b + t_i^* \hat{e}_i / \sqrt{1-h_i}$, sendo h_i o i -ésimo elemento diagonal da 'matriz chapéu' $X(X'X)^{-1}X'$ e X_i , como antes, sendo a i -ésima linha de X .

Os passos de 3 a 5 permanecem os mesmos. O mecanismo deste esquema de *bootstrap* consiste do seguinte: obtém-se um conjunto de n seleções independentes de uma distribuição com média zero e variância unitária; multiplicam-se estes valores pelos resíduos normalizados; forma-se uma amostra *bootstrap*; estima-se β usando a nova amostra; repete-se o procedimento um grande número de vezes (digamos, B); então, usam-se as estimativas das B réplicas *bootstrap* para se obter uma estimativa para $\text{var}(b)$. Note que a variância de $t_i^* \hat{e}_i / \sqrt{(1-h_i)}$ não é constante quando os erros originais não são homoscedásticos. Portanto, o esquema de *bootstrap* de Wu leva em consideração a não-constância potencial da variância do erro. Este esquema de reamostragem é freqüentemente chamado de *ponderado* ou *externo*.

Há diferentes maneiras de reamostrar t^* . Uma possibilidade é reamostrar de a_1, \dots, a_n , onde

$$a_i = \frac{\hat{e}_i - \bar{\hat{e}}_i}{\sqrt{n^{-1} \sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}}_i)^2}}, \quad (3.1)$$

com $\bar{\hat{e}}_i = n^{-1} \sum_{i=1}^n \hat{e}_i$; quando o modelo de regressão especificado contém um intercepto sabemos que $\bar{\hat{e}}_i = 0$. Esta é uma implementação não-paramétrica do esquema *bootstrap* de Wu, uma vez que a reamostragem é feita a partir da função de distribuição empírica dos resíduos normalizados. Em uma implementação paramétrica do *bootstrap* de Wu, os t_i^* 's podem ser selecionados aleatoriamente de uma distribuição específica, em geral, a distribuição normal padrão.

Uma alternativa diferente é o *bootstrap* ponderado inversamente ajustado. A idéia é amostrar t_i^* no esquema de *bootstrap* ponderado não com probabilidade $1/n$, como no esquema de Wu, mas com probabilidade que é inversamente proporcional a h_i , como proposto por Cribari-Neto & Zarkos (2004). Seja

$$\delta = \sum_{i=1}^n 1/h_i.$$

Então, o esquema pode ser descrito da seguinte forma:

1.c Para cada i , $i = 1, \dots, n$, retire t_i^* , de forma independente, de a_1, \dots, a_n , ou seja, de (3.1). Aqui, a_i é selecionado com probabilidade p_i , $i = 1, \dots, n$; ver (3.2) abaixo.

2.c Forme a amostra *bootstrap* (y^*, X) , onde $y_i^* = X_i b + t_i^* \hat{e}_i / \sqrt{(1-h_i)^{f_i}}$, $i = 1, \dots, n$, onde

$$f_i = \begin{cases} 1, & \text{se } h_i \leq 2p/n \\ 3, & \text{se } h_i > 2p/n \end{cases}$$

Seja $p_i = \gamma / h_i$. Então, $\sum_{i=1}^n p_i = 1$ implica

$$\sum_{i=1}^n \gamma h^{-1} = \gamma \sum_{i=1}^n h^{-1} = \gamma \delta = 1.$$

Logo,

$$p_i = \frac{1}{\delta h_i}.$$

A idéia é que observações com alta alavancagem serão selecionadas no esquema de reamostragem com probabilidade mais baixa, mas mesmo assim positiva, desta forma reduzindo seu impacto sobre a inferência resultante. Isto é, quanto mais alto for o grau de alavancagem de uma observação, com menos frequência, seu correspondente resíduo será selecionado na reamostragem *bootstrap*. Aqui, as probabilidades de reamostragem são controladas pelos graus de alavancagem das diferentes observações.

4. Avaliação numérica

O experimento de Monte Carlo de nossa avaliação numérica é baseado no modelo de regressão linear

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \quad i = 1, \dots, n$$

O conjunto de dados utilizado consiste de observações sobre despesas *per capita* em escolas públicas e renda *per capita* por estado nos Estados Unidos em 1979, estes dados tendo sido extraídos de Greene (1997). As covariáveis, x_1 e x_2 , são a renda *per capita*

(reescalada por 10^{-4}) e seu quadrado, respectivamente, com um total de 50 observações. Os valores utilizados para os erros são provenientes de simulações estocásticas, considerando-se diferentes graus de heteroscedasticidade. Cada uma das observações de x_1 e x_2 é repetida duas e três vezes para formar amostras de tamanhos $n = 100, 150$, que são mantidas constantes ao longo do experimento. Isto garante que $\lambda = \max(\delta_i^2) / \min(\delta_i^2)$ uma medida do grau de heteroscedasticidade mantém-se constante para os diferentes tamanhos amostrais, e que os resultados da simulação são afetados apenas pelo crescente número de observações. No que se refere aos valores dos parâmetros, nós utilizamos $\beta_0 = \beta_1 = 1$ e $\beta_2 = 0$. Os experimentos de simulação são baseados em 5 000 réplicas de Monte Carlo e 500 réplicas de *bootstrap* (isto é, um total de 2 500 000 de réplicas por experimento).

Utilizaremos dois conjuntos de dados; o primeiro é o conjunto de dados retirado de Greene (1997), citado anteriormente e o segundo é obtido do primeiro, retirando-se os pontos de alta alavancagem. A função cedástica foi especificada como

$$\sigma_i^2 = \exp\{c_1 + c_2 x_{i1}\}.$$

Simulações sob homoscedasticidade foram realizadas com $c_1 = c_2 = 0$, ao passo que para dados heteroscedásticos consideramos dois valores para λ ; primeiro usamos $c_1 = c_2 = 5, 30$, que resulta em $\lambda = 15, 04$, depois usamos $c_1 = c_2 = 0, 93$, que resulta em $\lambda = 50, 05$ para o primeiro conjunto de dados ('dados1'). Para o segundo conjunto de dados ('dados2'), também consideramos dois valores para λ : utilizamos $c_1 = 0$ e $c_2 = 7, 65$, que produz $\lambda = 15, 02$, e $c_1 = 0$ e $c_2 = 13, 45$, que resulta em $\lambda = 50, 30$. Os valores das constantes que indexam a função cedástica foram escolhidos de forma a fornecer graus de heteroscedasticidade próximos de 15 e 50, o último representando heteroscedasticidade muito forte (o maior desvio padrão sendo pouco mais que vezes maior que o menor desvio padrão), além, é claro, da situação onde não há heteroscedasticidade. Dessa forma, diferentes cenários são considerados na avaliação. Os desempenhos dos vários estimadores da matriz de covariâncias considerados no experimento de Monte Carlo foram avaliados com base no tamanho do teste quasi-*t* associado para a hipótese nula $H_0 : \beta_2 = 0$ contra $H_1 : \beta_2 \neq 0$.

Seis experimentos de Monte Carlo, portanto, foram realizados, dois sob homoscedasticidade e quatro sob heteroscedasticidade. Como mencionado anteriormente, o conjunto de dados originalmente usado é uma amostra de tamanho 50, retirada de Greene (1997); isto é o que chamamos de 'dados1'. O segundo conjunto de dados é obtido do primeiro retirando

três observações identificadas como pontos de alta alavancagem (seus h_i 's excedem $3p/n$); estas observações correspondem aos estados do Alasca e Mississipi e a Washington, D.C. Em particular, o Alasca é um ponto de alavancagem muito alta, pois $h_{\text{Alasca}} = 0,651$ ao passo que $3p/n = 0,180$. (Tipicamente, usa-se $3p/n$ como um limiar crítico na identificação de pontos de alavanca).

Os resultados relativos às taxas de rejeição dos diferentes testes quasi- t , aos níveis nominais de 10%, 5% e 1%, sob homoscedasticidade e sob heteroscedasticidade, para o conjunto de dados originais de Greene ('dados1'), encontram-se apresentados no Quadro 1. O Quadro 2 apresenta resultados para o conjunto de dados sem alavancagem ('dados2'). Os principais resultados encontram-se resumidos a seguir.

Primeiro, o teste que usa o estimador MQO apresenta desempenho satisfatório apenas sob homoscedasticidade. Sob heteroscedasticidade, os percentuais de rejeição para o teste não se aproximam do nível assintótico considerado; por exemplo, para $n = 150$ e ao nível assintótico de 5%, este teste apresenta rejeição de 4,76% quando $\lambda = 1,00$; na presença de forte heteroscedasticidade ($\lambda = 50,05$) e de pontos influentes, o mesmo teste apresenta rejeição de 34,72%, como revela o Quadro 1.

Segundo, o teste construído com base no estimador HC0 não apresenta comportamento confiável tanto sob homoscedasticidade quanto sob heteroscedasticidade; suas taxas de rejeição são consideravelmente superiores àquelas verificadas para os testes associados aos demais estimadores. O teste que usa o estimador de Wu apresenta taxas de rejeição superiores às esperadas, porém inferiores às taxas de rejeição apresentadas pelo teste que usa o estimador de White, como mostra o Quadro 1.

Terceiro, o teste baseado no estimador HC3 apresenta taxas de rejeição inferiores às do teste baseado no estimador HC0, sendo estas taxas, contudo, superiores aos níveis nominais selecionados (Quadro 1).

Quarto, quando pontos de alta alavancagem são introduzidos nos dados, as distorções de tamanho dos testes baseados nos diversos estimadores crescem substancialmente. A presença destes pontos exerce um efeito negativo sobre os desempenhos de quase todos estimadores avaliados quando estes são utilizados conjuntamente com testes quasi- t , elevando as taxas de rejeição. Por exemplo, o teste baseado no estimador HC4, aos níveis nominais considerados (10%, 5% e 1%), para $n = 50$ e $\lambda = 50,05$, apresenta taxas de rejeição 19,52%, 14,50% e 8,50%, enquanto, sem considerar as observações influentes ($n = 47$) e com $\lambda = 50,30$, temos taxas de rejeição mais próximas dos níveis nominais que as anteriores: 14,04%, 8,58%, 3,08%, respectivamente, como revela o Quadro 2.

Quinto, o teste baseado no estimador *invwu* apresenta taxas de rejeição relativamente próximas aos níveis nominais tanto na presença quanto na ausência das observações influentes. Por exemplo, para $n = 100$, $\lambda = 15,04$ e ao nível de 10%, a taxa de rejeição do teste baseado no estimador *invwu* é de 11,86%; sem considerar os pontos de alavanca (Quadro 2) ao mesmo nível nominal, a taxa de rejeição é 11,68%.

Sexto, o teste que usa o estimador HC4 apresenta o melhor desempenho na ausência das observações influentes dentre todos os testes considerados. A título de exemplo do comportamento desse teste, aos níveis nominais de 10%, 5% e 1%, para $n = 100$ (com pontos de alavanca - Quadro 1) e $\lambda = 15,04$, suas taxas de rejeição são 15,42%, 10,34% e 4,62%; sem considerar as observações influentes ($n = 94$) e com $\lambda = 15,02$, as correspondentes taxas são: 11,56%, 6,22%, 1,64%. Nesse último caso, as distorções de tamanho são relativamente pequenas, como pode ser visto no Quadro 2.

Sétimo, o teste baseado no estimador *naïve* apresenta comportamento satisfatório sob homoscedasticidade, o mesmo não acontecendo sob heteroscedasticidade. Por exemplo, considerando $n = 100$ (Quadro 1) e $\lambda = 1$, ao nível de 5%, o teste baseado no estimador *naïve* apresenta taxa de rejeição igual a 5,40%; sob heteroscedasticidade forte, $\lambda = 50,05$, e ao mesmo nível de significância, o teste baseado neste estimador apresenta taxa de rejeição de 35,58%.

Oitavo, na presença dos pontos de alavanca e sob heteroscedasticidade, o teste baseado no estimador *invwu* apresenta taxas de rejeição mais próximas dos níveis nominais do que o teste baseado no estimador HC4. Por exemplo, para $n = 50$, $\lambda = 15,04$ e ao nível nominal de 5%, o teste baseado no estimador *invwu* apresenta taxa de rejeição de 6,94%, enquanto o teste baseado no estimador HC4 apresenta quase o dobro desta taxa de rejeição (12,52%), como pode ser visto no Quadro 1.

Nono, sob homoscedasticidade e para os três tamanhos amostrais estudados, o teste baseado no estimador HC4 teve melhor desempenho que o teste baseado no estimador *invwu*, exceto para $\alpha = 1\%$ e para $n = 100$, ao nível nominal de 5%.

Décimo, de acordo com o Quadro 2, na ausência dos pontos de alavanca ($n = 47$), sob homoscedasticidade e heteroscedasticidade, o teste baseado no estimador *invwu* teve melhor desempenho que o teste baseado no estimador HC4 para todos os níveis nominais.

Por exemplo, $n = 47$ e $\lambda = 15,02$, ao nível de 10%, o teste baseado no estimador HC4 apresenta taxa de rejeição de 12,86%, enquanto o teste baseado no estimador *invwu* apresenta taxa de rejeição de 11,22%. O mesmo não acontece quando aumentamos o

tamanho amostral ($n = 94, 141$), ou seja, o teste baseado no estimador HC4 tem melhor desempenho que o teste baseado no estimador *invwu*, em quase todos os casos.

Com o propósito de ilustrar algumas conclusões anteriores, foram construídos gráficos das diferenças entre quantis exatos de algumas estatísticas de teste, estimados por simulação, e quantis assintóticos para $n = 50, 100$, sob homoscedasticidade e sob heteroscedasticidade.

As discrepâncias de quantis são apresentadas graficamente contra quantis assintóticos. Estes gráficos encontram-se na Figura 1. Quanto mais próximas as linhas estiverem da linha horizontal (zero), melhor será a aproximação pela distribuição assintótica $N(0, 1)$ da distribuição da estatística de teste sob a hipótese nula. A análise dessas figuras evidencia que os testes baseados nos estimadores HC4 e *invwu* se comportam melhor do que o teste baseado no estimador HC3. De fato, nota-se que suas distribuições exatas, em geral, são melhor aproximadas pela distribuição limite do que as distribuições exatas das demais estatísticas.

Um enfoque alternativo ao que foi apresentado anteriormente é usar um erro-padrão consistente para construir estatísticas quasi- t e então reamostrar esta quantidade, que é assintoticamente pivotal. O esquema de reamostragem é realizado impondo a restrição em teste na geração das amostras de *bootstrap*, a estatística de teste sendo calculada em cada réplica de *bootstrap*. No final do esquema de reamostragem de *bootstrap*, obtemos ou um valor crítico para o teste (a ser usado em substituição ao valor crítico assintótico obtido da distribuição normal padrão) ou um p -valor *bootstrap*. O teste *bootstrap* pode ser realizado como descrito a seguir. De início, calcule a estatística quasi- t , digamos τ .

Então:

1. Para cada $i, i = 1, \dots, n$, retire um número aleatório t_i^* de uma população que tem média zero e variância um;
2. Construa uma amostra *bootstrap* (y^*, X) , onde $y_i^* = X_i \tilde{b} + t_i^* \tilde{e}_i / (1 - h_i)$. Aqui, \tilde{b} e \tilde{e}_i são as estimativas restritas dos parâmetros e os resíduos de MQO sob a hipótese nula. Novamente, X_i denota a i -ésima linha de X ;
3. Calcule o EMQO de β , $b^* = (X'X)^{-1} X'y^*$, e calcule a estatística de teste quasi- t associada, τ^* ;
4. Repita os passos 1 a 3 um grande número de vezes (digamos, B);
5. Calcule o quantil de interesse da distribuição empírica de $(B + 1)$ realizações da estatística de teste;

6. Realize o teste usando a estatística *quasi-t* calculada inicialmente (τ) conjuntamente com o valor crítico *bootstrap* obtido no passo 5 acima.

Observe que no teste *bootstrap* não utilizamos valores críticos obtidos da distribuição assintótica da estatística de teste sob a hipótese nula, isto é, não são utilizados valores críticos da distribuição normal padrão. Aqui são usados valores críticos obtidos através da reamostragem *bootstrap*.

A regra de decisão pode ser mais convenientemente expressa usando o *p*-valor *bootstrap*. A aproximação do *p*-valor obtida a partir do esquema de *bootstrap*, para um teste bilateral, é dada por

$$p = \frac{1 + \#\{\tau_b^* \geq |\tau|\}}{B + 1},$$

onde τ_b^* , $b = 1, \dots, B$, são as realizações de *bootstrap* da estatística de teste. Rejeitamos a hipótese nula se o *p*-valor *bootstrap* for menor do que o tamanho nominal selecionado para o teste.

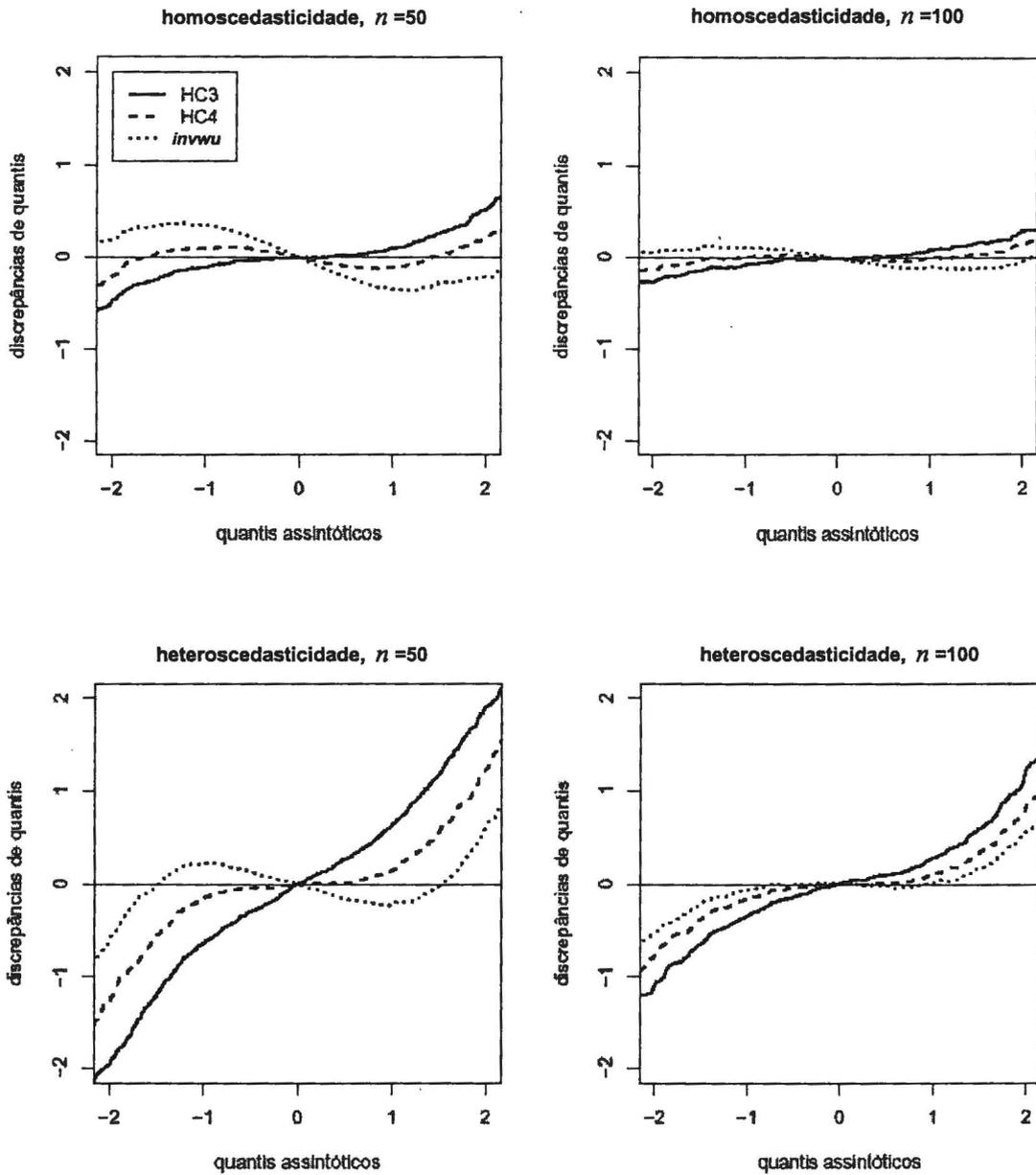
Quadro 1: Percentagens de rejeição de testes quasi-f baseados nos estimadores MQO, *white*, HC3, HC4, *naïve*, *wu* e *invwu* no modelo $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$, $e_i \sim (0, \exp\{c_1 + c_2 x_i\})$, $i = 1, \dots, n$, para $n = 50, 100, 150$. Os níveis nominais considerados são $\alpha = 0, 10, 0, 05, 0, 01$ (com pontos de alavanca) - 'dados1'

n = 50									
λ	1,00			15,04			50,05		
teste	10%	5%	1%	10%	5%	1%	10%	5%	1%
MQO	10,64	6,16	1,42	33,34	24,80	13,22	45,20	37,32	24,20
white	20,70	14,04	6,26	37,60	29,64	18,10	44,66	36,56	24,22
HC3	15,34	9,74	4,20	26,60	19,94	11,24	31,64	24,18	14,66
HC4	10,68	6,48	2,72	17,18	12,52	6,80	19,52	14,50	8,50
naïve	10,70	6,12	1,48	33,38	29,94	13,46	45,04	37,54	24,30
wu	15,28	9,98	4,20	26,76	20,04	11,48	31,84	24,14	14,68
invwu	5,86	3,40	1,32	10,32	6,94	3,34	11,42	8,16	4,66
n = 100									
λ	1,00			15,04			50,05		
teste	10%	5%	1%	10%	5%	1%	10%	5%	1%
MQO	10,62	5,28	1,04	32,22	24,10	12,44	43,70	35,86	22,52
white	16,12	9,92	3,36	24,70	17,52	8,98	27,92	20,54	11,22
HC3	13,50	7,84	2,40	19,62	13,58	6,58	22,18	15,62	8,30
HC4	11,08	6,10	1,92	15,42	10,34	4,62	16,80	11,64	5,68
naïve	10,70	5,40	1,12	32,36	24,38	12,60	44,02	35,58	22,80
wu	13,66	7,86	2,54	19,82	13,74	6,48	22,36	15,88	8,32
invwu	8,22	4,48	1,24	11,86	7,96	3,26	14,16	9,48	4,40
n = 150									
λ	1,00			15,04			50,05		
teste	10%	5%	1%	10%	5%	1%	10%	5%	1%
MQO	10,28	4,76	0,98	32,04	23,80	11,64	43,16	34,72	22,34
white	14,28	8,28	2,18	19,70	13,06	6,04	21,08	14,66	7,24
HC3	12,26	6,72	1,86	16,50	10,72	4,62	17,42	12,12	5,74
HC4	10,80	5,58	1,66	13,52	8,62	3,56	14,40	9,66	4,40
naïve	10,44	4,82	1,00	32,00	23,88	12,10	43,46	34,82	22,50
wu	12,50	6,84	1,78	16,50	10,72	4,62	17,72	12,08	5,84
invwu	8,56	4,22	1,10	11,92	7,30	2,86	13,74	8,76	3,78

Quadro 2: Percentagens de rejeição de testes quasi-t baseados nos estimadores MQO, *white*, HC3, HC4, *naïve*, *wu* e *invwu* no modelo $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$, $e_i \sim (0, \exp\{c_1 + c_2 x_i\})$, $i = 1, \dots, n$, para $n = 47, 94, 141$. Os níveis nominais considerados são $\alpha = 0, 10, 0, 05, 0, 01$ (sem pontos de alavanca) - 'dados2'

<i>n</i> = 47									
λ	1,00			15,02			50,30		
teste	10%	5%	1%	10%	5%	1%	10%	5%	1%
MQO	10,92	5,76	1,34	17,60	10,12	3,38	21,44	13,48	5,40
<i>white</i>	13,30	8,00	2,16	16,04	9,48	3,38	17,30	10,98	4,12
HC3	11,64	6,54	1,70	13,42	7,78	2,64	14,64	8,96	3,18
HC4	11,52	6,48	1,76	12,86	7,52	2,64	14,04	8,58	3,08
<i>naïve</i>	10,94	5,82	1,36	17,54	10,56	3,34	21,58	13,88	5,36
<i>wu</i>	11,80	6,88	1,70	13,74	8,16	2,92	14,92	9,28	3,40
<i>invwu</i>	9,42	4,84	1,08	11,22	6,08	1,92	12,86	7,52	2,54
<i>n</i> = 94									
λ	1,00			15,02			50,30		
teste	10%	5%	1%	10%	5%	1%	10%	5%	1%
MQO	9,90	5,08	0,96	16,88	10,10	2,84	20,52	13,34	4,52
<i>white</i>	11,40	6,32	1,44	12,90	6,22	1,64	13,64	8,18	2,44
HC3	10,32	5,50	1,18	11,80	6,44	1,68	12,36	7,12	1,88
HC4	10,36	5,48	1,18	11,56	6,22	1,64	12,00	6,86	1,82
<i>naïve</i>	10,14	5,16	0,98	17,00	10,12	2,90	20,78	13,62	4,82
<i>wu</i>	10,64	5,74	1,16	11,94	6,64	1,78	12,50	7,38	1,96
<i>invwu</i>	9,34	4,68	0,94	11,68	6,24	1,54	13,44	7,88	2,22
<i>n</i> = 141									
λ	1,00			15,02			50,30		
teste	10%	5%	1%	10%	5%	1%	10%	5%	1%
MQO	9,42	4,78	1,10	16,20	9,48	2,84	19,86	12,32	4,44
<i>white</i>	10,38	5,66	1,50	11,60	6,02	1,78	11,98	6,72	2,16
HC3	9,80	5,24	1,28	10,78	5,50	1,60	11,16	6,30	1,74
HC4	9,78	5,18	1,30	10,64	5,48	1,56	10,94	6,20	1,64
<i>naïve</i>	9,52	4,98	1,06	11,30	6,00	1,68	20,30	12,28	4,52
<i>wu</i>	10,10	5,10	1,34	10,98	5,72	1,66	11,18	6,48	1,94
<i>invwu</i>	9,04	4,62	1,06	11,30	6,00	1,68	13,00	7,36	2,32

Figura 1: Discrepâncias de quantis para o conjunto de 'dados1' (com pontos de alavanca). Diferenças entre quantis exatos (estimados por simulação) e os quantis assintóticos (computados da distribuição $N(0, 1)$) dispostos graficamente contra quantis assintóticos de três estatística de teste: HC3, HC4 e *invwu*



Com o objetivo de comparar numericamente os desempenhos dos testes quasi- t baseados nos estimadores wu e $invwu$ com os desempenhos de dois testes *bootstrap*, um que usa o esquema de reamostragem de *bootstrap* ponderado e outro que utiliza o esquema de reamostragem ponderado com probabilidades de amostragem inversamente proporcionais às medidas de alavancagem, utilizamos o conjunto de dados retirado de Greene (1997), 'dados1', e também 'dados2', que não contém pontos de alavanca.

Os desempenhos dos testes são avaliados com base nas taxas de rejeição da hipótese nula, sendo a hipótese alternativa bicaudal e a geração dos dados realizada sob a hipótese em teste. Os resultados apresentados a seguir são, como os anteriores, baseados em 5 000 réplicas de Monte Carlo e em 500 réplicas de *bootstrap*.

Os Quadros 3 e 4 apresentam percentagens de rejeição de testes quasi- t baseados nos estimadores wu e $invwu$ e também percentagens de rejeição de dois testes *bootstrap*, a saber: o teste que usa o esquema de reamostragem de *bootstrap* ponderado e o teste que usa o esquema de reamostragem ponderado com probabilidades de amostragem inversamente proporcionais às medidas de alavancagem. Nestes testes *bootstrap*, a reamostragem foi realizada para estatísticas de teste construídas com base no estimador *white*, que são assintoticamente pivotais, e objetiva obter valores críticos de *bootstrap*, que são usados em substituição aos valores críticos normais (assintóticos). Nós denotaremos o primeiro teste por wu^* e o segundo por $invwu^*$ na análise que segue.

O teste *bootstrap* realizado com base no esquema de *bootstrap* ponderado (wu^*) apresenta muito pouco ganho em amostras finitas relativamente ao uso do *bootstrap* para estimação de variâncias, como pode ser visto no Quadro 3. Para exemplificar, considere $\lambda = 1$ e $n = 100$, ao nível nominal de 10%. O teste quasi- t que usa o esquema de *bootstrap* ponderado para estimação de variância (wu) rejeita a hipótese nula 13,66% das vezes ao passo que o teste *bootstrap* wu^* o faz 13,58% das vezes. Sob heteroscedasticidade ($\lambda = 15,04$), as conclusões são similares e até há casos em que a distorção de tamanho do teste cuja estatística usa o estimador wu é menor que a do teste *bootstrap* wu^* ; por exemplo, ao nível de 5% e $n = 150$, o primeiro teste apresenta taxa de rejeição de 10,80% enquanto o segundo apresenta taxa de rejeição de 10,86%.

O teste *bootstrap* realizado com base no esquema de reamostragem ponderado com probabilidades de amostragem inversamente proporcionais às medidas de alavancagem ($invwu^*$) apresenta taxas de rejeição mais baixas do que se esperava. O resultado central favorece o uso do teste quasi- t baseado no estimador $invwu$. Por exemplo, sob homoscedasticidade, ao nível nominal de 5% e $n = 100$ a distorção de tamanho do teste

bootstrap *invwu** é aproximadamente cinco vezes maior que a distorção de tamanho do teste baseado no estimador *invwu*. Sob heteroscedasticidade ($\lambda = 15,04$), ao nível nominal de 10% e $n = 50$, o teste baseado no estimador *invwu* apresenta distorção de 0,32 enquanto para o teste *bootstrap invwu** esta distorção é 5,10. Podemos notar ainda que quando os dados não contêm observações influentes (Quadro 4), o teste baseado no estimador *invwu* apresenta menor distorção de tamanho do que o teste *bootstrap invwu**, exceto quando $n = 94,141$ e $\lambda = 15,02$.

5. Aplicação empírica

A seguir, apresentamos um exemplo empírico onde as variâncias dos estimadores de MQO dos coeficientes de um modelo de regressão linear são estimadas a partir dos distintos procedimentos considerados neste artigo. Os dados foram extraídos de Myers (1990, Tabela 5.2, p. 218). A variável dependente (y) representa o número de homens-hora mensais em instalações da marinha americana e as variáveis independentes (x_1, x_2) correspondem à ocupação média diária da instalação e ao número de alas construídas na instalação.

Inicialmente, é útil investigar a presença de heteroscedasticidade, o que pode ser feito aplicando-se um teste de heteroscedasticidade. O teste empregado foi o de Koenker (1981), considerando-se a hipótese nula de homoscedasticidade. O teste indica que há heteroscedasticidade aos níveis usuais de significância. Assim, testes de hipótese sobre os parâmetros lineares dos modelos devem ser realizados a partir de estatísticas de teste quasi- t associadas a estimadores para as matrizes de covariâncias dos estimadores de MQO que sejam consistentes sob heteroscedasticidade de forma desconhecida.

O modelo proposto inicialmente para descrever a relação entre a variável dependente e os regressores, é da forma

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \quad (5.1)$$

para $i = 1, \dots, 25$. Os parâmetros lineares em (5.1) foram estimados por MQO. As estimativas pontuais de β_0, β_1 e β_2 são $b_0 = 610,83, b_1 = 4,23, b_2 = 89,71$, respectivamente.

No Quadro 5 encontram-se os erros-padrão dos elementos de b , obtidos a partir dos diferentes estimadores consistentes apresentados anteriormente. Os erros-padrão diferem bastante quando calculados com base nos diferentes estimadores considerados. Observa-se

que os maiores erros-padrão são provenientes de HC4 e *invwu*, seguidos por HC3, *wu*, *white*, *naïve* e *MQO*, nesta ordem. Por exemplo, o erro-padrão de b_2 obtido a partir do estimador HC4 é quase 26 vezes maior do que aquele obtido a partir do estimador *white*.

Quadro 3: Percentagens de rejeição de testes quasi-*t* baseados nos estimadores *wu* e *invwu* e percentagens de rejeição dos testes *bootstrap wu e *invwu** no modelo $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$, $e_i \sim (0, \exp\{c_1 + c_2 x_i\})$, $i = 1, \dots, n$, para $n = 50, 100, 150$. Os níveis nominais considerados são $\alpha = 0, 10, 0, 05, 0, 01$ (com pontos de alavanca) - 'dados1'**

<i>n</i> = 50						
λ	1,00			15,04		
α	10%	5%	1%	10%	5%	1%
<i>wu</i>	15,28	9,98	4,20	26,76	20,04	11,48
<i>wu*</i>	15,22	9,44	4,32	26,92	19,60	11,54
<i>invwu</i>	5,86	3,40	1,32	10,32	6,94	3,34
<i>invwu*</i>	2,70	1,78	0,60	4,90	3,34	1,72
<i>n</i> = 100						
λ	1,00			15,04		
α	10%	5%	1%	10%	5%	1%
<i>wu</i>	13,66	7,86	2,54	19,82	13,74	6,48
<i>wu*</i>	13,58	8,04	2,80	20,04	13,58	6,72
<i>invwu</i>	8,22	4,48	1,24	11,86	7,96	3,26
<i>invwu*</i>	4,32	2,18	0,52	5,74	3,52	1,20
<i>n</i> = 150						
λ	1,00			15,04		
α	10%	5%	1%	10%	5%	1%
<i>wu</i>	12,50	6,84	1,78	16,50	10,72	4,62
<i>wu*</i>	12,60	7,00	2,06	16,58	10,86	4,62
<i>invwu</i>	8,56	4,22	1,10	11,92	7,30	2,86
<i>invwu*</i>	5,12	2,56	0,62	7,04	4,04	1,40

Quadro 4: Percentagens de rejeição de testes quasi-*t* baseados nos estimadores *wu* e *invwu* e percentagens de rejeição dos testes *bootstrap wu** e *invwu** no modelo $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$, $e_i \sim (0, \exp\{c_1 + c_2 x_i\})$, $i = 1, \dots, n$, para $n = 47, 94, 141$. Os níveis nominais considerados são $\alpha = 0, 10, 0, 05, 0, 01$ (sem pontos de alavanca) - 'dados2'

<i>n</i> = 47						
λ	1,00			15,04		
α	10%	5%	1%	10%	5%	1%
<i>wu</i>	11,80	6,88	1,70	13,74	8,16	2,92
<i>wu*</i>	11,74	6,80	1,80	13,66	7,90	2,66
<i>invwu</i>	9,42	4,84	1,08	11,22	6,08	1,92
<i>invwu*</i>	6,58	2,86	0,72	7,72	3,90	1,20
<i>n</i> = 94						
λ	1,00			15,04		
α	10%	5%	1%	10%	5%	1%
<i>wu</i>	10,64	5,74	1,16	11,94	6,64	1,78
<i>wu*</i>	10,68	5,72	1,34	12,10	6,48	1,76
<i>invwu</i>	9,34	4,68	0,94	11,68	6,24	1,54
<i>invwu*</i>	7,80	3,62	0,84	10,66	5,38	1,26
<i>n</i> = 141						
λ	1,00			15,04		
α	10%	5%	1%	10%	5%	1%
<i>wu</i>	10,10	5,10	1,34	10,98	5,72	1,66
<i>wu*</i>	10,06	5,20	1,52	10,96	5,66	1,74
<i>invwu</i>	9,04	4,62	1,06	11,30	6,00	1,68
<i>invwu*</i>	8,06	4,06	0,94	11,60	5,96	1,68

Considere o teste da hipótese nula $H_0 : \beta_2 = 0$ contra uma hipótese alternativa bicaudal. Uma vez que β_2 é o coeficiente do termo x_{i2} , o teste de tal hipótese avalia se há relação entre a variável dependente 'número de homens-hora mensais' e a covariável número de alas construídas'. O Quadro 6 apresenta os diferentes *p*-valores para este teste. Mesmo a tamanhos nominais muito pequenos, verifica-se que os testes cujas estatísticas usam os estimadores *MQO*, *white* e *naïve* rejeitam a hipótese nula, isto é, sugerem que o número de alas construídas é importante no sentido de que variações neste regressor levam a variações significantes, em média, no número de homens-hora mensais. O teste quasi-*t* baseado no estimador HC4 tem o maior *p*-valor dentre todos os testes, seguido pelos testes baseados em *invwu* e HC3, nesta ordem.

Quadro 5: Erros-padrão para b_0 , b_1 e b_2 em um modelo de regressão linear da forma

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \quad i = 1, \dots, 25$$

estimador	<i>MQO</i>	<i>white</i>	<i>HC3</i>	<i>HC4</i>	<i>naïve</i>	<i>wu</i>	<i>invwu</i>
$\sqrt{\hat{\text{var}}(b_0)}$	340,21	233,55	691,59	3553,80	330,32	358,86	1128,70
$\sqrt{\hat{\text{var}}(b_1)}$	1,67	2,90	17,79	123,76	1,62	6,79	35,41
$\sqrt{\hat{\text{var}}(b_2)}$	23,53	28,28	127,46	743,53	23,14	56,03	226,68

Uma análise de influência das observações correspondentes aos locais 22 e 23 revela que estas observações são pontos de alta alavancagem, uma vez que, para o modelo (5.1), o valor de referência $3p/n$ é igual a 0,360 e os elementos diagonais da matriz H referentes a estas observações são 0,724 e 0,857, respectivamente. Quando estas observações são removidas da amostra e o teste da hipótese $H_0: \beta_2 = 0$ é feito, a hipótese nula não é rejeitada aos níveis de significância usuais, qualquer que seja o estimador empregado para a matriz de covariâncias. Neste contexto, os estimadores *MQO*, *white* e *naïve* mostram-se pouco confiáveis, pois os testes a eles associados conduzem a conclusões provavelmente enganosas sobre a estrutura do modelo, o que não acontece com os estimadores *HC3*, *HC4*, *wu* e *invwu*, cujos testes associados apresentam as mesmas conclusões independente dos dados incluírem ou não observações influentes. Ou seja, as inferências realizadas a partir de testes construídos com base nos estimadores *invwu*, *HC3* e *HC4* não são sensíveis à presença nos dados de observações de extrema alavancagem (locais 22 e 23). Em particular, os p -valores dos testes realizados a partir de *invwu* e *HC4* são bastante elevados.

Com o objetivo de examinar o impacto que os locais 22 e 23 têm sobre a inferência resultante, estimamos o modelo (5.1) 25 vezes, cada vez retirando uma observação diferente.

As estimativas dos parâmetros resultantes estão apresentadas no Quadro 7. O grande impacto que as observações 22 e 23 têm sobre as estimativas é evidente. Quando a observação 22 não está na amostra, a estimativa de β_2 torna-se próxima do dobro da média (85,236); quando a observação 23 não está na amostra a estimativa de β_2 se torna negativa (-7,698). As inferências realizadas a partir dos testes que usam os estimadores *MQO*, *white* e *naïve* são dominadas por estas duas observações. As inferências realizadas através de outros testes consistentes, por outro lado, não sofrem do mesmo mal.

Quadro 6: Inferência quasi-t, p-valores

com pontos de alavanca, $n = 25$		sem pontos de alavanca, $n = 25$	
teste	p-valor	teste	p-valor
<i>MQO</i>	0,000	<i>MQO</i>	0,598
<i>white</i>	0,002	<i>white</i>	0,594
<i>HC3</i>	0,482	<i>HC3</i>	0,691
<i>HC4</i>	0,904	<i>HC4</i>	0,756
<i>naïve</i>	0,000	<i>naïve</i>	0,602
<i>wu</i>	0,109	<i>wu</i>	0,637
<i>invwu</i>	0,692	<i>invwu</i>	0,726

Quadro 7: Estimativas dos parâmetros sem a i -ésima observação

obs	b_0	b_1	b_2	obs	b_0	b_1	b_2
1	654,228	4,182	88,400	14	610,357	4,233	89,720
2	654,326	4,183	88,388	15	501,441	4,114	93,512
3	659,727	4,193	88,087	16	591,615	4,219	89,275
4	647,179	4,195	88,576	17	611,899	4,233	89,715
5	663,787	4,151	88,520	18	616,763	4,233	89,715
6	656,709	4,186	88,372	19	519,889	4,246	89,495
7	610,132	4,233	89,730	20	612,553	4,232	89,706
8	650,265	3,904	94,790	21	584,884	4,285	88,712
9	648,719	4,129	90,178	22	107,907	2,590	171,154
10	663,300	4,197	89,413	23	176,036	21,876	-7,698
11	569,820	4,213	90,942	24	680,015	2,702	83,432
12	591,936	4,240	90,269	25	609,566	4,233	89,725
13	624,558	4,201	90,263				

7. Conclusões

No presente artigo, nós utilizamos simulações de Monte Carlo para comparar os méritos relativos de inferências realizadas sob heteroscedasticidade de forma desconhecida a partir de testes cujas estatísticas usam erros-padrão obtidos de dois estimadores recentemente propostos: HC4 e *bootstrap* ponderado inversamente ajustado. Outros estimadores consistentes foram também incluídos na análise comparativa. Os resultados mostraram que os dois estimadores recentemente propostos tipicamente conduzem a

inferências mais confiáveis do que aquelas realizadas a partir dos estimadores mais tradicionais, especialmente quando os dados incluem pontos de alavanca, com alguma vantagem para o estimador baseado em reamostragem.

No que se refere à inferência via *bootstrap*, os resultados revelaram que há muito pouco ganho em amostras finitas em se usar o esquema de reamostragem de *bootstrap* ponderado para realizar testes *bootstrap*, estimando-se valores p ou valores críticos, ao invés de se utilizar o *bootstrap* ponderado para estimação de erros-padrão a serem utilizados em estatísticas de teste convencionais.

Referências Bibliográficas

- Chesher, A. & Jewitt, I. (1987). The bias of heteroskedasticity consistent covariance matrix estimator. *Econometrica*, 55, 1217–1222.
- Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis*, a aparecer.
- Cribari-Neto, F. & Zarkos, S.G. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews*, 18, 211–228.
- Cribari-Neto, F. & Zarkos, S.G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *Journal of Statistical Computation and Simulation*, 68, 391–411.
- Cribari-Neto, F. & Zarkos, S.G. (2004). Leverage-adjusted heteroskedastic bootstrap methods. *Journal of Statistical Computation and Simulation*, 74, 215–232.
- Davidson, R. & MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- Greene, W.H. (1997). *Econometric Analysis*, 3a ed. Upper Saddle River: Prentice Hall.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Hinkley, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics*, 19, 285–292.
- Horn, S.D., Horn, R.A. & Duncan, D.B. (1975). Estimating heteroskedastic variances in linear models. *Journal of the American Statistical Association*, 70, 380–385.
- Horowitz, J.L. (1997). Bootstrap methods in econometrics: theory and numerical performance. Em Kreps, D.M. & Wallis, K.F., eds., *Advances in Economics and Econometrics: Theory and Applications (Seventh World Congress)*, 3, 188–222. New York: Cambridge University Press.
- Jeong, J. & Maddala, G.S. (1993). A perspective on application of bootstrap methods in econometrics. Em Maddala, G.S., Rao, C.R. & Vinod, H.D., eds., *Handbook of Statistics: Econometrics*, 11, 573–610. Amsterdam: North-Holland.
- Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H. & Lee, T.-C. (1988). *Introduction to the Theory and Practice of Econometrics*. 2a ed. New York: Wiley.

- Koenker, R. (1981). A note on Studentizing a test for heteroscedasticity. *Journal of Econometrics*, 17, 107–112.
- Li, H. & Maddala, G.S. (1996). Bootstrapping time series models. *Econometric Reviews*, 15, 115–158.
- Long, J.S. & Ervin, L.H. (2000). Using heteroscedasticity-consistent standard errors in the linear regression model. *The American Statistician*, 54, 217–224.
- MacKinnon, J.G. & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite-sample properties. *Journal of Econometrics*, 29, 305–25.
- Myers, R.H. (1990). *Classical and Modern Regression with Applications*. Belmont: Duxbury Press.
- Shao, J. (1988). On resampling methods for variance and bias estimation in linear models. *Annals of Statistics*, 16, 986–1008.
- Vinod, H.D. (1993). Bootstrap methods: Applications in econometrics. In Maddala, G.S., Rao, C.R. and Vinod, H.D., eds., *Handbook of Statistics: Econometrics*, 11, 629–661. Amsterdam: North-Holland.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.
- Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261–1295.

Abstract

The chief goal of this paper is to evaluate the finite-sample behavior of different heteroskedasticity-consistent covariance matrix estimators, under both constant and unequal error variances. We consider the estimator proposed by Halbert White (HC0), its variant known as HC3, and Wu's (1986) weighted *bootstrap* estimator. Recently proposed estimators, such as Cribari-Neto's (2004) HC4 and Cribari-Neto & Zarkos's (2004) inversely adjusted weighted *bootstrap* (*invwu*), are also considered. Our numerical (Monte Carlo) results show that quasi-*t* inference based on HC4 and *invwu* are typically more reliable than inference based on other covariance matrix estimators. We also present one application to real data.

Keywords: Covariance matrix estimation, Heteroskedasticity, High leverage points, Linear Regression.

Agradecimentos

Os autores agradecem à CAPES e ao CNPq pelo auxílio financeiro parcial e a dois pareceristas por comentários e sugestões.

Modelos de riscos multiplicativos e aditivos para analisar dados de sobrevivência: Uma abordagem paramétrica

Vera L.D. Tomazella*
Camila Bertini Martins

Resumo

O objetivo da análise de sobrevivência está centrado na relação entre o tempo de sobrevivência e algumas variáveis explicativas de interesse. Uma questão de interesse é investigar a influência das covariáveis nos tempos de sobrevivência e um modo de fazer isso é por meio de modelos de regressão, em que a investigação da influência de covariáveis nos tempos de sobrevivência em populações heterogêneas é feita usualmente por meio da função de risco. A estruturação desta função é feita principalmente em termos aditivos ou multiplicativos. A aplicação do modelo multiplicativo tem levado quase que, exclusivamente, a considerações de riscos proporcionais (COX, 1972), assim, considerando situações onde não se impõe que as funções de riscos sejam proporcionais, temos os modelos de riscos aditivos originalmente sugerido por Aalen (1980). Usualmente, os métodos inferências freqüentistas para o modelo aditivo são não-paramétricos baseados em processo de contagem, uma versão paramétrica dos modelos aditivos foi apresentado por Andersen *et al.* (1993, cap. 7) e Tomazella (2003). Neste artigo, estas duas classes de modelos serão revisitadas com uma abordagem paramétrica. Para ilustrar a metodologia, consideramos um exemplo de dados de sobrevivência para a estimação e construção de intervalos de confiança assintóticos para os parâmetros dos modelos.

Palavras-chave: Modelos de regressão, Modelo de risco multiplicativo, Modelo de risco aditivo, Dados de sobrevivência

* Endereço para correspondência: Departamento de Estatística, Universidade Federal de São Carlos – UFSCar, CP 676 –13565-905 – São Carlos – SP. E-mail: vera@power_ufscar.com.br.

1. Introdução

Os dados de sobrevivência compreendem tanto tempos de sobrevivência quanto um conjunto de variáveis observáveis que podem estar relacionadas com esses tempos. Essas variáveis são conhecidas por covariáveis, variáveis explicativas ou variáveis explanatórias. A investigação da influência das covariáveis nos tempos de sobrevivência é feita, usualmente, por meio da função de risco. Estas funções são divididas em duas classes: modelo de riscos multiplicativos e modelo de riscos aditivos. O modelo de risco multiplicativo mais conhecido é o modelo de risco proporcional de Cox (COX, 1972) e o modelo aditivo mais conhecido é o modelo de risco aditivo de Aalen (AALEN, 1989). A aplicação do modelo de Cox tem levado quase que exclusivamente a consideração de riscos proporcionais. Isso é relevante em muitas situações reais, mas, em muitos casos, o risco não-proporcional é também de interesse. Em particular, esse é o caso quando a proposta é predição, isto é, determinar a probabilidade de algum evento ou eventos acontecerem para um único grupo ou indivíduo, com base em informações anteriores (que podem ser uma probabilidade geral ou baseada em algumas covariáveis) ou durante o estudo (isto é, probabilidades condicionais dada a resposta para outros indivíduos, ou observações mais recentes para o real indivíduo). Por exemplo, para uma companhia de seguro é muito interessante se o risco de fumantes e não-fumantes é o mesmo, qual é a comparação de risco, porque ele determina se eles deveriam ter diferentes prêmios para os dois grupos. Contudo, o risco não-proporcional é essencial para calcular o prêmio real.

Enquanto o modelo multiplicativo de Cox tornou-se muito popular, o modelo aditivo de Aalen passou quase despercebido desde seu aparecimento em 1980 até 1986, quando Mau (1986) utilizou os gráficos das suas funções de regressão, na apresentação de um método gráfico para detectar diferenças na influência das covariáveis ao longo do tempo. Apesar das características úteis do modelo de Aalen (1980), este ainda não é amplamente utilizado em análise de sobrevivência. Acredita-se que isso se deva ao fato de esse modelo não estar implementado nos principais *softwares* estatísticos. Atualmente, o modelo de Aalen já começa a estar mais difundido em análise de sobrevivência como uma boa alternativa ao modelo de Cox, sobretudo quando as funções de risco não são proporcionais (AALEN, 1989). Recentemente, alguns autores elaboraram macros executáveis em *softwares*, tais como: SAS (HOWELL & KLEIN, 1996) e S-Plus (FEKJAER & AALEN, 1998), e também muitos já são os trabalhos com aplicações e desenvolvimentos teóricos do modelo de Aalen (SILVA, 2001). Outros modelos de regressão em análise de sobrevivência foram considerados com o objetivo

de apresentar uma formulação geral para os modelos multiplicativo e aditivo (LIN e YING, 1995). O objetivo deste artigo é revisar as duas classes de modelos de risco com uma abordagem paramétrica, sobretudo o modelo aditivo que usualmente é não-paramétrico, onde consideramos estimativas dos parâmetros das funções de risco de base e das funções de regressão.

Na seção 2, apresentamos o modelo de risco multiplicativo; na seção 3, apresentamos o modelo de risco aditivo; na seção 4, apresentamos uma aplicação para os dois modelos; e na seção 5, concluímos esse artigo.

2. Modelo de risco multiplicativo

A função de risco em geral depende do tempo e de um conjunto de covariáveis. Assim, dado um conjunto de dados de sobrevivência com n unidades independentes e p covariáveis observadas, os modelos de risco multiplicativos são definidos por:

$$\lambda(t_i; \mathbf{x}_i) = \lambda_0(t_i)G(\mathbf{x}_i; \boldsymbol{\beta}), \quad (1)$$

em que $\lambda_0(\cdot)$ é uma função positiva, chamada função de risco de base, sendo comum a todos os indivíduos, $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ é um vetor ($1 \times p$) de covariáveis e $\boldsymbol{\beta}$ é um vetor p -dimensional de parâmetros desconhecidos de efeitos fixos associados a \mathbf{x}_i .

Existem várias formas para a função $G(\cdot)$ em (1). Neste artigo, assumimos que G é uma função exponencial. O modelo (1) implica que a razão de riscos para dois indivíduos é constante no tempo, desde que as covariáveis não mudem com o tempo, isto é, para diferentes valores de \mathbf{x} , as funções de risco são assumidas como proporcionais, e que estes modelos pertencem à classe dos modelos de riscos proporcionais.

Quando a função de risco de base $\lambda_0(t)$ é indexada por um parâmetro θ , estes modelos são conhecidos como modelos de risco paramétrico. Neste artigo, é assumido que existe um modelo paramétrico que descreve, adequadamente, a distribuição dos tempos de vida.

Considerando (1) e $G(\cdot) = \exp(\cdot)$, podemos escrever que para o i -ésimo indivíduo a função de risco é dada por:

$$\lambda(t_i) = \lambda_0(t_i) \exp\{\mathbf{x}'_i \boldsymbol{\beta}\}, \quad (2)$$

em que $\eta_i = \mathbf{x}'_i\beta$ é chamado de preditor linear. O modelo (2) implica que a razão de riscos para dois indivíduos depende das diferenças entre os preditores lineares, em qualquer tempo.

Integrando (2), temos a função de risco acumulada é dada por:

$$\begin{aligned}\Lambda(t_i) &= \left(\int_0^{t_i} \lambda_0(u_i) du_i \right) \exp\{\mathbf{x}'_i\beta\} \\ &= \Lambda_0(t_i) \exp\{\mathbf{x}'_i\beta\},\end{aligned}\tag{3}$$

em que $\Lambda_0(t_i) = \int_0^{t_i} \lambda_0(u_i) du_i$ é a função de risco acumulada da função de risco de base.

Assim, considerando (3) a função de sobrevivência para o i -ésimo indivíduo do modelo (2) é dada por:

$$S(t_i) = \exp\{-\Lambda_0(t_i) \exp\{\mathbf{x}'_i\beta\}\}, \quad i = 1, 2, \dots, n.\tag{4}$$

Considerando que a presença de tempos censurados é comum, no contexto de análise de sobrevivência, suponha que alguns dos tempos observados sejam censurados à direita. Suponha, também, um esquema de censuras onde estas ocorrem em tempos diferentes, de indivíduo para indivíduo, e tal esquema sempre possa ser considerado estatisticamente independente do mecanismo que causa a morte do indivíduo (ou ocorrência da característica em observação), ou seja, um esquema de censuras independentes dos tempos de sobrevivência e das covariadas. Em tal esquema, ou o tempo exato de ocorrência Y_i da característica para o i -ésimo indivíduo é conhecido, ou apenas temos a informação de que este tempo é maior do que o tempo pré-fixado de censura (L_i). Desta forma, podemos resumir os dados observados no par de variáveis (T_i, γ_i) , onde γ_i é uma variável indicadora definida por:

$$\gamma_i = \begin{cases} 1, & \text{se } Y_i \leq L_i \\ 0, & \text{se } Y_i > L_i \end{cases}\tag{5}$$

e $T_i = \min(Y_i, L_i)$.

Considerando (2) e a presença de censuras, de acordo com o esquema descrito acima, a função de verossimilhança é dada por:

$$L(\beta) = \prod_{i=1}^n \left[\lambda_0(t_i) e^{\mathbf{x}'_i\beta} \right]^{\gamma_i} \exp\{-\Lambda_0(t_i) e^{\mathbf{x}'_i\beta}\}.\tag{6}$$

A estimação dos parâmetros envolvidos no modelo pode ser feita via maximização direta de (6) ou pela utilização de métodos numéricos, como, por exemplo, o método de Newton-Raphson.

3. Modelo de risco aditivo

Considerando situações onde não se impõe que as funções de risco sejam proporcionais, temos a segunda classe de modelos de risco em análise de sobrevivência, conhecida por modelos aditivos. Assim, quando o efeito das covariáveis é expresso aditivamente na função de risco, os modelos de risco aditivos são definidos pela seguinte função de risco na unidade i :

$$\lambda(t_i; \mathbf{x}_i) = \lambda_0(t_i) + G(\mathbf{x}_i, \boldsymbol{\beta}), \quad (7)$$

em que $\lambda_0(\cdot)$ é a função de risco de base, $G(\cdot)$ é uma função positiva e $\boldsymbol{\beta}$ é um vetor p -dimensional de parâmetros desconhecidos de efeitos associados a \mathbf{x}_i . Mesmo que as covariáveis não sejam observadas ao longo do tempo, o uso de funções de regressão no modelo (7) permite averiguar a influência das covariáveis variando com o tempo. Este modelo, originalmente proposto por Aalen (1980), e apresentado de forma mais simples em Aalen (1989), é um modelo de risco aditivo para análise de regressão de dados de tempo de sobrevivência, que permite a verificação da influência das covariáveis variando de acordo com o tempo.

A forma mais geral da função de risco para o modelo aditivo (Aalen, 1989) é dada por:

$$\lambda_i(t) = \beta_0(t) + \sum_{j=1}^p \beta_j(t)x_{ij}(t) \quad (8)$$

A forma matricial do modelo (8) é dada por:

$$\lambda(t) = \boldsymbol{\beta}(t)Y(t), \quad (9)$$

em que $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \dots, \beta_p(t))'$ é um vetor de funções do tempo desconhecidas, cujo primeiro elemento $\beta_0(t)$ é interpretado como uma função de parâmetro básica, enquanto os $\beta_j(t)$'s, $j = 1, \dots, p$, chamados de funções de regressão que medem a influência das respectivas covariáveis. A matriz $Y(t)$ de ordem $n \times (p+1)$ é uma matriz de observações de processo estocástico. Esta matriz é construída da seguinte maneira: se o evento considerado ainda não ocorreu para o i -ésimo indivíduo e ele não é censurado, então a i -ésima linha da

matriz é o vetor $x_i(t) = (1, x_{i1}(t), \dots, x_{ip}(t))'$. Caso contrário, se o indivíduo não está sob risco no tempo t , então a linha correspondente de $Y(t)$ contém apenas zeros.

A função $G(\cdot)$ em (7), é uma função positiva escolhida usualmente como $G(\cdot) = \mathbf{x}'\boldsymbol{\beta}$. Assim, considerando o modelo aditivo (7), a função de risco pode ser reescrita como:

$$\lambda(t) = \lambda_0(t) + \mathbf{x}'\boldsymbol{\beta}, \quad (10)$$

em que $\mathbf{x}'\boldsymbol{\beta} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ e $\lambda_0(t)$ é a função de risco de base. Similarmente ao modelo de risco multiplicativo, a função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [\lambda_0(t_i) + \mathbf{x}_i'\boldsymbol{\beta}]^{y_i} \exp\{-\Lambda_0(t_i) - \mathbf{x}_i'\boldsymbol{\beta}t_i\}. \quad (11)$$

Usualmente, os métodos inferenciais freqüentistas para o modelo aditivo são não-paramétricos, pois nenhuma forma paramétrica particular é assumida para a função de risco de base e para a função de regressão. Estas funções podem variar arbitrariamente com o tempo, revelando mudanças na influência das covariáveis. Esta é uma das vantagens do modelo (10), bem como a exigência de tamanho de amostra extremamente grande. Uma desvantagem deste modelo é permitir valores estimados negativos para a função de risco.

4. Aplicação

Considere os tempos de sobrevivência (em semanas) e contagem do número de glóbulos brancos de dois grupos de pacientes, Ag positivo e negativo, com diagnóstico de leucemia (LOUZADA-NETO, 2002). Os dados encontram-se na Tabela 1, a seguir:

Tabela 1 – Dados de pacientes com leucemia

Tempos	Ag+	log ₁₀ (wbc)	Tempos	Ag-	log ₁₀ (wbc)
63	2 300	3,36	56	4 400	3,64
156	750	2,88	65	3 000	3,48
134	2 600	3,41	7	1 500	3,18
16	6 000	3,78	16	9 000	3,95
108	10 000	4,02	22	5 300	3,72
121	10 000	4,00	3	10 000	4,00
4	17 000	4,23	4	19 000	4,28
39	5 400	3,73	2	27 000	4,43
143	7 000	3,85	3	28 000	4,45
56	9 400	3,97	8	31 000	4,49
26	32 000	4,51	4	26 000	4,41
22	35 000	4,54	3	21 000	4,32
1	100 000	5,00	30	79 000	4,90
1	100 000	5,00	4	10 000	5,00
5	52 000	4,72	43	10 000	5,00
65	100 000	5,00			

4.1 Modelo de risco multiplicativo

Considerando uma forma funcional paramétrica para a função de risco de base no modelo (2), podemos definir o modelo de risco multiplicativo de forma paramétrica. Mostraremos isso assumindo um modelo exponencial para o tempo de vida, onde a função de risco de base é constante, isto é, $\lambda_0(t) = \mu$. Assim, podemos reescrever o modelo de risco proporcional (2) como:

$$\lambda(t | \mathbf{x}) = \mu \exp(\boldsymbol{\beta}'\mathbf{x}). \quad (12)$$

Considerando (6) e o modelo (12) com a função de risco acumulada $\Lambda_0(t_i) = \mu t_i$, a função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}, \mu) = \prod_{i=1}^n \mu \exp(\boldsymbol{\beta}'\mathbf{x}_i) \exp\{-[\mu t_i \exp(\boldsymbol{\beta}'\mathbf{x}_i)]\}. \quad (13)$$

O logaritmo da função de verossimilhança (13) é dado por:

$$l(\boldsymbol{\beta}, \mu) = n \ln \mu + \sum_{i=1}^n \boldsymbol{\beta}'\mathbf{x}_i - \sum_{i=1}^n \mu t_i \exp(\boldsymbol{\beta}'\mathbf{x}_i). \quad (14)$$

As Estimativas de Máxima Verossimilhança – EMV dos parâmetros são encontradas maximizando $l(\boldsymbol{\beta}, \mu)$.

Definindo $e^{\beta_0} = \mu$, fazendo a covariável $x_0 = 1$ para todos os indivíduos, $i = 1, 2, \dots, n$ o modelo de risco proporcional (12) para os dados da Tabela 1 é dado por:

$$\lambda(t | \mathbf{x}) = \exp[\beta_0 + \beta_1 \text{ag} + \beta_2 (\log_{10}(\text{wbc}))]. \quad (15)$$

A Tabela 2 mostra as Estimativas de Máxima Verossimilhança – EMV, desvio-padrão (DP) e intervalo de 95% de confiança (IC(95%)) para o modelo (12), e a Figura 1 mostra a função de risco considerando os grupos Ag+ e Ag-. Considerando os intervalos de confiança, observamos que o efeito das covariáveis é significativo.

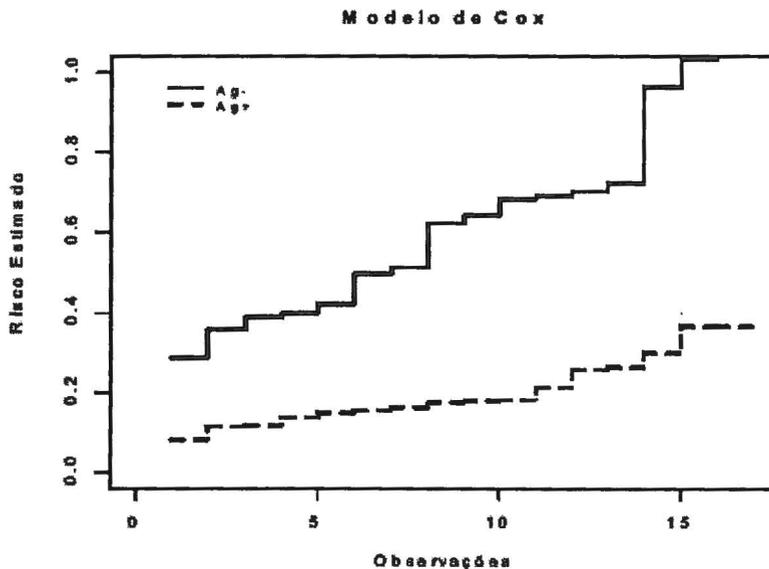
Tabela 2 – EMV, DP, IC(95%) para o modelo (15)

Parâmetros	EMV	DP	I.C.(95%)
β_0	-6,8271	1,158	[-9,161; -4,594]
β_1	1,0178	0,363	[0,295; 1,733]
β_2	0,6995	0,286	[0,135; 1,263]

Considerando que $e^{\beta_0} = \mu \rightarrow \mu = 0,0025$, o modelo de risco multiplicativo (12) pode ser expresso por:

$$\lambda(t | \mathbf{x}) = 0,0025 \exp[1,0178ag + 0,6995 \log(wbc)]. \quad (16)$$

Figura 1 – Gráfico do modelo de risco multiplicativo (15)



4.2 Modelo de risco aditivo

No modelo aditivo de Aalen, assume-se que as covariáveis atuam de maneira aditiva na função de risco de base, bem como que as funções de regressão são desconhecidas, a

aproximação para a estimação destas funções dependem das suposições sobre a forma funcional de tais funções. Neste artigo, assumimos uma forma funcional paramétrica para a função de risco de base, descrevendo, assim, um modelo de risco aditivo paramétrico.

Similarmente ao modelo multiplicativo, assumimos um modelo exponencial para o tempo de vida, onde a função de risco de base é constante, isto é, $\lambda_0(t) = \mu$. Assim, podemos reescrever o modelo de risco aditivo (10) como:

$$\lambda(t) = \mu + \mathbf{x}'\boldsymbol{\beta}. \quad (17)$$

A função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}, \mu) = \prod_{i=1}^n [\mu + \mathbf{x}'_i\boldsymbol{\beta}] \exp\{-\mu t_i - \mathbf{x}'_i\boldsymbol{\beta} t_i\}. \quad (18)$$

O logaritmo da função de verossimilhança (18) é dado por:

$$l(\boldsymbol{\beta}, \mu) = \sum_{i=1}^n \log(\mu + \mathbf{x}'_i\boldsymbol{\beta}) - \sum_{i=1}^n (\mu t_i + \mathbf{x}'_i\boldsymbol{\beta} t_i). \quad (19)$$

As Estimativas de Máxima Verossimilhança – EMV dos parâmetros são encontradas maximizando $l(\boldsymbol{\beta}, \mu)$.

Considerando os dados da Tabela 1, o modelo (17) é reescrito como:

$$\lambda(t | \mathbf{x}) = \mu + \beta_1 ag + \beta_2 (\log_{10}(wbc)). \quad (20)$$

Para garantir que a função de risco de base seja não negativa, utilizamos a seguinte reparametrização:

$$\mu = \exp(\phi_1) \rightarrow \phi_1 = \log(\mu). \quad (21)$$

A Tabela 3 mostra as EMVs, desvio-padrão (DP) e intervalos de confiança (IC(95%)) para o modelo (20), e a Figura 2 mostra a função de risco, considerando os grupos Ag+ e Ag-. Observamos, também, através dos intervalos de confiança, que os efeitos das covariáveis são significativos.

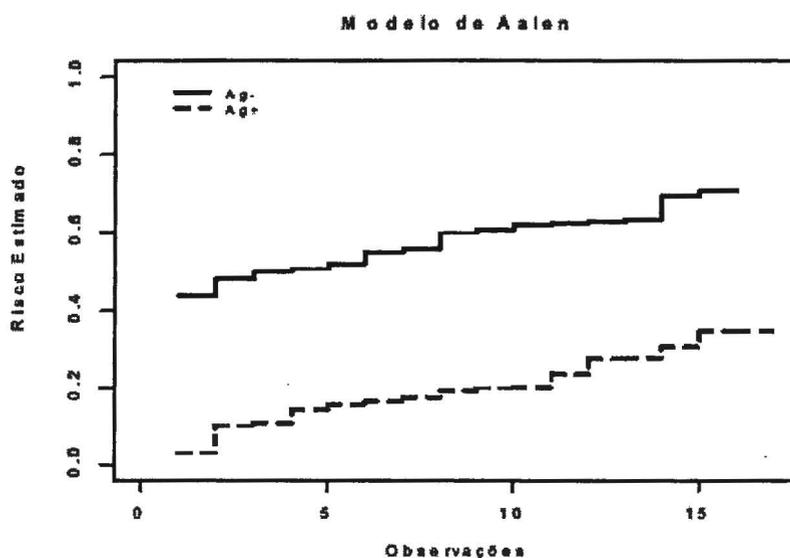
Tabela 3 – EMV, DP, IC(95%) para o modelo (20)

Parâmetros	EMV	DP	I.C.(95%)
μ	0,028	0,58	[0,010; 0,0867]
β_1	-0,043	0,014	[-0,0707; -0,0161]
β_2	0,082	0,004	[0,0005; 0,0158]

O modelo aditivo (20) pode ser expresso por:

$$\lambda(t | \mathbf{x}) = 0,028 - 0,043ag + 0,0082 \log(wbc). \quad (22)$$

Figura 2 – Gráfico do modelo de risco aditivo (20)



5. Conclusão

O modelo de risco proporcional de Cox apresenta vantagens de possuir uma interpretação simples dos resultados, ser facilmente estendido para incorporar covariáveis dependentes do tempo e estar disponível em vários *softwares* estatísticos. Entretanto, Aalen citou algumas limitações desse modelo, como, por exemplo, não ser adequado para detecção de mudanças de efeitos das covariáveis ao longo do tempo. O modelo aditivo de Aalen é uma alternativa ao modelo de riscos proporcionais de Cox, pois permite que ambos os parâmetros e os vetores de covariáveis variem com o tempo, já que os efeitos temporais não são assumidos serem proporcionais para cada covariável. O modelo de Aalen é capaz de fornecer informações detalhadas a respeito da influência temporal de cada covariável. Usualmente, a função de risco de base para os dois modelos são não-paramétricas, entretanto já existe uma extensiva literatura em que o modelo de Cox aparece numa perspectiva paramétrica, porém o modelo de Aalen ainda tem sido estudado com uma estrutura completamente não-paramétrica no sentido de que as funções de risco de base e de regressão são ajustadas e não os parâmetros.

Neste artigo, apresentamos os dois modelos considerando uma forma paramétrica para a função de risco de base, e uma aplicação envolvendo dados de sobrevivência na presença de covariáveis (Tabela 1). Os modelos ajustados a estes dados permitiram observarmos que, em essência, os dois modelos apresentaram as mesmas conclusões. As

covariáveis grupos (Ag+ e Ag-) e contagem do número de globos branco (wbc) foram identificadas como sendo fatores de risco para a ocorrência de leucemia. Embora a função de risco de base considerada neste artigo tenha sido a distribuição exponencial, poderíamos utilizar, também, a distribuição Weibull, a qual varia com o tempo (Tomazella, 2003).

Referências Bibliográficas

- AALEN, O.O. Nonparametric inference for a family of counting processes, *The Annals of Statistics*, 6, 701–726, 1978.
- AALEN, O.O. A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics*, 2, 1–25, 1980.
- AALEN, O.O. A linear regression model for the analysis of life times. *Statistics in Medicine*, 8, 907–925, 1989.
- COX, D.R. Regression models and life times (with discussion), *Journal of the Royal Statistical Society B*, 34: 187–220, 1972a.
- COX, D.R.; HINKLEY, D. V. *Theoretical Statistics*. London: Chapman and Hall, 1974.
- DAVISON, A.C.; HINKLEY, D. V. *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press, 582p. 1997.
- EFRON, B.; TIBSHIRANI, R. J. *An Introduction to the Bootstrap*. New York: Wiley, 436p., 1983.
- FEKJAER, H.; AALEN, O.O. *S-Plus package for Aalen's additive regression model: Version 1.05*, Norway: Section of Medical Statistics, University of Oslo, 1998.
- HALL, P.; WILSON, S. R. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47, 757–762, 1991.
- HOWELL, A.M.; KLEIN, J.P. A SAS macro for the additive hazards regression model. *Technical Report 18*, Division of Biostatistics, Medical College of Wisconsin, 1996.
- LAWLESS, J. F. *Statistical Models and Methods for Life Time Data*. New York: John Wiley, 1982.
- LIN, D. Y.; YING, Z. Semiparametric analysis of general additive–multiplicative hazards models for counting processes. *The Annals of Statistics*, 23, 1712–1734, 1995.
- LOUZADA-NETO, F.; MAZUCHELI, J.; ACHCAR, J. A. *Introdução a Análise de Sobrevida e Confiabilidade*. IMCA, 2002.
- SILVA, G.L.; AMARAL TURKMAN, M.A. Bayesian analysis of an additive survival model with frailty. *Communications in Statistics – Theory and Methods*, 33, 2517–2533, 2004.

Abstract

The objective of the survival analysis is centered in the relation between the time of survival and some explicatives variables of interest. An interest question is to investigate the influence of the covariates in the survival times and a way to make this is by means of regression models, where the inquiry of the influence of covariates in the times of survival in heterogeneous populations is usually made by means of the risk function. The risk function for these models are assumed have a multiplicative and additive structure. The application of the multiplicative model

has led almost that exclusively the consideration of proportional risks (COX, 1972). Thus considering situations where is not imposed that the functions of risks are proportional we have the additive models originally suggested for Aalen (1980). Usually the methods frequentists inferences for the additive model are not-parametric based on counting process, a parametric version of the additive models were presented by Andersen *et al.* (1993) and Tomazella (2003). In this work these two classrooms of models will be presented with a parametric boarding. To illustrate the methodology, we consider an example of survival data and we construct intervals reliable asymptotic for the parametres of the models.

Key-words: Regression models, Multiplicative risk model, Additive risk model, Survival data

Agradecimentos

À FAPESP pelo financiamento do Projeto-Bolsa de Iniciação Científica.

Estimação Pontual em Regressão Beta: Uma Avaliação do Desempenho de Algoritmos de Otimização Não-Linear

Nátaly Adriana Jiménez Monroy^{*}
Francisco Cribari-Neto^{**}
Klaus L. P. Vasconcellos^{***}

Resumo

Consideramos o modelo de regressão beta proposto por Ferrari & Cribari-Neto (2004), que é útil em situações em que a resposta é restrita ao intervalo (0, 1) e possui estrutura de regressão envolvendo regressores e parâmetros desconhecidos. Nosso interesse reside na avaliação de vários métodos de otimização não-linear no contexto da maximização da função de log-verossimilhança do modelo de regressão beta. Evidência numérica de simulações de Monte Carlo e análises empíricas baseadas em dados reais favorecem os algoritmos de Newton e BFGS, os quais são rápidos, precisos e apresentam comportamento estável mesmo em situações desfavoráveis, tais como a existência de pontos de alavanca e de correlação elevada entre regressores.

Palavras-chave: Modelo de regressão beta, otimização não-linear, função de log-verossimilhança.

^{*} Endereço para correspondência: Departamento de Estatística, UFPE -- *e-mail*: nataly.jimenez@gmail.com.

^{**} Departamento de Estatística, UFPE -- *e-mail*: cribari@de.ufpe.br.

^{***} Departamento de Estatística, UFPE -- *e-mail*: klaus@de.ufpe.br.

1. Introdução

Em diferentes áreas de aplicação da Estatística, é prática comum analisar a influência de diversas variáveis sobre a esperança condicional de porcentagens, proporções ou razões. Para este tipo de dados, freqüentemente são ajustados modelos de regressão normal linear, os parâmetros que indexam o modelo sendo estimados pelo método de Mínimos Quadrados Ordinários – MQO. Esta aproximação não é adequada, pois, entre outras razões, a variância da resposta deve ser variável, aproximando-se de zero quando a média se aproxima de qualquer um dos pontos extremos do intervalo $(0,1)$. Em resposta a este problema, surgiu o modelo de regressão beta, onde a resposta, que é continuamente distribuída no intervalo $(0,1)$, segue lei beta¹ e depende de outras variáveis através de uma estrutura de regressão. Mais geralmente, se a variável dependente (y) estiver restrita ao intervalo (a,b) , em que $a < b$ são escalares conhecidos, pode-se modelar a variável transformada $(y-a)/(b-a)$.

Na literatura, têm sido sugeridas várias parametrizações desta classe de modelos. Por exemplo, Paolino (2001) mostrou, por simulação de Monte Carlo, que o ajuste de modelos lineares por mínimos quadrados pode levar a inferências erradas sobre os efeitos das covariáveis. O autor propôs estimar o efeito da variável explicativa, especificando os parâmetros da distribuição da resposta em termos da média e da dispersão, tornando simples a interpretação dos parâmetros e a inferência sobre os efeitos das covariáveis sobre a resposta esperada e sobre a dispersão. Buckley (2003) apresentou uma alternativa bayesiana ao modelo proposto por Paolino (2001) e usou amostragem MCMC (Markov Chain Monte Carlo) para extrair amostras de uma distribuição beta *a posteriori*. Recentemente, Ferrari & Cribari-Neto (2004) propuseram um modelo de regressão beta usando uma parametrização da distribuição beta indexada pelos parâmetros de locação e dispersão. Estes parâmetros podem ser interpretados em termos da média das observações, que é modelada usando um preditor linear que relaciona a resposta média a covariáveis e parâmetros desconhecidos através de uma função de ligação, como acontece nos modelos lineares generalizados. O comportamento das estimativas dos parâmetros deste modelo em amostras pequenas foi estudado por Ospina, Cribari-Neto & Vasconcellos (2005) através da avaliação de diferentes estratégias de estimação pontual e intervalar.

¹ A família beta é caracterizada por sua flexibilidade para produzir distribuições unimodais, uniformes ou bimodais simétricas e assimétricas; adicionalmente, esta família é uma das mais usadas em modelagens paramétricas (JOHNSON, KOTZ & BALAKRISHNAN, 1995).

Nosso principal interesse é estudar aspectos computacionais inerentes à estimação pontual dos parâmetros que definem a classe de modelos de regressão beta proposta por Ferrari & Cribari-Neto (2004). Nós avaliamos os desempenhos de alguns algoritmos de otimização não-linear sob diferentes condições tanto via simulação de Monte Carlo quanto através de aplicações que usam dados reais. O presente artigo encontra-se estruturado como descrito a seguir. Na seção 2, apresentamos o modelo de regressão beta e aspectos da estimação pontual de seus parâmetros. Na seção 3, apresentamos alguns métodos de otimização de grande utilidade para a maximização da função de log-verossimilhança do modelo. A seção 4 contém resultados de simulações de Monte Carlo realizadas para a avaliação dos métodos de otimização usados na estimação pontual dos parâmetros. Na seção 5, são avaliados os desempenhos dos métodos considerados a partir de aplicações com dados reais. Finalmente, as conclusões encontram-se agrupadas na seção 6.

2. O Modelo de Regressão Beta

No modelo de regressão beta, assume-se que a variável resposta (y) segue distribuição beta com função densidade de probabilidade

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad y \in (0,1), \quad (1)$$

onde $p > 0$, $q > 0$ e $\Gamma(\cdot)$ é a função gama. A média e a variância de y são, respectivamente, $E(y) = \frac{p}{p+q}$ e $\text{var}(y) = \frac{pq}{(p+q)^2(p+q+1)}$.

Ferrari & Cribari-Neto (2004) definem a estrutura de regressão usando uma reparametrização da densidade (1), indexando-a pelos parâmetros de locação e precisão μ e

ϕ , respectivamente, sendo $\mu = \frac{p}{p+q}$ e $\phi = p+q$. Assim,

$$E(y) = \mu \quad \text{e} \quad \text{var}(y) = \frac{V(\mu)}{1+\phi},$$

em que $V(\mu) = \mu(1-\mu)$, de tal forma que μ é a média da variável resposta e ϕ pode ser interpretado como um *parâmetro de precisão*, no sentido de que, para μ fixo, quanto maior for o valor de ϕ , menor será a variância de y . A densidade de probabilidade da variável

resposta, y , pode ser re-escrita como

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \text{ para } 0 < y < 1, \quad (2)$$

onde $0 < \mu < 1$ e $\phi > 0$. Esta densidade pode apresentar muitas formas diferentes dependendo dos valores assumidos pelos parâmetros μ e ϕ .

Sejam y_1, \dots, y_n variáveis aleatórias independentes, o modelo é definido assumindo que cada y_t , $t = 1, \dots, n$, segue distribuição beta com densidade (2), de média μ_t e precisão desconhecida ϕ , e que

$$g(\mu_t) = \sum_{i=1}^k x_{ti} \beta_i = \eta_t, \quad (3)$$

em que $\beta = (\beta_1, \dots, \beta_k)^T$ é um vetor de parâmetros desconhecidos ($\beta \in \mathfrak{R}^k$); as covariáveis x_{t1}, \dots, x_{tk} são assumidas fixas e conhecidas. Aqui, $g(\cdot)$ é uma função de ligação estritamente monótona e duas vezes diferenciável com domínio no intervalo $(0,1)$ e contra-domínio \mathfrak{R} . Cabe notar que a variância de y_t é função de μ_t e, como consequência, dos valores das covariáveis. Assim, variâncias não-constantes da resposta são naturalmente acomodadas no modelo. Podem ser usadas diferentes funções de ligação, tais como a função logit $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$, a função probit $g(\mu) = \Phi^{-1}(\mu)$, onde $\Phi(\cdot)$ representa a função de distribuição normal padrão, e a função log-log complementar $g(\mu) = \log\{-\log\{1-\mu\}\}$, entre outras. Para maiores detalhes sobre funções de ligação, ver McCullagh & Nelder (1989).

A função de log-verossimilhança baseada em uma amostra de n observações independentes é dada por

$$\ell(\beta, \phi) = \sum_{i=1}^n \ell_i(\mu_i, \phi), \quad (4)$$

onde

$$\ell_i(\mu_i, \phi) = \log\Gamma(\phi) - \log\Gamma(\mu_i\phi) - \log\Gamma((1-\mu_i)\phi) + (\mu_i\phi - 1)\log y_i + \{(1-\mu_i)\phi - 1\}\log(1-y_i),$$

com μ_t definido segundo (3). Como (4) é uma reparametrização invertível da log-verossimilhança beta, pode-se garantir que o estimador de máxima verossimilhança é único (Ospina *et al.*, 2006).

A função escore é dada por $(U_\beta(\beta, \phi)^T, U_\phi(\beta, \phi))^T$, onde

$$U_{\beta}(\beta, \phi) = \frac{\partial \ell_i(\beta, \phi)}{\partial \beta_i} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \phi)}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_i} = \phi \sum_{i=1}^n (y_i^* - \mu_i^*) \frac{1}{g'(\mu_i)} x_{it},$$

em que $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$ e $\mu_i^* = \psi(\mu_i\phi) - \psi((1-\mu_i)\phi)$, $\psi(\cdot)$ denotando a função digama.

Em notação matricial,

$$U_{\beta}(\beta, \phi) = \phi X^T T (y^* - \mu^*), \quad (5)$$

em que X é uma matriz $n \times k$ cuja t -ésima linha é x_t^T , $y^* = (y_1^*, \dots, y_n^*)^T$,

$\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$ e $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$. De forma similar, pode-se mostrar que a função escore para o parâmetro ϕ pode ser escrita como

$$U_{\phi}(\beta, \phi) = \sum_{i=1}^n \{\mu_i(y_i^* - \mu_i^*) + \log(1-y_i) - \psi((1-\mu_i)\phi) + \psi(\phi)\}. \quad (6)$$

Os estimadores de máxima verossimilhança (EsMV) de β e ϕ podem ser obtidos igualando (5) e (6) a zero e resolvendo o sistema de equações resultante. Dado que a solução destas equações não possui forma fechada, ela deve ser obtida por maximização numérica da função de log-verossimilhança usando algoritmos de otimização não-linear. Estes algoritmos requerem a especificação de um valor inicial para os parâmetros no esquema iterativo. Ferrari & Cribari-Neto (2004) sugerem utilizar

$$\beta_0 = (X^T X)^{-1} X^T z,$$

onde $z = (g(y_1), \dots, g(y_n))$, como valor inicial do vetor β . No que tange ao parâmetro de dispersão, os autores sugerem usar como valor inicial

$$\sigma_0 = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\mu}_i(1-\tilde{\mu}_i)}{\tilde{\sigma}_i^2} - 1,$$

onde $\tilde{\mu}_i = g^{-1}(x_i^T (X^T X)^{-1} X^T z)$ e $\tilde{\sigma}_i^2 = \frac{\tilde{e}^T \tilde{e}}{(n-k)\{g'(\tilde{\mu}_i)\}^2}$, sendo $\tilde{e} = z - (X^T X)^{-1} X^T z$ o

vetor de resíduos de mínimos quadrados da regressão linear sob a resposta transformada. Foi obtida uma expressão para a matriz de informação de Fisher. Sejam $W = \text{diag}\{w_1, \dots, w_n\}$, com

$$w_i = \phi \{\psi'(\mu_i\phi) + \psi'((1-\mu_i)\phi)\} \frac{1}{\{g'(\mu_i)\}^2},$$

$c = (c_1, \dots, c_n)^T$ com elemento típico $c_i = \phi \{\psi'(\mu_i\phi)\mu_i - \psi'((1-\mu_i)\phi)(1-\mu_i)\}$, em que $\psi'(\cdot)$ é

a função trigama, e $D = \text{diag}\{d_1, \dots, d_n\}$, onde $d_i = \psi'(\mu_i, \phi)\mu_i^2 + \psi'((1-\mu_i)\phi)(1-\mu_i)^2 - \psi'(\phi)$.

A matriz de informação de Fisher pode ser escrita como

$$K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}, \quad (7)$$

onde $K_{\beta\beta} = \phi X^T W X$, $K_{\beta\phi} = K_{\phi\beta}^T = X^T T c$ e $K_{\phi\phi} = \text{tr}(D)$. Em contraste ao que ocorre na classe dos modelos lineares generalizados, os parâmetros β e ϕ não são ortogonais. É possível obter aproximações para as quantidades w_i , c_i , d_i , que facilitam o cálculo da matriz de informação de Fisher quando $\mu_i\phi$ e $(1-\mu_i)\phi$ são grandes. Elas são dadas por

$$\begin{aligned} w_i &\approx \frac{1}{\phi\mu_i(1-\mu_i)} = \frac{1}{\phi V(\mu_i)}, \\ c_i &\approx \phi \left[\mu_i \frac{1}{\phi\mu_i(1-\mu_i)} - \frac{1}{\phi(1-\mu_i)} \right] = 0, \\ d_i &\approx \frac{(1-\mu_i)}{\phi} + \frac{\mu_i}{\phi} - \frac{1}{\phi} = 0. \end{aligned}$$

Logo, $K_{\beta\beta} \approx X^T W_{\beta\beta} X$, com $W_{\beta\beta} = \text{diag}\left\{\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{\phi}{V(\mu_i)}\right\}$, $K_{\beta\phi} = K_{\phi\beta}^T \approx 0$ e $K_{\phi\phi} \approx 0$, onde 0

é um vetor de zeros de dimensão $k \times 1$. Desta forma, à medida que ϕ aumenta, os blocos $K_{\beta\phi}$ e $K_{\phi\beta}$ tendem a zero, assim como $K_{\phi\phi}$. Porém, quando $\phi \rightarrow \infty$, a derivada em (6) tende a zero e a curvatura tende para zero, isto é, a função de log-verossimilhança torna-se menos curva, o que faz com que ϕ seja de difícil estimação.

Usando a expressão padrão para a inversa de matrizes particionadas (ver RAO (1973) e HARVILLE (1997), entre outros), é possível mostrar que a inversa da matriz de informação de Fisher (7) é

$$K(\beta, \phi)^{-1} = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix},$$

com

$$\begin{aligned} K^{\beta\beta} &= (X^T W_{\beta\beta} X)^{-1} \left\{ I_k + \frac{X^T T c c^T T^T X (X^T W_{\beta\beta} X)^{-1}}{\gamma} \right\} \\ &= K_{\beta\beta}^{-1} \left\{ I_k + \frac{K_{\beta\phi} K_{\phi\beta} K_{\beta\beta}^{-1}}{\gamma} \right\}, \end{aligned}$$

em que $\gamma = \text{tr}(\text{diag}(d_i))K_{\phi\beta}K_{\beta\beta}^{-1}K_{\beta\phi}$. Aqui, I_k representa a matriz identidade de dimensão k . Adicionalmente,

$$K^{\beta\phi} = (K^{\phi\beta})^T = -\frac{1}{\gamma}(X^T W_{\beta\beta} X)^{-1} X^T T c = -\frac{1}{\gamma} K_{\beta\beta}^{-1} K_{\beta\phi}$$

e $K^{\phi\phi} = \frac{1}{\gamma}$. Sob condições de regularidade, quando o tamanho amostral é grande,

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N_{k+1} \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K(\beta, \phi)^{-1} \right),$$

com \sim denotando aproximadamente distribuído, $\hat{\beta}$ e $\hat{\phi}$ sendo os EsMV de β e ϕ , respectivamente, e N_{k+1} denotando distribuição normal $(k+1)$ -variada. Usando aproximações obtidas a partir da normalidade assintótica do estimador de máxima verossimilhança, foram obtidas estatísticas úteis para a realização de testes de hipóteses. Adicionalmente, os autores apresentaram algumas medidas de diagnóstico para o modelo de regressão beta. Posteriormente, Espinheira, Ferrari & Cribari-Neto (2006) propuseram medidas de diagnóstico baseadas em resíduos derivados do algoritmo iterativo escore de Fisher na estimação do parâmetro β quando ϕ é fixo.

3. Métodos de Otimização Numérica

Métodos de otimização de funções são de grande importância na resolução de problemas estatísticos cujas soluções correspondem a pontos de máximo ou de mínimo de funções de interesse. Embora alguns dos métodos sejam de minimização de funções, eles também podem ser usados para encontrar o ponto de máximo apenas trocando o sinal da função e minimizando a função modificada.

3.1. Métodos Gradiente

Talvez o método mais conhecido de busca unidimensional de raízes de funções seja o de Newton, também conhecido como Newton-Raphson. Geometricamente, o método consiste em traçar a reta tangente à função no ponto x_i e tomar como x_{i+1} o ponto determinado pela intersecção desta reta com o eixo das abcissas, avaliando a função nesse ponto e traçando uma nova tangente. Encontra-se, assim, uma melhor aproximação da raiz a cada iteração. O

processo pára quando é alcançada uma precisão estipulada. Este método não está restrito a apenas uma dimensão e pode ser facilmente generalizado para múltiplas dimensões (Press, Teukolsky, Vetterling & Flannery, 1992). O algoritmo de maximização pode ser obtido a partir da expansão da função $f(\theta)$ em série de Taylor em torno de um ponto $\theta(k)$ até a segunda ordem, o que fornece a seguinte aproximação quadrática:

$$f(\theta) \cong f(\theta(k)) + (\theta - \theta(k))^T \nabla f(\theta(k)) + \frac{1}{2} (\theta - \theta(k))^T H(\theta(k)) (\theta - \theta(k)) = q(\theta),$$

onde $\nabla f(\theta(k))$ é o gradiente da função $f(\theta)$ avaliado em $\theta(k)$, $H(\theta(k))$ é a matriz hessiana avaliada no ponto $\theta(k)$ e θ pertence a uma vizinhança de $\theta(k)$. O próximo ponto da iteração, denotado por $\theta(k+1)$, é dado pelo máximo de $q(\theta)$. Com este fim, primeiro se determina o gradiente de $q(\theta)$, isto é,

$$\nabla q(\theta) = \nabla f(\theta(k)) + H(\theta(k))(\theta - \theta(k));$$

fazendo $\nabla q(\theta) = 0$ e supondo que a inversa de $H(\theta(k))$ existe, tem-se que

$$\theta = \theta(k) - [H(\theta(k))]^{-1} \nabla f(\theta(k)),$$

θ sendo uma aproximação para o ponto maximizador da função $f(\theta)$. Assim, aproximações para o ponto de máximo desta função são obtidas através da lei de recorrência geral

$$\theta(k+1) = \theta(k) - s(k) [H(\theta(k))]^{-1} \nabla f(\theta(k)),$$

onde $s(k)$ é um escalar determinado por um procedimento de busca linear a partir de $\theta(k)$ na direção de $-[H(\theta(k))]^{-1} \nabla f(\theta(k))$ de tal forma que $f(\theta)$ cresça ao longo desta direção. Portanto, H deve ser negativa-definida para todo k . Porém, quando θ está longe do ponto de máximo, não há garantia de que a matriz hessiana satisfaça esta condição. Assim, o incremento $-s(k) [H(\theta(k))]^{-1} \nabla f(\theta(k))$ pode mover o ponto $\theta(k)$ para um ponto $\theta(k+1)$ no qual o valor da função é menor, podendo acarretar não convergência do algoritmo. Um outro inconveniente deste método é a necessidade do uso da matriz hessiana, pois o cálculo explícito da matriz de segundas derivadas é, muitas vezes, um processo computacionalmente custoso e convidativo ao erro. Para mais detalhes sobre este método de otimização, ver Nocedal & Wright (1999) e Press *et al.* (1992). Uma opção é utilizar o algoritmo BFGS, que não sofre deste problema.

Os métodos do tipo Newton apresentam taxa rápida de convergência local, tipicamente quadrática. Quando uma vizinhança da solução é alcançada, freqüentemente ocorre a convergência com alta precisão em poucas iterações. Porém, longe do ponto ótimo, onde os termos de ordem superior na expansão de Taylor são importantes, o método pode fornecer so-

luções imprecisas. Pode haver não convergência global, pois pode haver convergência a diferentes soluções dependendo do ponto inicial escolhido. Para contornar este problema têm sido desenvolvidas modificações que combinam as propriedades de rápida convergência local do método Newton com estratégias globalmente convergentes que garantam progressos na direção da solução a cada iteração; ver Sherman (1978), Moré & Sorensen (1979), Goldfarb (1980), Denbo, Eisenstat & Steihaug (1982) e Schnabel & Eskow (1991), entre outros.

3.2. Métodos Quasi-Newton

Os métodos de *métrica variável* ou quasi-Newton derivam do método de Newton, tendo o mesmo objetivo: acumular informações das sucessivas minimizações em linha que levem ao ótimo exato de uma forma quadrática em d dimensões. A principal diferença entre estes métodos reside no cálculo da inversa da matriz hessiana, que é aproximada por uma matriz simétrica e positiva-definida $Q(k)$ tal que

$$\left[\lim_{k \rightarrow \infty} Q(k) \right] = -H^{-1}.$$

A matriz $Q(k)$ é atualizada depois de cada passo para levar em conta o conhecimento adicional ganho durante a iteração. As atualizações usam o fato de que mudanças no gradiente proporcionam informação sobre a segunda derivada da função ao longo da direção de busca. A matriz inicial mais comumente usada para $Q(k)$ é a matriz identidade de mesma ordem, pois ela é positiva-definida e simétrica, resultando, assim, em aproximações $Q(k)$ positivas-definidas e simétricas. O método quasi-Newton mais utilizado é o *BFGS* (*Broyden, Fletcher, Goldfarb e Shanno*), cuja forma recursiva de atualização da aproximação da matriz hessiana é dada pela expressão

$$Q(k+1) = Q(k) + \frac{(\theta(k+1) - \theta(k)) \otimes ((\theta(k+1) - \theta(k)))}{(\theta(k+1) - \theta(k))^T (\nabla f(\theta(k+1)) - \nabla f(\theta(k)))} \\ - \frac{[Q(k)(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))] \otimes [Q(k)(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))]}{(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))^T Q(k)(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))} \\ + (\nabla f(\theta(k+1)) - \nabla f(\theta(k)))^T Q(k)(\nabla f(\theta(k+1)) - \nabla f(\theta(k))) U \otimes U,$$

onde o operador \otimes denota produto de Kronecker e U é o vetor coluna dado por

$$U = \frac{\theta(k+1) - \theta(k)}{(\theta(k+1) - \theta(k))^T (\nabla f(\theta(k+1)) - \nabla f(\theta(k)))} \\ - \frac{Q(k)(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))}{(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))^T Q(k)(\nabla f(\theta(k+1)) - \nabla f(\theta(k)))}.$$

Dessa forma, mesmo quando θ está longe do máximo, as matrizes $Q(k)$ garantem que os pontos se movem em direção ascendente. De forma semelhante ao método de Newton, o esquema iterativo é dado por

$$\theta(k+1) = \theta(k) + s(k)Q(\theta(k))\nabla f(\theta(k)),$$

onde $s(k)$ é o tamanho do passo. Embora o método de Newton convirja mais rapidamente (ou seja, quadraticamente), seu custo por iteração é maior devido à necessidade da solução de um sistema linear. Uma vantagem mais notável do método BFGS é que ele não requer o cálculo de segundas derivadas. Embora seja raro, pode acontecer que erros de arredondamento produzam uma matriz $Q(\theta(k))$ quase singular ou não positiva-definida. Isto pode ser sério, pois as direções de busca neste caso não seriam em descida e, adicionalmente, matrizes quase singulares tendem a induzir matrizes quase singulares subsequentes. Para lidar com este inconveniente, têm sido implementadas variantes do método que consistem em construir uma aproximação de H ao invés de H^{-1} .

3.3. Método Simplex

O procedimento de otimização deste método começa pela escolha dos $d+1$ pontos onde será feita a avaliação da resposta, com o objetivo de forçar o simplex a mover-se para a região de resposta ótima através de operações de reflexão, expansão e contração. Este algoritmo não precisa das derivadas da função e tipicamente não é eficiente em termos do número de avaliações da função até convergência. Nelder & Mead (1965) desenvolveram o algoritmo conhecido como *simplex downhill*, que considera a minimização de uma função de d variáveis, sem restrições. O método pode ser descrito como a seguir:

Sejam P_0, P_1, \dots, P_d os $(d+1)$ pontos no espaço d -dimensional definindo o simplex atual. Denota-se por y_i o valor da função em P_i e define-se h como o subíndice tal que $y_h = \max_i(y_i)$ e l como o subíndice tal que $y_l = \min_i(y_i)$. Ainda, define-se \bar{P} como o centróide dos pontos com $i \neq h$ e $[P_i P_j]$ como a distância entre P_i e P_j . Em cada etapa do processo, P_h é trocado por um novo ponto usando três operações:

- Reflexão de P_h : Denotada por P^* , suas coordenadas são definidas pela relação

$$P^* = (1 + \alpha)\bar{P} - \alpha P_h,$$

em que α (coeficiente de reflexão) é uma constante positiva. Assim, P^* está

sobre a reta que une P_h e \bar{P} com $[P^*\bar{P}] = \alpha[P_h\bar{P}]$. Se y^* estiver entre y_h e y_l , então P_h é trocado por P^* e o processo é reiniciado com o novo simplex.

- Expansão de P_h : Se $y^* < y_l$, isto é, se a operação de reflexão produz um novo ponto mínimo, então expande-se P^* a P^{**} pela relação

$$P^{**} = \gamma P^* + (1 - \gamma)\bar{P}.$$

O coeficiente de expansão γ , que é maior do que 1, é a razão entre as distâncias $[P^{**}\bar{P}]$ e $[P^*\bar{P}]$. Se $y^{**} < y_l$, troca-se P_h por P^{**} e o processo é reiniciado; se $y^{**} > y_l$, então a expansão é falha e deve-se trocar P_h por P^* antes de recomeçar o processo.

- Contração de P_h : Se na reflexão de P a P^* se encontra $y^* > y_l$ para todo $i \neq h$, isto é, que a troca de P por P^* resulta em um y^* máximo, então define-se um novo P_h sendo igual ao P_h anterior ou a P^* (qualquer um deles possui menor valor y). Neste caso, toma-se

$$P^{**} = \beta P_h + (1 - \beta)\bar{P}.$$

O coeficiente de contração β está entre 0 e 1 e é a razão das distâncias $[P^{**}\bar{P}]$ e $[P\bar{P}]$. Assim, aceita-se P^{**} para P_h e o processo é reiniciado, a não ser que $y^{**} > \min(y_h, y^*)$, isto é, o ponto depois da operação de contração é "pior" do que o "melhor" entre P_h e P^* . Quando ocorre tal falha na contração, todos os P_i são trocados por $(P_i + P_l)/2$ e o processo é reiniciado.

Finalmente, o critério de parada do algoritmo consiste em comparar o "erro-padrão" dos y 's, $\sqrt{\{\sum (y_i - \bar{y})^2 / n\}}$, com um valor pré-fixado. O processo pára quando este valor é maior que o erro-padrão. O sucesso deste critério depende do simplex não chegar a ser muito pequeno em relação à curvatura da superfície até que o mínimo seja alcançado.

4. Avaliação Numérica

Através de simulações de Monte Carlo, avaliamos os desempenhos de diferentes métodos de otimização não-linear no que tange à maximização da função de log-verossimilhança do modelo de regressão beta. No estudo de simulação, foram comparadas as estimativas obtidas no processo de otimização através dos métodos BFGS, Newton e simplex, descritos na seção 3, sob diferentes cenários. Os dados foram gerados do modelo de regressão beta dado por

$$g(\mu_t) = \beta_0 + \beta_1 x_t, \quad t = 1, \dots, n, \quad (8)$$

onde $g(\cdot)$ é a função de ligação logit. Os valores da variável regressora x_t foram obtidos como realizações independentes da distribuição normal padrão. A matriz de variáveis regressoras X permaneceu constante ao longo do experimento e, em cada réplica de Monte Carlo, foi gerada uma amostra aleatória da variável resposta $y = (y_1, \dots, y_n)^T$, onde $y_t \sim \mathbf{B}(\mu_t, \phi)$, $t = 1, \dots, n$. Os valores considerados para os parâmetros de intercepto e inclinação foram $\beta_0 = 1,0$ e $\beta_1 = 0,1$, respectivamente; o valor do parâmetro de precisão foi fixado em $\phi = 35$; e os tamanhos amostrais utilizados foram $n = 20, 60, 100, 200$ e 500 . Para cada tamanho amostral, foram calculadas média, variância e os critérios de avaliação erro absoluto médio (MAE) e erro absoluto percentual médio (MAPE) nas 1 000 réplicas de Monte Carlo. A simulação foi realizada através de procedimentos implementados na linguagem de programação Ox; ver <http://www.doornik.com>.

4.1. Resultados

Inicialmente avaliamos as precisões dos diferentes métodos na estimação dos parâmetros do modelo (8). Os valores iniciais (β_{0_0} , β_{1_0} e ϕ_0) para o processo iterativo da estimação dos parâmetros foram obtidos como descrito na seção 2 e são apresentados na Tabela 1. Os valores β_{0_0} e β_{1_0} estão próximos dos verdadeiros valores dos parâmetros β_0 e β_1 , enquanto o valor do ϕ_0 não se encontra tão próximo do valor verdadeiro.

Em resumo, o valor inicial para a estimação do parâmetro de locação (μ_t) está próximo do valor verdadeiro, diferentemente do que ocorre com o parâmetro de precisão do modelo (ϕ). Isto pode afetar negativamente a eficiência dos métodos no que tange à estimação do

parâmetro ϕ , pois o algoritmo pode necessitar maior número de iterações, levando assim maior tempo para atingir o ponto ótimo.

Tabela 1 – Valores iniciais na estimação dos parâmetros $\beta_0 = 1,0$, $\beta_1 = 0,1$ e $\phi = 35$

n	β_1	β_0	ϕ_0
20	0,89446	0,12707	31,44047
60	1,05745	0,10087	23,63574
100	0,99091	0,12537	29,27570
200	1,08458	0,09968	29,06153
500	1,01183	0,10887	32,45599

Os resultados na Tabela 2 revelam que, para o parâmetro de intercepto do modelo, todos os métodos proporcionam valores semelhantes no que tange às estimativas. Os métodos BFGS e Newton forneceram os mesmos resultados para as quantidades calculadas, e o método simplex foi o que apresentou maior variabilidade das estimativas, mesmo para tamanhos de amostra grandes.

Tabela 2 – Resultados da simulação para as estimativas de máxima verossimilhança do parâmetro $\beta_0 = 1,0$

n	Medida	BFGS	Newton	simplex
20	Média	1,0032	1,0032	1,0028
	Variância	0,0275	0,0275	0,0279
	MAE	0,1314	0,1314	0,1314
	MAPE	13,1400	13,1400	13,1390
60	Média	1,0002	1,0002	1,0003
	Variância	0,0108	0,0108	0,0108
	MAE	0,0830	0,0830	0,0830
	MAPE	8,2950	8,2950	8,2980
100	Média	1,0017	1,0017	0,9999
	Variância	0,0059	0,0059	0,0075
	MAE	0,0611	0,0611	0,0691
	MAPE	6,1120	6,1120	6,9130
200	Média	1,0001	1,0001	0,9950
	Variância	0,0030	0,0030	0,0080
	MAE	0,0434	0,0434	0,0432
	MAPE	4,3420	4,3420	4,3240
500	Média	0,9999	0,9999	0,9676
	Variância	0,0013	0,0013	0,0681
	MAE	0,0284	0,0284	0,0264
	MAPE	2,8380	2,8380	2,6350

Realizando a mesma análise para o parâmetro de inclinação do modelo (β_1), achou-se que os resultados obtidos (que são omitidos, por brevidade) foram análogos aos referentes ao parâmetro de intercepto. Todos os métodos forneceram estimativas

semelhantes, tendo os métodos BFGS e Newton apresentado as mesmas variabilidades. Para tamanhos amostrais pequenos, o método simplex apresentou variância igual às dos métodos BFGS e Newton, enquanto para os maiores tamanhos de amostra ele apresentou a maior variabilidade. No que tange ao parâmetro de precisão (ϕ), observamos na Tabela 3 que, para tamanhos de amostra pequenos, existe viés significativo nas estimativas e que a variância de tais estimativas é consideravelmente alta. Os valores estimados fornecidos pelos diferentes métodos foram muito homogêneos, sendo o método simplex aquele que apresentou sempre a estimativa mais próxima do verdadeiro valor do parâmetro. Todos os métodos apresentaram valores parecidos dos critérios MAE e MAPE. Em geral, o método simplex apresentou as menores taxas de erro.

Em geral, pode-se deduzir das Tabelas 2 e 3 que, em princípio, todos os métodos avaliados são competitivos no que tange à obtenção das estimativas de máxima verossimilhança dos parâmetros do modelo de regressão beta. O fato de que os métodos de Newton e BFGS sempre fornecem as mesmas estimativas, assim como a mesma variabilidade e taxas de erro, indica que não está havendo dificuldade no que concerne à inversão da matriz hessiana no método de Newton. Para os valores de n considerados, observou-se que as taxas MAE e MAPE foram sempre maiores na estimação de ϕ , evidenciando a dificuldade dos métodos para estimar com precisão esse parâmetro.

Tabela 3 – Resultados da simulação para as estimativas de máxima verossimilhança do parâmetro $\phi = 35$

n	Medida	BFGS	Newton	simplex
20	Média	43,9741	43,9741	43,9715
	Variância	273,7639	273,7639	274,2863
	MAE	12,9606	12,9606	12,9506
	MAPE	37,0300	37,0300	37,0020
60	Média	37,5441	37,5441	37,4428
	Variância	51,7303	51,7303	51,7413
	MAE	5,7365	5,7365	5,7397
	MAPE	16,3900	16,3900	16,3990
100	Média	36,5494	36,5494	36,4109
	Variância	27,6131	27,6131	27,9915
	MAE	4,2398	4,2398	4,1821
	MAPE	12,1140	12,1140	11,9490
200	Média	35,7115	35,7115	35,5278
	Variância	12,8998	12,8998	19,3014
	MAE	2,8716	2,8716	2,8565
	MAPE	8,2040	8,2040	8,1610
500	Média	35,2698	35,2698	35,2085
	Variância	4,9738	4,9738	7,9740
	MAE	1,7785	1,7785	1,6524
	MAPE	5,0820	5,0820	4,7210

Um segundo experimento de simulação foi realizado, desta vez usando um modelo com duas covariáveis. As conclusões extraídas foram as mesmas que no caso de uma covariável, isto é, todos os métodos forneceram estimativas semelhantes para o parâmetro de locação, havendo uma certa superioridade dos métodos que usam a informação da matriz hessiana além do gradiente da função. No referente ao parâmetro de precisão, houve maior dificuldade na determinação do método que apresenta o melhor desempenho de acordo com os critérios de avaliação. Independentemente do tamanho de amostra, o método simplex foi o que tipicamente apresentou as menores taxas de erro, apesar de ter fornecido as estimativas com maior variabilidade.

Visando a verificar se o viés significativo nas estimativas de ϕ está relacionado com o ponto inicial usado no processo de otimização, foram comparados os resultados fornecidos por cada método para diferentes valores iniciais do parâmetro ϕ , a saber: o valor inicial descrito na seção 2 ($\phi_{0_{FC}}$), o verdadeiro valor do parâmetro ($\phi_0 = 35$), um valor menor e outro maior do que o verdadeiro valor do parâmetro ($\phi_0 = 15$ e $\phi_0 = 70$, respectivamente).

Na Tabela 4, apresentamos os resultados fornecidos pelos métodos BFGS e simplex; não apresentamos os resultados referentes ao método de Newton por serem iguais aos do método BFGS. As estimativas apresentam insensibilidade ao ponto inicial e o viés diminui com o aumento do n , como esperado. O método simplex fornece estimativas e valores dos critérios de avaliação semelhantes aos obtidos com os métodos BFGS e Newton. Como esperado, os métodos tendem a apresentar menores taxas de erro quando o ponto inicial coincide com o verdadeiro valor do parâmetro. As estimativas mais próximas do valor verdadeiro foram as fornecidas pelo método simplex. Observamos também que a variabilidade é grande para pequenos tamanhos amostrais, diminuindo rapidamente com o aumento do número de observações.

O tempo gasto por cada método no processo de maximização foi também usado como critério de avaliação. Observamos, na Tabela 5, que o método simplex precisou sempre de mais tempo para convergir. Mais importante ainda é observar que, embora as estimativas fornecidas pelos métodos BFGS e Newton sejam iguais, este último precisou de menos tempo para realizar o processo de otimização. Embora o método simplex conduza às estimativas mais precisas quando o ponto inicial está próximo do valor verdadeiro, este método não é computacionalmente eficiente, pois é muito lento, comportamento que piora quando o tamanho amostral aumenta. Do observado aqui, notamos a eficiência do método de Newton na maximização da logverossimilhança do modelo de regressão (8).

Tabela 4 – Resultados da simulação para as estimativas de máxima verossimilhança do parâmetro $\phi = 35$ com diferentes valores iniciais do parâmetro ϕ

<i>n</i>	Medida	BFGS			Simplex		
		$\phi_0 = 15$	$\phi_0 = 35$	$\phi_0 = 70$	$\phi_0 = 15$	$\phi_0 = 35$	$\phi_0 = 70$
20	Média	43,9741	43,9741	43,9741	43,9626	43,9572	43,9568
	Variância	273,7639	273,7639	273,7641	274,0156	274,1241	274,1579
	MAE	12,9606	12,9606	12,9606	12,9561	12,9536	12,9538
	MAPE	37,0300	37,0300	37,0300	37,0170	37,0100	37,0100
60	Média	37,5441	37,5441	37,5441	37,5482	37,5399	37,5405
	Variância	51,7303	51,7304	51,7303	52,0045	51,7003	51,7582
	MAE	5,7365	5,7365	5,7365	5,7516	5,7344	5,7388
	MAPE	16,3900	16,3900	16,3900	16,4330	16,3840	16,3960
100	Média	36,5494	36,5494	36,5494	36,2801	36,5382	36,5513
	Variância	28,2322	27,6131	27,6131	33,3614	27,6620	27,7911
	MAE	4,2461	4,2398	4,2398	4,2255	4,2450	4,2485
	MAPE	12,1140	12,1140	12,1140	12,0730	12,1280	12,1390
200	Média	35,7115	35,7115	35,7115	35,5183	35,4999	35,5089
	Variância	12,8998	12,8998	12,8998	19,4357	20,3108	19,7923
	MAE	2,8716	2,8716	2,8716	2,8559	2,8563	2,8544
	MAPE	8,2040	8,2040	8,2040	8,1600	8,1610	8,1550
500	Média	35,2698	35,2698	35,2698	35,3605	35,2382	35,4480
	Variância	4,9738	4,9738	4,9738	8,5441	8,9082	8,6196
	MAE	1,7785	1,7785	1,7785	1,6605	1,6527	1,6592
	MAPE	5,0820	5,0820	5,0820	4,7440	4,7220	4,7400

Tabela 5 – Tempo gasto pelos métodos no processo de maximização para diferentes valores iniciais do parâmetro de precisão

<i>n</i>	Método	$\phi_{0,c}$	$\phi_0 = 35$	$\phi_0 = 15$	$\phi_0 = 70$
20	BFGS	00:01:17	00:01:17	00:01:15	00:01:21
	Newton	00:00:48	00:00:50	00:01:00	00:00:58
	simplex	00:07:10	00:08:50	00:05:44	00:06:43
60	BFGS	00:03:59	00:05:58	00:04:25	00:04:08
	Newton	00:02:21	00:02:22	00:03:09	00:02:57
	simplex	00:23:06	00:22:00	00:19:39	00:20:38
100	BFGS	00:06:49	00:04:54	00:08:16	00:07:47
	Newton	00:03:57	00:03:51	00:03:14	00:04:46
	simplex	00:34:13	00:40:34	00:31:09	01:10:23
200	BFGS	00:15:20	00:13:38	01:00:48	01:05:57
	Newton	00:09:05	00:08:02	00:42:36	00:43:21
	simplex	01:14:49	01:19:40	01:16:36	03:35:50
500	BFGS	05:21:10	05:31:19	06:22:38	06:37:29
	Newton	02:44:40	02:50:18	05:17:50	04:54:08
	simplex	12:01:16	08:54:17	07:55:41	20:49:46

Um outro interesse reside em determinar a robustez dos métodos à presença de pontos de alavanca nos dados. Com este fim, foram gerados $n-1$ valores da variável regressora x , a partir da distribuição normal padrão e o n -ésimo ponto de cada amostra foi fixado em 2,5; 3,0; 4,0; 5,0; 6,0 ou 7,0. Neste caso, o critério de comparação foi o número de iterações realizadas por cada método para alcançar o ponto ótimo. A Tabela 6 apresenta os resultados para os tamanhos amostrais $n=20$ e $n=100$. Observamos que o método mais eficiente é o de Newton seguido de longe pelo BFGS; o método que precisa do maior número de iterações é o simplex, com mais de 100 iterações em média. Todos os métodos tendem a aumentar o número de iterações quando a magnitude do ponto de alavanca aumenta. O método que permaneceu mais estável às perturbações nos dados foi o de Newton.

Também verificamos as taxas de convergência dos métodos e encontramos que os métodos BFGS e de Newton convergiram em todos os casos, independentemente da magnitude do ponto de alavanca e do tamanho amostral. O método simplex foi afetado nas suas propriedades de convergência pela existência dos pontos de alavanca, sendo que a taxa de convergência decresceu de aproximadamente 100%, quando o ponto de alavanca teve a menor magnitude, até aproximadamente 80%, quando o ponto de alavanca apresentou a maior magnitude.

Tabela 6 – Número médio de iterações até convergência dos métodos na presença de pontos de alavanca

Método	n	Magnitude do ponto de alavanca					
		2,5	3,0	4,0	5,0	6,0	7,0
BFGS	20	12	12	13	14	16	21
	100	10	11	11	12	13	14
Newton	20	3	3	4	4	5	6
	100	3	3	3	3	4	4
simplex	20	141	144	147	148	151	157
	100	128	133	137	144	147	145

Finalmente, foi avaliada a robustez dos métodos estudados a correlação acentuada entre as variáveis regressoras. Com este fim, foram geradas amostras de tamanho n da distribuição normal bivariada com vetor de médias $\mu = (0,0)$ e matriz de covariâncias

$$\Sigma = \begin{pmatrix} 1 & \sigma_{x_1x_2} \\ \sigma_{x_2x_1} & 1 \end{pmatrix},$$

onde $\sigma_{x_1x_2} = \sigma_{x_2x_1} = 0,1; 0,3; 0,5; 0,7$ ou $0,9$. O modelo empregado foi

$$g(\mu_i) = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2},$$

onde $g(\cdot)$ é a função de ligação logit e os valores dos parâmetros foram fixados em $\beta_0 = 1,0$,

$$\beta_1 = 0,1, \beta_2 = 0,5 \text{ e } \phi = 35.$$

Foram medidos os tempos gastos por cada método no processo de maximização. Os resultados apresentados na Tabela 7 sugerem que o método de Newton continua sendo o mais rápido e que o grau de correlação não afeta o processo de otimização dos métodos BFGS e Newton, pois seus tempos de execução permanecem estáveis para todos os graus de correlação. No que diz respeito ao método simplex, para os maiores tamanhos amostrais há leve tendência de aumento do tempo gasto na otimização quando o grau de correlação é elevado.

Em termos gerais, observou-se que o método de Newton é uma ótima opção para a maximização da função de log-verossimilhança do modelo de regressão beta, devido a sua rapidez, robustez e precisão para atingir o ponto ótimo. O método BFGS é competitivo também, dadas as precisões das suas estimativas, sendo, porém, sempre mais lento do que o método de Newton. Esta pode ser uma indicação de que para o caso específico da função de log-verossimilhança do modelo de regressão beta, a matriz hessiana não é de difícil inversão e está geralmente distante da singularidade.

Tabela 7 – Tempo gasto no processo de maximização em presença de correlação entre as covariáveis (Formato: hh:mm:ss)

<i>n</i>	Método	$\rho = 0,1$	$\rho = 0,3$	$\rho = 0,5$	$\rho = 0,7$	$\rho = 0,9$
20	BFGS	00:00:16	00:00:15	00:00:15	00:00:17	00:00:18
	Newton	00:00:14	00:00:13	00:00:13	00:00:14	00:00:15
	simplex	00:01:23	00:01:09	00:01:09	00:01:23	00:01:23
60	BFGS	00:00:38	00:00:38	00:00:37	00:00:40	00:00:40
	Newton	00:00:29	00:00:31	00:00:31	00:00:28	00:00:32
	simplex	00:03:14	00:03:03	00:03:02	00:03:31	00:03:13
100	BFGS	00:01:09	00:01:11	00:01:15	00:01:16	00:01:13
	Newton	00:00:55	00:00:59	00:00:54	00:01:04	00:01:01
	simplex	00:06:08	00:07:46	00:08:26	00:12:32	00:08:10
200	BFGS	00:03:51	00:03:57	00:04:09	00:04:13	00:04:11
	Newton	00:03:23	00:03:20	00:03:27	00:03:00	00:03:21
	simplex	00:24:45	00:40:23	00:44:42	01:49:04	02:09:13
500	BFGS	00:38:13	00:37:33	00:39:36	00:37:23	00:40:00
	Newton	00:33:26	00:33:59	00:34:07	00:35:40	00:35:11
	simplex	07:03:33	14:28:53	20:10:40	18:39:46	22:19:55

Os métodos BFGS e de Newton se mostraram robustos a condições comumente verificadas na prática estatística, como elevada correlação entre variáveis regressoras e existência de pontos de alavanca nos dados. O método simplex sempre necessita de muitas iterações para atingir o ponto ótimo e, quando submetido a algumas condições adversas, como a presença de pontos de alavanca, teve sua taxa de convergência afetada.

5. Aplicações

Nesta Seção, são apresentadas e discutidas duas aplicações a dados reais do modelo de regressão beta descrito na seção 2. Para cada conjunto de dados, a função de log-verossimilhança foi maximizada numericamente através dos métodos de otimização não-linear apresentados nas seções 3 e 4. Com a finalidade de avaliar a eficiência e a precisão de cada método, foram comparados os valores obtidos para as estimativas pontuais e seus erros-padrão assim como os valores máximos das funções de log-verossimilhança.

5.1. Aplicação a Dados de Gasolina de Prater

A variável dependente é a proporção de petróleo cru transformado em gasolina após o processo de destilação e fracionamento. As covariáveis são: temperatura em que 10% do petróleo cru é vaporizado e temperatura ($^{\circ}$ F) em que toda a gasolina evapora. Essas covariáveis correspondem aos diferentes tipos de petróleo cru submetidos a experimentação sob diferentes condições de destilação. O conjunto de dados contém 32 observações. A primeira covariável assume dez valores diferentes, que são usados para definir os dez tipos de petróleo cru. A especificação do modelo para a resposta média é

$$g(\mu_t) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_9 x_{t9} + \beta_{10} x_{t10}, \quad t = 1, \dots, 32, \quad (9)$$

em que $g(\cdot)$ é a função de ligação logit, x_{t1}, \dots, x_{t9} representam nove variáveis *dummy* para os primeiros nove tipos de petróleo cru e x_{t10} é a temperatura em que a gasolina é vaporizada.

A Tabela 8 apresenta as estimativas dos parâmetros do modelo (9) e os respectivos erros-padrão assintóticos. Observamos que as estimativas obtidas através dos métodos BFGS e Newton são iguais, sendo o valor máximo encontrado da função de verossimilhança de 84,7976. Como observado na Sub-seção 4.1., os métodos BFGS e Newton tipicamente são capazes de fornecer com boa precisão uma estimativa para o parâmetro de locação, apresentando erros-padrão menores do que os do método simplex. Cabe ressaltar que estes métodos convergiram em menos de 15 iterações, enquanto o método simplex precisou de aproximadamente 12 000 iterações.

Dado que o valor máximo da função de log-verossimilhança fornecido pelos métodos

BFGS e Newton foi superior ao valor máximo encontrado pelo método simplex, podemos considerar que as estimativas obtidas através daqueles métodos são preferíveis. Cabe destacar a diferença dos valores máximos da função de log-verossimilhança e das estimativas dos parâmetros, especialmente ϕ , pelos métodos BFGS, Newton e simplex.

Tabela 8 – Estimativas obtidas para o conjunto de dados de gasolina de Prater

Parâmetro	BFGS	Newton	simplex	Parâmetro	BFGS	Newton	simplex
β_0	-6,1596 (0,1823)	-6,1596 (0,1823)	-6,4152 (0,2109)	β_6	1,0402 (0,106)	1,0402 (0,106)	1,0976 (0,1221)
β_1	1,7277 (0,1012)	1,7277 (0,1012)	1,8147 (0,1167)	β_7	0,5437 (0,1091)	0,5437 (0,1091)	0,5974 (0,1257)
β_2	1,3226 (0,1179)	1,3226 (0,1179)	1,3991 (0,1359)	β_8	0,4959 (0,1089)	0,4959 (0,1089)	0,521 (0,1257)
β_3	1,5723 (0,1161)	1,5723 (0,1161)	1,6445 (0,1340)	β_9	0,3858 (0,1186)	0,3858 (0,1186)	0,4434 (0,1360)
β_4	1,0597 (0,1024)	1,0597 (0,1024)	1,1336 (0,1179)	β_{10}	0,011 (0,0004)	0,011 (0,0004)	0,0115 (0,0005)
β_5	1,1338 (0,1035)	1,1338 (0,1035)	1,1719 (0,1195)	ϕ	440,2784 (110,0256)	440,2784 (110,0256)	338,5943 (84,609)
Máximo	84,7976	84,7976	83,2066				

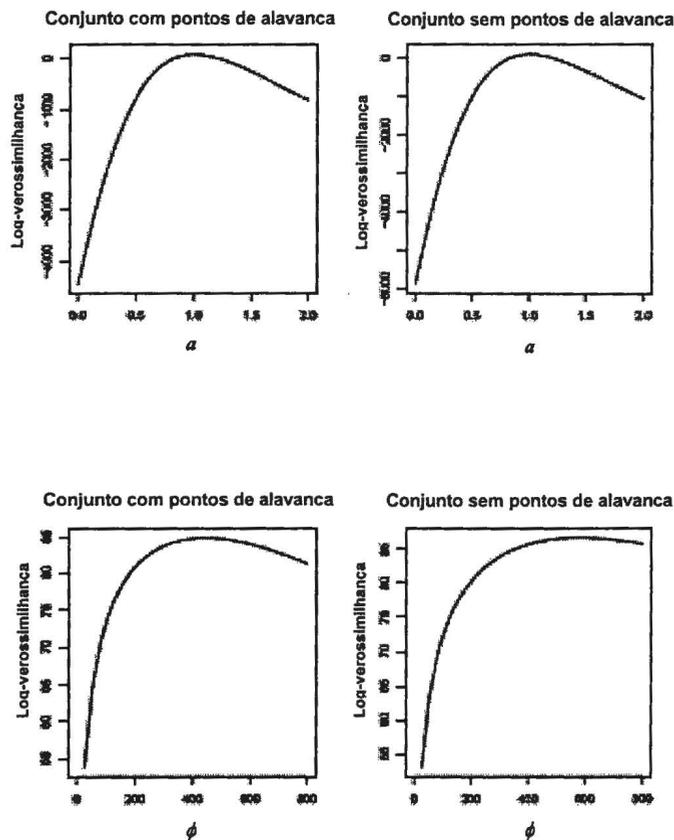
Neste conjunto de dados, foi detectado um ponto de alavanca, o qual foi removido, tendo sido os parâmetros do modelo então estimados novamente. Os resultados obtidos encontram-se apresentados na Tabela 9, onde é possível observar que o maior impacto do ponto de alavanca reside na estimação do parâmetro ϕ , cuja estimativa aumentou notavelmente com mudança relativa de 31,23% e 21,96% para os métodos BFGS (e Newton) e simplex, respectivamente. Adicionalmente, pode-se observar que os erros-padrão das estimativas dos parâmetros de locação não apresentaram mudanças significativas; entretanto, o erro-padrão da estimativa do parâmetro de precisão aumentou significativamente.

Tabela 9 – Estimativas obtidas para o conjunto de dados de gasolina de Prater sem pontos de alavanca

Parâmetro	BFGS	Newton	simplex	Parâmetro	BFGS	Newton	simplex
β_0	-6,3565 (0,1716)	-6,3565 (0,1716)	-6,5618 (0,205)	β_6	1,0577 (0,0929)	1,0577 (0,0929)	1,0922 (0,1094)
β_1	1,8869 (0,1002)	1,8869 (0,1002)	1,9206 (0,1186)	β_7	0,5652 (0,0956)	0,5652 (0,0956)	0,5315 (0,1135)
β_2	1,3704 (0,1042)	1,3704 (0,1042)	1,4047 (0,1233)	β_8	0,5007 (0,0953)	0,5007 (0,0953)	0,4751 (0,1128)
β_3	1,6251 (0,1028)	1,6251 (0,1028)	1,645 (0,1219)	β_9	0,3852 (0,1038)	0,3852 (0,1038)	0,3732 (0,1225)
β_4	1,0807 (0,0898)	1,0807 (0,0898)	1,1136 (0,1058)	β_{10}	0,0115 (0,0004)	0,0115 (0,0004)	0,012 (0,0005)
β_5	1,1516 (0,0907)	1,1516 (0,0907)	1,1476 (0,1071)	ϕ	577,791 (146,7204)	577,7907 (146,7204)	412,949 (104,8543)
Máximo	86,6187	86,6187	84,6132				

Dada a diferença notável entre as estimativas obtidas através dos três métodos para o parâmetro ϕ , foram construídos gráficos com o objetivo de examinar como variações em $\beta = (\beta_0, \beta_1, \dots, \beta_{10})^T$ e em ϕ alteram o valor da função de log-verossimilhança $\ell(\beta, \phi)$. Em primeiro lugar, foi investigado o impacto de variações no vetor de parâmetros de locação β ; este vetor, todavia, possui mais de dois parâmetros e não é possível fazer o gráfico desejado em três dimensões. Portanto, foi fixado o valor de ϕ como aquele estimado pelo método BFGS (e de Newton) e o valor de β variou em forma linear, com a função $\ell(\beta, \phi)$ substituída por $\ell(a\hat{\beta}, \phi)$. Os valores estabelecidos de $\hat{\beta}$ foram aqueles encontrados pelo método BFGS (e de Newton); o valor da constante a variou. Posteriormente, foi observada a variação da função de log-verossimilhança para diferentes valores do parâmetro de precisão ϕ , tendo sido considerados os valores $\hat{\beta}$ obtidos através dos métodos BFGS e Newton como fixos, tendo o parâmetro ϕ sido sujeito à variação. Este procedimento foi realizado tanto para os dados com ponto de alavanca quanto para o conjunto de dados sem este ponto. Os gráficos de avaliação são apresentados na Figura 1.

Figura 1 – Funções de log-verossimilhança para diferentes valores dos parâmetros de locação e precisão no conjunto de dados de gasolina de Prater.



No que se refere à variação do vetor β , pode-se notar a concavidade da curva que evidencia que o ponto de máximo está no valor $\alpha = 1$, ou seja, a otimização pelos métodos BFGS e Newton foi bem sucedida. No que tange ao parâmetro de precisão, observa-se que a concavidade da função se vê significativamente afetada pela eliminação do ponto de alavanca. Verifica-se que, no conjunto de dados que inclui o ponto de alavanca, há pouca variação na função para valores de ϕ entre 200 e 500, aproximadamente. No caso em que não há ponto de alavanca, o intervalo em que alterações em ϕ produzem pouca variação na função se estende aproximadamente de 400 a 700. Assim, pode-se concluir que a dificuldade dos métodos para fornecer estimativas precisas de ϕ deve-se ao fato de que uma variação no parâmetro ϕ não implica necessariamente uma variação significativa na função de log-verossimilhança.

5.2. Aplicação a Dados de Quociente Intelectual (QI)

Esta aplicação utiliza os dados de Pammer & Kevan (2004), posteriormente analisados por Smithson & Verkuilen (2006). A variável resposta são pontuações obtidas por 44 crianças

em um teste de leitura e as covariáveis são: condição de dislexia ou não-dislexia (x_1), quociente intelectual não-verbal convertido a pontuações z (x_2) e uma variável de interação (x_3). Participaram do estudo 19 crianças disléxicas e 25 controles com idades entre 8 anos e 5 meses e 12 anos e 3 meses, que foram selecionadas de escolas primárias na cidade de Canberra, capital da Austrália. A covariável x_1 assume valor 1 quando a criança é disléxica e -1 caso contrário. As pontuações observadas foram transformadas linearmente de sua escala original ao intervalo unitário aberto (0,1), tomando primeiro $y' = (y-a)/(b-a)$, onde a e b representam a maior e a menor pontuações possíveis, respectivamente; posteriormente, a escala foi comprimida para evitar zeros e uns tomando $y'' = [y'(n-1)+0,5]/n$, onde n representa o tamanho amostral. Para este conjunto de dados, a estrutura do modelo de regressão beta é

$$g(\mu_t) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad t = 1, \dots, 44, \quad (10)$$

onde $g(\cdot)$ é a função de ligação logit. As estimativas dos parâmetros do modelo (10) e os respectivos erros-padrão encontram-se apresentados na Tabela 10. Todos os métodos apresentaram convergência, sendo o método de Newton o mais eficiente, com apenas 4 iterações realizadas; o método BFGS executou 10 iterações e o método simplex foi o mais lento, com 287 iterações. Observamos que o valor máximo da função de log-verossimilhança, as estimativas pontuais e os erros-padrão obtidos foram iguais. É possível observar que o erro-padrão da estimativa do parâmetro de precisão é notavelmente maior do que os erros-padrão das estimativas dos parâmetros do preditor linear, os quais são muito semelhantes entre si.

Tabela 10 – Estimativas obtidas para o conjunto de dados de QI

Parâmetro	BFGS	Newton	simplex
β_0	1,3338 (0,1357)	1,3338 (0,1357)	1,3338 (0,1357)
β_1	-0,9736 (0,1335)	-0,9736 (0,1335)	-0,9736 (0,1335)
β_2	0,1608 (0,1344)	0,1608 (0,1344)	0,1608 (0,1344)
β_3	-0,2186 (0,1345)	-0,2186 (0,1345)	-0,2186 (0,1345)
ϕ	11,1332 (2,4435)	11,1332 (2,4435)	11,1332 (2,4435)
Máximo	51,3504	51,3504	51,3504

A análise dos resíduos do modelo revela que há um ponto de alavanca nos dados. Ele

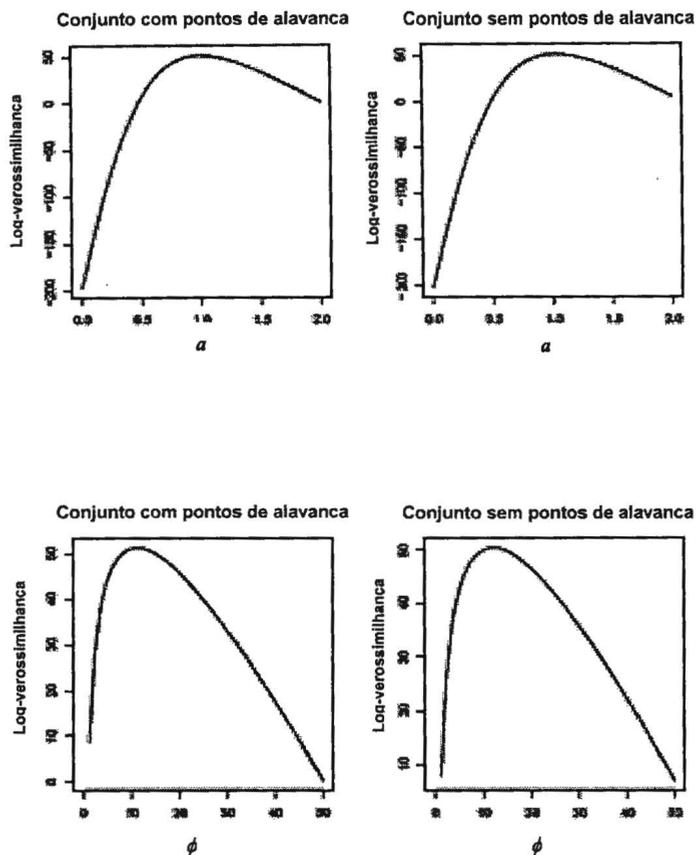
foi retirado e os parâmetros do modelo foram re-estimados. Verificou-se que o maior impacto exercido por esse ponto recai sobre as estimativas dos parâmetros β_2 e β_3 , que apresentaram mudanças relativas de 65,6% e 48,5%, respectivamente. Como pode ser observado na Tabela 11, o impacto do ponto de alavanca na estimação do parâmetro ϕ é leve, sendo a mudança relativa na estimativa de apenas 7,7%. A estimativa menos afetada pela retirada deste ponto é a do intercepto do modelo (β_0), com mudança relativa de 6,0%. Pode-se observar também que o erro-padrão da estimativa de ϕ aumentou consideravelmente em comparação com o aumento da mesma medida para os parâmetros de locação do modelo.

Tabela 11 – Estimativas obtidas para os dados de QI sem pontos de alavanca

Parâmetro	BFGS	Newton	simplex
β_0	1,2541 (0,1325)	1,2541 (0,1325)	1,2541 (0,1325)
β_1	-0,8916 (0,1306)	-0,8916 (0,1306)	-0,8916 (0,1306)
β_2	0,2663 (0,1368)	0,2663 (0,1368)	0,2663 (0,1368)
β_3	-0,3245 (0,1369)	-0,3245 (0,1369)	-0,3245 (0,1369)
ϕ	11,992 (2,6549)	11,992 (2,6549)	11,992 (2,6549)
Máximo	50,4688	50,4688	50,4688

Para este conjunto de dados, foi realizada a mesma análise gráfica efetuada na aplicação apresentada na Sub-seção 5.1. Da Figura 2, pode-se deduzir a razão pela qual todos os métodos atingiram o mesmo ponto: tanto variações no parâmetro de locação quanto variações no parâmetro de precisão conduzem a variações consideráveis na função de log-verossimilhança. Nota-se também que a concavidade da função de log-verossimilhança não apresentou mudança significativa após a eliminação do ponto de alavanca. Quando eliminado o ponto de alavanca, a velocidade de convergência do método de Newton não foi afetada, pois o número de iterações até convergência passou de 4 para 5, um aumento mínimo. Adicionalmente, o número de iterações realizadas pelo método BFGS aumentou de 10 para 12; no que se refere ao método simplex o aumento no número de iterações até convergência após a retirada do ponto de alavanca foi de 11 iterações.

Figura 2 – Funções de log-verossimilhança para diferentes valores dos parâmetros de locação e precisão no conjunto de dados de QI



6. Conclusões

No presente artigo, nós avaliamos a eficácia de alguns métodos de otimização não-linear como ferramentas a serem utilizadas na maximização numérica da função de log-verossimilhança do modelo de regressão beta. Considerando os resultados obtidos através das simulações de Monte Carlo realizadas e das aplicações apresentadas, pode-se concluir que os métodos de Newton e BFGS foram os mais eficazes nas avaliações realizadas, com ligeira vantagem para o primeiro, no que tange à estimação dos parâmetros do modelo. Levando em conta critérios como número de iterações, taxa de convergência e taxa de erro, verificamos que métodos que não utilizam informação referente à matriz hessiana da função de log-verossimilhança, como é o caso do método simplex, não são eficazes relativamente aos demais.

Constatou-se que o maior obstáculo para todos os métodos estudados reside na estimação do parâmetro de precisão ϕ do modelo, pois, para alguns conjuntos de dados, a curvatura da função de log-verossimilhança tende a ser pouco acentuada na região próxima à do ponto de máximo, fazendo com que pequenas variações nesse parâmetro não resultem em variações significativas no valor da função a ser maximizada. Um outro aspecto que pode impactar a eficácia dos métodos é o ponto inicial usado no processo de otimização; os métodos BFGS e Newton não são afetados pela escolha desse ponto, ao passo que o método simplex pode fornecer estimativas diferentes dependendo da distância do ponto inicial relativamente ao maximizador verdadeiro.

Na presença de pontos de alavanca, deve-se escolher o método de otimização a ser utilizado com cautela, pois alguns métodos, como, por exemplo, o simplex, podem apresentar desempenho falho em amostras de tamanho pequeno. Neste cenário, o método de Newton se mostrou eficiente, pois a magnitude do ponto de alavanca não impactou significativamente o número de iterações executadas para atingir o ponto ótimo. Analogamente, o desempenho do método BFGS também não foi fortemente afetado por pontos atípicos. Realizamos também uma avaliação numérica onde as covariáveis do modelo apresentavam fortemente correlação. Neste caso, o método de Newton continuou sendo o mais eficiente, pois o tempo gasto no processo de otimização foi o menor de todos para qualquer grau de correlação. Em grandes amostras, o método simplex apresentou leve tendência de aumento do tempo empregado quando o grau de correlação entre os regressores aumentou.

Referências

- Buckley, J. (2003), Estimation of models with beta distributed dependent variables: A replication and extension of Paolino (2001), *Political Analysis* 11 (1), 1-12.
- Denbo, R., Eisenstat, S. & Steihaug, T. (1982), Inexact Newton methods, *SIAM Journal on Numerical Analysis* 19, 400-408.
- Espinheira, P.L., Ferrari, S.L.P. & Cribari-Neto, F. (2006), On beta regression residuals, Technical Report, Universidade de São Paulo.
- Ferrari, S.L.P. & Cribari-Neto, F. (2004), Beta regression for modelling rates and proportions, *Journal of Applied Statistics* 31, 799-815.
- Goldfarb, D. (1980), Curvilinear path steplength algorithms for minimization which use directions of negative curvature, *Mathematical Programming* 18, 31-40.

- Harville, D.A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York: Springer.
- Johnson, N., Kotz, S. & Balakrishnan, N. (1995), *Continuous Univariate Distributions*, 2nd Ed., New York: John Wiley & Sons.
- McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, 2nd Ed., New York: Chapman and Hall.
- Moré, J. & Sorensen, D. (1979), On the use of directions of negative curvature in a modified Newton method, *Mathematical Programming* 16, 1-20.
- Nelder, J. & Mead, R. (1965), A simplex algorithm for function minimization, *Computer Journal* 7, 308-313.
- Nocedal, J. & Wright, S.J. (1999), *Numerical Optimization*, New York: Springer-Verlag.
- Ospina, R., Cribari-Neto, F. & Vasconcellos, K. (2006), Improved point and interval estimation for a beta regression model, *Computational Statistics and Data Analysis* 51, 960-981.
- Pammer, K. & Kevan, A. (2004), The contribution of visual sensitivity, phonological processing and non-verbal IQ to children's reading, *Technical Report*, The Australian National University, Canberra.
- Paolino, P. (2001), Maximum likelihood estimation of models with beta-distributed dependent variables, *Political Analysis* 9, 325-346.
- Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, 2nd Ed., New York: Cambridge University Press.
- Rao, C.R. (1973), *Linear Statistical Inference and its Applications*, 2nd Ed., New York: John Wiley & Sons.
- Schnabel, R. & Eskow, E. (1991), A new modified Cholesky factorization, *SIAM Journal on Scientific Computing* 11, 1136-1158.
- Sherman, A. (1978), On Newton-iterative methods for the solution of systems of nonlinear equations, *SIAM Journal on Numerical Analysis* 15, 755-771.
- Smithson, M. & Verkuilen, J. (2006), A better lemon squeezer? Maximum-Likelihood regression with beta-distributed dependent variables, *Psychological Methods* 11, 54-71.

Abstract

We consider the beta regression model proposed by Ferrari & Cribari-Neto (2004), which is tailored to situations where the response is restricted to the standard unit interval and has a regression structure involving regressors and unknown parameters. Our chief interest is the evaluation of several nonlinear optimization methods in the context of maximizing the beta regression log-likelihood function. The numerical evidence from Monte Carlo simulations and empirical analyses based on real data favors the Newton and BFGS algorithms, which are fast, accurate and behave well even in unfavorable situations such as the existence of leverage points and high correlation amongst regressors.

Key words. Beta regression model, non-linear optimization, log-likelihood function.

Agradecimentos

Os autores agradecemos à CAPES pelo apoio financeiro e as sugestões de dois pareceristas anônimos.

POLÍTICA EDITORIAL

A Revista Brasileira de Estatística - RBEs publica trabalhos relevantes em Estatística Aplicada, não havendo limitação no assunto ou matéria em questão. Como exemplos de áreas de aplicação, citamos as áreas de advocacia, ciências físicas e biomédicas, criminologia, demografia, economia, educação, estatísticas governamentais, finanças, indústria, medicina, meio ambiente, negócios, políticas públicas, psicologia e sociologia, entre outras. A RBEs publicará, também, artigos abordando os diversos aspectos de metodologias relevantes para usuários e produtores de estatísticas públicas, incluindo planejamento, avaliação e mensuração de erros em censos e pesquisas, novos desenvolvimentos em metodologia de pesquisa, amostragem e estimação, imputação de dados, disseminação e confiabilidade de dados, uso e combinação de fontes alternativas de informação e integração de dados, métodos e modelos demográfico e econométrico.

Os artigos submetidos devem ser inéditos e não devem ter sido submetidos simultaneamente a qualquer outro periódico.

O periódico tem como objetivo a apresentação de artigos que permitam fácil assimilação por membros da comunidade em geral. Os artigos devem incluir aplicações práticas como assunto central, com análises estatísticas exaustivas e apresentadas de forma didática. Entretanto, o emprego de métodos inovadores, apesar de ser incentivado, não é essencial para a publicação.

Artigos contendo exposição metodológica são também incentivados, desde que sejam relevantes para a área de aplicação pela qual os mesmos foram motivados, auxiliem na compreensão do problema e contenham interpretação clara das expressões algébricas apresentadas.

A RBEs tem periodicidade semestral e também publica artigos convidados e resenhas de livros, bem como incentiva a submissão de artigos voltados para a educação estatística.

Artigos em espanhol ou inglês só serão publicados caso nenhum dos autores seja brasileiro e nem resida no País.

Todos os artigos submetidos são avaliados quanto à qualidade e à relevância por dois especialistas indicados pelo Comitê Editorial da RBEs.

O processo de avaliação dos artigos submetidos é do tipo 'duplo cego', isto é, os artigos são avaliados sem a identificação de autoria e os comentários dos avaliadores também são repassados aos autores sem identificação.

INSTRUÇÃO PARA SUBMISSÃO DE ARTIGOS À RBEs

O processo editorial da RBEs é eletrônico. Os artigos devem ser submetidos via *e-mail* para: rbe@ibge.gov.br.

Após a submissão, o autor receberá um código para acompanhar o processo de avaliação do artigo. Caso não receba um aviso com este código no prazo de uma semana, fazer contato com a secretaria da revista no endereço:

Revista Brasileira de Estatística – RBES
IBGE – ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS
Rua André Cavalcanti, 106, sala 111
Centro, Rio de Janeiro – RJ
CEP: 20031-170
Tel.: 55 21 2142-4682 (Sandra Cavalcanti Barros – Secretária)
55 21 2142-4686 (Ismenia Blavatsky – Editor Executivo)
Fax: 55 21 2142-0501

INSTRUÇÕES PARA PREPARO DOS ORIGINAIS

Os originais enviados para publicação devem obedecer às normas seguintes:

1. Podem ser submetidos originais processados pelo editor de texto *Word for Windows* ou originais processados em LaTeX (ou equivalente) desde que estes últimos sejam encaminhados acompanhados de versões em pdf, conforme descrito no item 3, a seguir;
2. A primeira página do original (folha de rosto) deve conter o título do artigo, seguido do(s) nome(s) completo(s) do(s) autor(es), indicando-se, para cada um, a afiliação e endereço para correspondência. Agradecimentos a colaboradores e instituições, e auxílios recebidos, se for o caso de constarem no documento, também devem figurar nesta página;
3. No caso de a submissão não ser em *Word for Windows*, três arquivos do original devem ser enviados. O primeiro deve conter os originais no processador de texto utilizado (por exemplo, LaTeX). O segundo e terceiro devem ser no formato pdf, sendo um com a primeira página, como descrito no item 2, e outro contendo apenas o título, sem a identificação do(s) autor(es) ou outros elementos que possam permitir a identificação da autoria;
4. A segunda página do original deve conter resumos em português e inglês (*abstract*), destacando os pontos relevantes do artigo. Cada resumo deve ser digitado seguindo o mesmo padrão do restante do texto, em um único parágrafo, sem fórmulas, com, no máximo, 150 palavras;
5. O artigo deve ser dividido em seções, numeradas progressivamente, com títulos concisos e apropriados. Todas as seções e subseções devem ser numeradas e receber título apropriado;
6. Tratamentos algébricos exaustivos devem ser evitados ou alocados em apêndices;
7. A citação de referências no texto e a listagem final de referências devem ser feitas de acordo com as normas da ABNT;
8. As tabelas e gráficos devem ser precedidos de títulos que permitam perfeita identificação do conteúdo. Devem ser numeradas seqüencialmente (Tabela 1, Figura 3, etc.) e referidas nos locais de inserção pelos respectivos

números. Quando houver tabelas e demonstrações extensas ou outros elementos de suporte, podem ser empregados apêndices. Os apêndices devem ter título e numeração, tais como as demais seções de trabalho;

9. Gráficos e diagramas para publicação devem ser incluídos nos arquivos com os originais do artigo. Caso tenham que ser enviados em separado, devem ter nomes que facilitem a sua identificação e posicionamento correto no artigo (ex.: Gráfico 1; Figura 3; etc.). É fundamental que não existam erros, quer no desenho, quer nas legendas ou títulos;
10. Não serão permitidos itens que identifiquem os autores do artigo dentro do texto, tais como: número de projetos de órgãos de fomento, endereço, *e-mail*, etc. Caso ocorra, a responsabilidade será inteiramente dos autores; e
11. No caso do artigo ser aceito para a publicação após a avaliação dos pareceristas, serão encaminhadas as sugestões/comentários aos autores sem a sua identificação. Uma vez nesta condição, é de responsabilidade única dos autores fazer o *download* da formatação padrão da revista (em doc ou em LaTeX) para o envio da versão corrigida.