

Presidente da República  
**Fernando Henrique Cardoso**

Ministro do Planejamento, Orçamento e Gestão  
**Martus Antônio Rodrigues Tavares**

## **INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE**

Presidente  
**Sérgio Besserman Vianna**

Diretor-Executivo  
**Nuno Duarte da Costa Bittencourt**

### **ÓRGÃOS ESPECÍFICOS SINGULARES**

Diretoria de Pesquisas  
**Maria Martha Malard Mayer**

Diretoria de Geociências  
**Guido Gelli**

Diretoria de Informática  
**Paulo Roberto Ribeiro da Cunha**

Centro de Documentação e Disseminação de Informações  
**David Wu Tai**

Escola Nacional de Ciências Estatísticas  
**Kaizô Iwakami Beltrão**

Ministério da Fazenda  
Secretaria de Estado de Planejamento e Avaliação  
**Instituto Brasileiro de Geografia e Estatística - IBGE**

# **REVISTA BRASILEIRA DE ESTATÍSTICA**

volume 59 número 212 julho/dezembro 1998

ISSN 0034-7175

*R. bras. Estat.*, Rio de Janeiro, v. 59, n. 212, p. 1-94, jul./dez. 1998

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

© IBGE. 2000

**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:

Pedro Luis do Nascimento Silva  
Editor responsável – RBEs – IBGE.  
Av. República do Chile, 500 – Centro  
20031-170 – 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**

Pedro Luis do Nascimento Silva (IBGE)

**Editor de Estatísticas Oficiais**

Djalma Galvão Carneiro Pessoa (IBGE)

**Editor de Metodologia**

Hélio dos Santos Migon (UFRJ)

**Editores Associados**

Gilberto Alvarenga Paula (USP)

Kaizô Iwakami Beltrão (IBGE)

Lisbeth Kaiseliana Cordani (USP)

Renato Martins Assunção (UFMG)

Wilton de Oliveira Bussab (FGV-SP)

**Impressão**

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2000

**Capa**

Renato J. Aguiar – Divisão de Criação – DIVIC/CDDI

**Ilustração da Capa**

Marcos Balster – Divisão de Criação – DIVIC/CDDI

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

---

ANÁLISE DE ITENS E DA ESTRUTURA FATORIAL  
DE UM INVENTÁRIO PARA AVALIAÇÃO DE REPERTÓRIO  
DE HABILIDADES SOCIAIS

7

Maria Cecília Mendes Barreto  
Zilda A. P. Del Prette  
Almir Del Prette

---

PROCEDIMENTOS LINEARES PARA IDENTIFICAÇÃO  
E ESTIMAÇÃO DOS PARÂMETROS DE MODELOS

25

Paulo Roberto de Holanda Sales  
Basilio de Bragança Pereira  
Ailton de Mesquita Vieira

---

ESTUDO DA DIREÇÃO TOMADA POR POMBOS  
ATRAVÉS DE EQUAÇÕES DE ESTIMAÇÃO PARA  
DADOS CIRCULARES LONGITUDINAIS

53

Rinaldo Artes  
Gilberto Alvarenga Paula  
Ronald Ranvaud

---

ESTIMANDO O RISCO SISTÊMICO DE AÇÕES  
BRASILEIRAS USANDO REGRESSÃO ORTOGONAL

71

Heatriz Vaz de Melo Mendes  
Edel Alexandre da Silva Pontes

---

POLÍTICA EDITORIAL

93

## NOTA DO EDITOR

Apresentamos mais um número da RBEs, com 4 artigos de diferentes especialidades e tratando de aplicações em áreas distintas. Barreto e co-autores tratam da aplicação de técnicas de análise multivariada em psicometria, particularmente na avaliação de inventário de habilidades sociais. Sales e co-autores abordam aplicações de métodos de análise de séries temporais a dados de vazões de rios brasileiros. Artes e co-autores analisam dados longitudinais circulares de experimentos com pombos empregando equações de estimação generalizadas. Finalmente, Mendes e Pontes examinam a aplicação de métodos de regressão ortogonal para estimar risco sistemático de ações brasileiras.

Este conjunto de técnicas e assuntos tratados reflete a variedade da produção e especialidade da comunidade estatística brasileira, oferecendo um bom exemplo do que a Revista Brasileira de Estatística (RBEs) pode fazer pela divulgação dessa produção.

A editoração da revista passou a ser feita usando o processador de textos “*Word for Windows 97*”, e a tarefa é bastante agilizada quando os autores submetem suas contribuições nesse formato. Encorajamos, portanto, que artigos destinados à submissão para a RBEs sejam preparados nesse padrão. Em breve estaremos divulgando um modelo de documento Word que os autores poderão usar caso queiram preparar seus originais especificamente com o formato adotado na revista.

Ao sair este número, estará se avizinando a realização do SINAPE. Encorajamos todos os autores de artigos e resumos enviados para apresentação no SINAPE a considerarem a submissão de versões completas de tais artigos para publicação na RBEs.

Saudações,

Pedro Luis do Nascimento Silva

Editor Responsável

# Análise de Itens e da Estrutura Fatorial de um Inventário para Avaliação de Repertório de Habilidades Sociais

Maria Cecilia Mendes Barreto\*

Zilda A. P. Del Prette\*\*

Almir Del Prette\*\*

## RESUMO

A identificação de portadores de dificuldades em habilidades sociais permite auxiliar o desenvolvimento profissional de indivíduos com maior potencial de comportamentos interpessoais. Neste trabalho as propriedades psicométricas de um inventário de habilidades sociais (IHS) de 38 itens são estudadas quando de sua aplicação em um grupo de 527 universitários de ciências humanas, biológicas e exatas, de ambos os sexos. Os principais resultados da análise de itens e da análise fatorial alfa mostram: índices de discriminação positivos, variando de 3,0 a 16,7; índices de correlação positivos, com apenas dois não significativos e 10 abaixo de 0,30; e cinco fatores principais, identificados em termos comportamentais-situacionais, com coeficientes alfa variando de 0,74 a 0,96, que explicam 92,75% da variância total. Esses resultados apontam para a utilização segura do IHS tanto em pesquisa como na avaliação clínica.

**Palavras-chave:** Análise de Itens, Análise Fatorial Alfa, Competência Social, Correlação Item-total, Educação, Índice de Discriminação, Habilidades Sociais, Psicologia Clínica, Análise Multivariada.

## 1. INTRODUÇÃO

---

As sociedades constituem hoje uma rede de relações interpessoais bastante complexa. Cada vez mais se exige que as pessoas, independentemente de suas atividades profissionais, apresentem também desempenhos sociais aceitáveis e mais elaborados.

\* Endereço para correspondência: UFSCar - C.Postal 676 - CEP.13565-905 - São Carlos - SP - e-mail:cbarreto@power.ufscar.br.

\*\* Departamento de Psicologia - UFSCar.

A temática das habilidades sociais ganha, em nossa atualidade, um novo interesse, despertando a atenção dos pesquisadores e do público em geral. Uma questão particularmente crítica dessa área é a da avaliação. Ela remete ao conceito de habilidades sociais e seus correlatos e ao conjunto de procedimentos e critérios pelos quais se pode caracterizar o repertório de habilidades sociais das pessoas. Além disso, envolve questões próprias da psicometria em termos da construção e da avaliação dos instrumentos utilizados nessa caracterização.

Essas questões estão na base do objetivo do presente trabalho. Inicialmente apresenta-se uma breve introdução ao campo teórico das habilidades sociais e aos procedimentos estatísticos utilizados para avaliar alguns aspectos das propriedades psicométricas de inventários de auto-relato. Posteriormente descreve-se a construção de um instrumento, o Inventário de Habilidades Sociais - IHS -, destinado a universitários, os procedimentos de aplicação (a uma amostra de 527 estudantes) e o exame das propriedades psicométricas do instrumento, com ênfase na análise de itens e da estrutura fatorial alfa. O estudo da estrutura fatorial foi desenvolvido considerando-se que os 38 itens elaborados constituem uma amostra de todos os possíveis itens para se medir as habilidades sociais. A análise fatorial alfa (Harman, 1976) leva em consideração essas suposições e no entanto não está descrita na bibliografia básica de Análise Multivariada (Johnson & Wichern, 1992; Morrison, 1976; Anderson, 1984). Este trabalho relata as técnicas estatísticas utilizadas no estudo das propriedades psicométricas do inventário de habilidades sociais, em particular a análise de itens e a estrutura fatorial alfa. Essas análises foram importantes para a validação do instrumento e sinalizam para o aperfeiçoamento do instrumento, discutidos na seção final.

## **2. MATERIAL**

### **2.1. Conceitos de habilidades sociais**

O campo teórico-prático das habilidades sociais é constituído de vários modelos conceituais da Psicologia (Caballo, 1987; Hidalgo & Abarca, 1991), destacando-se a abordagem cognitiva (em particular o modelo de processamento de informação) e da aprendizagem (especialmente a aprendizagem social).

A área comporta algumas controvérsias, entre as quais a da diferenciação entre habilidades sociais, competência social e assertividade. Embora nem sempre a literatura apresente uma distinção clara, defende-se (Del Prette & Del Prette, 1999) que esses termos não devem ser tomados como equivalentes.

A definição apresentada por Caballo (1987, p. 14) exemplifica a sobreposição das noções de habilidades sociais e de competência social: "O comportamento socialmente habilidoso é esse conjunto de comportamentos emitidos por um indivíduo no contexto interpessoal, que expressa sentimentos, atitudes, desejos, opiniões ou direitos desse indivíduo, de um modo adequado à situação, respeitando esses comportamentos nos demais e que geralmente resolvem uma situação ao mesmo tempo em que minimizam a probabilidade de problemas futuros". Outros autores, como McFall (1982) diferenciam os dois termos, utilizando habilidades sociais para designar as diferentes classes de comportamentos (observáveis e não-observáveis) envolvidos no desempenho interpessoal, reservando a noção de competência para a avaliação externa do nível de proficiência com que esse desempenho ocorre, o que implica a identificação de suas conseqüências pessoais e sociais.

A diferenciação entre habilidades sociais e competência social apresenta, portanto, implicação direta para a avaliação. Enquanto a avaliação das habilidades sociais se orienta para uma caracterização mais descritiva dos desempenhos, a avaliação da competência supõe o acesso às conseqüências desse desempenho e a critérios (especialmente culturais) para qualificá-los em termos de proficiência.

A avaliação do repertório de habilidades sociais pode focalizar aspectos observáveis ou não observáveis. Os aspectos observáveis incluem dois conjuntos de classes comportamentais: as mais amplas ou molares (por exemplo, fazer e responder (a) cumprimentos e elogios; expressar opiniões e discordâncias; iniciar, manter e encerrar conversações; responder a críticas, etc.) e as moleculares que compõem cada classe molar (tom de voz, contato visual, gestos, postura, etc.). A dimensão não-observável inclui pensamentos, percepções, representações, etc., que precedem, acompanham ou seguem o desempenho interpessoal.

Essa definição de habilidades sociais permite também diferenciá-la da noção de assertividade. Embora alguns autores tomem os dois termos como sinônimos, Del Prette & Del Prette (1999) defendem uma diferenciação que coloca a assertividade como uma subárea das habilidades sociais. Entende-se que as habilidades sociais representam um universo mais amplo dos desempenhos interpessoais e que a assertividade refere-se aos desempenhos específicos de expressão de sentimentos e de defesa de direitos com controle da ansiedade e da agressividade.

Conforme se pode constatar, o conceito de habilidades sociais é multidimensional. A literatura tem enfatizado quatro dimensões desse conceito, apesar das controvérsias (Caballo, 1993) quanto à sua prioridade na avaliação da competência social: a dimensão pessoal (cognições e percepções), a comportamental



(classes molares e moleculares), a situacional (contextos e interlocutores) e a cultural (valores e práticas sociais). Algumas análises têm apontado para a dificuldade de isolar essas dimensões na avaliação e para a necessidade de maior investimento na dimensão situacional (Caballo, 1993).

## 2.2. A avaliação das habilidades sociais

Avaliar o repertório de habilidades sociais de uma pessoa é um requisito prévio à aplicação de qualquer treinamento ou terapia para resolver problemas interpessoais.

Considerando-se a complexidade das dimensões subjacentes aos conceitos de habilidades sociais e de competência social, justifica-se a existência de vários métodos de avaliação (Del Prette & Del Prette, 1996): a observação, os procedimentos de *role-playing*, as medidas fisiológicas, a avaliação feita por outros significantes e os auto-relatos, através de entrevistas e inventários.

O instrumento que mais se ajusta às avaliações em larga escala, tipo *survey*, é o inventário de auto-relato. Os inventários usualmente restringem-se às classes molares do desempenho, embora a necessidade de contextualizá-los acabe implicando sobreposição da dimensão situacional e, conforme o caso, permitindo inferências sobre a dimensão cultural. Caballo (1993) chama a atenção para essa sobreposição e para a importância da dimensão situacional, reconhecendo que, dos 12 fatores identificados na análise da estrutura fatorial da Escala Multidimensional de Expressão Social (EMES-M), oito deles implicavam diretamente a dimensão situacional.

A dimensão situacional tem sido bastante enfatizada na literatura. Embora no limite superior da competência social se busque a flexibilidade comportamental própria da capacidade adaptativa do ser humano, um indivíduo usualmente não é igualmente "habilidoso" nas diferentes situações interpessoais. Esse caráter situacional da competência social opõe-se à noção de unidimensionalidade própria dos chamados "traços de personalidade". Uma consequência dessa compreensão é que, mais do que buscar "fatores de personalidade" associados aos padrões de desempenho social apresentados pelos indivíduos, dever-se-ia examinar as características comuns aos grupos de situações que possivelmente afetam, de forma semelhante, esse desempenho. Para contemplar a dimensão situacional, um inventário de habilidades sociais deve incluir um conjunto diversificado de situações prováveis e significativas nas relações interpessoais dos respondentes aos quais se destina, aí incluindo diferentes contextos interpessoais e diferentes interlocutores.

A avaliação do repertório de habilidades sociais de estudantes universitários e, portanto, o aperfeiçoamento de um instrumento para efetuarla, adquire particular importância quando se considera que, de um lado, a formação de terceiro grau deveria incluir o desenvolvimento interpessoal como parte dos objetivos acadêmicos (Del Prette & Del Prette, 1983) e, de outro, que os padrões comportamentais por eles apresentados constituem uma amostra da influência da cultura, em termos dos valores e normas de convivência social que estão sendo assimilados, servindo, portanto, para análises transculturais desses padrões.

Muito embora existam itens de avaliação comuns a diversas culturas, certas especificidades aconselham a adaptação de instrumentos elaborados em outros contextos ou, mesmo, a reconstrução a partir de características próprias da subcultura para a qual ele se destina. Além disso, a passagem do tempo imprime modificações em certas práticas culturais, exigindo atualizações constantes. Um exemplo típico é a emergência de novos problemas interpessoais, que pode ser exemplificado, na atualidade, pela preocupação com o uso de preservativos nas relações sexuais, aspecto ainda não contemplado pela maioria dos inventários. O inventário examinado neste trabalho foi elaborado tendo em vista essas questões.

### **2.3. A análise de itens**

O estudo das propriedades psicométricas do inventário de habilidades sociais é iniciado com a análise de itens, onde as respostas dadas a cada item do instrumento pelos indivíduos são analisadas em relação ao seu escore, que corresponde à soma das respostas dadas a todos os itens do inventário.

Uma primeira medida utilizada é o índice de discriminação de item, calculado a partir da categorização do escore em três grupos, de maior, de médio e de menor escore. Para cada item monta-se uma tabela de dupla entrada, sendo a variável linha os grupos de maior, de médio e de menor escore, e a variável coluna, as categorias do referido item. A diferença entre as porcentagens de ocorrência de valores altos para os grupos de maior e de menor escore é definida como o índice de discriminação do item e ajuda a perceber se o item em questão contribui ou não para o valor do escore. Uma outra medida de discriminação do item é a correlação linear de Pearson existente entre o item e o escore. Essas duas medidas são sugeridas por Vianna (1973) para a análise de itens em um inventário.

## 2.4. A análise fatorial alfa

Outro estudo de propriedades psicométricas do inventário de habilidades sociais é o interesse em agrupar os itens em poucos fatores não-observáveis que possam explicar as relações de covariâncias/correlações entre os itens.

Em geral nos estudos estatísticos, especialmente os de ciências sociais, os elementos amostrais são pessoas. As medidas ou atributos dessas pessoas são as variáveis analisadas para a descrição e compreensão do grupo a ser estudado. Um importante aspecto de tais estudos envolve a generalização dos resultados encontrados em uma particular amostra para outros grupos de pessoas possivelmente pertencentes à mesma população da qual a amostra original foi retirada. Os conceitos de amostragem e generalização se referem às entidades ou pessoas e as variáveis são então predeterminadas.

Para essas situações, existem diversos métodos de agrupamento de itens em poucos fatores não-observáveis tais como as análises de componentes principais e as análises fatoriais pelo método de máxima verossimilhança. Neste caso específico em que as respostas aos itens são variáveis aleatórias ordinais em cinco categorias, a análise de correspondência é a mais indicada. Em cada um desses métodos de obtenção de fatores, os resultados, em geral, diferem quer no número de fatores obtidos, quer no conjunto de itens que compõem cada fator e, conseqüentemente, na interpretação de cada um deles.

Outro tipo de generalização que pode ser feita é quando as variáveis (itens) são os elementos a serem amostrados. Nesse caso elas são consideradas como pertencentes ao universo de conteúdos e as inferências são feitas a partir de tal amostra para a área de conceitos. Tem-se, assim, a inferência psicométrica.

A análise fatorial alfa, que aqui está sendo utilizada, parte do princípio de que, em uma amostra de testes, os fatores comuns são determinados de tal modo que tenham uma correlação máxima com o correspondente universo de fatores (Harman, 1976; Rummel, 1970; Tabachnick & Fidell, 1996).

O conceito de generalização máxima é quantificado nos coeficientes alfa associados a cada um dos fatores. De acordo com Peterson (1994) existem duas classes de coeficientes de fidedignidade de um instrumento: os que são baseados em dados longitudinais, por exemplo, o coeficiente de fidedignidade de teste e reteste; e aqueles que são baseados em dados com apenas uma intervenção no tempo, *cross-sectional*, por exemplo, os coeficientes de consistência interna e os coeficientes de fidedignidade de equivalência. O coeficiente alfa de Cronbach (1951) é o mais antigo e

mais difundido coeficiente de fidedignidade usado para estimar a consistência interna de um instrumento.

## 2.5. O modelo de análise fatorial alfa

Sejam as  $p$  variáveis de um estudo representadas por um vetor aleatório observável  $X$ , de média  $\mu$  e matriz de covariância  $\Sigma$ . O objetivo da análise fatorial é representar cada variável  $X_j$ , que compõe o vetor  $X$ , em termos de diversos fatores subjacentes ou construtos hipotéticos,  $F_1, F_2, \dots, F_m$ , chamados fatores comuns, e fontes adicionais de variação  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , chamadas erros ou fatores específicos, na forma

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p \end{aligned} \quad (1)$$

ou, em notação matricial

$$X_{px1} - \mu = L_{pxm}F_{mx1} + \varepsilon_{px1} \quad (2)$$

O coeficiente  $l_{ij}$  é chamado de *carga* da  $i$ -ésima variável no  $j$ -ésimo fator, de tal modo que  $L$  é a *matriz de cargas fatoriais*. O fator específico  $\varepsilon_i$  está associado apenas com a  $i$ -ésima resposta  $X_i$ . Os  $p$  desvios  $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$  são expressos em termos de  $p + m$  variáveis aleatórias  $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , que são *não-observáveis*.

Como o número de variáveis não observáveis é grande, uma verificação direta do modelo não é possível, entretanto, com algumas suposições adicionais é possível verificar certas relações entre as covariâncias. Assim, supondo-se que os vetores aleatórios não-observáveis  $F$  e  $\varepsilon$  são independentes entre si e  $E(F) = 0$ ,  $\text{Cov}(F) = I$ ,  $E(\varepsilon) = 0$  e  $\text{Cov}(\varepsilon) = \Psi$ , onde  $\Psi$  é uma matriz diagonal, tem-se que:

$$\text{Cov}(X) = LL' + \Psi, \quad \text{ou} \quad \begin{cases} \text{Var}(X_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i \\ \text{Cov}(X_i, X_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \dots + l_{im}l_{km} \end{cases} \quad (3)$$

e

$$\text{Cov}(X, F) = L \quad \text{ou} \quad \text{Cov}(X_i, F_j) = l_{ij} \quad (4)$$

A porção da variância da  $i$ -ésima variável que contribui para os  $m$  fatores comuns é chamada de *comunalidade*,  $h_i^2$ , e a porção da variância devida ao fator específico é chamada, *variância específica*,  $\psi_i$ , sendo  $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$ .

## 2.6. Métodos de estimação

Para ver se as observações da amostra multivariada podem ser representadas por um pequeno número de fatores não correlacionados basta verificar se as relações (3) e (4) são verdadeiras ou não.

O método dos componentes principais é baseado na decomposição espectral do estimador da matriz  $\Sigma$ . O método de máxima verossimilhança supõe que os fatores comuns e os fatores específicos têm distribuição normal, e desse modo obtêm-se as estimativas de máxima verossimilhança das cargas fatoriais e das variâncias específicas.

Para o desenvolvimento da análise fatorial alfa, o modelo (2) é adotado, levando-se em consideração que as variáveis aleatórias  $X_i$  são uma amostra das possíveis variáveis ou itens a serem observados. Assumindo que os indivíduos são uma população de casos, Kaiser & Caffrey (1965) desenvolveram o chamado método de análise fatorial alfa empregando o princípio básico, já citado, de que os fatores comuns da amostra de variáveis são determinados de tal modo que tenham correlação máxima com os correspondentes fatores do universo das variáveis. O quadrado dessas correlações pode ser chamado de generalização de um fator comum.

Ao maximizar a generabilidade de um fator, os autores chegaram à equação característica

$$\left| H^{-1}(R - U^2)H^{-1} - \lambda I \right| = 0 \quad (5)$$

onde  $H^2$  é a matriz diagonal das comunalidades das variáveis, e  $U^2 = I - H^2$ .

O coeficiente de generabilidade alfa associado a cada fator comum é uma função do

correspondente autovalor e para o fator  $F_j$ , associado ao autovalor  $\lambda_j$  e  $p$  variáveis, é dado por

$$\alpha_j = \frac{p}{p-1} \left[ 1 - \frac{1}{\lambda_j} \right] \quad (6)$$

Para valores grandes de  $\lambda_j$  e amostras pequenas de variáveis,  $\alpha_j$  está próximo da unidade, significando generabilidade perfeita. Ao resolver a equação (5) é possível obter a matriz dos fatores alfa por

$$F = HE\lambda^{1/2}$$

onde  $E$  é a matriz com os autovetores,  $\lambda$  é a matriz diagonal cujos elementos são os autovalores.

Na análise fatorial alfa, os fatores comuns a serem extraídos são aqueles que possuem generabilidade positiva, ou a expressão (6) é positiva. Isso corresponde a considerar como fatores aqueles cujos autovalores associados à equação característica (5) são maiores do que 1.

Como as comunalidades  $h_i$  são desconhecidas, o processo iterativo se inicia considerando como valor inicial para as comunalidades a correlação múltipla quadrada (*SMC*) que corresponde a

$$SMC = 1 - \frac{1}{r^{ii}}$$

onde  $r^{ii}$  é o  $i$ -ésimo elemento diagonal da inversa da matriz de correlação amostral.

Após a extração dos fatores é comum fazer sua rotação para melhorar a interpretação científica da solução. A rotação ortogonal dos fatores os mantém não correlacionados e, em geral, apresentam soluções que são mais fáceis de interpretar e descrever os resultados. Já na rotação oblíqua, os fatores podem se tornar correlacionados, o que, muitas vezes, pode ser importante para o pesquisador, mas a interpretação e descrição dos resultados podem apresentar desvantagens práticas.

No presente trabalho a rotação ortogonal varimax foi a que produziu melhores interpretações para descrever os resultados.

### **3. MÉTODO**

---

#### **3.1. Características da população**

A população constituiu-se de 527 alunos de graduação de uma instituição pública e uma particular da cidade de Uberlândia (MG), com proporções equivalentes de alunos de início e meados de curso e de ambos os sexos, das áreas de Ciências Humanas, Exatas e Biológicas. A idade foi predefinida entre 18 e 25 anos (os que excederam essa faixa foram excluídos da amostra), com a maioria situando-se na faixa dos 19 aos 22 anos (65,7%). Há uma maior proporção de alunos da instituição pública (cerca de dois terços) em relação aos da particular e maior proporção de respondentes de Ciências Humanas em relação às Exatas e Biológicas; apenas uma pequena parcela de respondentes relata estar inserida no mercado de trabalho ou receber ajuda institucional do tipo Bolsa de Estudos e/ou Crédito Educativo. Os cursos predominantes foram os de Administração, Direito, Engenharia Mecânica, Fisioterapia, Medicina e Psicologia (frequência maior que 10%). A maioria dos respondentes têm pelo menos um dos pais com no máximo o segundo grau, constatando-se que mais da metade da população (51%) é representada por dois subgrupos predominantes: os que têm ambos os pais com no máximo o primeiro grau e os que têm pelo menos um do pais com terceiro grau (os demais respondentes têm um dos pais com segundo grau, ambos com segundo e ambos com terceiro).

#### **3.2. Características do instrumento - o IHS**

O Inventário de Habilidades Sociais (IHS) é um instrumento de auto-relato para a avaliação da dimensão comportamental molar das habilidades sociais. O IHS compõe-se de duas partes. A primeira contém uma lista de 38 itens, cada um deles descrevendo uma relação interpessoal e uma reação possível ao interlocutor naquela situação. Nas instruções solicita-se que o respondente estime a frequência com que reage da forma sugerida em cada item. A segunda parte contém um cabeçalho para a caracterização do respondente e um quadro para a anotação das respostas, precedido por instruções e pelo modelo da escala de estimativa da frequência das respostas. A escala é do tipo Likert, de cinco pontos, variando de "nunca ou raramente (zero a 20% das vezes)" a "sempre ou quase sempre (81 a 100% das vezes)". Os itens do IHS

foram elaborados a partir da análise da literatura sobre as situações e demandas abrangidas pelo conceito de habilidades sociais e de estudos prévios de Del Prette et al (1992a; b) com estudantes universitários, de modo a contemplar situações, contextos e interlocutores diversificados e demandas interpessoais pertinentes à clientela universitária. A versão inicial do IHS foi submetida a uma testagem de validação semântica, efetuando-se as modificações consideradas necessárias.

### 3.3. A aplicação do IHS

A aplicação do IHS, com a versão atual, ocorreu nas salas de aula, em grupo, procurando-se completar pelo menos 30 alunos de início e 30 de meados de cada curso. O aplicador expunha os objetivos da pesquisa, lia e explicava as instruções, com ênfase na questão do anonimato e na ausência de respostas corretas ou incorretas. Depois supervisionava a atividade dos respondentes procurando verificar se as instruções tinham sido entendidas e se todos os itens estavam sendo respondidos.

## 4. TRATAMENTO DOS DADOS E RESULTADOS

### 4.1. A análise do escore no IHS

As respostas obtidas em cada uma das 38 questões do inventário para avaliação do repertório de habilidades sociais foram quantificadas de acordo com níveis de frequência que o indivíduo apresenta à reação sugerida em cada item, considerando o total de vezes que se encontrou na situação descrita. Na Tabela 1 encontram-se o nível de frequência e o valor atribuído.

Tabela 1 - Nível de frequência às respostas de cada item e valor atribuído

Nível de frequência	Valor
nunca ou raramente (zero a 20% das vezes)	0
com pouca frequência (21 a 40% das vezes)	1
com regular frequência (41 a 60% das vezes)	2
muito frequentemente (61 a 80% das vezes)	3
sempre ou quase sempre (81 a 100% das vezes)	4

Calcularam os escores para cada um dos indivíduos que respondeu a todos os itens, somando-se o valor atribuído a cada um dos itens de acordo com a Tabela 1. Entre os 527 universitários que responderam ao questionário, 472 (90%) o fizeram em sua totalidade. O valor mínimo do escore foi de 47 e o valor máximo, de 141. O valor mediano foi de 95, o primeiro quartil foi de 86, o terceiro , de 106, e a distância



interquartilica, de 20. O escore médio foi de 95,61, e o desvio padrão do escore, de 15,58.

A análise de algumas distâncias indica que a distribuição é simétrica: a média e a mediana estão bem próximas; a distância entre a mediana e o primeiro quartil (9) está bem próxima da distância entre o terceiro quartil e a mediana (11); e a distância entre o máximo e o terceiro quartil (35) está bem próxima da distância entre o primeiro quartil e o mínimo (39).

#### 4.2. A análise de itens do IHS

Após o cálculo dos escores os 472 valores foram ordenados montando-se dois subgrupos extremos formados pelos 27% dos indivíduos com os menores escores e os 27% dos indivíduos com os maiores escores. Como havia diversos indivíduos com os mesmos escores obtiveram-se três grupos conforme a Tabela 2.

**Tabela 2 - Distribuição de freqüência dos 472 indivíduos que responderam a todos os itens de acordo com os grupos de maior (A), médio (B), e menor (C) escore**

Grupo	Freqüência	Porcentagem	Acumulada
A	131	27,8	27,8
B	216	45,8	73,6
C	125	26,5	100,0
Total	472	100,0	-

Para analisar a relação existente entre cada item e o escore foi calculado o índice de discriminação e a correlação entre o item e o escore (Viana, 1973).

O cálculo do índice de discriminação do item é baseado na diferença entre as porcentagens de ocorrência dos valores 3 e 4 para os grupos de maior escore e de menor escore. O índice de discriminação para esta amostra pode variar no intervalo [-26,5 ; 27,8]. Quanto maior o índice mais discriminante é o item. Na Tabela 3 estão os resultados para os valores obtidos do índice de discriminação.

Examinando-se a Tabela 3 percebe-se que, em todos os itens, o índice de discriminação é positivo, indicando que a proporção de indivíduos que possuem um escore alto e escolhem as alternativas 3 e 4 é sempre maior que a proporção de indivíduos que possuem um escore baixo (grupo C) e escolhem valor 3 e 4. O item que apresenta o maior índice de discriminação (16,7%) é o de número 11, seguido do item 29, com índice 14,2%, o de número 14 (14,0%), 16 (13,8%), 36 (13,3%), 30 (12,9%), 31 (12,9%), 32 (11,4%), 27 (11,4%), 15 (10,8%), 6 (10,8%), 7 (10,2%), 21

(10,2%) e 20 (10,0%). Os demais têm índice de discriminação menor do que 10,0%. O item 2 apresenta o menor índice, que é de 3,0%.

A correlação entre cada item e o escore corresponde à correlação de Pearson calculada entre o item e o escore, e seus valores, para os 38 itens, estão na Tabela 3. O item que apresenta a maior correlação com o escore (0,55) é o de número 11, seguido do item 29, com correlação 0,49, o de número 16 (0,48), 14 (0,47), 7 (0,46), 1 (0,44), 12 (0,41), 31 (0,40), 36 (0,40), 13 (0,40), 19 (0,38), 30 (0,38), 37 (0,37), 23 (0,37), 21 (0,37), 31 (0,36). Os demais têm coeficiente de correlação inferior a 0,35. O item 34 apresenta o menor coeficiente de correlação, que é de 0,07.

Como o índice de discriminação e o coeficiente de correlação linear de Pearson medem a associação entre o valor do item e o valor do escore, os resultados obtidos são em grande parte concordantes.

Tabela 3 – Índice de discriminação e coeficiente de correlação de Pearson entre o item e o escore total

Item	Índice de Discriminação	Coefficiente de Correlação
S11	16,7 (1)	0,55 (1)
S29	14,2 (2)	0,49 (2)
S14	14,0 (3)	0,47 (4)
S16	13,8 (4)	0,48 (3)
S36	13,3 (5)	0,40 (9)
S30	12,9 (6)	0,38 (12)
S31	12,9 (7)	0,40 (8)
S13	11,5 (8)	0,40 (10)
S32	11,4 (9)	0,36 (16)
S27	11,4 (10)	0,33
S15	10,8 (11)	0,34
S12	10,8 (12)	0,41 (7)
S6	10,8 (13)	0,33
S7	10,2 (14)	0,46 (5)
S21	10,2 (15)	0,37 (15)
S20	10,0 (16)	0,36
S1	9,5	0,44 (6)
S19	9,1	0,38 (11)
S37	8,5	0,37 (13)
S23	10,0	0,37 (14)
S2	3,0	0,08
S3	6,8	0,28
S4	8,3	0,32
S5	8,7	0,29
S8	7,6	0,34
S9	9,3	0,31
S10	9,3	0,31
S17	7,4	0,30
S18	7,4	0,18
S22	5,9	0,18
S24	8,7	0,31
S25	8,3	0,26
S26	7,4	0,20
S28	9,5	0,34
S33	4,7	0,14
S34	3,4	0,07
S35	8,7	0,31
S38	5,5	0,21

Os números entre parênteses representam a ordem dos maiores valores para o índice de discriminação e o coeficiente de correlação item total.

### 4.3. A análise da estrutura fatorial do IHS

Para agrupar os 38 itens em poucos fatores não observáveis que possam explicar as suas relações de covariância ou correlações, foi feita a análise fatorial alfa, uma vez que, em sua formulação, o conjunto de itens corresponde a uma amostra da população de itens existentes.

Na Tabela 4 estão os principais resultados da análise fatorial alfa e a matriz fatorial após a rotação varimax, obtidos usando o sistema de análise estatística SAS.

Seis fatores produziram autovalores maiores do que um. Entretanto, apenas cinco deles apresentaram itens com cargas fatoriais com valores, em módulo, superiores a 0,30 (Harman, 1976; Tabachnick & Fidell, 1996). O primeiro fator explica 43,55% da variância total, o segundo, 16,93%, o terceiro, 12,97%, o quarto, 9,84%, e o quinto, 9,46%.

No primeiro fator contribuem expressivamente 11 itens, no segundo, 6 itens, no terceiro, 7 itens, no quarto, 4 itens e no quinto, 3 itens.

O coeficiente alfa, que corresponde à quantificação da generalização máxima de cada fator, foi de 0,9650 para o primeiro fator, 0,8673 para o segundo fator, 0,8187 para o terceiro fator, 0,7525 para o quarto fator, 0,7413 para o quinto fator, e 0,6543 para o sexto fator. Esses resultados, associados à consistência interna geral obtida (alfa de Cronbach de 0,75) sugerem um alto grau de validade do IHS.

Tabela 4 – Matriz das cargas fatoriais com valores iguais ou superiores a 0,30 (em módulo) da análise fatorial alfa após a rotação varimax, autovalores associados a cada fator, variância explicada pelo fator, porcentagem da variância total, número de itens com cargas significativas e coeficientes alfa

NÚMERO DO ITEM	FATOR 1	FATOR 2	FATOR 3	FATOR 4	FATOR 5
S7	0,558				
S12	0,532				
S11	0,523				
S16	0,477				
S5	0,459				
S20	0,354				
S15	0,351				
S14	0,338			0,353	
S21	0,325				
S1	0,324				
S29	0,323				
S28		0,638			
S35		0,423			
S6		0,407			
S3		0,405			
S10		0,397			
S30		0,353			
S8		0,302			
S36			0,508		
S17			0,433		
S24			0,405		
S19			0,392		
S13			0,383		
S37			0,332		
S22			0,303		
S9				0,431	
S26				0,373	
S23				0,300	
S18					0,396
S38					0,340
S31					-0,333
<b>AUTOVALORES</b>	<b>16,55</b>	<b>6,43</b>	<b>4,93</b>	<b>3,74</b>	<b>3,59</b>
<b>VARIÂNCIA EXPLICADA PELO FATOR</b>	<b>10,07</b>	<b>9,011</b>	<b>7,180</b>	<b>4,363</b>	<b>4,370</b>
<b>% VARIÂNCIA TOTAL</b>	<b>43,55</b>	<b>16,93</b>	<b>12,97</b>	<b>9,84</b>	<b>9,46</b>
<b>NÚMERO DE ITENS</b>	<b>11</b>	<b>7</b>	<b>7</b>	<b>4</b>	<b>3</b>
<b>COEFICIENTES ALFA</b>	<b>0,9650</b>	<b>0,8673</b>	<b>0,8187</b>	<b>0,7525</b>	<b>0,7413</b>

## 5. DISCUSSÃO E CONCLUSÃO

Algumas considerações podem ser feitas à guisa de conclusão, a partir das análises efetuadas.

A análise do escore mostrou que sua distribuição é aproximadamente simétrica. Esse resultado reflete, em parte, os cuidados com a coleta de dados, especialmente em termos da preocupação com a composição da população quanto a cursos e áreas de formação, distribuição etária, equilíbrio entre meados e início dos cursos, etc. O alto índice de formulários completos mostra que houve efetividade nos procedimentos de controle de preenchimento.

A análise de itens mostrou que parte dos que abordavam questões significativas para a população universitária foram pouco discriminativos, indicando a necessidade de reformulá-los, de modo a tornar mais evidente a demanda ou a resposta sugerida para a situação. Entre os itens mais discriminativos, dois deles não apareceram em nenhum dos fatores, sugerindo uma alteração futura de redação (análise a seguir) mas tiveram suas redações alteradas.

Na análise da estrutura fatorial, foram identificados cinco conjuntos de itens, agrupados em torno dos principais fatores detectados e apenas sete deles (02,04, 25, 27, 32, 33, 34) não entraram em nenhum fator.

Tomando-se como pressuposto a natureza situacional das habilidades sociais e a sua vinculação com a dimensão comportamental, o exame do conteúdo dos itens que tiveram suas cargas fatoriais significativas em cada um dos fatores permitiu nomear e caracterizar cada um deles como segue:

O fator 1 reúne itens (01, 05, 07, 11, 12, 14, 15, 16, 20, 21,29) que retratam situações interpessoais em que a demanda de reação ao interlocutor se caracteriza principalmente pela afirmação e defesa de direitos e de auto-estima. Pode-se afirmar, portanto, que, em termos comportamentais/situacionais, o fator 1 refere-se às habilidades de enfrentamento e auto-afirmação com risco potencial de reação indesejável (possibilidade de rejeição, de réplica ou de oposição) por parte do interlocutor. Relacionando-se à estrutura conceitual das habilidades sociais, pode-se afirmar que o fator 1 reflete principalmente o conceito de *assertividade*.

As situações agrupadas em torno do fator 2, que reúne os itens 03, 06, 08, 10, 28, 30, 35, retratam demandas interpessoais de expressão de afeto positivo e de afirmação da auto-estima, com risco mínimo de reação indesejável, podendo-se também relacioná-la a uma parte do conceito de assertividade. O fator 2 foi denominado, então, em termos comportamentais/situacionais, de habilidades de auto-afirmação na expressão de afeto positivo.

Os itens do fator 3 (13, 17, 19, 22, 24, 36, 37) retratam situações sociais, neutras em termos de afeto positivo ou negativo, de aproximação com risco mínimo de reação indesejável, demandando, principalmente, "traquejo social" na conversação, o que supõe conhecimento das normas de relacionamento cotidiano. Pode-se dizer, então, que, em termos comportamentais/situacionais, o fator 3 refere-se às habilidades de conversação ou de desenvoltura social.

O fator 4 inclui itens (09, 14, 23, 26) de situações que envolvem a abordagem a pessoas desconhecidas e pode ser definido, em termos comportamentais/situacionais, como habilidades de auto-exposição a desconhecidos ou a novas situações.

O fator 5 inclui itens (18, 31, 38) que envolvem reação a estimulações aversivas do interlocutor, demandando controle da raiva e da agressividade. Em termos situacionais, foi identificado como habilidades de autocontrole da agressividade a situações adversas.

Os fatores encontrados na análise são coerentes com a estrutura conceitual das habilidades sociais, entendida como mais ampla que o conceito de assertividade. Em relação aos sete itens que não entraram em nenhum fator, tomou-se a decisão de modificar a redação de modo a que eles expressem mais adequadamente a habilidade que deveriam aferir.

A análise das propriedades psicométricas do IHS, apresentada neste trabalho, é parcial na medida em que não inclui ainda indicadores da validade externa e da fidedignidade do instrumento. Esses aspectos serão objeto de uma reaplicação da nova versão do IHS, com teste reteste e aplicação simultânea de outros instrumentos (possivelmente nominação por colegas), porém o alto índice do coeficiente alfa encontrado sinaliza para a consistência interna do IHS e permite prever resultados animadores no próximo estudo, que deverá também fornecer dados para a padronização dos escores de IHS na população universitária brasileira de 18 a 25 anos.

## REFERÊNCIAS

- ANDERSON, T. W. (1984) *An Introduction to Multivariate Statistical Methods*. 2nd ed. John Wiley, Nova York.
- CABALLO, V. E. (1987) *Teoría, evaluación y entrenamiento de las habilidades sociales*. Promolibro, Valencia.
- CABALLO V. E. (1993) La multidimensionalidad conductual de las habilidades sociales: propiedades psicométricas de una medida de auto-informe, la EMES-M. *Psicología Conductual*, 1(2), 221-231.
- CRONBACH, L. J. (1951) Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.
- DEL PRETTE, Z. A. P. & DEL PRETTE, A. (1983) Análise de repertório assertivo em estudantes de Psicologia. *Revista de Psicologia*, 1(1), 15-24.
- DEL PRETTE, Z. A. P. & DEL PRETTE, A. (1996) Habilidades sociais: Uma área em desenvolvimento. *Psicologia: Reflexão e Crítica*, 9(2), 287-289.
- DEL PRETTE, Z. A. P., DEL PRETTE, A. & CORREIA, M. F. B. (1992) Competência social: Um estudo comparativo entre alunos de Psicologia, Serviço Social e Engenharia Mecânica. *Psicólogo Escolar: Identidade e Perspectivas*, 382-384.
- DEL PRETTE, A., DEL PRETTE, Z. A. P. & CASTELO BRANCO, U. V. (1992) Competência social na formação do psicólogo. *Paidéia: Cadernos de Educação*, 2, 40-50.
- HARMAN, H. H. (1976) *Modern Factor Analysis*. 3rd ed. The University of Chicago Press, Chicago. 487p.
- HIDALGO, C. H. & ABARCA, N. M. (1991) *Comunicación interpersonal - Programa de entrenamiento en habilidades sociales*. Editorial Universitaria, Santiago.
- JOHNSON, R. A.; WICHERN, D. W. (1992) *Applied Multivariate Statistical Analysis*. 3rd ed. Prentice-Hall, Eaglewood Cliffs. 642p.
- KAISER, H. F.; CAFFREY, J. (1965) Alpha Factor Analysis *Psychometrika*, 30, 1-14.

- MORRISON, D. F. (1990) *Multivariate Statistical Methods*. 3rd ed. Mc Graw Hill, Nova York. 495p.
- PETERSON, R. A. (1994) A Meta-analysis of Cronbach's Coefficient Alpha *Journal of Consumer Research*, 21, 381-391.
- RUMMEL, R. J. (1970) *Applied Factor Analysis* Evanston, Northwestern University Press. 617p.
- TABACHNICK, B.G.; FIDELL, L.S. (1996) *Using multivariate statistics*. 3<sup>rd</sup>. ed. Harper Collins College Publishers, Nova York. 880p.
- VIANNA, H. M. (1976) *Testes em Educação* 2a. ed. Ibrasa, São Paulo.

### ABSTRACT

The identification of social skills deficits is usefulness to promote professional development with emphasis in the interpersonal behavior. In this work we have studied the psychometrical properties of a Social Skills Inventory (IHS) applied to 527 college students, male and female, from humanities, biological and mathematical graduation courses. The major results of the item analysis and alpha factor analysis have showed: positive discrimination coefficients, ranging from 3,0 to 16,7; all correlation coefficients are positive, two not significant, and only ten lower than 0,30; and five factors identified by behavioral situational terms, with alpha coefficients ranging from 0,74 to 0,96, which explain 92,75 of the total variance. Theses results show the IHS usefulness in research and the clinical evaluation.

# Procedimentos Lineares para Identificação e Estimação dos Parâmetros de Modelos

Paulo Roberto de Holanda Sales\*

Basilio de Bragança Pereira\*\*

Ailton de Mesquita Vieira\*\*\*

## RESUMO

Este artigo apresenta procedimentos lineares para identificação e estimação de parâmetros de modelos ARMA, ARMAX (TF), SURARMA (CARMA), MARMA (VARMA) e MARMAX. Os algoritmos são baseados nos trabalhos de Hannan e co-autores e aplicados a séries temporais hidrológicas de vazões mensais em reservatórios do Brasil.

Palavras-chave: Modelos ARMA, ARMAX, SURARMA, MARMA e MARMAX, Regressão/auto-regressão. Critérios de Informação. Séries endógenas e exógenas, Séries de vazões.

## 1. INTRODUÇÃO

---

O sistema interligado de geração elétrica-hidrotérmica brasileiro tem capacidade instalada de 55020 MW das quais 93% são de usinas hidroelétricas. Desde 1973, GCOI - Grupo Coordenador para Operação Interligada do sistema, que tem representantes das 18 principais companhias distribuidoras e da Eletrobrás, é responsável em obter a mais eficiente utilização dos recursos hidro e térmicos do sistema. As atividades do GCOI variam desde o planejamento de operações para os próximos cinco anos até o controle em tempo-real do sistema. Nestas operações de planejamento, a previsão de vazões mensais tem um papel importante como visto em Terry et al. (1985).

---

\* Endereço para correspondência: ELETROBRÁS e Instituto de Matemática e Estatística - UERJ.

\*\* Faculdade de Medicina - NESC e COPPE/UFRJ.

\*\*\* ELETROBRÁS.



Até este trabalho, as previsões de vazões mensais do GCOI eram baseadas na metodologia de Box-Jenkins, Sales (1977) ou Sallas et al. (1980). Na seleção de uma metodologia para um conjunto de dados particular devemos ter em mente as seguintes questões:

- (a) porque a técnica foi usada em vez de outra técnica ?
- (b) o que levou a escolha da técnica nesta aplicação ?
- (c) que utilidade tem os modelos escolhidos ?

Baseado na dimensão do sistema elétrico brasileiro e também na natureza das necessidades do GCOI, combinado com diversas longas séries temporais hidrológicas, cada uma com mais de 600 observações mensais, os métodos lineares automáticos fornecem uma metodologia eficiente e rápida, e por conseqüência uma melhor performance do grupo do GCOI envolvido, do que se usarem métodos não-lineares e não-automáticos em que a função do analista de séries temporais é vital. Além disso, a cada ano alguns representantes das companhias distribuidoras são substituídos e o grupo tem que fornecer treinamento para os novos componentes, e neste caso métodos automáticos são bem-vindos.

Considere-se as relações de dependência entre duas ou mais séries temporais. Os três tipos básicos de dependência podem ser classificados como: contemporânea, unidirecional ou triangular, e de retroalimentação (feedback).

(i) Duas ou mais séries têm relação contemporânea quando os valores presentes são relacionados, isto é, as séries apresentam causalidade instantânea. As vazões em diferentes estações de medição de uma região fornecem um exemplo deste caso se a dependência é apenas contemporânea e, portanto, as mesmas não são relacionadas às modificações naturais ou do homem. A maior dependência então se deve à precipitação, isto é, às séries de entradas das bacias hidrográficas.

(ii) Causalidade unidirecional para duas séries  $x$  e  $y$  pode ser definida quando  $x$  causa  $y$ . Então valores de  $y$  são melhores explicados por seu passado e pelo passado de  $x$ . Como exemplo, vazões de entrada na bacia são importantes (causa) para explicar as vazões afluentes em um outro posto a jusantes do primeiro no mesmo rio, mas não vice-versa. Outro exemplo de causalidade unidirecional seria "chuva" X vazão.

(iii) Uma relação de feedback existe se e só se existe causalidade em ambas as direções. Neste caso a relação é simétrica, isto é,  $x$  causa  $y$  e  $y$  causa  $x$ . Como exemplo, poderíamos mencionar as séries de precipitação e evapotranspiração de uma bacia hidrográfica ou mais ainda, precipitação, evaporação e o armazenamento em um reservatório.

(iv) Finalmente, não é comum, porém possível, encontrar relações mistas entre três ou mais séries. Isto é, duas delas podem ser relacionadas contemporaneamente, porém elas são relacionadas unidirecionalmente com uma terceira e assim por diante.

Neste artigo uma metodologia para modelagem de séries temporais introduzidas no grupo GCOI é apresentada. Ela é combinada com a análise estrutural de Yevjevich e será chamada metodologia de Yevjevich/Hannan. Na Seção 2, modelos de séries temporais para a componente estocástica são apresentados. A Seção 3 apresenta os algoritmos para identificação e estimação destes modelos. Aplicações em séries temporais hidrológicas do Brasil são apresentadas na Seção 4 e alguns comentários na Seção 5.

## 2. MODELOS PARA SÉRIES TEMPORAIS

Um grande número de modelos para séries temporais univariadas e multivariadas tem sido proposto em hidrologia, e eles podem ser classificados de acordo com as relações de dependência mencionadas previamente.

Nesta seção, uma descrição das seguintes subclasses de modelos será apresentada: Auto-regressivo Média Móvel (ARMA), ARMA Exógeno ou Função de Transferência (ARMAX ou TF), ARMA Aparentemente Não Relacionado ou Contemporâneo ARMA (SÜRARMA ou CARMA), ARMA Multivariado ou Vetorial (MARMA ou VARMA) e MARMA Exógeno (MARMAX). Outros modelos como Regressão Múltipla Agregativos, Periódicos ARMA (PARMA) e ARMA Espaço-Temporal (STARMA) não serão abordados.

Considere-se  $m$  séries temporais representadas por:

$$\underline{W}_t = [W_{1t}, W_{2t}, \dots, W_{mt}] \quad (1)$$

e vetor de médias

$$\underline{\mu} = [\mu_1, \mu_2, \dots, \mu_m] \quad (2)$$

e

$$\underline{W}_t = \underline{W}_t^* - \underline{\mu} \quad (3)$$

O modelo MARMA (p, q) é descrito por:

$$\Phi_p(B)Z_t = \Theta_q(B)a_t \quad (4)$$

onde,

(i)  $B$  é o operador de retardo definido por  $B^h W_t = W_{t-h}$ ;

(ii)  $\Phi_p(B)$  e  $\Theta_q(B)$  são matrizes de coeficientes auto-regressivos e médias móveis respectivamente, cujos elementos  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$  são polinômios de graus  $p_{ij} (\leq p)$  e  $q_{ij} (\leq q)$  em  $B$ .

(iii)  $Z_t = [\nabla^{d_1} \nabla_{s_1}^{D_1} W_{1t}^{(\lambda_1)}, \dots, \nabla^{d_m} \nabla_{s_m}^{D_m} W_{mt}^{(\lambda_m)}]$ ;

(iv)  $\nabla^{d_i} = [1 - B]^{d_i}$  é o operador de diferenças de ordem  $d$  das séries  $i \in [1, m]$ ;

(v)  $\nabla_{s_i}^{D_i} = [1 - B^{s_i}]^{D_i}$  é o operador de diferenças sazonal de ordem  $D_i$  e período sazonal  $s_i$  das séries  $i \in [1, m]$ ;

(vi)  $W_{it}^{(\lambda_i)}$  e a transformação do tipo Box & Cox para as séries  $i \in [1, m]$ ; e

(vii)  $a_t$  é o vetor de variáveis aleatórias normais independentes e identicamente distribuídas com média zero e matriz de covariância  $\Omega$ . Isto é, as inovações  $a_{it} \in [1, m]$  são os erros de previsão um passo a frente quando os parâmetros do modelo são conhecidos.

O Processo (4) é dito estacionário e inversível, respectivamente, se e só se as raízes das equações determinantes  $|\Phi_p(B)| = 0$  e  $|\Theta_q(B)| = 0$  estão fora do círculo unitário.

Com relação ao modelo completo multivariado expresso por (4), observe-se que:

(i) durante o processo de construção do modelo os graus dos operadores  $\Phi_{ij}(B)$  e  $\Theta_{ij}(B)$  podem ser ajustados de forma que os modelos ARMA para cada série temporal descrevam precisamente o comportamento de cada série;

(ii) o modelo SURARMA é o resultado dos coeficientes de  $\Phi_{ij}(B)$  e  $\Theta_{ij}(B)$  serem nulos para  $i \neq j$ , isto é, as matrizes de parâmetros são diagonais;

(iii) o modelo ARMAX é obtido quando os coeficientes  $\Phi_{ij}(B)$  e  $\Theta_{ij}(B)$  são nulos para  $i < j$  ou, em outras palavras, quando as matrizes de parâmetros são triangulares; e

(iv) o modelo MARMAX é obtido quando se retira uma ou mais linhas da matriz de parâmetros auto-regressivos,  $\Phi_p(B)$ , e uma ou mais colunas da matriz de parâmetros médias móveis,  $\Theta_q(B)$ .

### 3. ALGORITMOS

Nesta seção quatro algoritmos lineares são apresentados para identificação e estimação de parâmetros de modelos ARMA, ARMAX, SURARMA e MARMA (p, q) como apresentados em Sales (1989) e Sales et al. (1987 e 1989 a, b) para séries temporais hidrológicas.

Os algoritmos foram originalmente propostos por Hannan e Rissanen (1982) e outros desenvolvimentos são apresentados também em Hannan e Deistler (1988).

#### 3.1 – Modelos ARMA (p, q) Univariados

Seja  $\{Z_t, t=1, \dots, T\}$  uma série temporal estacionária gerada por um modelo ARMA definido por:

$$\phi_p(B) Z_t = \theta_q(B) a_t \quad (5)$$

onde,

$$\phi_p(B) = 1 - \phi_1(B) - \dots - \phi_p B^p$$

$$\theta_q(B) = 1 - \theta_1(B) - \dots - \theta_q B^q.$$

Os parâmetros do modelo são denotados  $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ .

Um algoritmo linear para identificação e estimação de parâmetros do modelo ARMA (p, q) é definido como:

#### Passo I – Preliminar

Um modelo auto-regressivo de ordem n é ajustado para as observações  $Z_t$  (t = 1, ..., T). A estimativa de  $\{a_t\}$  no modelo,

$$\phi_n(B) Z_t = a_t \quad (6)$$

é dada por:

$$\hat{a}_t = \sum_{i=1}^n \hat{\phi}_i Z_{t-i} \quad (7)$$

onde,

$$\hat{\phi}_0 = 1 \text{ e } Z_t = 0, \quad \forall t \leq 0.$$

Neste caso,  $n$  é selecionado de forma a minimizar:

$$\text{BIC}(v) = \ln \hat{\sigma}_v^2 + v \frac{\ln T}{T} \quad (8)$$

Onde,  $\hat{\sigma}_v^2$  é a estimativa de  $\sigma_a^2$  na auto-regressão de ordem  $v$ . É suficiente que  $v \leq [\ln T]^{1+\delta}$ ,  $\delta > 0$ .

## Passo II – Identificação e Estimação Inicial

As ordens  $p$ ,  $q$  e a variância dos resíduos  $\sigma_a^2$  são estimadas por  $\tilde{p}$ ,  $\tilde{q}$  e  $\tilde{\sigma}_a^2$ , ao regressar  $Z_t$  em  $Z_{t-i}$  ( $i=1, \dots, \tilde{p}$ ) e  $\hat{a}_1$  ( $i=1, \dots, \tilde{q}$ ). A estimativa inicial  $\beta_{\sim 0} = \begin{pmatrix} \phi_{\sim 0} \\ \theta_{\sim 0} \end{pmatrix}$ , do vetor de parâmetros  $\beta = \begin{pmatrix} \phi_{\sim} \\ \theta_{\sim} \end{pmatrix}$  do modelo, é também estimada neste passo. Neste caso, as estimativas deverão minimizar:

$$\text{BIC}(p, q) = \ln \sigma_{p,q}^2 + (p + q) \frac{\ln T}{T} \quad (9)$$

## Passo III – Estimação Final

Dado  $(\tilde{p}, \tilde{q})$ ,  $\beta_0 = \begin{pmatrix} \phi_{\sim 0} \\ \theta_{\sim 0} \end{pmatrix}$  e usando

$$\hat{a}_{t,0} = \phi_{\tilde{p},0}(B) \theta_{\tilde{q},0}^{-1}(B) Z_t \quad (10)$$

e as séries auxiliares  $\{u_t\}$  e  $\{v_t\}$ , definidas como:

$$u_{i,t} = - \left. \frac{\partial \hat{a}_t}{\partial \phi_i} \right|_{\beta_0} = v_{t-i} \quad (11)$$

$$v_{i,t} = - \left. \frac{\partial \hat{a}_t}{\partial \phi_i} \right|_{\beta_0} = v_{t-i} \quad (12)$$

onde,

$$u_t = \theta_{\bar{q},0}^{-1}(B) Z_t = \phi_{\bar{p},0}(B) \hat{a}_{t,0} \quad (13)$$

$$v_t = \theta_{\bar{q},0}^{-2}(B) Z_t = \phi_{\bar{p},0}(B) Z_t = -\theta_{\bar{q},0}^{-1}(B) \hat{a}_{t,0} \quad (14)$$

A seguinte regressão linear fornece um conjunto de (segundas) estimativas:

$$\hat{a}_{t,0} = \sum_{i=1}^{\bar{p}} [\phi_i - \phi_{i,0}] u_{t-i} + \sum_{i=1}^{\bar{q}} [\theta_i - \theta_{i,0}] v_{t-i} + \hat{a}_t \quad (15)$$

Alternativamente (15) pode ser escrita como:

$$\hat{a}_{t,0} + u_t + v_t = \sum_{i=1}^{\bar{p}} \phi_i u_{t-i} + \sum_{i=1}^{\bar{q}} \theta_i v_{t-i} + \hat{a}_t \quad (16)$$

Em outras palavras, é suficiente regredir, em cada iteração,  $[\tilde{a}_{t,0} + u_t + v_t]$  nas variáveis auxiliares  $u_{i,t}$  ( $i = \overline{1, \bar{p}}$ ) e  $v_{i,t}$  ( $i = \overline{1, \bar{q}}$ ). O processo continua até obter convergência.

Observe que a busca de  $p$  e  $q$  no Passo II pode ser feita economicamente como em Hannan and Rissanen (1982) e Poskitt (1987), isto é, primeiro fixe  $p = q$ ,  $\tilde{p}$  então obtido é uma estimativa consistente de  $\max(p, q)$ . Portanto é necessário somente fazer a busca entre os modelos ARMA ( $\tilde{p}, r$ ) e ARMA ( $r, \tilde{p}$ ),  $r = 0, 1, \dots, \tilde{p}-1$  usando um critério de informação.

### 3.2 – Modelos de Função de Transferência/ARMAX

Sejam  $\{Z_t\}$ ,  $\{X_t\}$  e  $\{Y_t\}$ , respectivamente, as séries endógena e exógenas. Suponha que elas são relacionadas pelo modelo ARMAX:

$$C_p(B) Z_t = D_r(B) X_t + E_s(B) Y_t + F_q(B) a_t ; \quad (17)$$

onde,

$$C_p(B) = 1 - c_1 B - \dots - c_p B^p;$$

$$D_r(B) = d_0 - d_1 B - \dots - d_r B^r;$$

$$E_s(B) = e_0 - e_1 B - \dots - e_s B^s; e$$

$$F_q(B) = 1 - f_1 B - \dots - f_q B^q.$$

A série de resíduos  $\{a_t\}$  pode ser estimada por:

$$\hat{a}_{t,0} = F_{q,0}^{-1}(B) [C_{p,0}(B) Z_t - D_{r,0}(B) X_t - E_{s,0}(B) Y_t] \quad (18)$$

e as séries auxiliares  $\{m_t\}$ ,  $\{n_t\}$ ,  $\{u_t\}$  e  $\{v_t\}$  são dadas por

$$m_{i,t} = - \left. \frac{\partial \hat{a}_t}{\partial c_i} \right|_{\xi_0} = m_{t-i}; \quad m_t = F_{q,0}^{-1}(B) Z_t; \quad (19)$$

$$n_{i,t} = - \left. \frac{\partial \hat{a}_t}{\partial d_i} \right|_{\xi_0} = n_{t-i}; \quad n_t = F_{q,0}^{-1}(B) X_t; \quad (20)$$

$$u_{i,t} = - \left. \frac{\partial \hat{a}_t}{\partial e_i} \right|_{\xi_0} = u_{t-i}; \quad u_t = F_{q,0}^{-1}(B) Y_t; e \quad (21)$$

$$v_{i,t} = - \left. \frac{\partial \hat{a}_t}{\partial f_i} \right|_{\xi_0} = v_{t-i}; \quad v_t = F_{q,0}^{-1}(B) \hat{a}_t. \quad (22)$$

Logo, um algoritmo linear para identificação das ordens e a estimação dos parâmetros do modelo ARMAX, como definido por (17), pode ser sumarizado por:

### Passo I - Preliminar

Ajuste às observações  $\{Z_t\}$ ,  $\{X_t\}$  e  $\{Y_t\}$  o seguinte modelo regressivo-auto-regressivo:

$$\sum_{i=0}^h \hat{c}_i Z_{t-1} = \sum_{i=1}^h \hat{d}_i X_{t-1} + \sum_{i=1}^h \hat{e}_i Y_{t-1} + \hat{a}_t \quad (23)$$

onde,

$$\hat{\sigma}_h^2 = \frac{T_1}{T} \sum_{t=1}^T \hat{a}_{1,t} \quad (24)$$

e  $h$  é tal que

$$\text{BIC}(h) = \ln \hat{\sigma}_h^2 + 3h \frac{\ln T}{T} \quad (25)$$

é um mínimo.

### Passo II – Identificação e Estimação Inicial

Para cada  $p, q, r$  e  $s$ , regredir  $Z$  em

$$-Z_{t-i} \left( i = \overline{1, p} \right), X_{t-i} \left( i = \overline{1, r} \right), Y_{t-i} \left( i = \overline{1, s} \right) \text{ e } \hat{a}_{1,t-i} \left( i = \overline{1, q} \right).$$

Seja  $\xi$  um vetor cujos elementos  $k = p + q + r + s$  são os parâmetros do modelo.  $k$  é tal que minimiza:

$$\text{BIC}(k) = \ln \hat{\sigma}_k^2 + k \frac{\ln T}{T}. \quad (26)$$

### Passo III – Estimação Final

Com as segundas estimativas dos resíduos obtidos no passo anterior, isto é:

$$\hat{a}_{z_t} = \sum_{i=1}^p \hat{c}_{z_t} - \sum_{i=1}^r \hat{d}_{z_t} X_{t-i} - \sum_{i=1}^s \hat{e}_{z_t} Y_{t-i} - \sum_{i=1}^q \hat{f}_{z_t} \hat{a}_{z_{t-i}} \quad (27)$$

onde

$$\hat{a}_{z_t} = Z_t = X_t = Y_t = 0, \quad \forall t \leq 0$$

e com as séries auxiliares dadas por (19), (20), (21) e (22), use séries de Taylor em  $\xi_{-0}$  para

obter novas estimativas dos parâmetros



$$\langle \hat{a}_t | \xi_t, Z_t, X_t, Y_t \rangle = \langle \hat{a}_{t,0} | \xi_t, Z_t, X_t, Y_t \rangle - \sum_{i=1}^k (\xi_i - \xi_{i,0}) \eta_{i,t} \quad (28)$$

onde

$$\eta_{i,t} = - \left. \frac{\partial \langle \hat{a}_t | \xi_t, Z_t, X_t, Y_t \rangle}{\partial \xi_i} \right|_{\xi=\xi_0} \quad (29)$$

onde  $\langle . \rangle$  denota valor esperado.

O procedimento é repetido até a convergência final.

Aqui, como no caso ARMA, uma busca consistente e econômica para  $(p, q, r, s, q)$  é dada pelo Teorema 3 de Poskitt (1989).

### 3.3 – Modelos SURARMA/CARMA $(p, q)$

Considere-se o sistema de  $m$  equações ARMA  $(p_i, q_i)$ :

$$\phi_i(B) \tilde{Z}_{i,t} = \theta_i(B) a_{i,t}, \quad 1 \leq i \leq m \quad (30)$$

onde,

$$\tilde{Z}_{i,t} = Z_{i,t} - \mu_i.$$

O modelo definido em (30) é chamado CARMA  $(p, q)$  onde  $p = \max(p_1, \dots, p_m)$  e  $q = \max(q_1, \dots, q_m)$  se as inovações  $\{a_{i,t}\}$  são somente contemporaneamente relacionadas. Neste caso, as matrizes dos parâmetros são diagonais. Como conseqüência, esta classe de modelos pode ser vista como uma extensão natural do caso univariado e de acordo com Camacho et al. (1985), Camacho e McLeod (1987), como um caso particular do caso multivariado. Por outro lado ela pode ser vista como um conjunto de  $m$  modelos ARMA aparentemente não correlacionados ou modelo SURARMA de acordo com Nelson (1976).

Um algoritmo linear para estimação e identificação do modelo SURARMA  $(p, q)$  pode ser resumido como:

#### Passo I - Preliminar

Para cada uma das  $n$  equações do modelo SURARMA definido por (30) aplica-se o algoritmo de identificação e estimação proposto na seção 3.1.

## Passo II – Identificação e Estimação Inicial

Com as  $m$  séries de resíduos  $\{\hat{a}_{i_t}\}$  do passo anterior, obtenha a matriz de correlação cruzada  $\tilde{\Omega}$  para as séries  $\{\hat{a}_{i_t}\}$  e  $\{\hat{a}_{j_t}\}$ ,  $i, j = 1, m$ , isto é:

$$\hat{\omega}_{i,j} = \frac{\hat{a}_{i_t} \hat{a}_{j_t}}{T} \quad i, j = 1, \dots, m. \quad (31)$$

Se o modelo SURARMA é adequado, só as correlações no lag zero devem ser significativamente diferentes de zero.

Expandindo  $\{\hat{a}_{i_t}\}$  em série de Taylor em  $\beta_{-i_0} = \begin{bmatrix} \phi & \theta \\ -i_0 & -i_0 \end{bmatrix}$  obtém-se aproximadamente,

$$\hat{a}_{i_{t,0}} = \sum_{j=1}^{k_i} \begin{bmatrix} \beta & -\beta \\ -i_j & -i_{j,0} \end{bmatrix} X_{i_{t,j}} + \hat{a}_{i_t} \quad (32)$$

onde

$$X_{i_{t,j}} = - \left. \frac{\partial \hat{a}_{i_t}}{\partial \beta_{i_j}} \right|_{\beta_{-i_0}} \quad (33)$$

e  $k = p_i + q_i, i = 1, \dots, m$ .

A equação (32) pode ser escrita como

$$\hat{a}_{i_{t,0}} = X_t \beta + \hat{a}_{i_t}, \quad t = 1, \dots, T \quad (34)$$

onde  $\hat{a}_{i_{t,0}}$  é um vetor ( $m \times 1$ ) e  $X_t$  é uma matriz ( $m \times k$ ):

$$X_t = \begin{bmatrix} X_1' & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ 0 & 0 & \dots & X_{m_t} \end{bmatrix}. \quad (35)$$

O estimador MQG (mínimos quadrados generalizados viável) é dado por:

$$\hat{\beta} = \left[ \sum_{t=1}^T \mathbf{X}_t' \hat{\Omega}^{-1} \mathbf{X}_t \right]^{-1} \sum_{t=1}^T \mathbf{X}_t' \hat{\Omega}^{-1} \hat{a}_{t,0} \quad (36)$$

e a matriz de covariância de  $\hat{\beta}$  por

$$\text{VAR}(\hat{\beta}) = \left[ \sum_{t=1}^T \mathbf{X}_t' \hat{\Omega}^{-1} \mathbf{X}_t \right]^{-1}. \quad (37)$$

### Passo III – Estimação Final

Obtenha um novo conjunto de  $m$  séries de resíduos de (34) e então aplique (31), (33), (35), (36) e (37) até que a convergência seja obtida.

### 3.4 – Modelos MARMA (p, q)

Considere-se um conjunto de  $m$  séries temporais representadas por (1) e suponha que elas podem ser modeladas por um modelo ARMA (p, q) multivariado como definido em (1). Um algoritmo para identificação da ordem (p, q) e da estimação dos parâmetros correspondentes é:

#### Passo I - Preliminar

Considere-se, inicialmente, modelos estritamente auto-regressivos de ordem  $k$ ,

$$\tilde{Z}_{-t} = \phi_1 \tilde{Z}_{-t-1} + \dots + \phi_k \tilde{Z}_{-t-k} + \hat{a}_{-k,t} \quad (38)$$

onde,

$$\langle \hat{a}_{k,t}, \hat{a}_{k,s} \rangle = \begin{cases} \hat{\Omega}, & t = s \\ 0, & \text{de outro modo.} \end{cases}$$

Estime (38) para todos os valores possíveis de  $k$ . A ordem da auto-regressão é tal que minimiza

$$\text{AIC}(k) = \ln |\hat{\Omega}_k| + \frac{2km^2}{T} \quad (39)$$

onde  $T$  é o menor número de observações entre as  $m$  séries.

O critério AIC é conservador pois sobrestima a ordem do modelo. Entretanto, o principal objetivo do Passo I é estimar as  $m$  séries de resíduos e não o modelo verdadeiro.

As inovações podem ser estimadas por

$$\hat{a}_{\tilde{t}} = \tilde{Z}_{\tilde{t}} - \hat{\phi}_1 \tilde{Z}_{\tilde{t}-1}, \dots, \hat{\phi}_k \tilde{Z}_{\tilde{t}-k} \quad (40)$$

## Passo II – Identificação e Estimação Inicial

Obtenha, para os diversos valores de  $p$  e  $q$ , as regressões  $\tilde{Z}_{\tilde{t}}$  em  $\tilde{Z}_{\tilde{t}-i}$  ( $i = 1, \dots, q$ ) e,  $\hat{a}_{\tilde{t}}$  onde  $\hat{a}_{\tilde{t}}$  são obtidos de (40).

Os modelos estimados neste passo são:

$$\tilde{Z}_{\tilde{t}} = \hat{\phi}_1 \tilde{Z}_{\tilde{t}-1} + \dots + \hat{\phi}_p \tilde{Z}_{\tilde{t}-p} - \hat{\theta}_1 \hat{a}_{\tilde{t}-1}, \dots, -\hat{\theta}_q \hat{a}_{\tilde{t}-q} + \hat{a}_{\tilde{t},q} \quad (41)$$

onde  $\hat{a}_{\tilde{t},q}$  é o novo vetor de resíduos.

A vantagem do passo anterior é que a equação (41) pode ser facilmente estimada por mínimos quadrados ordinários.

Finalmente, selecione  $p$  e  $q$  tal que

$$\text{BIC}(p, q) = \ln |\hat{\Omega}_{p,q}| + (p+q)m^2 \frac{\ln T}{T}. \quad (42)$$

Então, após a seleção das ordens  $p$  e  $q$ , obtém-se

$$\hat{a}_{\tilde{t}} = \tilde{Z}_{\tilde{t}} - \hat{\phi}_1 \tilde{Z}_{\tilde{t}-1} - \dots - \hat{\phi}_p \tilde{Z}_{\tilde{t}-p} + \hat{\theta}_1 \hat{a}_{\tilde{t}-1} + \dots + \hat{\theta}_q \hat{a}_{\tilde{t}-q} \quad (43)$$

## Passo III – Estimação Final

O procedimento consiste em linearizar  $\hat{a}_{\tilde{t}}$  em  $\beta_{0_i}$  ( $i = 1, \dots, k$ ), onde  $\beta_{0_i}$  é uma estimativa inicial do vetor de parâmetros auto-regressivos e médios móveis,  $\beta_i$  ( $i = 1, \dots, k$ ), obtido no passo anterior. Então a expressão do estimador de  $\beta_i$  pode ser escrita como

$$\hat{\beta}_{\sim i} = \left[ \sum_{t=\sim i}^T \hat{a}^{i'} \hat{\Omega}_{p,q}^{-1} \hat{a}^i \right]^{-1} \sum_{t=\sim i}^T \hat{a}^{i'} \hat{\Omega}_{p,q}^{-1} \hat{a}_{\sim t,0}^i \quad (44)$$

onde  $\hat{a}_{\sim t}$  é a derivada parcial de  $\hat{a}_{\sim t}$  com respeito a  $\beta_i$ .

O estimador da variância assintótica de  $\hat{\beta}_{\sim i}$  é dado por

$$\text{VAR}(\hat{\beta}_{\sim i}) = \left[ \sum_{t=\sim i}^T \hat{a}^{i'} \hat{\Omega}_{p,q}^{-1} \hat{a}^i \right]^{-1} \quad (45)$$

e o estimador da covariância entre  $\hat{\beta}_{\sim i}$  e  $\hat{\beta}_{\sim j}$  é dado por

$$\text{COV}(\hat{\beta}_{\sim i}, \hat{\beta}_{\sim j}) = \left[ \sum_{t=\sim i}^T \hat{a}^{i'} \hat{\Omega}_{p,q}^{-1} \hat{a}^j \right]^{-1} \quad (46)$$

Se o modelo não tem termos médias-móveis,  $a_{\sim t}$  é linear em  $\beta_{\sim i}$  e a solução é dada simplesmente por:

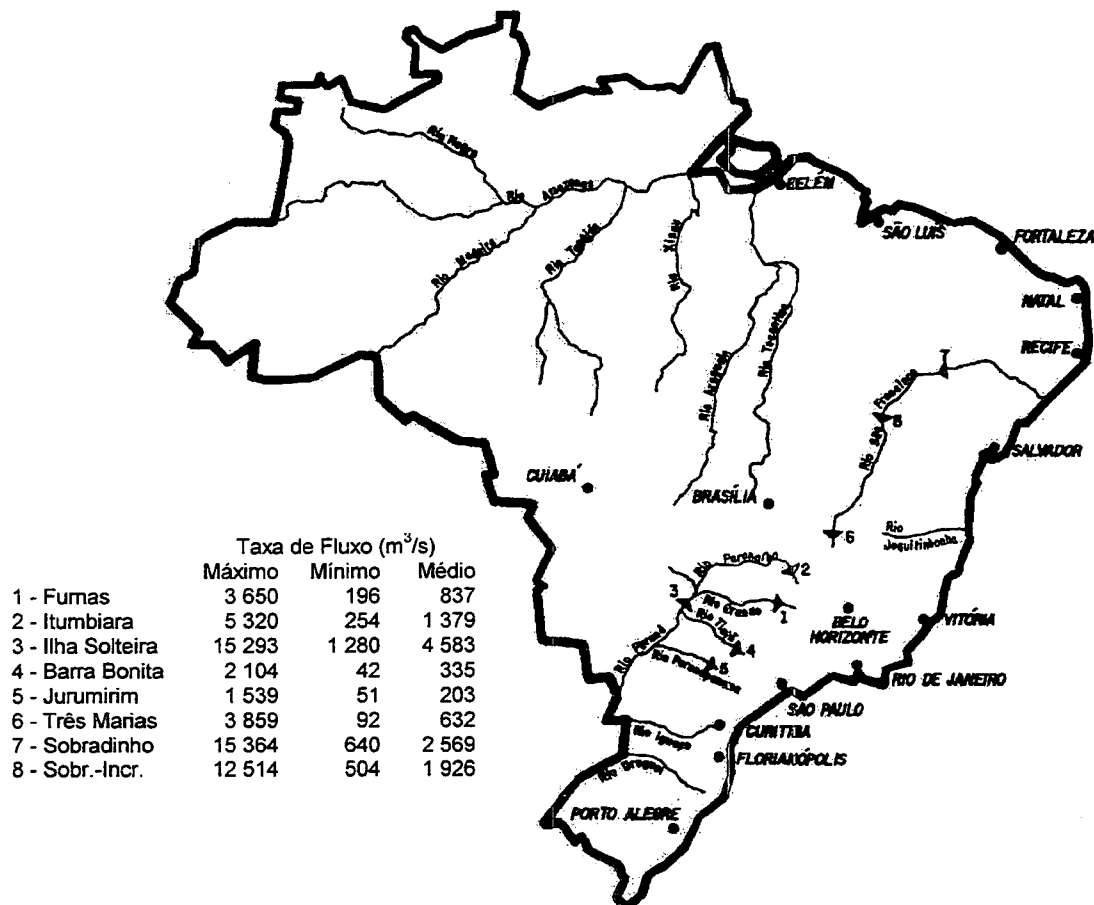
$$\hat{\beta}_{\sim} = \hat{\beta}_{\sim 0} + \left[ \hat{\beta}_{\sim} - \hat{\beta}_{\sim 0} \right] \quad (47)$$

Caso contrário, um novo vetor de estimadores do parâmetro dado por (47) é usado no lugar de  $\hat{\beta}_{\sim 0}$  e o processo de linearização é repetido até a convergência.

#### 4. APLICAÇÕES

Os algoritmos propostos foram usados em oito séries de vazões naturais mensais nos reservatórios de Furnas no Rio Grande, Itumbiara no Rio Paranaíba, Ilha Solteira no Rio Paraná, Barra Bonita no Rio Tietê, Jurumirim no Rio Paranapanema, Três Marias e Sobradinho no Rio São Francisco, bem como a série incremental de Sobradinho. Cada uma das séries temporais hidrológicas analisadas tem 648 observações. Os dados cobrem o período de janeiro de 1931 a dezembro de 1984 e foram obtidos das Centrais Elétricas Brasileiras S. A. – ELETROBRÁS – Brasil – Figura 1.

Figura 1- Localização das séries temporais usadas nos algoritmos lineares



Inicialmente, pela natureza das séries temporais hidrológicas, foi necessário transformar os dados usando a transformação de Box & Cox, dada por:

$$Z_t^{(\lambda_1, \lambda_2)} = \begin{cases} \lambda_1^{-1} [(Z_t + \lambda_2)^{\lambda_1} - 1] & , \lambda_1 \neq 0 \\ \ln(Z_t + \lambda_2) & , \lambda_1 = 0 \end{cases} \quad (49)$$

Este tipo de transformação pode evitar problemas com heteroscedasticidade e não-normalidade nas séries. De acordo com a Tabela 1 todas as séries foram transformadas tomando o logaritmo.

Tabela 1 – Transformação de Box-Cox para as séries mensais das taxas de vazões mensais

Séries	$\lambda_1$	$\lambda_2$
Furnas	0	- 179
Itumbiara	0	- 179
Ilha Solteira	0	- 1238
Barra Bonita	0	- 35
Jurumirim	0	- 45
Três Marias	0	- 90
Sobradinho	0	- 529
Bacia Intermediária	0	- 468

Em seguida, para obter estacionariedade de segunda ordem, cada série foi dessazonalizada subtraindo a média mensal e dividindo pelo desvio padrão mensal cada observação da série dos logaritmos, Sales (1977), Sallas et al. (1980).

#### 4.1 – Modelo univariado ARMA (p, q) para série de Furnas

No Passo I, modelos auto-regressivos foram ajustados. O critério BIC(p) indicou  $p = 2$ . Passando para o Passo II onde “regressões” ARMA são ajustadas para a série transformada e para a série de resíduos obtida no primeiro passo. O critério BIC indicou  $p = q = 1$ .

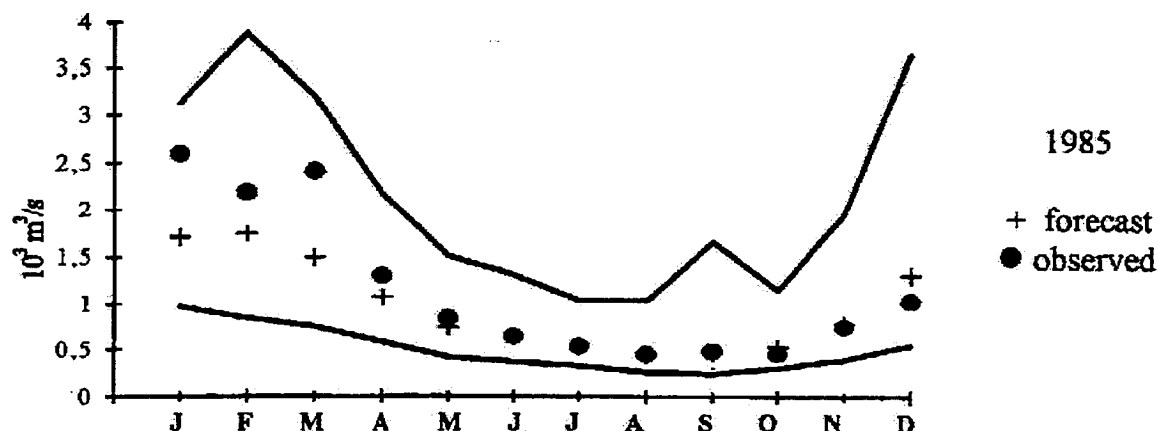
A partir das estimativas iniciais dos parâmetros do modelo ARMA (1,1) identificado no passo anterior,  $\hat{\phi}_1 = 0.8426$ ,  $\hat{\theta}_1 = -0.2278$  e  $\hat{\sigma}_a^2 = 0.4351$  movemos para o Passo III do algoritmo proposto. As estimativas finais na Tabela 2 foram obtidas após dez iterações com precisão  $1 \times 10^{-4}$ . Erros padrão das estimativas também são dados na Tabela 2.

Tabela 2 – Estimativa final dos parâmetros  $\phi_1$ ,  $\theta_1$  e  $\sigma_a^2$  do modelo ARMA (1,1) para Furnas

Parâmetro	Estimativa	Erro padrão
$\phi_1$	0.8421	0.0237
$\theta_1$	- 0.2398	0.0426
$\sigma_a^2$	0.4343	

Previsões para um ano a frente com intervalos de dois erros padrão são apresentadas na Figura 2.

Figura 2 – Previsões para 1985 com dois erros padrão (EP) modelo ARMA de Furnas



#### 4.2 – Modelo ARMAX (p, r, s, q) para as séries Três Marias, Sobradinho e Incremental de Sobradinho

Diversos modelos de regressão/auto-regressão foram testados no Passo I de acordo com (23) e (24), considerando diversos valores de  $h$ . O critério  $BIC(h)$ , dado em (25) indicou  $h = 2$ .

Partindo para o Passo II onde regressões/auto-regressões são ajustadas para às séries  $\{Z_t\}$ ,  $\{X_t\}$ ,  $\{Y_t\}$  e à série de resíduos  $\{a_t\}$  obtida no passo anterior. A Tabela 3 apresenta os valores de  $BIC(k)$  e as variâncias dos resíduos das regressões/auto-regressões consideradas no Passo II. Neste passo as ordens do modelo ARMAX identificado foram  $p = 1$ ,  $r = 2$ ,  $s = 0$  e  $q = 1$ . As estimativas iniciais dos parâmetros são  $\hat{c}_1 = 0.8075$ ,  $\hat{d}_1 = 0.6227$ ,  $\hat{d}_2 = -0.5202$ ,  $\hat{f}_1 = -0.4130$  e  $\hat{\sigma}_a^2 = 0.2946$ .

Após as estimativas iniciais do Passo II do algoritmo, passamos ao Passo III. As estimativas finais foram obtidas após oito iterações com precisão  $1 \times 10^{-4}$  e são apresentadas na Tabela 4 com os erros padrão correspondentes.



Tabela 3 – Valor de BIC(k),  $K=p+r+s+q$  e as variâncias dos resíduos dos modelos regressão/auto-regressão do Passo II

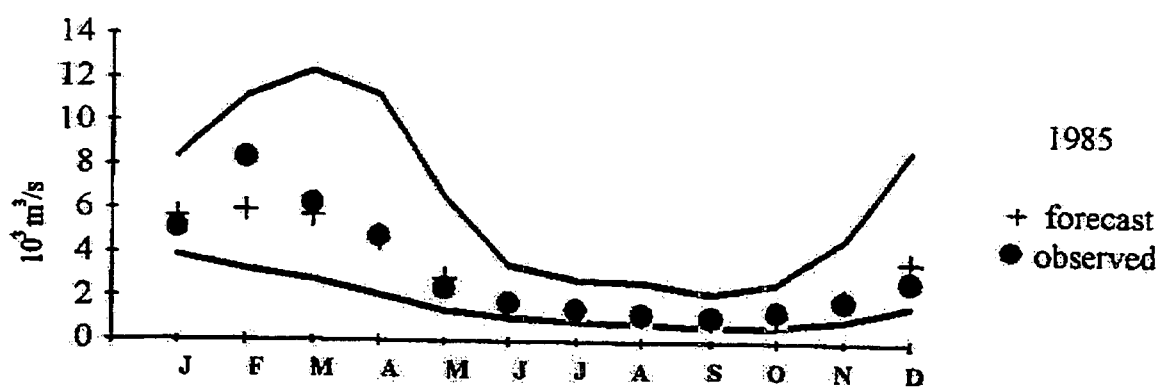
p	q											
	0				1				2			
	r	s			r	s			r	s		
0		1	2	0		1	2	0		1	2	
0	0		-0,6675 (0,5079)	-0,6708 (0,5012)	0	-0,2858 (0,8059)	-0,6708 (0,5012)	-0,6939 (0,4849)	0	-0,3147 (0,7156)	-0,6463 (0,5085)	-0,688 (0,4829)
	1	-0,9308 (0,3903)	-1,1122 (0,3223)	-1,1115 (0,3194)	1	-0,9913 (0,3637)	-1,1051 (0,3214)	-1,1418 (0,3067)	1	-1,0258 (0,3479)	-1,0951 (0,3214)	-1,1307 (0,3071)
	2	-0,9308 (0,3909)	-1,1442 (0,3091)	-1,1710 (0,2979)	2	-0,9938 (0,3592)	-1,1596 (0,3013)	-1,1600 (0,2982)	2	-1,0183 (0,3471)	-1,1486 (0,3016)	-1,1505 (0,2981)
1	0	-0,8599 (0,4190)	0,9260 (0,3883)	0,9159 (0,3883)	0	-0,8522 (0,4180)	-0,9525 (0,3744)	-1,0061 (0,3513)	0	-0,88442 (0,4172)	-0,9412 (0,3749)	-0,9970 (0,3508)
	1	-1,0940 (0,3283)	-1,1045 (0,3216)	-1,1081 (0,3173)	1	-1,0954 (0,3245)	-1,0949 (0,3215)	-1,1304 (0,3072)	1	-1,0864 (0,3242)	-1,0850 (0,3214)	-1,1631 (0,2943)
	2	-1,1557 (0,3055)	-1,1448 (0,3058)	-1,1723 (0,2946)	2	-1,1819 (0,2947)	-1,1709 (0,2950)	-1,1649 (0,2938)	2	-1,1740 (0,2941)	-1,1631 (0,2943)	-1,1585 (0,2928)
2	0	-0,8485 (0,4196)	-0,9155 (0,3885)	-1,0045 (0,3519)	0	-0,8498 (0,4149)	-0,9730 (0,3632)	-1,0355 (0,3377)	0	-0,8437 (0,4133)	-0,9730 (0,3632)	-1,0567 (0,3774)
	1	-1,0831 (0,3286)	-1,0980 (0,3205)	-1,1446 (0,3028)	1	-1,1137 (0,3155)	-1,1074 (0,3143)	-1,1454 (0,0296)	1	-1,1030 (0,3157)	-1,0965 (0,3146)	-1,1484 (0,2957)
	2	-1,1762 (0,2964)	-1,1654 (0,2966)	-1,1660 (0,2935)	2	-1,1732 (0,2943)	-1,1623 (0,2940)	-1,1570 (0,2932)	2	-1,1652 (0,2937)	-1,1545 (0,2939)	-1,1030 (0,2899)

Tabela 4 – Estimativas finais dos parâmetros  $c_1, d_1, d_2, f_1$  e  $\sigma_a^2$  do modelo ARMAX para Três Marias, Sobradinho e Bacia Intermediária

Séries	Variáveis	Parâmetro	Estimativas	Erro Padrão
Sobradinho	$Z_{t-1}$	$c_1$	0.8469	0.0866
Três Marias	$X_{t-1}$	$D_1$	0.5996	0.0358
	$X_{t-2}$	$D_2$	-0.4626	0.0742
Residual	$a_{t-1}$	$f_1$	-0.3536	0.0899
		$\sigma_a^2$	0.2939	

Previsões ex-ante para um ano à frente com dois erros padrão são apresentadas na Figura 3.

Figura 3 – Previsão ex-ante para 1985 com dois erros padrão. Modelo ARMAX, de Sobradinho com Três Marias como input



#### 4.3 – Modelos SURARMA (p, q) para as Séries: Ilha Solteira, Barra Bonita e Jurumirim

Correlações cruzadas entre os resíduos de modelos univariados foram calculadas para Ilha Solteira, Barra Bonita e Jurumirim. Padrão típicos são apresentados na Figura 4. O comportamento das correlações cruzadas sugere adequação do modelo SURARMA para estas séries.

Figura 4a - Correlação cruzada dos resíduos: Ilha Solteira e Barra Bonita

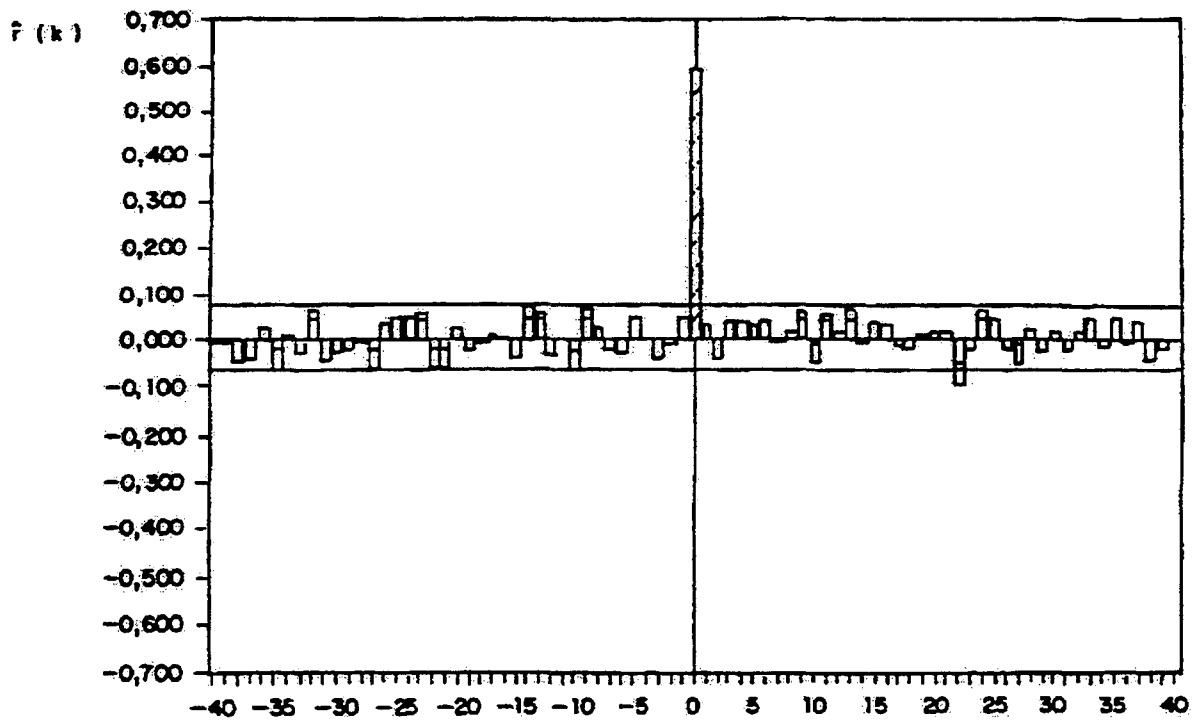


Figura 4b - Correlação cruzada dos resíduos: Barra Bonita e Jurumirim

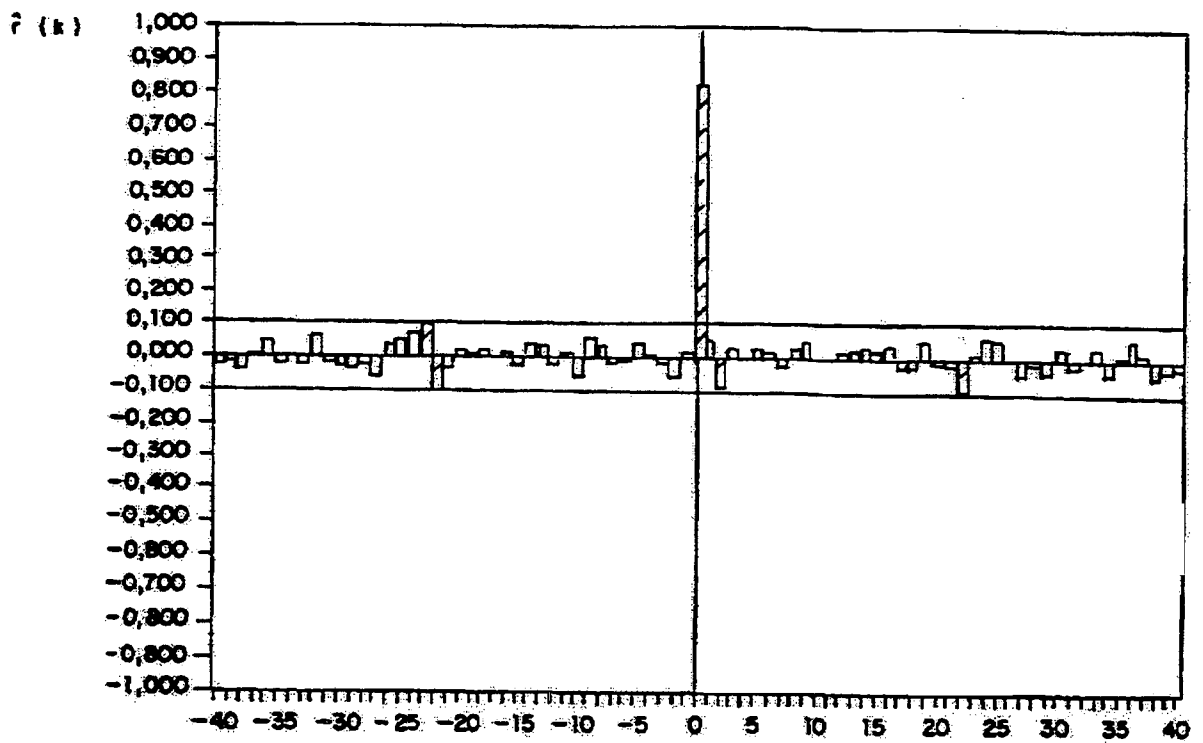
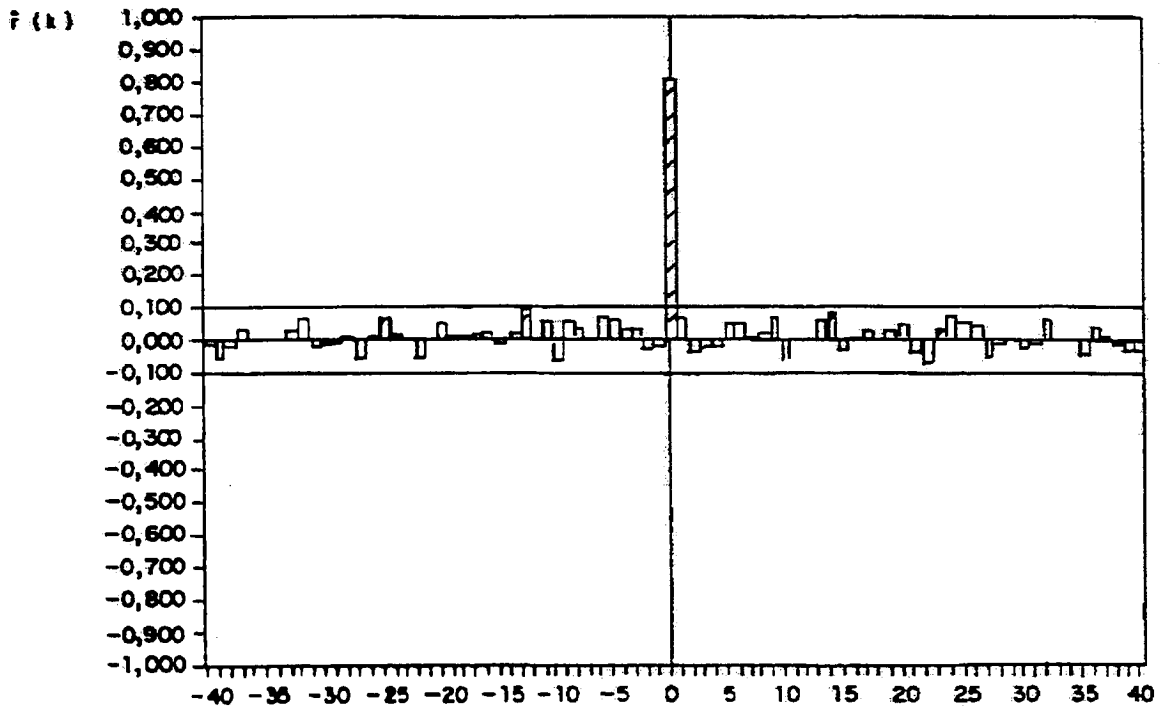


Figura 4c - Correlações cruzadas dos resíduos: Ilha Solteira e Jurumirim



O Passo I do algoritmo consiste em identificar modelos univariados,  $ARMA [p_i, q_i]$  para cada uma das  $i = 1, 2, 3$  séries. Neste estágio, obtêm-se também, as estimativas dos modelos selecionados. Deve-se enfatizar que estes são os valores iniciais do processo iterativo do Passo III e são obtidos aplicando o algoritmo proposto na Seção 3.1 a cada uma das séries. A Tabela 5 apresenta os modelos selecionados.

Tabela 5 - Modelos  $ARMA [p_i, q_i]$  selecionados para Ilha Solteira, Barra Bonita e Jurumirim

Séries	(p, q)	Estimativas
Ilha Solteira	(1,1)	$\hat{\phi}_1 = 0.8229 (0.0258)$
		$\hat{\theta}_1 = -0.6120 (0.0449) \quad \hat{\sigma}_a^2 = 0.4104$
Barra Bonita	(1,0)	$\hat{\phi}_1 = 0.8214 (0.0251)$
		$\hat{\phi}_1 = -0.2948 (0.0420) \quad \hat{\sigma}_a^2 = 0.5136$
Jurumirim	(1,1)	$\hat{\phi}_1 = 0.5747 (0.0291) \quad \hat{\sigma}_a^2 = 0.5414$

No Passo II, com as séries dos resíduos obtidos no passo anterior, aplicamos (31) e (36) para obter as estimativas iniciais da matriz de covariância  $\Omega$  e o vetor de parâmetros do modelo, respectivamente.

O Passo III considera de forma iterativa, as estimativas da matriz  $\Omega$  para obter estimativas do vetor  $\beta$  e os erros padrão correspondentes.

A Tabela 6 apresenta os resultados de convergência do algoritmo após quatro iterações com precisão  $1 \times 10^{-4}$ . Comparando os resultados apresentados na Tabela 5 notamos que os erros padrão do modelo SURARMA são menores que aqueles dos modelos ARMA.

Tabela 6 – Estimativa final dos parâmetros do modelo SURARMA (1,1) para Ilha Solteira, Barra Bonita e Jurumirim

Séries	Estimativa Final
Ilha Solteira	$\hat{\phi}_1 = 0.7954 (0.0234)$
	$\hat{\theta}_1 = -0.1610 (0.0404)$ $\hat{\sigma}_a^2 = 0.4099$
Barra Bonita	$\hat{\phi}_1 = 0.7616 (0.0188)$
	$\hat{\phi}_1 = -0.3324 (0.0245)$ $\hat{\sigma}_a^2 = 0.5151$
Jurumirim	$\hat{\phi}_1 = 0.6588 (0.0231)$ $\hat{\sigma}_a^2 = 0.5421$

Previsões para um ano a frente para as três séries e intervalos de confiança com dois erros padrão são apresentados na Figura 5.

Figura 5a - Previsão para 1985 com dois erros padrão. Modelo SURARMA de Ilha Solteira

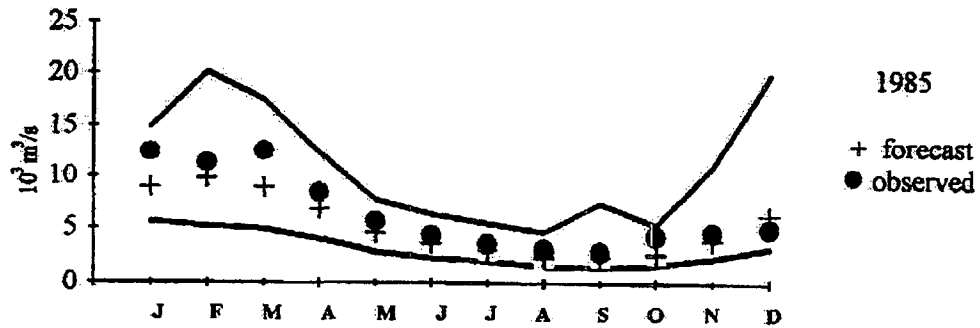


Figura 5b - Previsão para 1985 com dois erros padrão. Modelo SURARMA de Barra Bonita

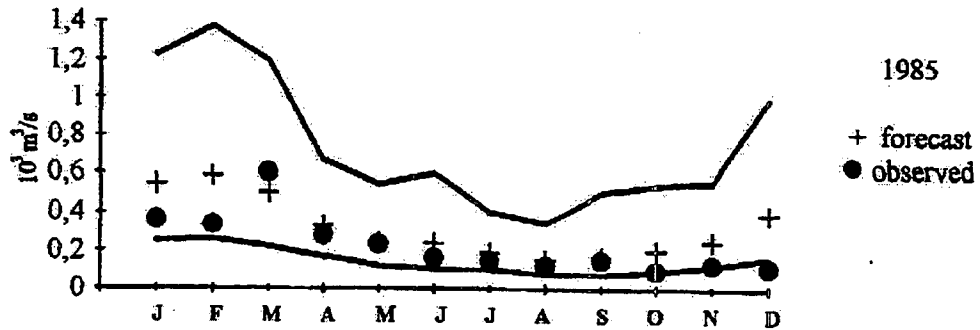
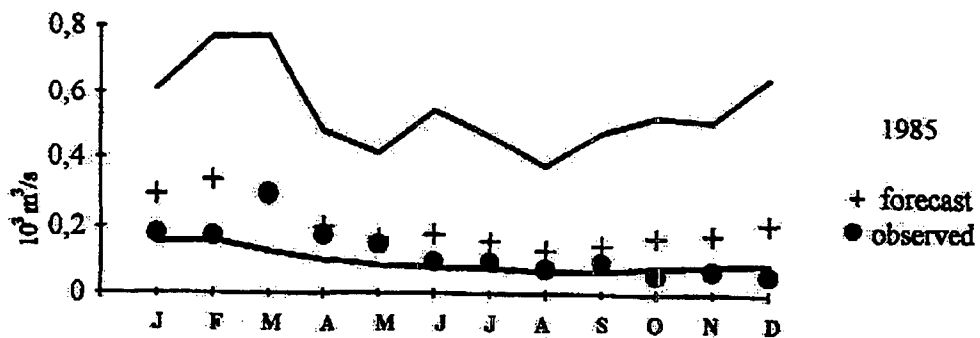


Figura 5c - Previsão para 1985 com dois erros padrão. Modelo SURARMA de Jurumirim



#### 4.4 – Modelo MARMA (p, q) para as Séries: Itumbiara e Três Marias

No Passo I, modelos auto-regressivos são ajustados aos dados. O critério AIC(k) dado em (39) sugere um modelo MAR(3).

Usando os resíduos das três séries obtidas no passo anterior, estima-se diversos modelos MARMA (p, q). Com o critério BIC (p, q) dado em (42) escolheu-se um modelo MARMA (1,1). Estimativas iniciais dos parâmetros do modelo são obtidos no Passo II com os respectivos erros padrão e são apresentados na Tabela 7.

Tabela 7 – Estimativa final do modelo MARMA (1,1) de Furnas, Itumbiara e Três Marias

Séries	$\phi$ Matriz			$\theta$ Matrix		Variação Residual	
	Furnas	Itumbiara	Três Marias	Furnas	Itumbiara	Três Marias	
Furnas	0.8693 (0.0355)	-0.0516 (0.0382)	0.0151 (0.0429)	-0.3296 (0.0509)	-0.0814 (0.0509)	0.0513 (0.0577)	0.4301
Itumbiara	0.0188 (0.0362)	-0.7517 (0.0390)	0.0199 (0.0438)	0.0466 (0.0506)	-0.2004 (0.0520)	0.0939 (0.0589)	0.4479
Três Marias	0.0452 (0.0362)	-0.0617 (0.0391)	0.8201 (0.0439)	-0.0423 (0.0362)	-0.1055 (0.0521)	0.2094 (0.0590)	0.4498

Uma análise cuidadosa dos resultados do Passo II permite o máximo uso do algoritmo proposto para as séries de Furnas. Em vista disso, tentamos dois modelos MARMA (1,1) no Passo III.

Inicialmente, o processo iterativo do terceiro passo considera o modelo multivariado completo, isto é, sem restrições impostas aos parâmetros, as estimativas finais obtidas após cinco iterações com precisão  $1 \times 10^{-4}$ . A Tabela 8 apresenta os resultados obtidos.

Tabela 8 – Estimativa final do modelo MARMA (1,1) sem restrições, para Furnas, Itumbiara e Três Marias

Séries	$\phi$ Matriz			$\theta$ Matrix		Variação Residual	
	Furnas	Itumbiara	Três Marias	Furnas	Itumbiara	Três Marias	
Furnas	0.8587 (0.0209)	-0.0477 (0.0243)	0.0204 (0.0260)	-0.3769 (0.0359)	-0.0994 (0.0366)	0.0643 (0.0415)	0.4297
Itumbiara	0.0094 (0.0214)	-0.7444 (0.0248)	0.0288 (0.0266)	0.0106 (0.0366)	-0.2072 (0.0373)	0.1141 (0.0424)	0.4477
Três Marias	0.0359 (0.0214)	-0.0548 (0.0248)	0.8232 (0.0267)	-0.0868 (0.0367)	0.1324 (0.0374)	0.2093 (0.0425)	0.4493

Em seguida, impõe-se restrições aos parâmetros do modelo MARMA (1,1). Em outras palavras, a hipótese de que nem todos os parâmetros diferem de zero é considerada. De fato, o modelo SURARMA parece adequado, neste caso, porém, para efeito de ilustração do algoritmo MARMA, deletamos somente parâmetros dentro do intervalo de um erro padrão.

No caso específico das séries analisadas, quando as restrições  $\phi_{19} = \phi_{21} = \phi_{29} = 0$  e  $\theta_{19} = \theta_{21} = \theta_{29} = 0$  foram impostas, um novo conjunto de estimativas para os outros parâmetros foi obtido, bem como a série de resíduos. A Tabela 9 mostra essas novas estimativas para o modelo MARMA (1,1) restrito, bem como os erros padrão correspondentes.

Tabela 9 – Estimativa inicial do modelo restrito MARMA (1,1) de Furnas, Itumbiara e Três Marias

Séries	$\phi$ Matriz			$\theta$ Matrix		Variação Residual	
	Furnas	Itumbiara	Três Marias	Furnas	Itumbiara	Três Marias	
Furnas	0.8688 (0.0205)	-0.0555 (0.0384)	-	-0.3382 (0.0366)	0.0815 (0.0523)	-	0.4299
Itumbiara	-	-0.7389 (0.0327)	-	-	-0.2214 (0.0569)	0.0967 (0.0449)	0.4478
Três Marias	0.0458 (0.0213)	-0.0678 (0.0363)	0.8213 (0.0242)	-	0.1052 (0.0566)	-0.2177 (0.0436)	0.4497

As estimativas finais dos parâmetros do modelo MARMA (1,1) restrito foram obtidas após quatro iterações no Passo III do algoritmo proposto com  $1 \times 10^{-4}$  de precisão. A Tabela

10 resume os principais resultados das estimativas, bem como os erros padrão mostrados entre parênteses.

Tabela 10 – Estimativa final do modelo restrito MARMA (1,1) de Furnas, Itumbiara e Três Marias

Séries	$\phi$ Matriz			$\theta$ Matrix		Variação Residual	
	Furnas	Itumbiara	Três Marias	Furnas	Itumbiara	Três Marias	
Furnas	0.86858 (0.0206)	-0.05515 (0.0354)	-	-0.3397 (0.0363)	0.0827 (0.0363)	-	0.4299
Itumbiara	-	-0.7380 (0.0333)	-	-	-0.2239 (0.0565)	0.0971 (0.0445)	0.4478
Três Marias	0.0459 (0.0215)	-0.0665 (0.0369)	0.8202 (0.0243)	-	0.1093 (0.0562)	-0.2218 (0.0431)	0.4497

Previsões para as três séries para um ano a frente com intervalos de dois erros padrão são apresentadas na Figura 6.

Figura 6a - Previsão para 1985 com dois erros padrão. Modelo MARMA de Furnas

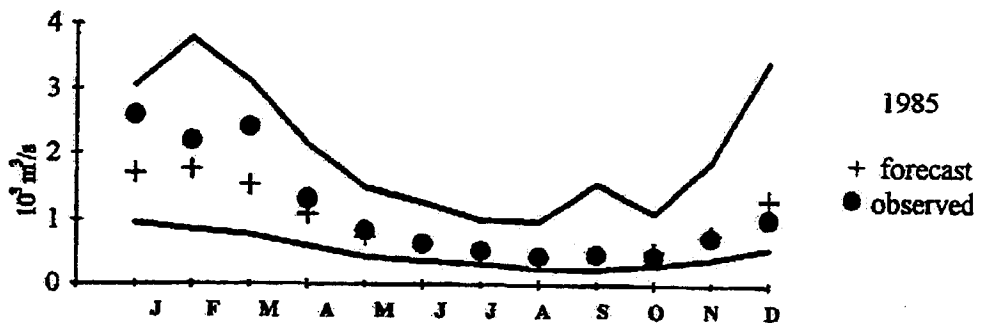


Figura 6b - Previsão para 1985 com dois erros padrão. Modelo MARMA de Itumbiara

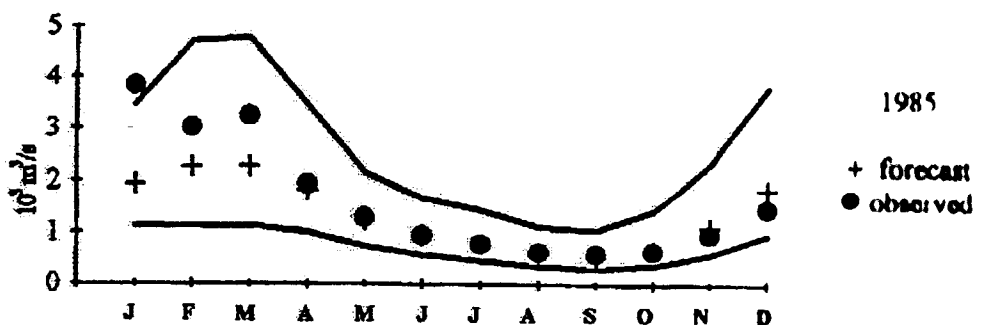
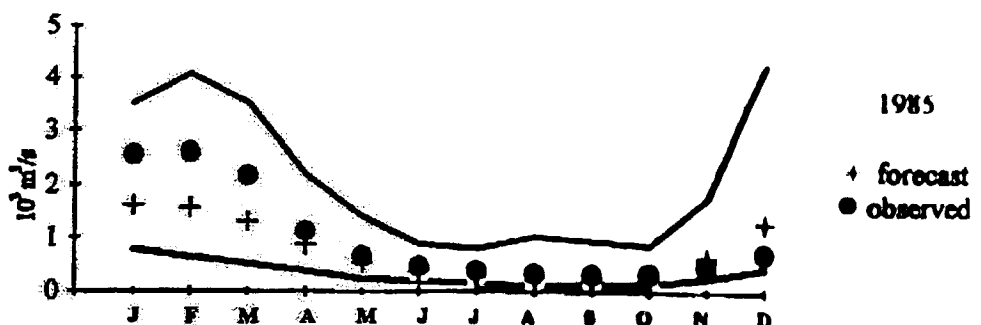


Figura 6c - Previsão para 1985 com dois erros padrão. Modelo MARMA de Três Marias





## 5. COMENTÁRIOS FINAIS

Propriedades teóricas dos procedimentos de identificação e estimação apresentados neste artigo são apresentadas em Hannan e Deistler (1988). Resultados de simulação e aplicações destes métodos e outros relacionados com ele, são apresentados em Saikkonen (1986), Newbold e Hotopp (1986), Hannan e McDougall (1988), Poskitt (1980), Koreisha e Pukkila (1989, 1990 a, b), Pukkila et al. (1990) e Galbraith e Linde-Walsh (1997).

Os tempos de computação em um IBM 4381 R14 foram respectivamente 3.06 s para o ARMA, 10.52 s para o ARMAX,  $(4.67 + 3 \times 3,06)$  s para o SURARMA e 19.78 s para o MARMA.

### AGRADECIMENTOS

Os autores são gratos ao falecido Professor E. J. Hannan, por seu incentivo e por tornar disponível muitos de seus artigos na época ainda não publicados.

### REFERÊNCIAS

- CAMACHO, F.; MCLEOD, A. I. E HIPEL, K. W. (1985) – Contemporaneous autorregressive-moving average (CARMA) modelling in water resources. *Water Resources Bulletin*, 21, 709-720.
- CAMACHO, F. E MCLEOD, A. I. (1987) – Contemporaneous bivariate time series. *Biometrika*, 74, 103-113.
- GALBRAITH, J. W. E LINDE-WALSH, V. (1997) – On some simple, autorregression-based estimation and identification techniques for ARMA models. *Biometrika*, 84, 685-696.
- HANNAN, E. J. E DEISTLER, M. (1988) – *The Statistical Theory and Linear Systems*. John Wiley & Sons, N. Y.
- HANNAN, E. J. E MCDOUGALL, A. J. (1988) – Regression procedures for ARMA estimation. *J. Am. Statist. Ass.*, 83, 490-498.
- HANNAN, E. J. E RISSANEN, J. (1983) – Recursive estimation of mixed autorregressive-moving average estimator. *Biometrika*, 69, 81-94. Correction. *Biometrika*, 70, 303.
- KOREISHA, S. E PUKKILA, T. (1989) – Fast linear estimation methods for vector autorregressive moving-average models. *J. of Time Series An.*, 10, 325-339.
- KOREISHA, S. E PUKKILA, T. (1990 a) – Linear methods for estimating ARMA and regression models with serial correlation. *Commun. Statist. Simula*, 19, 71-102.
- KOREISHA, S. E PUKKILA, T. (1990 b) – A generalized least-squares approach for estimation of autorregressive moving average models. *J. of Time Series An.*, 11, 139-151.
- NELSON, C. R. (1976) – Gains in efficiency from joint estimation of system of autorregressive-moving average processes. *J. of Econometrics*, 4, 331-348.
- NEWBOLD, P. E HOTOPP, S. M. (1986) – Testing causality using efficiently parametrized vector ARMA models. *App. Math. And Comp.*, 20, 329-348.
- POSKITT, D. S. (1987) – A modified Hannan-Rissanen strategy for mixed autorregressive-moving average order determination. *Biometrika*, 74, 781-790.
- POSKITT, D. S. (1989) – A method for the estimation and identification of transfer function models. *J. Roy. Statist. Soc. B*, 51, 29-46.
- PUKKILA, T.; KOREISHA, S. E KALLINEN, A. (1990) – The identification of ARMA models. *Biometrika*, 73, 537-548.
- SAIKKONEN, P. (1986) – Asymptotic properties of some preliminary estimators for autorregressive moving average time series models. *J. of Time Series An.*, 7, 133-155.

- SALES, P. R. H. (1977) – A stochastic deterministic model for hydrological time series forecasting, *M. Sc. Thesis*, COPPE/UFRJ.
- SALES, P. R. H. (1989) – Linear procedures for identification and parameter estimation of models for uni and multivariate time series, *D. Sc. Thesis*, COPPE/UFRJ.
- SALES, P. R. H.; PEREIRA, B. DE B. E VIEIRA, A. M. (1987) – Linear procedures for identification and estimation of ARMA models for hydrological time series *Annals of the VII Brazilian Symposium of Hydrology and Hydric Resources*, Salvador, Bahia, 605-615.
- SALES, P. R. H.; PEREIRA, B. DE B. E VIEIRA, A. M. (1989 a) – A linear procedure for identification of transfer function models for hydrological time series. *Annals of the IV Lusitanian-Brasilian Symposium on Hydraulics and Hydric Resources*, Lisboa, Portugal, 321-336.
- SALES, P. R. H.; PEREIRA, B. DE B. E VIEIRA, A. M. (1989 b) – A linear procedure for identification and estimation of SURARMA models applied to multivariate hydrological time series. *Annals of the IV Lusitanian-Brasilian Symposium of Hydraulics and Hydric Resources*. Lisboa, Portugal, 283-248.
- SALLAS, J. D.; DELLEUR, J. W.; YEVJEVICH, V. E LANE, W. L. – Applied Modelling of Hydrologic Time Series. *Water Resources Publication*.
- TERRY, L. A.; PEREIRA, M. V. F.; ARARIPE NETO, T. A.; SILVA, L. F. C. A. E SALES, P. R. H. (1986) – Coordinating the energy generation of the Brazilian national hydrothermal electrical generating system. *Interfaces*, 16, 16-38.

## ABSTRACT

This paper presents linear procedures for identification and parameter estimation of ARMA, ARMAX (TF), SURARMA (CARMA), MARMA (VARMA) and MARMAX models. The algorithms were based on Hannan and co-authors works and applied to hydrological time series of monthly average natural inflows of reservoirs in Brazil.

**Keywords:** ARMA, ARMAX, SURARMA, MARMA and MARMAX models. Natural inflow series Regression/Autorregressions. Natural inflow series. Information criteria. Endogenous, exogenous series.

# Estudo da Direção Tomada por Pombos Através de Equações de Estimação para Dados Circulares Longitudinais

Rinaldo Artes\*

Gilberto Alvarenga Paula\*

Ronald Ranvaud\*\*

## RESUMO

Apresentamos neste trabalho uma nova metodologia para a análise da direção tomada por pombos em diferentes instantes de tempo após a soltura. A metodologia proposta é uma extensão do método de quase-verossimilhança (Wedderburn, 1974) e da técnica das equações de estimação generalizadas (Liang e Zeger, 1986), podendo ser aplicada a uma grande variedade de dados circulares longitudinais. Discutimos formas para a modelagem dos parâmetros de posição e dispersão em dados circulares bem como apresentamos estimadores robustos para a matriz de covariância dos estimadores obtidos das equações de estimação propostas. Ilustramos a metodologia analisando parte dos dados de um experimento sobre a orientação de pombos realizado próximo à cidade de Camocim, no Estado do Ceará.

## 1. INTRODUÇÃO

---

Neste artigo, analisamos parte dos dados de uma série de experimentos realizados, de 1982 a 1991, próximo à cidade de Camocim, localizada a 300 km de distância ao noroeste da cidade de Fortaleza, no Estado do Ceará. Os experimentos visavam a estudar mecanismos de bússola na orientação de pombos. A escolha dessa cidade deveu-se ao fato de que durante o período de realização desses experimentos o equador magnético da Terra atravessava aquela região, dessa forma, eliminando a influência da bússola magnética na orientação das aves. A escolha da época do ano e da hora do dia, também, foi criteriosa. Os experimentos foram executados ao redor do equinócio, de forma a eliminar, durante a passagem do sol nas proximidades do zênite (sol a pino), também a pista ambiental que viabiliza a bússola solar.

---

\* Endereço para correspondência: Instituto de Matemática e Estatística - USP - C.Postal 66281 - CEP 05315-970 - São Paulo.

\*\* Instituto de Ciências Biomédicas - USP - Av. Prof. Lineu Prestes, 1524 - CEP 05508-900 - São Paulo - SP.

Os experimentos foram realizados em três momentos do dia: pela manhã e à tarde, quando os pombos podiam se apoiar na bússola solar, dado que o sol indicava nestes momentos claramente uma direção cardeal, e ao meio-dia, momento em que o sol a pino não fornecia um referencial confiável para que os pombos pudessem embasar a escolha do rumo a ser tomado. Em resumo, todos os pombos no experimento foram privados da bússola magnética pela escolha da região do experimento; o grupo liberado ao meio-dia também não tinha à sua disposição a bússola solar (Schmidt-Koenig et al., 1991). Estes são os únicos dois mecanismos de bússola evidenciados na literatura. Os pombos liberados de manhã e à tarde, por outro lado, tinham à disposição a bússola solar, formando assim grupos de controle para os pombos liberados ao meio-dia.

Os pombos eram retirados de suas gaiolas e liberados individualmente e independentemente um do outro. Tomando-se por referência o Norte geográfico da Terra, observou-se o azimute (posição) das aves (em graus) após 30, 60 e 90 segundos da soltura e no momento de desaparecimento no horizonte. Maiores detalhes sobre essa série de experimentos podem ser obtidos em Ranvaud et al. (1991).

Para o estudo da influência dos fatores horário e local de soltura e da variação dos rumos ao longo do tempo, iremos, neste artigo, apresentar uma nova técnica dirigida à análise de dados circulares longitudinais. Essa técnica é uma extensão da teoria de quase-verossimilhança de Wedderburn (1974) e, mais especificamente, desenvolve-se a partir do método das equações de estimação generalizadas de Liang e Zeger (1986). Uma das motivações para o desenvolvimento deste formalismo foi o fato de que até o presente a literatura que relata pesquisa sobre a orientação de pombos-correio baseia conclusões levando em conta apenas os rumos de sumida, ignorando o comportamento das aves entre a soltura e o sumir. Há muitos dados, portanto, que nunca foram analisados, e que podem conter informações úteis.

Na Seção 2, caracterizamos os dados circulares e apresentamos brevemente algumas medidas descritivas, distribuições de probabilidades e modelos de análise de regressão adequados à análise dessas variáveis. Na Seção 3, descrevemos, resumidamente, o método da quase-verossimilhança e o método das equações de estimação generalizadas. Em seguida, na Seção 4, introduzimos nosso método, cuja aplicação encontra-se na Seção 6. Por fim, na Seção 7, apresentamos nossos comentários finais.

## **2. DADOS CIRCULARES**

Seja  $z$  uma variável aleatória absolutamente contínua; uma variável aleatória circular contínua  $y$  pode ser definida como  $y = a + \{z(\bmod m)\}$ , onde  $a \in \mathfrak{R}$ ,  $m > 0$  e o operador está indicando o resto da divisão de  $z$  por  $m$ . Um exemplo importante é o caso em que  $y$

corresponde a um ângulo (variável angular). Nessa situação, todos os valores  $y + 2k\pi, k = \dots, -2, -1, 0, 1, 2, \dots$  correspondem ao mesmo ângulo.

Uma variável quantitativa absolutamente contínua circular difere de uma não-circular (doravante denominada linear), por possuir uma métrica bastante singular. Para ilustrar isso, suponha que se tenha observado dois ângulos: o primeiro sendo  $15^\circ$  e o segundo  $35^\circ$ . Pode-se questionar sobre qual seria o valor do ângulo dessas observações. Ao se calcular a média aritmética chega-se a  $25^\circ$ , o que parece ser bastante razoável. Suponha agora que a origem dos ângulos seja modificada através da aplicação de uma rotação de  $-25^\circ$ . Os valores transformados passam a ser  $350^\circ$  e  $10^\circ$ , que têm  $180^\circ$  como média aritmética. Esse valor está muito distante dos observados. Baseando-se nos dados, pode-se intuir que o valor "médio" deveria ser  $0^\circ$ <sup>1</sup>. Na verdade, a distância entre duas observações de uma variável circular,  $a < b$ , é dada por  $\min \{ |a - b| \pmod{m}; |a + m - b| \pmod{m} \}$ .

O estudo de dados circulares é bastante antigo e tem sua origem no estudo da posição de corpos celestes (Bernoulli, 1734, por exemplo). Como exemplos de variáveis circulares temos: a direção tomada por animais após certo tratamento, a direção do vento, a direção de correntes marítimas, o horário de falha de um sistema, o horário de entrada de pacientes em uma UTI, etc. Fisher (1993) apresenta 24 conjuntos de dados circulares reais.

O procedimento padrão na análise de dados com essa natureza é convertê-los em ângulos, através de uma transformação linear e analisá-los nessa escala. A partir de agora, não se fará distinção entre os termos "circular" e "angular" e, doravante, considerar-se-á a variável já convertida em ângulos. Em Mardia (1972), Batschelet (1981) e Fisher (1993), encontra-se uma grande variedade de técnicas e modelos probabilísticos para a análise de dados circulares.

## 2.1 Representação Gráfica e Conceitos Básicos

Usualmente, uma observação  $y$  de uma variável circular pode ser representada graficamente em um círculo de raio unitário na posição  $(1, y)$ , quando se utilizam coordenadas polares, ou  $(\cos(y), \sin(y))$ , quando se utilizam coordenadas cartesianas.

A média circular de um conjunto de dados circulares,  $y_i, i = 1, 2, \dots, n$ , é dada por

$$\hat{\mu} = \begin{cases} \arctan(S / C), & \text{se } S \geq 0 \text{ e } C > 0, \\ \arctan(S / C) + \pi, & \text{se } C < 0, \\ \arctan(S / C) + 2\pi, & \text{se } S < 0 \text{ e } C > 0, \end{cases} \quad (1)$$

<sup>1</sup> Valor que corresponde à denominada média circular dos dados, cujo conceito será apresentado na Seção 2.1.

onde,  $S = \sum_{i=1}^n \text{sen}(y_i)$  e  $C = \sum_{i=1}^n \text{cos}(y_i)$ . Essa medida corresponde à direção do

baricentro dos dados, considerados como massas pontuais posicionadas na circunferência de um círculo de raio unitário. Outra interpretação possível pode ser obtida quando se considera cada observação como um vetor de comprimento  $um$  e direção  $y_i$ . Nesse caso, a média circular corresponde à direção do vetor resultante.

Uma medida de concentração bastante utilizada na análise de dados circulares é o comprimento do vetor resultante,

$$R = \sqrt{S^2 + C^2}.$$

No caso de todas as observações serem coincidentes, tem-se que  $R = n$ ; esse é o caso de concentração máxima dos dados (variabilidade mínima). Outro caso limite se dá quando os ângulos encontram-se uniformemente distribuídos no círculo, onde  $R = 0$ ; trata-se do caso de concentração mínima (variabilidade máxima). Note que o conceito de variabilidade máxima não existe para dados lineares. Cabe salientar que quando a variabilidade é máxima, a média circular não está definida. Usualmente utiliza-se o comprimento da resultante média, definido por

$$\bar{R} = \frac{R}{n}, \quad (2)$$

que tem a vantagem de variar no intervalo  $[0,1]$ .

## 2.2 Modelos Probabilísticos

Seja  $y$  uma variável aleatória circular definida no intervalo semi-aberto de amplitude  $2\pi, S_{2\pi}$ . Inicialmente, definiremos a média circular e o comprimento da resultante média de  $y$ , para tanto, é necessário introduzir o conceito de função característica circular. A função característica circular de  $y$ , no ponto  $r, r = 1, 2, \dots$  (Mardia, 1972, por exemplo), é uma função complexa, dada por

$$T_r = E\{\cos(ry)\} + iE\{\text{sen}(ry)\}.$$

Sejam  $\mu$  e  $\rho$ , respectivamente, a média circular e o comprimento da resultante média, então

$$T_1 = \rho e^{i\mu}.$$

Dentre os vários modelos probabilísticos adequados à análise de dados circulares, dois serão destacados: as distribuições uniforme circular e von Mises. Outros modelos podem ser encontrados em Mardia (1972), Watson (1982) e Fisher (1993).

### 2.2.1 Distribuição Uniforme Circular

A distribuição uniforme circular possui função densidade de probabilidades dada por

$$p(y) = \frac{1}{2\pi}, y \in S_{2\pi}.$$

Uma característica que merece destaque é o fato dessa distribuição não possuir média circular e ter comprimento da resultante média igual a zero. Esse fato é bastante importante uma vez que a distribuição uniforme circular tem um papel central na teoria assintótica desenvolvida para dados circulares (Mardia, 1972, por exemplo).

### 2.2.2 Distribuição von Mises

A distribuição von Mises é uma das mais utilizadas na modelagem de dados reais. Isso se deve ao fato de sua densidade ter uma forma relativamente simples, possuir boas propriedades (regularidade, parâmetros interpretáveis e simetria) e parâmetros que são facilmente estimáveis. Se  $y \in S_{2\pi}$  segue uma distribuição von Mises, com média circular  $\mu \in S_{2\pi}$  e parâmetro de concentração  $\lambda > 0$ , representa-se  $y \sim M(\mu, \lambda)$ , então sua função densidade de probabilidades é dada por

$$p(y; \mu, \lambda) = \frac{1}{2\pi I_0(\lambda)} \exp\{\lambda \cos(y - \mu)\}, \quad (3)$$

onde  $I_0(\lambda)$  é a função modificada de Bessel de primeiro tipo e ordem zero, avaliada no ponto  $\lambda$  (Abramowitz e Stegun, 1970, por exemplo). De um modo geral, tem-se

$$I_r(\lambda) = \sum_{k=0}^{\infty} \frac{\{(k+r)!\ k!\}}{k!} \left(\frac{1}{2}\lambda\right)^{2k+r}, r = 0, 1, 2, \dots$$

Os momentos trigonométricos de uma distribuição von Mises com parâmetros  $\mu$  e  $\lambda$  são

$$E\{\cos(ry)\} = A_r(\lambda) \cos(r\mu) \text{ e } E\{\sin(ry)\} = A_r(\lambda) \sin(r\mu),$$

onde,  $A_r(\lambda) = I_r(\lambda) / I_0(\lambda)$ .

O comprimento da resultante média é dado por  $A_1(\lambda)$ .

### 2.3 Relação entre as Distribuições Uniforme Circular e von Mises

A distribuição uniforme circular surge naturalmente, a partir de uma von Mises, quando  $\lambda$  se aproxima de zero. Como a distribuição uniforme circular não possui média circular e tem

comprimento da resultante média zero, tem-se que, quando  $\lambda$  for muito pequeno, as técnicas baseadas na função densidade de probabilidades (estimação por máxima verossimilhança, teste da razão de verossimilhança, etc) não produzem bons resultados. Regras empíricas que estipulam valores de  $\lambda$  para os quais essas técnicas levam vantagem sobre procedimentos não paramétricos são propostas em Fisher (1993) ( $\lambda > 2$ , na maioria das vezes ou mesmo  $\lambda > 1$ , em alguns casos). Outra propriedade importante é o fato da distribuição von Mises se aproximar de uma distribuição normal para valores grandes de  $\lambda$ . No limite, quando  $\lambda$  tende a infinito, essa distribuição tende a uma degenerada no ponto  $\mu$ .

## 2.4 Modelos de Regressão

Há na literatura algumas sugestões de modelos de regressão para dados circulares (ver Gould, 1969; Johnson e Wehrly, 1978 e Rivest, 1989); no entanto, apresentar-se-á, por sua generalidade, o modelo sugerido por Fisher e Lee (1992). Considere uma amostra aleatória  $(y_i, x_i), i = 1, 2, \dots, n$ , onde  $y_i \sim M(\mu, \lambda)$  e  $x_i$  é um vetor  $p$ -variado de covariáveis fixas.

Seja

$$\mu_i = \alpha + h(\eta_i), \eta_i = x_i^T \beta \quad (4)$$

onde,  $\beta$  é um vetor paramétrico,  $h(\cdot)$  é uma função  $h: \mathcal{R} \rightarrow [-\pi, \pi]$ , duplamente diferenciável e inversível, com  $h(0) = 0$  e  $\alpha$  é a origem. Uma possível escolha para  $h$  é  $h(a) = 2 \arctan(a)$ .

A função score associada aos parâmetros desse modelo conduz às seguintes equações:

$$\begin{aligned} X^T H u &= 0, \\ \bar{R} \text{sen}(\hat{\alpha}) &= 0, \\ \bar{R} \text{cos}(\hat{\alpha}) &= 0, \end{aligned} \quad (5)$$

$$\text{e } A_1(\hat{\lambda}) = \bar{R},$$

onde,

$$u = (u_1, u_2, \dots, u_n)^T, u_i = \text{sen}\{y_i - \hat{\alpha} - h(x_i^T \hat{\beta})\}, X = (x_1, \dots, x_n)^T, H = \text{diag}\left\{\frac{\partial n}{\partial \eta_1}(\eta_1), \dots, \frac{\partial n}{\partial \eta_n}(\eta_n)\right\}$$

e  $\bar{R} = (1/n)(S^2 + C^2)^{1/2}$ , sendo

$$S = \sum_{i=1}^n \text{sen}\{y_i - \hat{\alpha} - h(x_i^T \hat{\beta})\} \text{ e}$$

$$C = \sum_{i=1}^n \text{cos}\{y_i - \hat{\alpha} - h(x_i^T \hat{\beta})\}.$$



Fisher e Lee (1992) sugerem a utilização do método de mínimos quadrados ponderados iterativo (Green, 1984) para a obtenção da estimativa de  $\beta$ . Em resumo, o valor atualizado da estimativa na  $k$ -ésima iteração ( $\hat{\beta}^{(k)}$ ) seria obtido a partir de

$$X^T H^2 X (\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)}) = X^T H^2 y,$$

sendo  $H^2 = HH$ .

Em alguns casos, pode-se estar interessado na modelagem do parâmetro de dispersão. Uma situação real onde isso é importante, é o estudo da direção tomada por animais após algum tratamento, ou durante o processo migratório; a variabilidade dessa direção tende a diminuir com o passar do tempo. Em geral, a direção é representada por um ângulo que usualmente se supõe seguir uma distribuição von Mises. No caso de independência entre as observações há opções para a modelagem do parâmetro de concentração  $\lambda$  (Fisher, 1993 e suas referências). Considere, por exemplo,  $y_i \sim M(\mu_i, \lambda_i)$  e suponha a modelagem de  $\mu_i$  proposta anteriormente. Fisher e Lee (1992) propõem o seguinte modelo para a modelagem do parâmetro de concentração:

$$f(\lambda_i) = x_i^T \gamma,$$

onde,  $f(\cdot)$  é uma função de ligação positiva, por exemplo a logarítmica, e  $\gamma$  é um vetor de parâmetros desconhecidos.

Fisher e Lee (1992) apresentam um algoritmo para a obtenção da estimativa de  $\gamma$  baseado no método de mínimos quadrados ponderados iterativo.

### 3. QUASE-VEROSSIMILHANÇA

O método de quase-verossimilhança, Wedderburn (1974), permite a estimação de parâmetros de um modelo de regressão em situações onde a distribuição da variável dependente é desconhecida, mas há informações sobre seus dois primeiros momentos. Mais especificamente, conhecemos o comportamento da média em função das variáveis independentes e da variância em função da média.

Sejam  $y_1, y_2, \dots, y_n$  as observações da variável dependente  $Y$  e  $x_1, x_2, \dots, x_n$  seus respectivos vetores coluna  $p$ -dimensionais de covariáveis. Suponha que  $E(y_i) = \mu_i = h(x_i^T \beta)$  onde  $h(\cdot)$  é uma função contínua duplamente diferenciável e  $\beta$  é um vetor  $p \times 1$  de parâmetros. Assuma que  $\text{Var}(y_i) = \phi V(\mu_i)$  onde  $V(\cdot)$  é uma função conveniente. O estimador de quase-verossimilhança de  $\beta$  é obtido através da raiz de

$$\Psi(\beta) = D^T W^{-1} (y - \mu) = 0$$

onde  $y = (y_1, y_2, \dots, y_n)^T$ ,  $\mu = (\mu_1, \mu_2, \dots, \mu_n)^T$ ,  $D = \partial \mu / \partial \beta$  e  $W = \text{diag}(V_1, V_2, \dots, V_n)$ , McCullagh (1983). Essa função é denominada quase-escore. Sob condições gerais de regularidade, prova-se que a raiz de  $\Psi(\beta) = 0$  é um estimador consistente e assintoticamente normal de  $\beta$ .

### 3.1 Observações Dependentes

Seja  $y_i = (y_{i1}, y_{i2}, \dots, y_{it})^T$ ,  $i = 1, 2, \dots, n$  uma amostra de vetores aleatórios mutuamente independentes, tal que associado à  $y_{ij}$  exista um vetor  $\hat{p} \times 1$  de covariáveis fixas,  $x_{ij}$ . Suponha que  $E(y_{ij}) = \mu_{ij}$ ,  $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{it})^T$ ,  $\mu_{ij} = h(x_{ij}^T \beta)$  e  $\text{Var}(y_{ij}) = \phi V_{ij}(\mu_{ij})$  com  $h(\cdot)$  e  $\beta$ , como na seção anterior. Suponha também que a matriz de covariância de  $y_i$  seja dada por

$$\text{Cov}(y_i) = W_i = \phi V_i^{1/2} R_i V_i^{1/2}$$

com,  $V_i = \text{diag}(V_{i1}, V_{i2}, \dots, V_{it})$  e tal que  $R_i$  a matriz de correlação de  $y_i$ , possa ser definida como função exclusivamente de  $\mu_i$ ,  $R_i = R_i(\mu_i)$ . Nesse caso, o estimador de quase-verossimilhança de  $\beta$  é dado pela raiz de

$$\Psi(\beta) = \sum_{i=1}^n D_i^T W_i^{-1} (y_i - \mu_i) = 0 \quad (6)$$

onde,  $D_i = \partial \mu_i / \partial \beta$ . Sob condições gerais de regularidade, demonstra-se que tal estimador é consistente e assintoticamente normal.

No entanto, geralmente não se deseja modelar  $R_i$  em função de  $\mu_i$ . Na verdade, tal modelagem traz uma série de complicações de ordem prática, uma vez que o coeficiente de correlação está definido em um intervalo limitado e, mais ainda, que  $R_i$  deve ser uma matriz positiva definida. Dessa forma, em termos aplicados a função (6) tem pouca utilidade. Liang e Zeger (1986) desenvolveram uma metodologia para a estimação de parâmetros de um modelo de regressão multivariado que contorna esses problemas. Este é o assunto da próxima seção.

### 3.2 Equações de Estimação Generalizadas

A idéia central das equações de estimação generalizadas (EEGs) é a utilização de parâmetros de perturbação ao invés da média, na modelagem de  $R_i$ . Considere a notação da Seção 3.1.

As EEGs são construídas a partir de

$$\sum_{i=1}^n D_i^T W_i^{-1}(\alpha)(y_i - \mu_i),$$

onde  $D_i = H_i X_i$ ,  $X_i = (x_{i1}, \dots, x_{it_i})^T$ ,

$H_i = \text{diag}\{\dot{h}_{i1}, \dots, \dot{h}_{it}\}$ ,  $\dot{h}_{ij} = dh(w) / dw \Big|_{w=x_{ij}^T \beta}$ ,  $W_i = \phi V_i^{1/2} R_i(\alpha) V_i^{1/2}$  e  $\alpha$  é um vetor  $s \times 1$

que define  $R_i$ .

Note que ao definir a matriz de pesos  $W_i$ , os autores introduzem um novo vetor de parâmetros,  $\alpha$ . O problema passa a ser a estimação dessa quantidade. Liang e Zeger (1986), demonstram que, sob condições bem gerais, ao se substituir  $\alpha$  por um estimador,  $\hat{\alpha}_n$ ,  $\sqrt{n}$ -consistente dados  $\beta$  e  $\phi$ , a solução de

$$\sum_{i=1}^n D_i^T W_i^{-1}(\hat{\alpha}_n)(y_i - \mu_i) = 0 \quad (7)$$

é um estimador consistente e assintoticamente normal de  $\beta$ . A expressão (7) é denominada equações de estimação generalizadas.

Outro ponto importante é o fato de que  $W_i$  não precisa necessariamente ser a matriz de covariância de  $y_i$ . Para que as propriedades assintóticas se verifiquem, basta que  $\hat{\alpha}_n$  satisfaça as condições expostas no último parágrafo.

Uma hipótese importante nas demonstrações das propriedades assintóticas dos estimadores apresentados nesta seção é que  $E_{\beta}(\Psi) = 0$ .

## 4. METODOLOGIA PROPOSTA

Nesta seção, apresentamos o método desenvolvido para dados circulares.

### 4.1 Modelagem da Média Circular

Considere a notação da Seção 3.1, sendo  $y_{ij}$  uma variável aleatória circular, que segue uma distribuição biparamétrica com média circular  $\mu_{ij}$  e comprimento da resultante média  $\rho$ .

Seja

$$\mu_{ij} = h(\eta_{ij}), \eta_{ij} = x_{ij}^T \beta.$$

Sabe-se que  $E_{\beta} \{\text{sen}(y_{ij} - \mu_{ij})\} = 0$ . De modo análogo às equações de estimação generalizadas, definindo  $u_i = (u_{i1}, u_{i2}, \dots, u_{it})^T$ , com  $u_{ij} = \text{sen}(y_{ij} - \mu_{ij})$  e tomando-se  $W_i = \text{Cov}(u_i)$ ,  $\beta$  pode ser consistentemente estimado através das raízes de

$$\Psi_n^W(\beta) = \sum_{i=1}^n \psi_i^W(\beta) = 0$$

onde,  $\psi_i^W(\beta) = D_i W_i u_i$ , com  $D_i = X_i^T H_i$ ,  $H_i = \frac{\partial \mu_i}{\partial \eta_i} = \text{diag} \left\{ \frac{\partial \mu_{i1}}{\partial \eta_{i1}}, \dots, \frac{\partial \mu_{it}}{\partial \eta_{it}} \right\}$ . No caso em que

$W_i$  é conhecida, as raízes de  $\Psi_n^W$  são, sob condições gerais, estimadores de  $\beta$  consistentes e assintoticamente normais com matriz de covariância assintótica dada por

$$n \left( - \sum_{i=1}^n S_i^{-1} \right) \quad (8)$$

onde

$$S_i = D_i^T W_i^{-1} D_i.$$

Em termos práticos a função  $\Psi_n^W$  tem pouca importância, uma vez que, em geral, as matrizes  $\text{Cov}(u_i)$  são desconhecidas. Isto pode ser contornado, substituindo-se  $W_i$  por um estimador, de modo análogo ao desenvolvido por Liang e Zeger (1986).

Considere a existência de um vetor paramétrico  $\alpha$  que define a matriz de covariâncias de  $u_i$ ,  $W_i = W_i(\alpha)$ ,  $i = 1, 2, \dots, n$ . Dessa forma, o problema reduz-se a encontrar um estimador de  $\alpha$ ,  $\hat{\alpha}_n$ , de tal modo que o estimador obtido a partir de

$$\Psi_n(\beta) = \sum_{i=1}^n D_i^T W_i^{-1}(\hat{\alpha}_n) u_i \quad (9)$$

continue sendo consistente e assintoticamente normal. Para isso, além de condições gerais de regularidade (Artes, 1997 e Artes e Jørgensen, 1998 e Artes, Paula e Ranvaud, 1998, para detalhes), basta que  $\hat{\alpha}_n = \hat{\alpha}_n(\beta, \rho)$  seja um estimador  $\sqrt{n}$ -consistente de  $\alpha$ , dados  $\beta$  e  $\rho$ . Note que nesse caso,  $W_i$  não precisa necessariamente ser a matriz de covariâncias de  $u_i$ . As propriedades assintóticas dos estimadores serão válidas mesmo no caso de assumida

uma estrutura de covariâncias incorreta para  $u_i$ . A matriz de covariância assintótica é dada por

$$n \left( \sum_{i=1}^n S_i \right)^{-1} \left( \sum_{i=1}^n V_i \right) \left( \sum_{i=1}^n S_i \right)^{-IT}$$

onde

$$V_i = D_i^T W_i^{-1} \text{Cov}(u_i) W_i^{-1} D_i.$$

Note que essa matriz reduz-se a (8) quando  $W_i = \text{Cov}(u_i)$ . Sendo  $\hat{\beta}_n$  um estimador de  $\beta$  obtido a partir de (9) temos que um estimador consistente de sua matriz de covariância assintótica é dado por

$$n \left( \sum_{i=1}^n \hat{S}_i \right)^{-1} \left( \sum_{i=1}^n \hat{V}_i \right) \left( \sum_{i=1}^n \hat{S}_i \right)^{-IT},$$

onde

$$\hat{S}_i = D_i^T (\hat{\beta}_n) W_i^{-1}(\hat{\alpha}_n) D_i (\hat{\beta}_n)$$

e

$$\hat{V}_i = D_i^T (\hat{\beta}_n) W_i^{-1}(\hat{\alpha}_n) u_i u_i^T W_i^{-1}(\hat{\alpha}_n) D_i (\hat{\beta}_n).$$

Outros detalhes técnicos sobre a metodologia proposta são apresentados em Artes (1997), Artes e Jørgensen (1998) e Artes, Paula e Ranvaud (1998).

## 4.2 Modelagem Mista

Em muitas situações práticas, em dados circulares, observa-se uma variação na variabilidade da resposta ao longo do tempo, desse modo o procedimento apresentado na seção anterior pode não ser suficiente para a modelagem de dados reais. Nesta seção, adaptamos a metodologia proposta de modo a permitir que a variabilidade possa se alterar longitudinalmente. Admita a existência de vetores de covariáveis  $z_{ij}$  associados às observações  $y_{ij}$ , que supostamente poderiam vir a influenciar a variabilidade dos dados.

Uma medida de variabilidade bastante comum em dados circulares é o comprimento da resultante média, definido na Seção 2.2. Da definição, temos que o comprimento da resultante média associado à variável circular  $y_{ij}$ , com média circular  $\mu_{ij}$ , é dado por

$$\rho_{ij} = E \left\{ \cos(y_{ij} - \mu_{ij}) \right\}$$

Como  $\rho_{ij}$  assume valores em um intervalo restrito  $([0,1])$  convém, a fim de evitar-se problemas na estimação desse parâmetro, que sua parametrização seja feita através de uma função real que assuma valores nesse intervalo. Como exemplo, poderíamos utilizar, dentre outras, a função:

$$\rho_{ij} = \frac{\exp(z^T \gamma)}{1 + \exp(z^T \gamma)} \quad (\text{regressão logística}),$$

onde  $\gamma$  é um vetor paramétrico. Quando  $y_{ij}$  segue uma distribuição von Mises temos que

$$\rho_{ij} = A_1(\lambda_{ij}), \quad (10)$$

onde  $\lambda_{ij}$  é um parâmetro de concentração. Como a distribuição von Mises é a distribuição mais utilizada na prática para a modelagem de dados circulares, parece-nos natural utilizar a função de ligação definida por (10) na modelagem da variabilidade de dados desse tipo, mesmo nos casos em que a distribuição de  $y_{ij}$  não seja a von Mises. A partir deste ponto, adotaremos a transformação definida por (10). A extensão dos resultados seguintes a outras funções de ligação não é complicada, podendo ser obtida através da teoria geral descrita em Artes(1997).

Defina  $v_{ij} = \cos(y_{ij} - \mu_{ij}) - A_1(\lambda_{ij})$  e assumamos o seguinte modelo:

$$\lambda_{ij} = f(\xi_{ij}), \quad \xi_{ij} = z_{ij}^T \gamma$$

onde,  $f(\cdot)$  é uma função positiva, duplamente diferenciável e  $\gamma$  é um vetor de parâmetros.

Denomine  $\theta = (\beta^T, \gamma^T)^T$ ,  $s_i = (u_i^T, v_i^T)^T$ ,  $v_i = (v_{i1}, \dots, v_{it})$  e  $u_i$  e  $\beta$  como definidos na Seção 4.1. Temos que

$$E_{\theta}(v_i) = 0 \text{ e } \text{Cov}_{\theta}(v_i) = K_i.$$

Adotando-se a mesma metodologia utilizada nas seções anteriores, sob condições gerais de regularidade (ver Artes, 1997),  $\gamma$  pode ser consistentemente estimado através das raízes de

$$\gamma_n^P(\theta) = \sum_{i=1}^n G_i P_i s_i, \quad (11)$$

onde

$$G_i = \begin{pmatrix} D_i^{*T} & 0 \\ 0 & E_i^T \end{pmatrix}, \quad P_i = \begin{pmatrix} W_i^{-1} & 0 \\ 0 & K_i^{-1} \end{pmatrix},$$

$$D_i^{*T} = D_i N_i, \quad N_i = E \left\{ \text{diag} \left( \frac{\partial u_{i1}}{\partial \eta_{i1}}, \dots, \frac{\partial u_{it}}{\partial \eta_{it}} \right) \right\} = \text{diag}(\rho_{i1}, \dots, \rho_{it}),$$

$$E_i^T = Z_i^T F_i M_i, \quad Z_i = (z_{i1}, \dots, z_{it})^T, \quad F_i = \text{diag} \left( \frac{\partial \lambda_{i1}}{\partial \xi_{i1}}, \dots, \frac{\partial \lambda_{it}}{\partial \xi_{it}} \right), \quad M_i = E \left\{ \text{diag} \left( \frac{\partial v_{i1}}{\partial \lambda_{i1}}, \dots, \frac{\partial v_{it}}{\partial \lambda_{it}} \right) \right\}.$$

O estimador obtido a partir da raiz de (11) será, sob condições gerais de regularidade (ver Artes, 1997), assintoticamente normal com matriz de covariância assintótica dada por

$$\lim_{n \rightarrow \infty} n \left( \sum_{i=1}^n S_i \right)^{-1} \left( \sum_{i=1}^n V_i \right) \left( \sum_{i=1}^n S_i \right)^{-1T}$$

com

$$S_i = G_i P_i G_i^T$$

e

$$V_i = G_i P_i \text{Cov}(s_i) P_i G_i^T.$$

De modo análogo ao ocorrido na seção anterior, a utilização da função (11) requer o conhecimento de  $W_i$  e de  $K_i$ , o que não ocorre em situações práticas. Para contornar esse problema, admita a existência de um vetor paramétrico  $\alpha$  tal que  $W_i = W_i(\alpha)$  e  $K_i = K_i(\alpha)$  e defina a seguinte função:

$$\gamma_n(\theta) = \sum_{i=1}^n G_i P_i (\hat{\alpha}_n) s_i.$$

Demonstra-se, sob condições gerais de regularidade (ver Artes, 1997), que se  $\hat{\alpha}_n = \hat{\alpha}(\theta)$  for um estimador  $\sqrt{n}$ -consistente de  $\alpha$  dado  $\theta$ , então o estimador obtido a partir da raiz de  $\gamma_n$  será consistente e assintoticamente normal com matriz de covariância assintótica dada por (12), essa matriz pode ser consistentemente estimada através de

$$J = n \left( \sum_{i=1}^n \hat{S}_i \right)^{-1} \left( \sum_{i=1}^n \hat{V}_i \right) \left( \sum_{i=1}^n \hat{S}_i \right)^{-1T},$$

com

$$\hat{S}_i = G_i(\hat{\theta}_n) P_i(\hat{\alpha}_n) G_i^T(\hat{\theta}_n)$$

e

$$\hat{V}_i = G_i(\hat{\theta}_n) P_i(\hat{\alpha}_n) s_i s_i^T P_i(\hat{\alpha}_n) G_i^T(\hat{\theta}_n)$$

onde  $\hat{\theta}_n$  é um estimador de  $\theta$  obtido a partir de  $\gamma_n$ .

## 5. TESTES DE HIPÓTESES

Nesta seção, descrevemos brevemente a construção de testes de hipóteses do tipo Wald para verificação de hipóteses lineares para o modelo misto. A extensão para o modelo proposto na Seção 4.1 é imediata e não será abordada neste texto.

Suponha que se deseja testar a hipótese  $H_0: C\theta = 0$  versus  $H_1: C\theta \neq 0$ , onde  $C$  é uma matriz de posto completo  $h \times q$ ,  $h \leq q$  e  $q$  é a dimensão de  $\theta$ . Utilizando-se as propriedades assintóticas de  $\hat{\theta}_n$ , demonstra-se que a estatística de teste

$$Q = \hat{\theta}_n^T C^T (CJC^T)^{-1} C \hat{\theta}_n$$

tem, sob  $H_0$ , uma distribuição assintótica qui-quadrado com  $h$  graus de liberdade.

## 6. APLICAÇÃO

Considerando-se o experimento descrito na introdução, analisaremos os dados coletados durante o ano de 1982; uma análise inicial desses dados pode ser encontrada em Ranvaud et al. (1983). No experimento de 1982 duas precauções foram tomadas para se evitar viéses. Primeiramente, os pombos foram soltos de dois locais diferentes para reduzir a possibilidade de que alguma característica espúria do ponto de soltura polarizasse a direção dos pombos, portanto mascarando o efeito da remoção dos mecanismos de bússola no experimento. As solturas foram realizadas em diferentes momentos do dia, pois as condições ambientais não são as mesmas ao longo do dia. Os pombos soltos à tarde permaneceram por mais tempo no local de soltura do que os soltos pela manhã, e aves poderiam assim vir a ter informações adicionais, o que resultaria em uma vantagem em relação às demais. Schmidt-Koenig (1979) menciona uma análise de dados não publicada que indicou que os últimos cinco de vinte pombos soltos em um determinado dia orientam-se na direção correta mais rapidamente que os demais soltos no mesmo dia, sugerindo uma melhor orientação após uma permanência mais demorada no local de soltura.

A fim de verificar a existência de diferenças significativas entre as posições dos pombos ao longo do tempo e considerando a variação dos fatores local de soltura (com dois níveis: CMC1 e CMC2) e horário da soltura (com 3 níveis manhã, meio-dia e tarde) definimos o seguinte vetor de resposta:  $Y_{ijk} = (y_{ijk1}, y_{ijk2}, y_{ijk3})^T$ , onde as componentes do vetor correspondem às diferenças entre as posições da ave  $k$  solta no local  $i$ , no horário  $j$ , respectivamente, entre os instantes 60s e 30s, 90s e 60s e de desaparecimento e 90s,  $i = 1, 2$ ,  $j = 1, 2, 3$ ,  $k = 1, \dots, n_{ij}$ .

Partimos do seguinte modelo misto (no sentido de Fisher e Lee, 1992):

$$\mu_{ijkl} = 2a \tan \left\{ \mu_0 + \alpha_{0i} + \beta_{0j} + \alpha\beta_{0ij} + (1-2) \left( \mu_1 + \alpha_{1i} + \beta_{1j} + \alpha\beta_{1ij} \right) \right\},$$

$$v_{ijkl} = A_1(\lambda_{ijkl}), \text{ com}$$

$$\lambda_{ijkl} = \exp \left\{ \lambda_0 + \eta_{0i} + k_{0j} + \eta k_{0ij} + (1-2) \left( \lambda_1 + \eta_{1i} + k_{1j} + \eta k_{1ij} \right) \right\}.$$



Tratam-se de modelos análogos a modelos de análise de covariância quando se permite a interação entre os fatores e as covariáveis. Nesse modelo consideramos a presença dos fatores local e horário de soltura; a dimensão longitudinal surge como uma covariável, assumindo o valor -1 quando se considera a diferença de posição entre os instantes 60 e 30 segundos ( $l=1$ ); 0 entre 90 e 60 segundos ( $l=2$ ) e 1 entre o ângulo de desaparecimento e 90 segundos ( $l=3$ ). Adotamos a parametrização da casela de referência, sendo que  $\mu_0$  corresponde à metade da tangente da média circular observada no período da manhã, no local de soltura 1 para  $l$  igual a 2;  $\alpha_{02}$  indica a variação dessa função da média ao se passar do local 1 para o local 2 de coleta ( $\alpha_{01} = 0$ );  $\beta_{0j}$  corresponde à variação da função ao se passar do período de soltura matutino para meio-dia ( $j=2$ ) ou para a tarde ( $j=3$ ),  $\beta_{01} = 0$ ;  $\alpha\beta_{0ij}$  corresponde aos efeitos de interação entre local e período de soltura ( $\alpha\beta_{01j} = \alpha\beta_{0i1} = 0$ , para qualquer  $i$  e  $j$ );  $\mu_1$  indica a variação da função da média circular ao longo do tempo para os pombos liberados no local 1 e no período da manhã; os parâmetros  $\alpha_{1i}$ ,  $\beta_{1j}$  e  $\alpha\beta_{1ij}$  têm uma interpretação análoga a  $\alpha_{0i}$ ,  $\beta_{0j}$  e  $\alpha\beta_{0ij}$  para representar a variação longitudinal da média circular. Os parâmetros do modelo para  $\lambda_{ijkl}$  têm uma interpretação análoga aos do modelo para  $\mu_{ijkl}$  que possuem os mesmos índices, com a diferença que indicam a variação do logaritmo neperiano do parâmetro de concentração,  $\lambda_{ijkl}$  ao invés da metade da tangente da média circular.

As matrizes de covariâncias utilizadas na função de estimação  $\gamma_{ii}$  foram do tipo não-estruturada.

Realizamos sucessivos testes de Wald para retirada de efeitos não significativos.

Para a realização das análises, foram desenvolvidos programas computacionais no pacote S-Plus.

Tabela 1 - Estimativas dos parâmetros do modo final

Parâmetro	estimativa	erro padrão	estimativa / erro padrão	nível descritivo
$\mu_0$	0,03	0,01	2,87	0,004
$\lambda_0$	1,88	0,13	13,97	<0,001
$\eta_{02}$	-0,80	0,23	-2,21	0,027
$k_{01}$	-0,01	0,10	-0,20	0,001
$\eta k_{021}$	1,30	0,44	2,96	0,003

Chegamos a um modelo final cujas estimativas dos parâmetros são dadas na Tabela 1. Como pode ser observado, efeitos não significativos dos fatores e covariável foram encontrados para as médias circulares. A direção média comum foi estimada por  $3,34^\circ$ . Em termos dos parâmetros de concentração, detectamos um efeito significativo de interação entre o período e local de soltura.

A Tabela 2 traz as estimativas dos parâmetros de concentração e do comprimento da resultante média, estimadas através do modelo final. O fato de no local de soltura 2 o parâmetro de concentração ser maior à tarde reflete provavelmente uma melhora na orientação das aves com a permanência no local de soltura, o que está de acordo com Schmidt-Koenig (1979). Os resultados do local 1, por outro lado, podem ser interpretados como evidência de algum efeito polarizador, que orientaria inicialmente as aves numa direção espúria. Correções para o rumo correto seriam então mais evidentes nos pombos liberados à tarde, pois eles teriam melhores condições de orientação de forma a se opor com maior eficiência ao viés local.

Essa conclusão é reforçada pelo fato de que em outros estudos o comportamento dos pombos nesse local foi diferente dos demais. A possibilidade de um local de soltura acarretar algum viés está bem documentada na literatura (Keeton, 1973, por exemplo).

Tabela 2 - Estimativas dos parâmetros de concentração e dos comprimentos das resultantes médias

Local de soltura	Horário de soltura	Parâmetro de concentração	Resultante média
1	manhã / meio-dia	6,40	0,92
1	tarde	3,88	0,86
2	manhã / meio-dia	3,49	0,84
2	tarde	7,75	0,93

## 10. CONCLUSÃO

Apresentamos um método apropriado para a análise de dados circulares longitudinais. Um grande atrativo da técnica é que seu uso dispensa o conhecimento exato da distribuição de probabilidades geradora dos dados, necessitando apenas de suposições sobre o comportamento das médias circulares, do comprimento da resultante média e sobre a estrutura de covariância existente entre senos e, eventualmente cossenos das variáveis originais centradas. Estudos de simulação para a verificação do desempenho do método em situações controladas são encontrados em Artes (1997) e Artes e Jørgensen (1998). Análises de dados reais podem ser vistas em Artes (1997), Artes e Jørgensen (1998) e Artes et al. (1998).

## ABSTRACT

We describe in this paper a new methodology for the analysis of the direction taken by pigeons in different instants of time after their release. The proposed method is an extension of the quasi-likelihood method (Wedderburn, 1974) and of the generalized estimating equations method (Liang e Zeger, 1986), that may be applied to a wide variety of longitudinal circular data. We discuss position and dispersion parameters modelling and we present robust estimators of the covariance matrix of the estimators obtained by the proposed estimating equations. We illustrate the methodology with an analysis of part of the data from an experiment on homing pigeon conducted near the city of Camocim, at Ceará state in Brazil.

## AGRADECIMENTOS

Os dois primeiros autores foram parcialmente financiados pela FAPESP e pelo CNPq.

## BIBLIOGRAFIA

- ABRAMOWITZ, M. e STEGUN, I.A. (1970) *Handbook of mathematical functions*. New York: Dover Pub.
- ARTES, R. (1997) *Extensões da teoria das equações de estimação generalizadas a dados circulares e modelos de dispersão*. Tese (Doutorado). São Paulo: IME-USP.
- ARTES, R. e JØRGENSEN, B. *Longitudinal data estimating equations for dispersion models*. Scand. J. of Statist. A sair.
- ARTES, R., PAULA, G.A. E RANVAUD, R. (1998) *Strategy for the analysis of circular data based on generalized estimating equations*. São Paulo: IME-USP.
- BATSCHLET, E. (1981) *Circular statistics in biology*. London: Academic Press.
- BERNOULLI, D. (1734) *Recherches Physiques et Astronomiques, Sur le Problème Proposé Pour la Seconde Fois Par L'Académie Royale des Sciences des Paris. Récuil des Pièces Qui Ont Remporté le prix de L'Académie Royale des Sciences*, Tome III, p.95-134.
- FISHER, N.I. (1993) *Statistical analysis of circular data*. Cambridge: University Press.
- FISHER, N.I. e LEE, A.J. (1992) Regression models for an angular response. *Biometrics*, v.48, p.665-77.
- GOULD, A.L. (1969) A regression technique for angular variates and related regression models. *Biometrics*, v.25, p.683-700.
- GREEN, P. J. (1984) Iteratively reweighted least squares for maximum likelihood estimation and some robust and resistant alternatives. *Journal of the Royal Statistical Society, Serie B*, v. 40, p. 149-92.
- JOHNSON, R.A. e WEHRLY, T.E. (1978) Some angular-linear distributions and related regression models. *JASA*, v.73, p.602-6.
- JØRGENSEN, B., LUNDBYI-CHRISTENSEN, S., SONG, X-K., SUN, L. (1996) State space models for multivariate longitudinal data of mixed types. *Canad. J. Statist.* v.24, p.385-402.
- KEETON, W.T. (1973) Release site bias as a possible guide to the 'map' component in pigeon homing. *J. Comp. Physiol.* v.86, p.1-16.
- LIANG, K.-Y. e ZEGER, S.L. (1996) Longitudinal analysis using generalized linear models. *Biometrika*, v.73, p.13-22.
- MARDIA, K.V. (1972) *Statistics of directional data*. London: Academic Press.
- MCCULLAGH, P. (1983) Quasi-likelihood functions. *Annals of Statistics*, v.11, p.59-67.
- RANVAUD, R., SCHMIDT-KOENIG, K., KIEPENHEUER e J. GASPAROTTO, O.C. (1983) Initial orientation of homing pigeons at the magnetic equator with and without sun compass. *Behav. Ecol. Sociobiol.* v.14, p.77-79.
- RANVAUD, R., SCHMIDT KOENIG, K., GANZHORN, J.U., KIEPENHEUER, J., GASPAROTTO, O.C. e BRITTO, L.R.G. (1991) The initial orientation of homing pigeons at the magnetic equator: compass mechanisms and the effect of applied magnets. *J. Exp. Biol.* v.299, p.299-314.
- RIVEST, L-P. (1989) Spherical regression for concentrated Fisher-von Mises distributions. *The Annals of Statistics*, v.17, p.307-17
- SCHMIDT-KOENIG, K. (1979) *Avian orientation and navigation*. London: Academic Press.
- SCHMIDT-KOENIG, K., GANZHORN, J.U. e RANVAUD, R. (1991) The sun compass. Em *Orientation in birds*, Ed. P. Berthold, Birkhäuser, Basel, p. 1-15.
- WATSON, G.S. (1982) Distributions on the circle and sphere. In Gani, J. e Hannan, E.J. (Ed.), *Essays in statistical science*. Sheffield: Applied Probability Trust. p.265-80. (Journal of Applied Probability, special volume 19A).
- WEDDERBURN, R.W.M. (1974) Quasi-likelihood function, generalized linear models, and the Gauss-Newton method. em *Biometrika*, v.61, p.439-47.

# Estimando o Risco Sistemático de Ações Brasileiras Usando Regressão Ortogonal

Beatriz Vaz de Melo Mendes\*

Edel Alexandre da Silva Pontes\*\*

## RESUMO

O risco sistemático de uma ação pode ser entendido como um quantificador da sua tendência de acompanhar os movimentos de um determinado índice de mercado. Propomos o método de regressão ortogonal para estimar o risco sistemático de ações. O pressuposto principal deste método, que o faz atraente e adequado para o problema abordado, é o de que as variáveis envolvidas, isto é, os retornos de uma ação e de um índice de mercado, podem ser observadas com erro. Dada a característica volátil do nosso mercado, que tende a produzir *outliers* com bastante frequência, estimamos o risco sistemático de ações brasileiras usando estimadores robustos para o método de regressão ortogonal. Introduzimos o R-R Plot dos resíduos ortogonais como instrumento para detectar *outliers*.

Palavras-chaves: Risco Sistemático; Modelo com Erros nas Variáveis; Estimação Robusta.

## 1. INTRODUÇÃO

O modelo CAPM - Capital Asset Pricing Model, Lintner (1965), Sharpe (1964) relaciona o retorno esperado de uma ação, o retorno de uma aplicação financeira sem risco e o retorno esperado de um índice de mercado através de uma relação linear. Retornos são definidos como as mudanças relativas do preço de uma ação ou índice de mercado sobre um determinado horizonte. Neste trabalho definimos como retorno diário de um ativo no dia  $t$  o valor  $\log(v_t / v_{t-1})100$ , onde  $v_t$  representa o preço de fechamento deste ativo no dia  $t$ . De acordo com o CAPM, o valor esperado  $E[Y]$  do retorno  $Y$  de uma ação pode ser descrito como

$$E[Y] - r_f = \beta(E[M] - r_f), \quad (1)$$

\* Endereço para correspondência: DME-IM, Universidade Federal do Rio de Janeiro - e-mail: bmmendes@dme.ufrj.br.

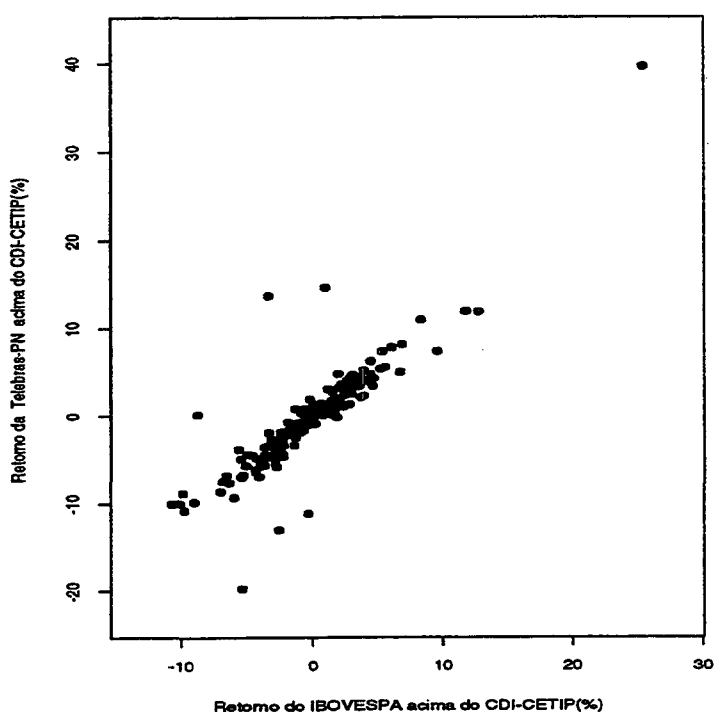
\*\* Endereço para correspondência: DM, Universidade Federal de Alagoas.

onde  $r_f$  é o retorno de uma aplicação caracterizada pela absoluta certeza quanto a sua apreciação,  $\beta$  é o risco sistemático do ativo considerado, e onde  $E[M]$  é o retorno esperado do índice  $M$  escolhido para medir a performance do mercado. A quantidade  $\beta(E[M] - r_f)$  representa o prêmio pelo risco assumido.

O risco sistemático de uma ação, ou o beta desta ação, pode ser entendido como um quantificador da capacidade inerente desta ação de acompanhar (em maior ou menor grau) os movimentos de um determinado índice de mercado. O risco sistemático é uma quantidade muito importante para os analistas de mercado, gerenciadores de risco, *traders*, etc. Seu cálculo pode ser considerado o primeiro passo no processo de análises diárias de uma instituição financeira.

A estimação do risco sistemático de ações brasileiras a partir do CAPM tem sido recentemente objeto de vários estudos. Por exemplo, a Figura 1 mostra o gráfico dos retornos diários do Índice da Bolsa de Valores de São Paulo -IBOVESPA- versus os retornos diários de uma das ações mais negociadas no mercado financeiro brasileiro, a Telebrás - PN. O interesse aqui é estimar a inclinação da reta a ser ajustada aos dados. Carmona (1996) estudou a relação entre o prêmio de risco de mercado (o lado direito de (1)) e a volatilidade (uma medida da variabilidade das séries, em geral o desvio padrão). Soares et al. (1997) abordaram o problema da seleção do índice de mercado comparando as performances do IBOVESPA e do FGV100 (índice da Fundação Getulio Vargas). Mendes (1998) estimou os betas de ações brasileiras usando um método de reamostragem por importância.

Figura 1 - Gráfico dos retornos diários do IBOVESPA versus os retornos diários da Telebrás - PN



A procura de técnicas alternativas para a estimação do beta sob suposições adicionais tem sido também objeto de várias pesquisas. Merton (1973) supôs a não estacionariedade dos betas e propôs ajustes na estimação. Chang e Weiss (1991) assumiram que o parâmetro  $\beta$  segue um processo auto-regressivo média móvel e Medeiros et al. (1997), assumindo um processo auto-regressivo de ordem 1, usaram um processo de estimação baseado no filtro de Kalman. Mendes e Duarte (1998) estimaram os betas de ações brasileiras usando um modelo de regressão com erros ARCH (Engle (1982)). Neste trabalho usamos o método de regressão ortogonal (MRO) para estimar os betas das 130 ações brasileiras mais negociadas. A característica principal deste método, que o faz atraente e adequado para este problema, reside na sua suposição básica de que ambas as variáveis envolvidas no modelo (1), isto é, os retornos de uma ação e de um índice de mercado, podem ser observadas com erro. Esta suposição baseia-se no fato de que o processo de coleta e distribuição dos dados é o mesmo para os retornos diários de ações e índices. Além disto, índices em geral são obtidos a partir de médias ponderadas de uma seleção de ações.

As estimativas produzidas pelo MRO coincidem com as estimativas de máxima verossimilhança de um modelo de regressão com erros nas variáveis (MREV) sob a suposição de normalidade (Bolfarine et al. (1992)). Contudo, é fato bastante conhecido hoje que dados em finanças caracterizam-se por possuírem *outliers* devido aos dias atípicos e por apresentarem caudas mais pesadas que as de uma distribuição normal. A Figura 1 ilustra essas duas características. Nesse caso, as estimativas clássicas para o MREV obtidas a partir da suposição de normalidade tornam-se distorcidas e ineficientes (Carrol e Gallo (1982)). Somente estimadores robustos, isto é, aqueles capazes de resistir à presença de contaminações nos dados, podem produzir estimativas confiáveis. Existem na literatura propostas alternativas robustas para o MRO (Zamar (1989), Carrol e Gallo (1982), Brown (1982), entre outros) capazes de lidar com dados não normais e com a presença de observações atípicas.

O uso de estimadores robustos na análise de problemas oriundos do mercado financeiro, em particular dos mercados emergentes, tem ganho bastante popularidade ultimamente. Por exemplo, Badrinath e Chatterjee (1993) mostraram que *outliers*, pontos de alavanca e erros grosseiros, podem distorcer as estimativas dos betas caso sejam usados métodos não robustos. Duarte e Mendes (1997) provaram empiricamente a não normalidade de dados do mercado financeiro brasileiro, e estimaram o risco sistemático de ações brasileiras usando estimadores robustos. Dada a característica volátil do nosso mercado, temos que o principal objetivo deste trabalho é o de estimar o risco sistemático de ações brasileiras usando estimadores robustos para o MRO.

Este artigo está organizado da seguinte forma. Na Seção 2 explicamos o modelo de regressão com erros nas variáveis, o método de regressão ortogonal, e os estimadores robustos a serem usados neste trabalho. Na Seção 3 estimamos os betas de mais de 130

ações brasileiras usando estimadores clássicos e robustos, analisamos os resultados obtidos, e introduzimos um novo gráfico para a identificação de *outliers*. Na Seção 4 damos nossos comentários finais e resumimos nossas conclusões.

## 2. O MODELO DE REGRESSÃO COM ERROS NAS VARÁVEIS E O MÉTODO DE REGRESSÃO ORTOGONAL

Para estimar o risco sistemático de um ativo em geral, usa-se uma extensão do CAPM original (1) dada por

$$E[Y] - r_f = \alpha + \beta(E[M] - r_f), \quad (2)$$

onde temos a introdução do coeficiente alfa. A expressão (2) sugere o uso de um modelo linear de regressão para estimar o alfa e o beta das ações, bastando para isto que tenhamos um conjunto de  $n$  observações  $\{(r_{f_1}, y_1, m_1), \dots, (r_{f_n}, y_n, m_n)\}$  dos retornos do investimento livre de risco, dos retornos da ação e do índice de mercado considerados, durante um certo período de tempo, isto é, temos um modelo linear de regressão usual que assume que

$$y_t - r_{f_t} = \alpha + \beta(m_t - r_{f_t}) + \varepsilon_t, \quad (3)$$

$t = 1, 2, \dots, n$ , onde  $m_1, m_2, \dots, m_n$  são valores fixos, e onde os erros  $\varepsilon_t$  são variáveis aleatórias não correlacionadas e identicamente distribuídas, com esperança zero e variância constante. Por simplicidade, continuaremos a denotar por  $y_t$  e  $m_t$  os retornos da ação e do índice considerados já corrigidos pela subtração do retorno livre de risco. Podemos então reescrever o modelo linear de regressão (3) como

$$y_t = \alpha + \beta m_t + \varepsilon_t, \quad (4)$$

ou alternativamente como

$$y_t = Y_t + \varepsilon_t \quad b' \begin{pmatrix} Y_t \\ m_t \end{pmatrix} = a, \quad (5)$$

para  $t = 1, 2, \dots, n$ , onde  $b = (b_1, b_2) \in \mathcal{R}^2$ ,  $\|b\| = 1$ ,  $\|\cdot\|$  é a norma Euclidiana,  $a \in \mathcal{R}$ ,  $v$  representa o transposto de  $v$ , e onde  $\varepsilon_t$  é um erro aleatório com as características descritas acima. Notemos que (5) pode ser reescrita como  $y_t = \frac{a}{b_1} - \frac{b_2}{b_1} m_t + \varepsilon_t$ .

Contudo, como o próprio modelo CAPM sugere, a variável índice de mercado  $M$  poderia ter sido também medida com erro. Esta é a suposição básica do modelo (clássico) de regressão com erros nas variáveis que pode ser descrito como

$$x_t = X_t + \varepsilon_t \quad , \quad b' X_t = a \quad , \quad (6)$$

onde  $b, x_t, X_t$  e  $\varepsilon_t$  são vetores em  $\mathfrak{R}^{p+1}$ ,  $p$  é o número de variáveis explanatórias,  $\|b\| = 1$ ,  $a \in \mathfrak{R}$ , e onde  $X_1, \dots, X_n, \varepsilon_1, \dots, \varepsilon_n$  são vetores aleatórios não observáveis. Neste trabalho lidamos com um modelo linear de regressão simples ((4) ou (5)), sendo pois  $p = 1$ .

Uma característica interessante deste modelo é o fato de tratar as variáveis simetricamente, não fazendo distinção entre variável resposta e variável explanatória, já que ambas são medidas com erro. Isto justifica as notações simplificadas  $x_t$  e  $X_t$ .

Em (6), em geral, assume-se que os erros  $\varepsilon_t$  tenham uma distribuição normal bivariada com  $E[\varepsilon_t] = 0$  e  $\text{cov}(\varepsilon_t) = \sigma^2 I$ , onde  $I$  é a matriz identidade. Sob estas suposições, o procedimento para obter os estimadores de máxima verossimilhança é o método de regressão ortogonal, MRO (Fuller (1987)). Independentemente dos dados serem ou não serem normais, o MRO consiste em achar  $\hat{b} \in \mathfrak{R}^2$  e  $\hat{a} \in \mathfrak{R}$  com  $\|\hat{b}\| = 1$  soluções do problema :

$$\underset{\{ \|b\|=1, a \}}{\text{minimize}} \sum_t (b' x_t - a)^2 \quad (7)$$

Os estimadores clássicos obtidos de (7) ( veja Bolfarine (1992) para uma apreciação das dificuldades envolvidas neste processo de estimação) não são robustos, perdendo eficiência estatística e apresentando vício considerável sob ligeiros desvios quanto às suposições feitas (Carroll e Gallo (1982) e Brown (1982)). Zamar (1989) propôs os M-estimadores para regressão ortogonal, definidos como as soluções do problema

$$\underset{\{ \|b\|=1, a \}}{\text{minimize}} \sum_t \rho \left( \frac{b' x_t - a}{S_n} \right) \quad (8)$$

onde  $\rho(\cdot)$  é uma função de perda limitada, simétrica, não decrescente, e  $S_n$  é um estimador robusto da escala dos resíduos ortogonais. Essa formulação geral (8) engloba (7) como caso particular, bastando para isso fazer  $S_n = 1$  e  $\rho(u) = u^2$ . O papel da função  $\rho$  é o de dar



diferenciados às observações, reduzindo a influência daqueles resíduos considerados *is* depois da padronização feita através do estimador de escala  $S_n$ .

Zamar (1989) sugere usar o estimador de escala obtido a partir de um S-estimador de regressão ortogonal como o estimador robusto para o cálculo de  $S_n$ . Para definir o estimador para regressão ortogonal consideremos primeiro a definição de um M-estimador de escala (Hampel et al. (1986), Huber (1981)). Dados um vetor  $b \in \mathfrak{R}^2$ ,  $\|b\| = 1$ , e  $a \in \mathfrak{R}$ , um estimador de escala  $S(b,a)$  dos dados  $b'x_1, \dots, b'x_n$  é uma solução do sistema de equações implícitas

$$\frac{1}{n} \sum_t \rho \left( \frac{b'x_t - a}{S} \right) = k, \quad (9)$$

$\rho(\cdot)$  tem as características dadas acima e  $k$  é uma constante, em geral  $k = E[\rho(z)]$ , onde  $z$  tem distribuição normal padrão.

Os S-estimadores para regressão ortogonal são definidos como os valores  $\tilde{b}$  e  $\tilde{a}$  que minimizam  $S(b,a)$ . Isto é,

$$(\tilde{b}, \tilde{a}) : S(\tilde{b}, \tilde{a}) = \min_{\{ \|b\|=1, a \}} S(b,a). \quad (10)$$

O estimador robusto da escala dos resíduos de uma regressão ortogonal, associado aos estimadores de regressão ortogonal, é então

$$S(\tilde{b}, \tilde{a}). \quad (11)$$

Zamar (1989) provou que estes estimadores robustos têm alto ponto de ruptura (em grandes proporções de *outliers* na amostra) o qual depende da função  $\rho$ . Além disso são consistentes (se a distribuição dos erros for esférica) e assintoticamente normais. Neste trabalho a função  $\rho$  de Tukey dada por

$$\rho\left(\frac{r}{c}\right) = 3 \begin{cases} \frac{r^2}{c^2} - 3\left(\frac{r^2}{c^2}\right)^2 + \left(\frac{r^2}{c^2}\right)^3 & \text{se } \frac{r^2}{c^2} \leq 1 \\ 1 & \text{se } \frac{r^2}{c^2} > 1, \end{cases} \quad (12)$$

onde  $c$  é uma constante reguladora responsável pela eficiência dos estimadores.

### 3. ESTIMAÇÃO DOS BETAS DE AÇÕES BRASILEIRAS

Nesta seção usaremos alguns estimadores clássicos e robustos para estimar o risco sistemático das ações brasileiras. Os dados contêm os retornos diários das 130 ações mais negociadas na Bolsa de Valores de São Paulo entre 1994 e 1996. O índice de mercado usado é o IBOVESPA e o investimento livre de risco é o retorno da taxa interbancária *overnight* de juros CDI-CETIP (Certificado de Depósito Interfinanceiro, o qual é publicado diariamente pela Central de Custódia e Liquidação de Títulos Privados). Estes dados foram previamente usados por Duarte e Mendes (1997) onde suas características (caudas pesadas, presença de *outliers* e de pontos de alavanca) foram exaustivamente investigadas. Para estimar os betas dessas ações usaremos os seguintes procedimentos:

- S-RO : S-estimadores para regressão ortogonal  $(\tilde{b}, \tilde{a})$ , obtidos de acordo com (9), (10) e (11). Aqui o estimador de escala é o  $S(\tilde{b}, \tilde{a})$ . A função de perda é a função de Tukey (12) com  $c = 1.56$ . Estes estimadores possuem alto ponto de ruptura, porém não são estatisticamente eficientes.

- M-RO : M-estimadores para regressão ortogonal  $(\hat{b}, \hat{a})$ , obtidos a partir de (9) e da função de perda de Tukey com  $c = 4.67$ , a qual produz estimativas 95% eficientes sob normalidade. O estimador de escala auxiliar  $S_n$  é o S-estimador de escala obtido no item anterior,  $S(\tilde{b}, \tilde{a})$ .

- M-RO\* : obtido como no item anterior, mas tendo como estimador de escala o 1.483 Median Absolute Deviation - MAD - dos resíduos ortogonais (Hampel et al. (1986)), o qual é um estimador de escala consistente, com ponto de ruptura 0.5 e fácil de calcular.

- MQ-RO : estimadores clássicos (7) de regressão ortogonal (estimadores de máxima verossimilhança do modelo de regressão com erros nas variáveis sob a suposição de normalidade dos erros) obtidos supondo ser igual a 1 a razão entre as variâncias dos erros. Suposições extras para o processo de estimação são em geral necessárias (Bolfarine et al. (1992) e Pontes (1998)), e neste caso parece ser bastante razoável supor que as variabilidades dos erros de medida dos retornos das ações e índices de mercado sejam iguais.

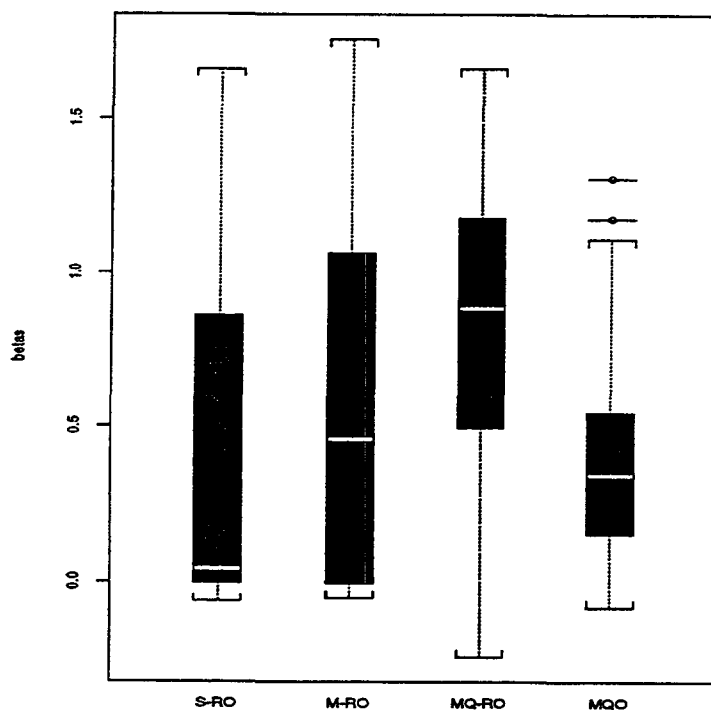
- MQO : Mínimos Quadrados Ordinários para o modelo usual de regressão linear (4). A inclusão deste estimador deve-se somente ao fato de ser o método mais conhecido e usado pela grande maioria dos analistas financeiros brasileiros. Contudo frisamos que não é adequado para estimar os betas de ações brasileiras conforme provado empiricamente por Duarte e Mendes (1997).

Tabela 1 - Risco sistemático de seis ações brasileiras fornecido pelos cinco métodos de estimação

Procedimento	S-RO	M-RO	M-RO*	MQ-RO	MQO
Ação					
ELETOBRÁS - ON	1,57	1,51	1,55	1,57	1,31
TELEBRÁS - PN	1,20	1,15	1,32	1,40	1,17
PETROBRAS - PN	1,22	1,30	2,48	1,39	1,11
VALE RIO DOCE - PN	1,24	0,95	1,87	1,13	0,79
TELESP - PN	1,30	1,48	1,93	1,54	1,09
BRADESCO - PN	0,61	1,06	1,25	1,23	0,81

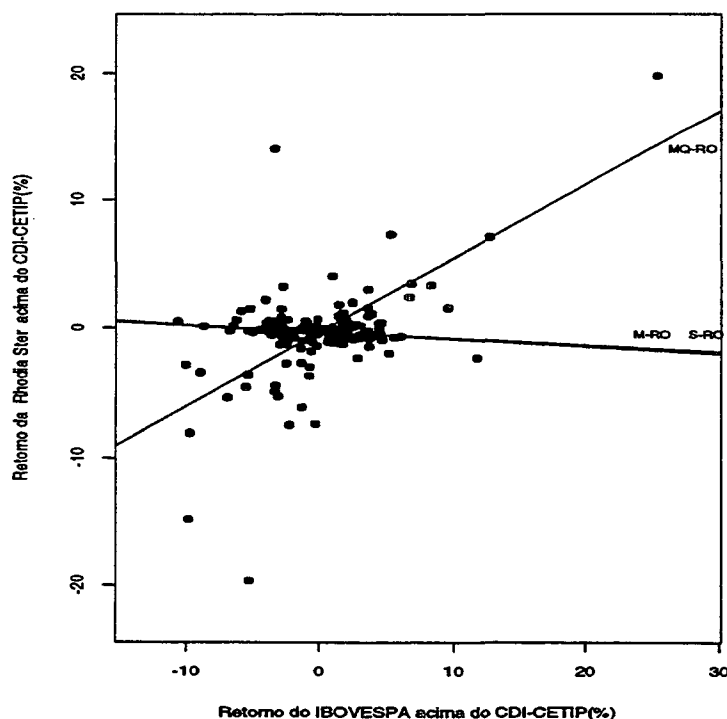
Detalhes sobre o algoritmo usado para o cálculo desses estimadores podem ser obtidos em Pontes (1998) ou diretamente com os autores. Parte dos dados está disponível em Mendes (1999). Damos na Tabela 1 os valores para os betas das seis ações mais negociadas (na época considerada) na Bolsa de Valores de São Paulo, estimados de acordo com os cinco métodos.

Figura 2 - Box-plots das estimativas dos betas segundo quatro métodos de estimação



A Figura 2 mostra os *Box-plots* das estimativas dos betas, segundo quatro métodos de estimação. O método M-RO\* não aparece neste gráfico, pois revelou-se um método computacionalmente instável, produzindo algumas estimativas duvidosas, como se pode observar na Tabela 1, para as ações Petrobras-PN, Vale do Rio Doce-PN e Telesp-PN. Os *Box-plots* da Figura 2 revelam que os métodos produziram resultados bastante diferentes, o que nos leva a uma análise mais cuidadosa dos mesmos. Para nos certificar da influência dos *outliers* nas estimativas dos betas, fizemos os gráficos dos ajustes obtidos para várias ações, incluindo tanto aquelas com maior liquidez de mercado, como aquelas negociadas infreqüentemente. Os dados relacionados com as ações ilíquidas (os quais são regulados de acordo com os princípios de contabilidade brasileiros), muitas vezes contêm uma grande quantidade de zeros, não apresentando uma tendência bem definida. Nestes casos, apenas alguns pontos influentes podem determinar o ajuste, como ilustra a Figura 3, onde vemos as retas ajustadas pelos métodos M-RO, S-RO, e MQ-RO para a ação Rhodia Ster.

Figura 3 - Gráfico das retas ajustadas pelos métodos M-RO, S-RO, e MQ-RO, para a ação Rhodia Ster



Gráficos de ajuste como o da Figura 3 são bastante ilustrativos e de fato auxiliam bastante o analista no processo de decisão sobre a qualidade do ajuste. Entretanto, a análise dos resíduos padronizados é uma ferramenta ainda mais poderosa pois pode ser usada no caso de regressão múltipla e também pode indicar outros tipos de desvios. Contudo, se o método de estimação não for robusto, a análise dos resíduos pode não revelar a presença de *outliers* e pontos de alavanca. As Figuras 4 e 5 ilustram este fato e apresentam os gráficos dos resíduos ortogonais padronizados obtidos dos ajustes M-RO e MQ-RO para a ação Bradesco-PN. Notamos que, diferentemente do estimador clássico MQ-RO (Figura 5), o ajuste robusto (Figura 4) foi capaz de identificar seis observações, possuindo resíduos ortogonais 2.5 vezes maiores que a unidade de escala robusta. Os outros três *outliers* não foram identificados pelo ajuste clássico provavelmente porque "atraíram" a reta ajustada ao mesmo tempo que "inflaram" a estimativa da escala dos resíduos (conforme explicado em Rousseeuw e Leroy (1987))

Figura 4 - Dispersão dos resíduos ortogonais padronizados obtidos do ajuste M-RO para a ação Bradesco - PN

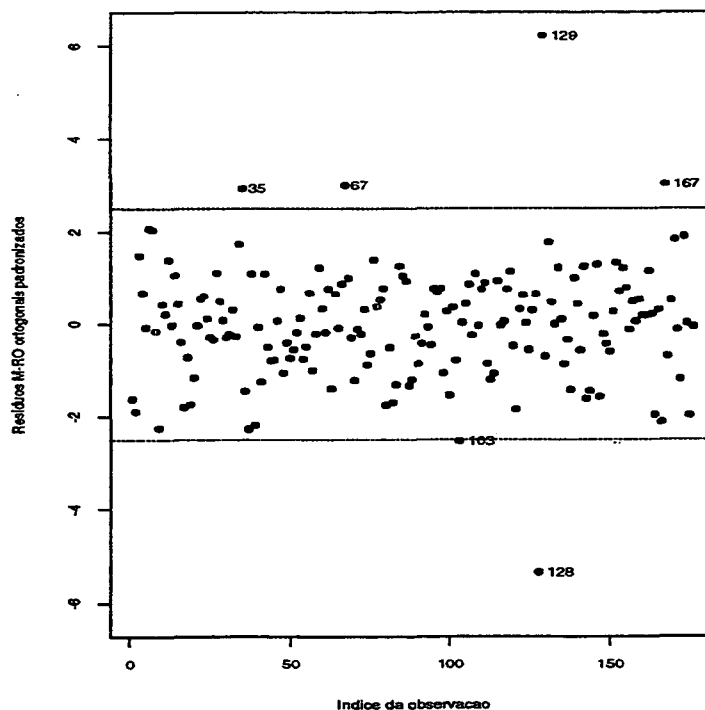
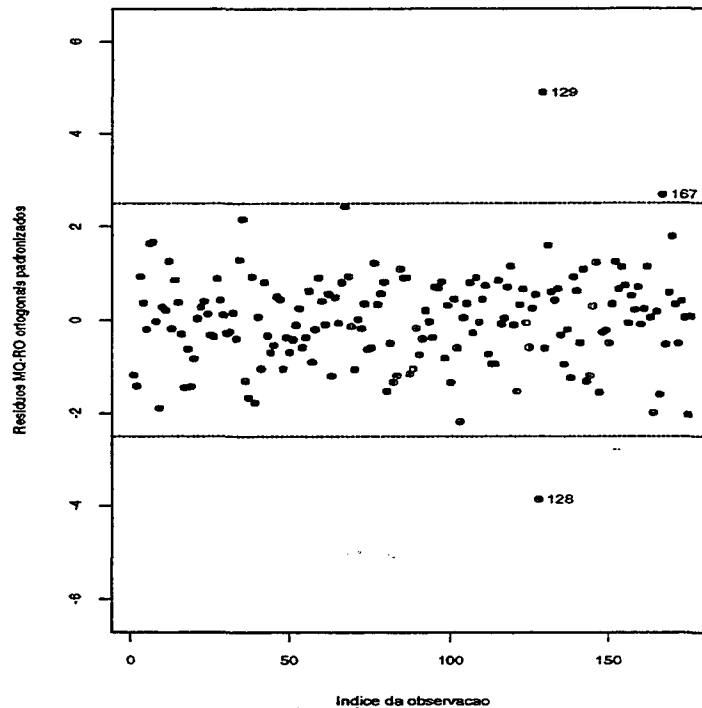


Figura 5 - Dispersão dos resíduos ortogonais padronizados obtidos do ajuste MQ-RO para a ação Bradesco - PN



Testes estatísticos foram empregados na análise dos resultados obtidos, para decidirmos se houve diferença significativa entre as soluções apresentadas pelos quatro métodos usados. Supondo amostras emparelhadas, fizemos um teste t clássico e o teste robusto de Wilcoxon (baseado nos *ranks* das observações), para testar  $H_0$ : A média da distribuição dos betas de acordo com o procedimento M-RO é igual à média da distribuição dos betas de acordo com o procedimento MQ-RO versus  $H_1$ : As médias não são iguais. Estes testes foram repetidos para todos os pares de procedimentos diferentes. A Tabela 2 resume os resultados destes testes. Ambos os testes clássico e robusto rejeitaram a hipótese nula (nível de significância de 5%) ao testar S-RO versus M-RO, M-RO versus MQ-RO, MQ-RO versus MQO, e M-RO versus MQO, tendo este último ambos p-valores iguais a zero. Por outro lado, os testes empregados discordaram ao testar os procedimentos S-RO versus MQ-RO, e S-RO versus MQO, o que pode ser consequência de algumas estimativas instáveis produzidas pelo S-RO.

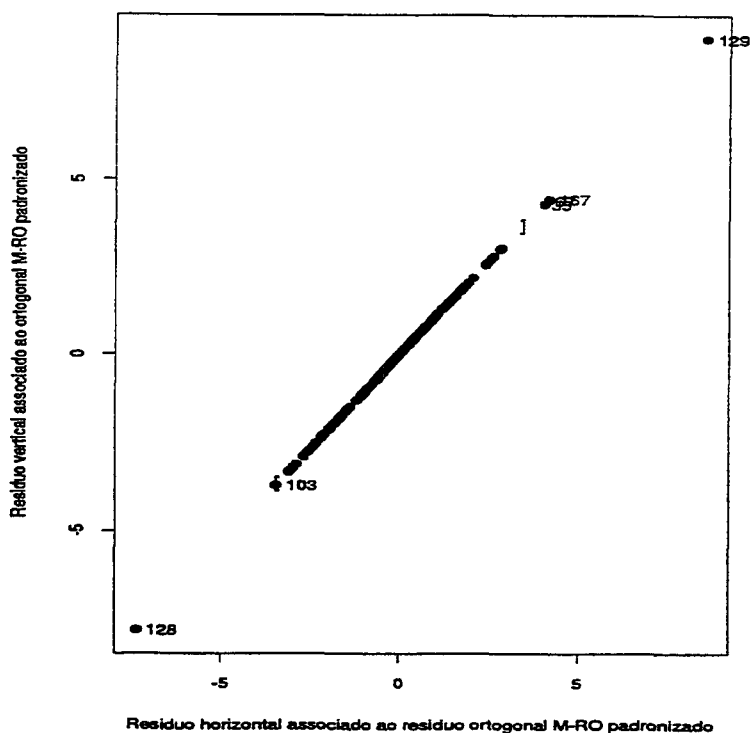
Para complementarmos estas análises, introduzimos aqui um novo gráfico, o *R-R Plot*, baseado nos resíduos ortogonais padronizados. As ordenadas dos pontos grafados em um *R-R Plot* são os resíduos padronizados usuais (verticais) de um modelo de regressão linear. As abscissas desses mesmos pontos são os análogos dos anteriores, agora tomados

paralelamente ao eixo horizontal. Dois pontos de corte [e] definem um intervalo que contém as observações possuindo resíduos ortogonais iguais ou menores que 2.5 vezes a unidade de escala robusta. Observações fora deste intervalo podem ser consideradas *outliers*. Este gráfico é bastante interessante pois tira partido da característica simétrica do modelo de regressão com erros nas variáveis, isto é, os pontos extremos são identificados sem haver a necessidade de se fazer a distinção entre *outliers* de regressão e pontos de alavanca. Isto se deve ao fato de não ser preciso definir quais seriam as variáveis explanatória e dependente. Para facilitar o entendimento deste gráfico, a Figura 6 mostra o R-R Plot para os resíduos M-RO para a mesma ação usada na Figura 4. Notemos que as mesmas observações detectadas como *outliers* na Figura 4 são identificadas na Figura 6.

Tabela 2 - P-valores dos testes clássico (esquerda) e robusto (direita) para testar as hipóteses:  $H_0$ : "A média da distribuição dos betas de acordo com o procedimento da linha é igual à média da distribuição dos betas de acordo com o procedimento da coluna versus  $H_1$ : "As médias não são iguais"

Procedimento	M-RO	MQ-RO	MQO
	Clássico e Robusto	Clássico e Robusto	Clássico e Robusto
S-RO	0,03 e 0,00	0,09 e 0,00	0,12 e 0,05
M-RO	-	0,05 e 0,00	0,00 e 0,00
MQ-RO	-	-	0,05 e 0,00

Figura 6 - R-R Plot : Gráfico dos resíduos horizontais versus os resíduos verticais para o ajuste M-RO e a ação Bradesco - PN



Um *R-R Plot* leva em consideração as distâncias nos dois sentidos de cada observação até a reta. Notemos ainda que os "bons" pontos de alavanca (pontos distantes dos demais, porém situados na direção definida pela maioria) terão ambas as distâncias pequenas e portanto os valores a serem grafados também pequenos. Por outro lado, *outliers* de regressão terão esses mesmos valores grandes, e ficarão fora do intervalo definido.

As Figuras 7, 8, 9 e 10 mostram uma análise gráfica e completa para a ação Telebrás - PN (dados no Apêndice). O primeiro gráfico mostra a dispersão dos dados e os ajustes M-RO e MQ-RO. Em seguida, temos na Figuras 8 e 9, respectivamente, um gráfico de dispersão e um histograma dos resíduos M-RO ortogonais padronizados. Finalmente, a Figura 10 mostra o *R-R Plot*.

Figura 7 - Ajustes clássico (MQ-RO) e robusto (M-RO) para a Telebrás - PN

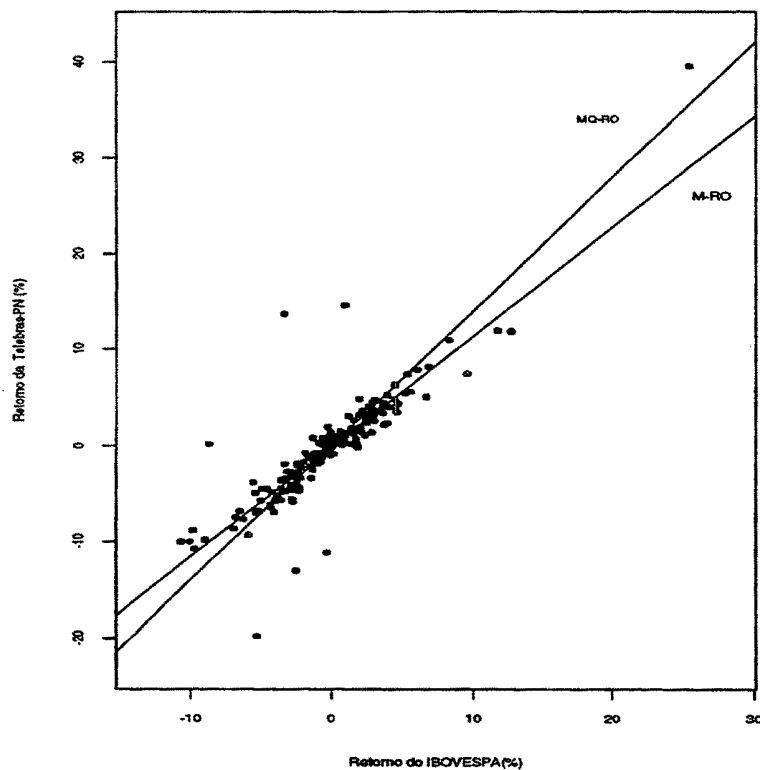




Figura 8 - Gráfico dos resíduos M-RO ortogonais padronizados no tempo para a Telebrás - PN

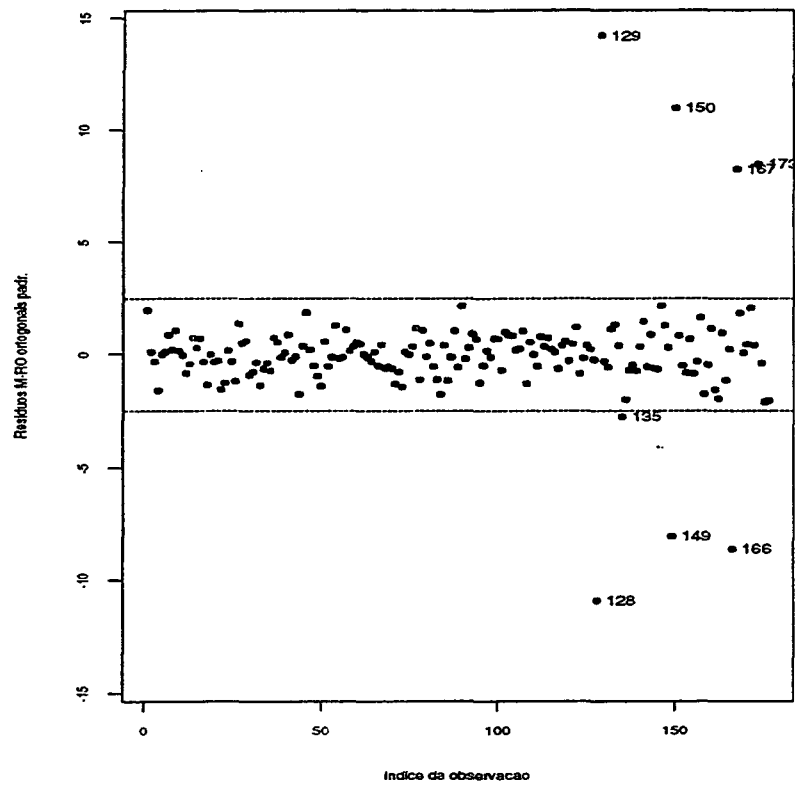


Figura 9 - Histograma dos resíduos M-RO ortogonais padronizados para a Telebrás - PN

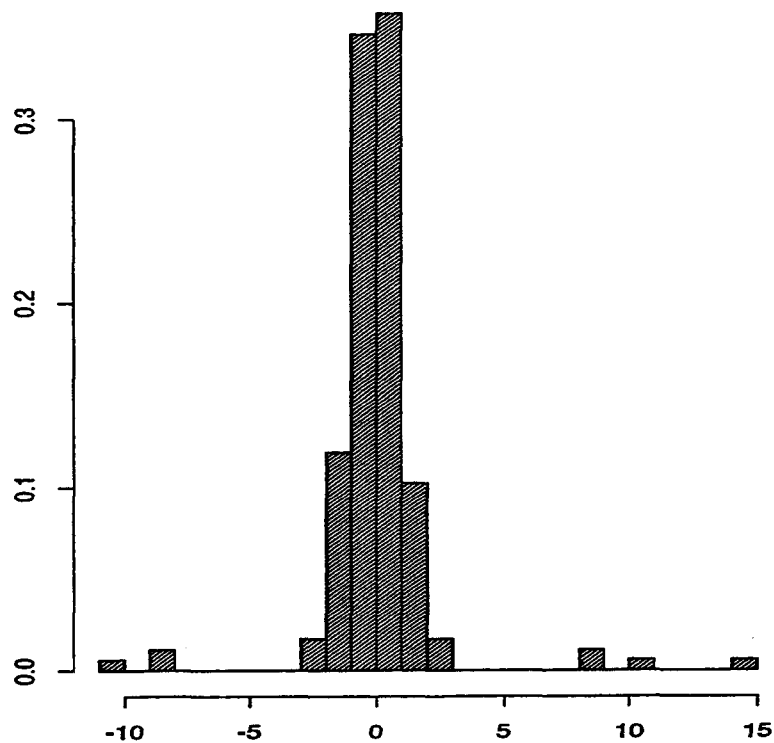
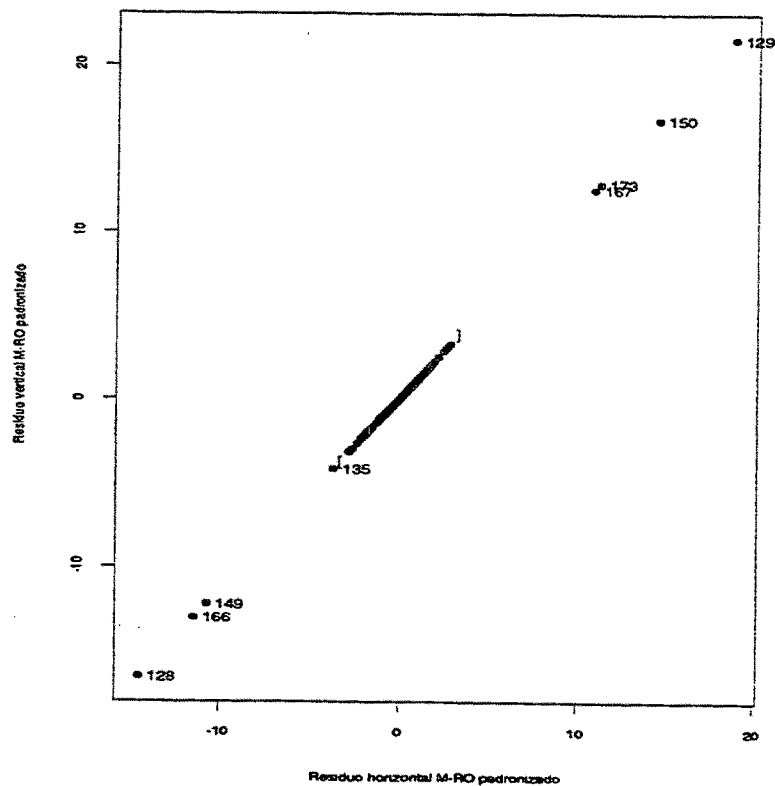


Figura 10 - R-R plot para a Telebrás - PN



#### 4. CONCLUSÕES

Neste trabalho propusemos o método de regressão ortogonal para estimar o risco sistemático de ações brasileiras e experimentamos vários estimadores clássicos e robustos.

Os métodos produziram resultados diferentes, o que nos levou a uma análise mais cuidadosa dos mesmos para nos certificarmos da sua eficiência computacional e da possível influência dos *outliers* nas estimativas dos betas. Foram feitos testes estatísticos e análises gráficas. Introduzimos um novo gráfico, o *R-R Plot*, baseado nos resíduos ortogonais padronizados, para a identificação de *outliers*. Este gráfico leva em consideração as distâncias nos dois sentidos de cada observação até a reta e tira partido da simetria do modelo.

Os algoritmos usados para o cálculo dos S-RO e dos M-RO\* revelaram-se muito instáveis, produzindo por várias vezes estimativas duvidosas. Os estimadores MQ-RO foram por diversas vezes influenciados por pontos extremos e produziram estimativas claramente

distorcidas, isto é, valores não justificados por teorias financeiras. Não recomendamos, portanto, o uso destes três estimadores ao estimar os betas das ações brasileiras. O resultado conclusivo dos testes foi o de que o procedimento M-RO é significativamente diferente clássicos.

Dada a característica volátil do mercado financeiro brasileiro, que tende a produzir outliers com bastante frequência, sugerimos o uso do estimador robusto M-RO para o modelo regressão ortogonal para estimar o risco sistemático de ações brasileiras. As estimativas RO para as 130 ações encontram-se no Apêndice.

## REFERÊNCIAS

- BADRINATH, S.G., e CHATTERJEE, S. (1993). Systematic Risk Estimation in the Presence of Large and Many Outliers. *Review of Quantitative Finance and Accounting*, 3, 5-28.
- BOLFARINE, H. RODRIGUES, J., e CORDANI, L. (1992). O Modelo de Regressão com Erros nas Varáveis. 10o. Simpósio Nacional de Probabilidade e Estatística, ABE.
- BROWN, M. (1982). Robust Line Estimation with Errors in Both Variables. *Journal of the American Statistical Association*, 77, 71-79.
- CARMONA, C.M. (1996). Relação entre o Prêmio de Risco de Mercado e a Volatilidade. Anais do XXVII SBPO - VIII CLAIO, Rio de Janeiro.
- CARROLL, R., e GALLO, P. (1982). Some Aspects of Robustness in the Functional Errors-in-Variables Model. *Comm. Statistics*, A, 11, 2573-2585.
- CHANG, W. e WEISS, D.E. (1991). An Examination of the Time Series Properties of Beta in the Market Model. *Journal of the American Statistical Association*, 86, 416, 883-890.
- DUARTE, A.M., JR., e MENDES, B. V. M. (1997). Robust Estimation of Systematic Risk in Emerging Stock Markets. *Emerging Markets Quarterly*, Vol. 1, No. 4, 85-95.
- ENGLE, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 4, 987-1007.
- FULLER, W.A. (1987). *Measurement Errors Models*. Wiley, New York.
- HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J., e STAHEL, W.A. (1986). *Robust Statistics: The Approach Based On Influence Functions*. Wiley, New York.
- LINTNER, J. (1965). The Valuation of Risk Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47, 1, 13-37.
- HUBER, P.J. (1981). *Robust Statistics*. Wiley, New York.
- MEDEIROS, M.C., VEIGA, A., FERNANDES, C., e OLIVEIRA, F. (1997). Extensões do CAPM : Uma Análise Comparativa. Anais do XXX SBPO, 165, Salvador, BA.
- MENDES, B.V.M. (1998). Financial Modeling Using Sampling-Importance Resampling. Aceito para publicação na *Revista de Econometria*.
- MENDES, B.V.M. e DUARTE, JR., A. M. (1998). Modelos Estatísticos Aplicados ao Mercado Financeiro Brasileiro. Associação Brasileira de Estatística, 135 pg.
- MENDES, B.V.M. (1999). *Regressão Robusta: Conceitos, Aplicações e Aspectos Computacionais*. Associação Brasileira de Estatística, 132 pg.
- MERTON, R.C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica*, 867-887.

- PONTES, E. A. (1998). Estimação Robusta no Modelo de Regressão com Erros nas Varáveis e uma Aplicação em Finanças. Dissertação de Mestrado, DME-IM-UFRJ.
- ROUSSEEUW, P.J. e LEROY, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- SHARPE, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*, 19, 3, 425-442.
- SOARES, L.J., MEDEIROS, M.C., e VEIGA, A. (1997). Análise Comparativa de Índices de Mercado Através do CAPM-VAR. Anais do XXIX SBPO, 135, Salvador, BA.
- ZAMAR, R. H. (1989). Robust Estimation in the Errors in Variables Model. *Biometrika*, 76, 1, 149-160.

## AGRADECIMENTOS

Mendes agradece o auxílio da FAPERJ e PRONEX. Pontes agradece o auxílio do CNPq. Os autores agradecem ao Editor Responsável pelos comentários e sugestões valiosas.

## 5. APÊNDICE

### I. Dados da Telebrás - PN

[1] "Telebras-PN" "TEL4"				
[1,]	0.81400557	-1.3135358	-1.6652186	-10.0101327
[2,]	-0.67697871	-10.0310896	4.8110106	8.1786667
[3,]	5.40475542	4.3397234	1.9526318	7.4199901
[4,]	2.35479348	5.5821535	-0.8901311	5.0365276
[5,]	4.00497989	0.2215124	0.7264201	-2.5352016
[6,]	-1.32641459	1.3789076	1.4603815	3.3967014
[7,]	-1.81584395	0.4036593	-4.5389863	-4.8195423
[8,]	-2.32884782	-1.6824392	-4.7372191	-7.4684068
[9,]	6.26756031	-4.3992965	-5.7043357	3.0708557
[10,]	1.69992025	1.4177438	-0.1920634	-4.1063739
[11,]	-1.29486081	-0.5648452	-2.0495715	3.2467851
[12,]	-6.14542807	-2.9286042	5.2112085	1.4812088
[13,]	4.03341610	4.7288160	-5.5767713	-5.3243189
[14,]	0.05610413	0.9970933	0.8903767	-3.8272189
[15,]	-0.86151840	1.3671043	7.8486108	10.9705276
[16,]	0.78629740	-3.5789779	0.3019188	3.0891605
[17,]	1.69249445	-0.7599691	-0.1779693	-13.0302279
[18,]	0.09445022	1.0123145	-3.1387024	14.6507172
[19,]	2.59474884	3.1416228	3.6373033	-2.6781172
[20,]	-2.16881018	0.2057448	-3.3640342	-6.7669471
[21,]	-1.96868519	-0.1809391	0.9691504	-3.9441033
[22,]	2.17425254	-4.0347073	3.9489915	-4.5244960
[23,]	0.30702159	-3.4057745	-1.8668678	-4.6067819
[24,]	-1.04785825	-5.2320461	2.6322491	-1.2256280

[25.]	2.33539200	-3.3740840	0.6542881	-1.9207138
[26.]	1.09763254	2.5647410	3.9423725	-6.9674558
[27.]	0.84079767	-4.3529608	-3.1196722	-8.6360008
[28.]	-0.79872949	-0.8524545	1.5698884	7.4117339
[29.]	-0.80543198	-4.6774218	-0.9869168	-4.6650246
[30.]	0.73447883	-1.3616609	-2.7201692	-9.3292824
[31.]	1.37918074	2.6751431	-6.8403643	4.0463940
[32.]	1.57260228	-3.4248150	-7.6627643	11.9600356
[33.]	3.44555973	0.2921365	3.7842416	1.7947453
[34.]	0.25807090	-5.6594469	-0.7668776	-11.1622294
[35.]	1.78280949	-4.4673587	-5.7806643	0.1573154
[36.]	4.33558260	4.8129124	1.7425482	1.9185410
[37.]	-1.1112544	3.32425484	4.5115490	-4.889873
[38.]	-1.6946900	0.05607056	-0.3007994	-9.827780
[39.]	2.7922102	3.67663493	-1.2770324	-8.827268
[40.]	-3.1983543	-5.53491413	-19.7599205	-10.750614
[41.]	3.1882637	-2.15335904	13.6947445	39.550438
[42.]	0.6102047	-6.45361220	1.2633093	-1.475190
[43.]	-1.4722003	3.55494282	-6.9797263	11.900320
[44.]	-0.1656983	4.45035536	-4.9352966	-5.818760

## II. Estimativas M-RO para as 130 ações usadas

	alfa	beta
Acesita ON	-0.064388	0.820615
Albarus ON	-0.206183	-0.009000
Alpargatas PN	-0.026935	0.627208
Sid Aconorte PNA	-0.183471	-0.003407
Antártica ON	-0.206946	-0.032419
Aracruz PNB	-0.151461	0.531582
Artex PN	0.006245	0.191897
Avipal ON	0.083398	0.624424
Bahema	-0.217242	-0.013001
Brasil PN	0.078538	1.263230
Bradesco PN	0.294154	1.064419
Bamerindus ON	0.061214	0.385777
Econômico PN	-0.208914	0.206453
BCN PN	-0.099692	0.449196
Bardella PN	0.456579	0.808968
Belgo Mineir ON	0.153460	0.728491
Francês e Bras ON	0.000000	0.000000
Brasinca PN	-0.158734	-0.002407
Merc S Paulo PN	-0.191907	-0.035422
Brasmotor PN	0.530027	1.284209
Nacional PN	0.005016	0.945108
Bombriil PN	0.040969	1.210121
Braspérola PNA	-0.190655	-0.005000

Petrobras BR PN	0.237450	0.966154
Brahma PN	0.093076	0.871057
Bahia Sul PNA	-0.155382	0.005593
Nitrocarbono PNA	-0.241815	-0.025413
CBV Ind Mec PN	-0.200985	-0.006000
Cesp PN	0.081320	1.271047
Cacique PN	-0.187255	0.000185
Celesc PNB	-0.090550	0.948902
Caemi Metal PN	0.147400	0.472728
Confab PN	-0.098368	-0.018002
Const Beter PNB	-0.184684	-0.013408
Cosigua PN	-0.154749	0.354404
Copene PNA	0.421470	1.014494
Copesul ON	0.249790	1.131891
Cremer PN	-0.118369	0.023189
Souza Cruz ON	0.027242	0.729399
Cosipa PNB	0.192263	1.085092
Sid Nacional ON	-0.074193	1.137398
Sid Tubarão PNB	0.267443	1.044552
Karsten PN	-0.223624	-0.008000
Continental PN	-0.155433	0.338722
Coteminas PN	-0.246092	0.292533
Dixielalekia PN	-0.134155	0.025191
Ipiranga Dist PN	0.143202	0.217948
Duratex PN	-0.010458	0.495680
Embraco PN	-0.210482	-0.009000
Eletróbrás ON	0.252193	1.509135
Weg PN	-0.123928	0.008185
Eletropaulo PNB	-0.309280	-0.041023
Ericsson PN	-0.000855	0.799757
Estrela PN	-0.123293	1.125071
Eternit ON	-0.193791	-0.006000
Eucatex PN	-0.198809	-0.026006
Cofap PN	0.048811	0.952710
Ficap-Marvin PN	-0.161083	-0.007407
Ferbasa PN	-0.166439	0.004185
Forja Taurus PN	-0.074933	1.281563
F Cataguazes PNA	0.182979	1.055924
Frigobrás PN	-0.171582	0.064274
Met Gerdau PN	-0.208361	0.118739
Granoleo PN	-0.178810	-0.014001
Cia Hering PN	0.388733	1.123729
Cim Itaú PN	-0.393340	0.393156
Gradiente PNA	-0.178599	-0.002000
Iguaçu Café PNA	-0.172746	0.031195
Império PN	-0.330954	-0.022004
Inepar PN	1.115065	1.359484
Itaubanco PN	0.360906	0.836480

Itaúsa PN -0.301574 0.452092  
Klabín PN 0.009882 0.895985  
Lacta PN -0.153955 -0.021003  
Lojas Americanas PN -0.221378 0.714804  
Metal Leve PN -0.069580 0.464191  
Light ON 0.353900 1.232545  
Brasilit ON -0.173977 -0.004000  
Lix da Cunha PN -0.224442 -0.017409  
Cimaf ON -0.206625 -0.007000  
Magnesita PNA -0.188478 -0.001407  
Manah PN 0.553586 1.189162  
Mannesmann ON 0.121697 1.089456  
Santista Alm ON -0.197254 0.027600  
Mangels PN 0.193458 0.541360  
Onteiro Aranha ON -0.182156 -0.015001  
Oinho Santista ON -0.189607 -0.004000  
Multibras PN 0.056406 1.256522  
Iochp-Maxion PN -0.187538 0.584823  
Nakata PN -0.165908 -0.010000  
OSA PN -0.128338 0.056245  
OXITENO PN -0.101338 0.013593  
Paul F Luz ON -0.031881 0.726961  
Petrobras PN -0.098507 1.301324  
Pirelli ON -0.177767 -0.003407  
Parapanema PN -0.294106 1.123729  
Pronor PNA -0.303881 -0.029008  
Marcopolo PN -0.219689 -0.010000  
Papel Simão PN 0.061511 1.047499  
Ipiranga Pet PN 0.533694 1.128265  
Randon Part PN 0.170305 1.183367  
Recrusul PN -0.198570 -0.011000  
Real PN -0.267534 0.062675  
Rheem PN -0.152404 0.068700  
Rhodia Ster -0.295079 -0.052047  
Ren Hermann PN 0.280538 0.540067  
Sid Riogrand PN 0.279119 0.879167  
Ipiranga Ref PN -0.072155 0.038204  
Inds Romi PN -0.181847 -0.016001  
Ripasa PN -0.103140 0.217522  
Agrocere PN -0.101876 0.035608  
Samitri ON 0.183624 0.608427  
Sadia Concord PN 0.171612 0.768419  
Sibra PNC -0.181392 -0.014408  
Suzano PN 0.138994 0.233284  
Telebahia PNB -0.191552 -0.009000  
Teka PN 0.701358 1.381336  
Telebrás-PN -0.119805 1.145672  
Telepar PN 0.585110 1.352650

Telerj PN 0.476337 1.462917  
Tibrás PNA -0.209243 -0.007000  
Telesp PN 0.087956 1.475551  
Unibanco PN 0.604743 1.361174  
Ucar Carbon ON 0.424654 1.062289  
Unipar PNB 0.210276 1.139694  
Usiminas PN 0.025770 1.020600  
Vale Rio Doce PN 0.156494 0.954620  
Varga Freios PN -0.189956 0.001593  
Vidr S Marina ON -0.178544 0.017187  
White Martins ON 0.374019 0.956533  
Wembley PN -0.132702 -0.00600

### ABSTRACT

The systematic risk of a stock may be understood as a measure of its tendency to follow the movements of a chosen market index. We propose the method of orthogonal regression to estimate the systematic risk of stocks. The main assumption behind this procedure, what makes it attractive and suitable for the problem treated here, is that the variables (the stock and market index returns) may be observed with errors. Given the high volatility of our financial market, which frequently tends to produce outliers, we estimate the market risk of Brazilian stocks using robust estimates for the method of orthogonal regression. We introduce the *R-R Plot* of the orthogonal residuals as a tool to detect *outliers*.

**Key Words:** Systematic Risk; Errors in Variables Model; Robust Estimation.



## Política Editorial

A Revista Brasileira de Estatística - RBEs - objetiva promover a estatística relevante para aplicação em questões sociais, interpretadas amplamente para incluir questões educacionais, de saúde, demográficas, econômicas, legais, de políticas públicas e de estatísticas oficiais, entre outras. A revista apresenta artigos num formato que permite fácil assimilação pelos membros da comunidade científica em geral. Os artigos devem incluir aplicações práticas como assunto central. Essas aplicações devem ter conteúdo estatístico substancial. As análises deverão ser exaustivas e bem apresentadas, mas o emprego de métodos estatísticos inovadores não é essencial para publicação.

Artigos contendo exposição de métodos são aceitáveis, desde que estes sejam relevantes para as áreas cobertas pela revista, auxiliem na compreensão do problema e contenham interpretação clara das expressões matemáticas apresentadas. A apresentação de aplicações ilustrativas envolvendo dados adequados é requerida. Tratamentos algébricos extensos devem ser evitados.

A revista tem periodicidade semestral e publica também artigos escritos à convite e resenha de livros, bem como artigos abordando os diversos aspectos de metodologias relevantes para órgãos produtores de estatísticas, incluindo:

- a) planejamento de pesquisas;
- b) avaliação e mensuração de erros em pesquisas;
- c) uso e combinação de fontes alternativas de informações e integração de dados;
- d) novos desenvolvimentos em metodologia de pesquisa;
- e) crítica e imputação de dados;
- f) amostragem e estimação;
- g) disseminação e confiabilidade de dados;
- h) análise de dados;
- i) análise de séries temporais;
- j) modelos e métodos demográficos; e
- k) modelos e métodos econométricos.

Todos os artigos submetidos são avaliados quanto à qualidade e relevância por dois especialistas indicados pelo Comitê Editorial da Revista Brasileira de Estatística. Os artigos submetidos devem ser inéditos e não ter sido, simultaneamente, submetidos a qualquer outro periódico nacional. O processo de avaliação é do tipo duplo cego, isto é, os artigos são avaliados sem identificação da autoria, e os comentários dos avaliadores também são repassados aos autores sem identificação.

### Instruções para submissão de artigos

Os artigos submetidos para publicação devem ser remetidos em três vias, que não serão devolvidas, para:

Pedro Luis do Nascimento Silva  
Editor responsável  
Revista Brasileira de Estatística  
Av. República do Chile, 500 10º andar  
Rio de Janeiro - RJ - 20031-170  
Tel.: xx 55-21-5144548  
Fax: xx 55-21-514 0039  
E-mail: pedrosilva@ibge.gov.br

Para cada artigo publicado, são fornecidas gratuitamente 20 separatas.

## Instruções para Preparo de Originais

Os originais entregues para publicação devem obedecer às seguintes normas:

- 1 - O texto deve ser editado, preferencialmente, em Word, sem formatação (*default*), configurado em A4.
- 2 - As páginas do original devem ser numeradas seqüencialmente.
- 3 - A primeira página do original (folha de rosto) deve conter o título do artigo, nome completo do(s) autor(es), com indicação das instituições a que está(ão) vinculado(s) e endereço para correspondência. Agradecimentos a colaboradores e instituições e auxílios recebidos devem figurar também nessa página.
- 4 - A segunda página do original deve conter um resumo informativo de no máximo 150 palavras, em português e inglês (Abstract), destacando os pontos relevantes do artigo. Deve seguir o mesmo padrão do texto, em um único parágrafo, sem inclusão de fórmulas. Ver a respeito a norma da Associação Brasileira de Normas Técnicas - ABNT -, *Resumos*: NBR 6028, de julho de 1988.
- 5 - As notas explicativas devem ser numeradas numa seqüência única, listadas no pé-de-página onde elas se encontram.
- 6 - As tabelas e gráficos devem ser precedidos de títulos que permitam a identificação do conteúdo. Devem ser numerados seqüencialmente (Tabela 1, Figura 3, etc.) e com ordem de indicação de entrada no texto.

Toda tabela e gráfico deve ter fonte. Recomenda-se a indicação dos documentos publicados que foram utilizados na sua elaboração, identificados por referências bibliográficas completas, com as páginas ou volumes específicos de onde foram extraídas as informações. No caso de publicação que contenha tabelas com dados numéricos resultantes de uma única fonte, já identificada na própria publicação, é dispensável a apresentação da fonte em cada uma das tabelas.

No caso de tabelas e demonstrações extensas ou outros elementos de suporte, podem ser incluídos anexos, que devem ter título e numeração.
- 7 - As citações bibliográficas no texto devem ser feitas de acordo com a norma da ABNT, *Apresentação de citações em documentos*: NB 896, de maio de 1990.
- 8 - As referências bibliográficas devem ser redigidas segundo a norma da International Organization for Standardization - ISO -, *Referência bibliográfica e documentação*: n. 690, de 1987, contendo os elementos necessários à identificação da publicação. Devem ser organizadas em ordem alfabética.

Se o assunto é Brasil,  
procure o IBGE

<http://www.ibge.gov.br>

<http://www.ibge.org>

atendimento  
0800 21 81 81