MARCELO KRIEGER

UM MÉTODO AUTOMÁTICO PARA PREVISÃO DE
SÉRIES TEMPORAIS USANDO A METODOLOGIA BOX & JENKNIS

Dissertação apresentada ao Departamento de Engenharia Elétrica da PUC/RJ como parte dos requisitos para obtenção do título de Mestre em Ciências de Engenharia Elétrica.

Orientador: Reinaldo Castro Souza.

DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

Rio de Janeiro, 29 de julho de 1983.

A todos que venham a usá-la.

Meus agradecimentos

- Ao Prof. Reinaldo, pelo apoio pessoal e técnico.
- À Biblioteca e às Bibliotecárias da DATAPREV.
- À Sra. Edsael Maria Falcão, pela datilografia.

RESUMO

Apesar do reconhecimento amplo da qualidade das previsões obtidas na aplicação de um modelo ARIMA à previsão de séries temporais univariadas, seu uso tem permanecido restrito pela falta de procedimentos automáticos, computadorizados. Neste trabalho este problema é discutido e um algoritmo é proposto.

ABSTRACT

In spite of general recognition of the good forecasting ability of ARIMA models in predicting time series, this approach is not widely used because of the lack of automatic, computerized procedures. In this study this problem is discussed, and an algorithm is proposed.

SUMÁRIO

LISTA DE ILUSTRAÇÕES.....	V
I - INTRODUÇÃO.....	1
II - A METODOLOGIA BOX & JENKINS.....	3
II.1 - Introdução.....	3
II.2 - Os modelos ARIMA.....	4
II.2.1 - Modelos Autorregressivos (AR).....	5
II.2.2 - Modelos de Médias Móveis (MA).....	6
II.2.3 - Modelos não Estacionários.....	6
II.2.4 - Modelos Sazonais.....	7
II.3 - A Metodologia Box & Jenkins.....	9
II.3.1 - Identificação.....	11
II.3.1.1 - Função autocorrelação.....	11
II.3.1.2 - Função de autocorrelação parcial.....	12
II.3.1.3 - O processo de identificação.....	14
II.3.2 - Estimação e Testes de Hipótese.....	15
II.3.2.1 - Teste de Portmanteau.....	15
II.3.2.2 - Teste do periodograma acumulado.....	16
II.3.2.3 - Teste de sobrefixação.....	17
II.3.3 - Previsão.....	17
II.4 - O Problema da Identificação.....	19
II.4.1 - A Função de Autocorrelação Inversa.....	19
II.4.2 - O Critério de Informação de Akaike.....	21
II.4.3 - A Função de Autocorrelação Extendida.....	22
III - MÉTODOS AUTOMÁTICOS.....	24
III.1 - Introdução.....	24

III.3 - Um Algoritmo de Seleção Automática de Modelos Combinando um Método de Reconhecimento de Padrão com um Procedimento de Reexame e Libe- ração guiado por um critério.....	25
III.3 - Uma Abordagem Simplificada do Método de Box & Jenkins para a Análise e Previsão de Séries Temporais Unidimensionais.....	30
III.4 - ARIMAID.....	32
III.5 - Comentários.....	35
IV - O MÉTODO PROPOSTO.....	37
IV.1 - Introdução.....	37
IV.2 - A Função de Identificação Escolhida.....	38
IV.2.1 - FAC-FACP.....	38
IV.2.2 - FAC-FACI.....	38
IV.2.3 - Os Critérios de Akaike.....	38
IV.2.4 - FACE.....	39
IV.3 - As "Figuras" da FACE.....	40
IV.4 - O Algoritmo.....	42
IV.5 - O Problema da Sazonalidade.....	44
V - APLICAÇÃO.....	45
VI - CONCLUSÕES.....	48
APÊNDICE - Listagem do Programa de Identificação.....	49
REFERÊNCIAS BIBLIOGRÁFICAS.....	54

LISTA DE ILUSTRAÇÕES

Um Filtro Linear Como Gerador da Série.....	04
O Filtro Completo.....	07
Fluxograma da Metodologia Box & Jenkins.....	10
Comportamento da FAC e da FACP para os Modelos ARMA.....	14
Espectro Acumulado de Potência de um Ruído Branco.....	16
Comportamento da FAC e da FACI para os Modelos ARMA	20
Quadro das FACE Estimadas	23
Fluxograma do SIFT.....	28
Principais Pacotes B&J Automáticos.....	35
Resultados de Ray	36
Resultados de Souza e Squadri.....	39
O Triângulo.....	40
O Trapézio.....	41
O Pentágono	41
Modelo AR(2).....	45
Modelo ARMA(2, 1).....	46
Modelo ARMA(1, 2).....	46

I - INTRODUÇÃO

Apesar do interesse pelo assunto previsão existir desde o início da civilização, sua utilização sistemática através da aplicação de técnicas científicas, estatísticas data do século passado. Este redirecionamento da função de previsão, causado pelo surgimento da necessidade da previsão no ambiente produtivo e administrativo ao invés da utilização especulativo-religiosa anterior, causou o aumento do número de cientistas e técnicos envolvidos na pesquisa e aplicação de novos métodos à previsão.

Uma das correntes surgidas advoga a utilização das séries temporais ao estudo de fenômenos que envolvem no tempo. Uma série temporal é um conjunto de observações de um determinado fenômeno tomadas ao longo do tempo, em intervalos regulares, por exemplo, coleta semanal do número de camisas vendidas em uma loja. Por esta corrente, as observações coletadas contêm informações úteis para a previsão do fenômeno em tempos futuros.

Box & Jenkins (1970) [1], formalizaram a teoria da utilização de componentes autorregressivas e de médias móveis na previsão de séries temporais utilizando-se de duas idéias básicas na criação de sua metodologia de construção de modelos:

- 1) Parcimônia que consiste na utilização do menor número possível de parâmetros para obter uma representação adequada do fenômeno em estudo e
- 2) Construção iterativa do modelo em que a informação empírica é analisada teoricamente sendo o resultado deste estágio confrontado com a prática e assim sucessivamente até a obtenção de um modelo satisfatório.

Na análise das vantagens e desvantagens do uso da metodologia de Box e Jenkins, a conclusão comum a diversos autores é que a metodologia apesar de poderosa e precisa, é também custosa e complexa, exigindo para sua aplicação a disponibilidade de um técnico com profundo conhecimento do método. Para superar esta dificuldade têm-se buscado a automatização da metodologia através de técnicas computadorizadas.

O objetivo deste trabalho é adicionar um passo a este caminho, com a procura de um algoritmo automático à modelagem Box & Jenkins alternativo aos existentes, que não são satisfatórios na opinião do autor e sua implementação em um microcomputador, com o intuito de tornar o método mais acessível aos administradores e técnicos de outras áreas que necessitem da previsão.

O restante do trabalho está organizado da seguinte maneira: no capítulo II é apresentado um resumo da metodologia de Box & Jenkins, no III são apresentados alguns dos métodos automáticos existentes, no IV é proposto um algoritmo automático alternativo que é analisado no capítulo V e finalmente, o capítulo VI contém as conclusões.

II - A METODOLOGIA BOX & JENKINS

II.1 - INTRODUÇÃO

Os modelos autorregressivos integrados médias móveis (ARIMA) foram estudadas exaustivamente por Box & Jenkins [1], e seus nomes são frequentemente usados como sinônimo da aplicação de um processo ARIMA à análise, previsão e controle de séries temporais. Os modelos autorregressivos foram introduzidos por Yule (1926) e posteriormente generalizados por Walker (1931), enquanto os modelos médias móveis foram usados inicialmente por Slutsky (1937). Wold (1938) forneceu fundamento teórico aos processos mistos AR-MA. Construídos sobre o trabalho de Wold, os modelos ARIMA desenvolveram-se em duas direções: procedimentos eficientes de identificação e estimação e, extensão dos resultados para incluir Séries Temporais Sazonais. [3]

Box & Jenkins (1970) apresentaram as informações necessárias à compreensão e o uso dos modelos ARIMA univariados às séries temporais.

No restante do capítulo serão introduzidos os modelos ARIMA, a metodologia Box & Jenkins propriamente dita e será discutido o problema da identificação.

II.2 - OS MODELOS ARIMA

Consideremos uma série temporal em que observações sucessivas podem ser representadas por uma combinação linear de variáveis aleatórias independentes, $e_t, e_{t-1}, e_{t-2}, \dots$ retiradas de uma distribuição estacionária com média zero e variância constante (γ_e^2). Se, além disto assumirmos que a distribuição geradora tem distribuição normal, então a sequência de variáveis aleatórias é chamada de ruído branco.

A combinação linear dos e_t pode ser escrita como:

$$x_t = \psi_0 e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$$

ou

$$x_t = \sum_{k=0}^{\infty} \psi_k e_{t-k}$$

ou seja, o processo estocástico gerador da série temporal é do tipo obtido pela passagem de um processo ruído branco por um filtro linear. (Figura II.1)

Figura II.1

Definimos o operador de atraso B , como tendo o seguinte efeito sobre uma variável discreta:

$$B e_t = e_{t-1}, \text{ o que generalizando resulta em:}$$

$$B^k e_t = e_{t-k}$$

Podemos escrever o processo em função de B , como:

$$x_t = \psi(B) e_t$$

Box & Jenkins definem:

$$\psi(B) = \frac{\theta_q(B)}{\phi_p(B)}$$

O que implica:

$$\phi_p(B)x_t = \theta_q(B) e_t$$

Onde p e q representam a ordem dos polinômios ϕ (componentes autorregressivas e θ (componentes médias móveis). Obtemos, portanto, um polinômio infinito a partir da razão de dois polinômios finitos, o que torna tratável o problema, pois para caracterizar o filtro linear, precisamos identificar $p + q$ parâmetros ($\phi_1, \phi_2, \dots, \phi_p$ e $\theta_1, \theta_2, \dots, \theta_q$) ao invés de infinitos parâmetros ($\psi_1, \psi_2, \dots$) como anteriormente.

II.2.1 - Modelos Autorregressivos (AR)

Um modelo autorregressivo puro é aquele em que $\theta_q(B) \equiv 1$, o que resulta em:

$$x_t = e_t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p}$$

Neste modelo, o valor presente do processo depende de um conjunto finito de valores passados e de uma componente de erro do tipo ruído branco. O nome provem do fato de que modelos em que uma variável depende linearmente de outras mais uma componente de erro são chamados modelos de regressão.

II.2.2 - Modelos de Médias Móveis (MA)

São assim chamados aqueles em que $\phi_p(B) = 1$, resultando em:

$$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

O valor atual do processo depende de uma soma ponderada dos e_t que se movem no tempo, partindo daí o nome. Na verdade, como o somatório dos θ 's é em geral diferente de 1, o nome média não é exato, porém é adotado por todos os autores quando tratam destes modelos.

II.2.3 - Modelos Não Estacionários

Até o momento tratamos de modelos estacionários (AR-MA), o que em termos algébricos significa que as raízes do polinômio ϕ estão localizadas fora do círculo unitário. Porém, na prática muitas séries apresentam um comportamento não estacionário de uma forma homogênea, em que apesar do nível da série variar, as oscilações em torno do nível são semelhantes para todos os níveis. Em termos algébricos, isto significa que pelos menos um dos zeros do polinômio autorregressivo localiza-se sobre o círculo unitário.

Obtemos então o modelo ARIMA (p, d, q):

$$\phi(B) (1 - B)^d x_t = \theta_q(B) e_t$$

A razão para a inclusão da palavra integrado pode ser compreendida a partir do diagrama em blocos que representa a obtenção da série temporal através da passagem de um ruído branco por uma sequência de filtros lineares. (Figura II.2)

Figura II.2

II.2.4 - Modelos Sazonais

A formulação de modelos ARIMA pode ser estendida para representar séries temporais que apresentem sazonalidade. Em geral, dizemos que uma série exibe comportamento sazonal de período S , quando observamos similaridades na mesma após S intervalos de tempo e, portanto, apresentam forte correlação nas defasagens relativas ao período sazonal.

O modelo sazonal multiplicativo é denotado por SARIMA $(p, d, q) (P, D, Q)_S$, definido formalmente como:

$$\phi(BS) \phi(B) \nabla^{DS} \delta^d X_t = \theta(BS) \theta(B) e_t$$

onde: $\phi(B)$: polinômio AR simples de grau P .

$\theta(B)$: polinômio MA simples de grau Q .

$\phi(BS) = (1 - \phi_1 B^S - \dots - \phi_P B^{SP})$: polinômio AR múltiplo de grau SP .

$\theta(BS) = (1 - \theta_1 B^S - \dots - \theta_Q B^{SQ})$: polinômio MA múltiplo de grau SQ .

δ^d : filtro não linear homogêneo simples.

$\nabla^{DS} = (1 - B^S)^D$: filtro não linear sazonal.

Assim, o modelo sazonal é multiplicativo no sentido que os dados observados, resultam da filtragem sucessiva do ruído de entrada, e_t , através do filtro não sazonal e então pelo filtro sazonal.

II.3 - A METODOLOGIA BOX & JENKINS

Para modelarmos uma série temporal através de um modelo ARIMA precisamos caracterizar completamente o filtro linear que deu origem a série, a partir de um ruído branco, como na Figura II.2.

Na prática, entretanto, o problema é o inverso. A partir de uma realização do processo, tentamos chegar ao ruído branco, através da passagem de x_t pelos filtros no sentido contrário.

Box & Jenkins dividiram o processo de utilização de um modelo ARIMA à modelagem de uma série temporal em três fases:

- 1) Identificação;
- 2) Estimação e testes de validade e;
- 3) Previsão;

e desenvolveram uma metodologia iterativa para sua aplicação cujo fluxograma é mostrado na figura II.3 e que será apresentada a seguir:

FIGURA II.3 - FLUXOGRAMA DA METODOLOGIA BOX & JENKINS

II.3.1 - Identificação

A identificação para Box & Jenkins consiste na determinação da ordem p , q e d dos polinômios constituintes do filtro linear. É realizada através da análise da série histórica, que precisa ter comprimento maior ou igual a 50 para que os resultados sejam satisfatórios.

Box & Jenkins sugeriram a comparação do comportamento das funções de autocorrelação e autocorrelação parcial empíricas com o comportamento teórico delas para os modelos ARIMA, como ferramenta para a identificação.

II.3.1.1 - Função Autocorrelação

Como consequência da suposição de estacionariedade (e ergodicidade) do processo gerador da série, a autocovariância entre duas observações depende somente do número de períodos separando as mesmas.

Com isto, a função de autocovariância, definida por:

$$\gamma_k = \text{COV}(x_t, x_{t+k}) = E [(x_t - E[x_t]) (x_{t+k} - E[x_{t+k}])]$$

será, devido a $E[x_t] = E[x_{t+k}] = \mu$ (média constante do processo estacionário), reduzida a [2]:

$$\gamma_k = E [(x_t - \mu) (x_{t+k} - \mu)]$$

A análise da série poderia ser feita a partir do estudo do comportamento das diversas autocovariâncias, que

nos revelaria a dependência entre os diversos termos da série. Porém, a função γ_k , conforme definida acima não permite que séries diferentes sejam comparadas por ser uma medida absoluta. Para contornar este problema, definimos a função autocorrelação ρ_k .

$$\rho_k = \frac{\gamma_k}{\text{VAR}(x_t)} \quad k = 0, 1, 2, \dots$$

Para qualquer que seja a série em estudo, temos ρ_k como uma medida padrão da dependência entre os intervalos de tempo (atrasos), onde, para todo k , $|\rho_k| \leq 1$.

"A função de autocorrelação decai exponencialmente com o atraso para os modelos AR, sofre um corte brusco após o atraso q para os modelos MA e decai exponencialmente após o atraso q para os modelos ARMA." Souza (1981) [2]. Com isto a função de autocorrelação não é suficiente para determinar o grau p do polinômio autorregressivo.

II.3.1.2 - Função de Autocorrelação Parcial

Consideramos um modelo AR(P) para o qual a função de autocorrelação é como mostrada abaixo:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}$$

Fazendo $k = 1, 2, \dots, p$ e levando-se em consideração que $\rho_i = \rho_{-i}$, obtemos o sistema de equações abaixo, conhecido como equações de Yule - Walker:

$$\begin{array}{rcl} \rho_1 & = & \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 & = & \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ \hline \rho_p & = & \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{array}$$

Portanto, conhecendo o valor de p e as autocorrelações $\rho_1, \rho_2 \dots \rho_p$, poderíamos resolver o sistema para os p valores $\phi_1, \phi_2, \dots, \phi_p$ desconhecidos.

Um método simples para a determinação da ordem p do polinômio autorregressivo, sugerido pelo comentário acima, consiste em aplicar sucessivamente as equações de Yule-Walker, utilizando os valores estimados dos ρ 's pela função autocorrelação em pírica, da seguinte forma:

$$\begin{bmatrix} 1 & \rho_1 \dots \rho_{k-1} \\ \rho_1 & 1 \dots \rho_{k-2} \\ \hline \rho_{k-1} & \rho_{k-2} \dots 1 \end{bmatrix} \times \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \hline \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \hline \rho_k \end{bmatrix}$$

A autocorrelação parcial de um processo é definida como a sequência dos ϕ_{kk} obtidos pela resolução do sistema acima para $k = 1, 2, \dots$

A autocorrelação parcial decai exponencialmente para os modelos MA, sofre um corte brusco após o intervalo de tempo (atrasos) p para os modelos AR e decai exponencialmente após o atraso p para os modelos ARMA.

O Quadro II.1 resume as características das funções autocorrelação e autocorrelação parcial, para os diversos tipos de processo.

	AR	MA	ARMA
Função de Autocorrelação (FAC) ρ_k	Infinita (exponenciais amortecidas e/ou senoídes amortecidas) não se anulam bruscamente	Finita. Anulam-se bruscamente	Infinita (exponenciais e/ou senoídes amortecidas para $k > q - p$) não se anulam bruscamente
Função de Autocorrelação parcial (FACP) ϕ_{kk}	Anulam-se bruscamente	Infinita (exponenciais amortecidas e/ou senoídes amortecidas) não se anulam bruscamente	Infinita (exponenciais e/ou senoídes amortecidas para $k > p - q$) não se anulam bruscamente

Quadro II.1 - Comportamento de FAC e da FACP para os modelos ARMA [2].

II.3.1.3 - O Processo de Identificação

O Processo de Identificação consiste de dois estágios: o primeiro de identificação do filtro não estacionário (d) e em seguida identificação do filtro estacionário (p, d).

A identificação de d é feita diferenciando-se sucessivamente a série original até que as funções autocorrelação e autocorrelação parcial apresentem um comportamento semelhante ao do Quadro II.1 quando então teremos os valores de p e q.

Na prática, as funções empíricas não se comportam tão bem quanto desejável, o que fez com que os estudiosos no assunto voltassem suas atenções para alternativas a este procedimento. Voltaremos a isto mais adiante.

II.3.2 - Estimação e Testes de Hipótese

Uma vez que o modelo esteja identificado, a estimação é feita de uma forma simples, através de estimadores de mínimos quadrados que coincidem com os estimadores de máxima verossimilhança, desde que os resíduos sejam normais e independentes. Como a obtenção de resíduos do tipo ruído branco e a preocupação do processo de modelagem como um todo, os testes de adequabilidade do modelo estimado são geralmente, comparações entre as propriedades da série de resíduos obtida a partir do modelo identificado e estimado e as propriedades de um processo ruído branco.

II.3.2.1 - Teste de Portmanteau

A partir da série dos resíduos estimados podemos calcular as k primeiras autocorrelações dos resíduos:

$$r_i(\epsilon); \quad i = 1, 2, \dots, k$$

O teste proposto assume que, se o modelo fixado é correto, então a estatística:

$$\xi = M \sum_{j=1}^k r_j^2(\epsilon)$$

tem uma distribuição qui-quadrado com $(k - p - q)$ graus de liberdade e $M = N - d$. Em geral $k \in (20, 25)$.

Assim, referindo-se o valor de ξ na tabela da distribuição qui-quadrado com M graus de liberdade, podemos testar a validade do modelo dentro de um determinado nível de significância. [2]

II.3.2.2 - Teste do Periodograma Acumulado

Neste teste, comparamos os periodogramas acumulados da série dos resíduos estimados e_t e de um ruído branco a_t para verificarmos a existência de componentes periódicas na série dos e_t .

Para o ruído branco a_t , o espectro acumulado de potência (integral da transformada de Fourier da função autocorrelação da série) é como mostrado na Figura II.4. O teste consiste em verificar a linearidade do periodograma acumulado da série dos resíduos, no intervalo $(0, 0)$ e $(0,5, 1)$. Os limites de aceitação são traçados de acordo com o teste de Smirnov-Kolmogorov na distribuição de frequências. [2]

Figura II.4 - Espectro Acumulado de Potência de um Ruído Branco

II.3.2.3 - Teste de Sobrefixação

Este teste não é aplicado aos resíduos, consistindo na elaboração de um modelo com um número de parâmetros superior ao do modelo fixado. O modelo sobrefixado será aceito caso o parâmetro ou parâmetros sejam significantes. Caso contrário o modelo fixado será aceito. [2]

II.3.3 - Previsão

Uma vez selecionado, com sucesso, um dado modelo ARIMA (p, d, p) ou SARIMA (p, d, q) (P, D, Q)s desejamos obter os valores futuros da série.

Está demonstrado em Souza (1981) [2], que a previsão $x_t(L)$ previsão L passos à frente, feita na origem dos tempos T da série x, que apresenta erro médio quadrático mínimo, é aquela em que explicitando o modelo da seguinte forma:

Utilizamos os seguintes valores para as expectâncias:

- a) Expectância condicional dos x_t já realizados são as próprias realizações

$$E_t [x_{t+j}] = E [x_{t+j} | x_t, x_{t-1} \dots] = x_{t+j} \quad j=1, -2 \dots$$

- b) Expectância condicional dos x_t ainda não realizados são as respectivas previsões:

$$E_t [x_{t+j}] = E [x_{t+j} | x_t, x_{t-1}] = x_t(j) \quad j=1, 2, 3 \dots$$

c) Expectância condicional dos e_t são:

$$E_t [e_{t+j}] = E [e_{t-j} | e_t, e_{t-1}, \dots] = 0 \quad j=1, 2, \dots$$

$$E_t [e_{t-j}] = E [e_{t-j} | e_t, e_{t-1}, \dots] = e_{t-j} = x_{t-j} - x_{t-j-1} \quad (1) \quad j=0, 1, 2, \dots$$

II.4 - O PROBLEMA DA IDENTIFICAÇÃO

A identificação é considerada a etapa mais difícil da modelagem B&J, além de possuir um forte componente subjetivo. Segundo B&J (1970) [1], "O processo de identificação é necessariamente inexato. É inexato porque a questão de quais tipos de modelos ocorrem na prática e em que circunstâncias, é uma propriedade do comportamento do mundo físico e não pode, por conseguinte, ser deduzida por argumentos puramente matemáticos". Pukkila (1980) [4], diz: "Como a construção de modelos em geral, também a construção de um modelo ARIMA seguindo a metodologia B&J requer habilidade e muito trabalho".

Realmente, conseguir identificar um decaimento exponencial ou diferenciar um valor significativo do início de um decaimento exponencial nas funções de autocorrelação e autocorrelação parcial empíricas de uma série temporal, não são tarefas fáceis para alguém que não esteja habituado com este tipo de observação.

O mundo físico não pode ser mudado, mas o processo de criação de modelos pode e, com este raciocínio em mente, diversos estudiosos lançaram-se a procura de métodos alternativos para o estágio de identificação, com o intuito de torná-lo menos trabalhoso e mais objetivo, de forma a permitir que até mesmo um computador seja hábil o suficiente para construir um modelo ARIMA.

A seguir serão apresentadas algumas destas propostas.

II.4.1 - A Função de Autocorrelação Inversa (FACI)

A FACI de um modelo ARIMA (p, d, q) é definida da seguinte forma:

"A FACI de um modelo ARIMA (p, d, q) é idêntica a função de autocorrelação do modelo inverso ARIMA (p, d, q) ." Hipel (1977). [5]

O uso da FACI no estágio da identificação se faz, comparando o comportamento dos estimadores das funções FAC e FACI com o comportamento teórico destas funções, conforme mostra o quadro 2.2.

	AR	MA	ARMA
Função de Autocorrelação (FAC) ρ_k	Infinita (exponenciais amortecidas e/ou senóides amortecidas) não se anulam bruscamente	Finita. Anulam-se bruscamente	Infinita (exponenciais e/ou senóides amortecidas para $k > q-p$) não se anulam bruscamente
Função de Autocorrelação inversa (FACI)	Finita. Anulam-se bruscamente	Infinita (exponenciais amortecidas e/ou senóides amortecidas) não se anulam bruscamente	Infinita (exponenciais e/ou senóides amortecidas para $k > q-p$) não se anulam bruscamente

Quadro II.2 - Comportamento de FAC e da FACI para os modelos ARMA. [6]

As vantagens obtidas pelo uso da FACI em substituição a FACP são as seguintes segundo Pereira e Souza (1981) [6]:

- 1) Estimação bem mais simples e menos onerosa em termos computacionais. (A estimação da FACP requer sucessivas inversões de matrizes, enquanto a da FACI requer apenas uma inversão de matriz).

- 2) A FACL é o DUAL da FAC tendo, como consequência interpretação mais simples que a FACP.
- 3) É particularmente útil na identificação de modelos com parâmetros intermediários identicamente nulos. O uso da FACP para este tipo de problema só permite a identificação da ordem do modelo, mas não fornece nenhuma indicação sobre os parâmetros identicamente nulos.

II.4.2 - O Critério de Informação de Akaike

O critério de Akaike é definido como o mínimo da função AIC (Akaike Information Criterion) expresso por:

$$AIC = -2\text{LOG} [\text{FUNÇÃO DE MÁXIMA VEROSSIMILHANÇA}] + 2[\text{N. DE PARÂMETROS DO MODELO}]$$

Para modelos ARIMA (p, d, q) o critério AIC é dado por [7]:

$$AIC = \frac{\text{LOG } se^2 + 2(p + q + 1 + \delta)}{N - d}$$

onde: N = número de observações da série
 se^2 = variância estimada dos resíduos

$$\delta = 1; \quad d = 0$$

$$\delta = 0; \quad d \neq 0$$

As seguintes considerações foram feitas a respeito do critério de Akaike [6]:

- 1) É uma das mais importantes alternativas para a identificação de modelos B&J, pelo fato de permitir uma automatização, via computador, sem intervenção do analista.

- 2) É uma formalização do princípio de parcimônia, pois procura um modelo adequado com o mínimo de parâmetros.

Critérios partidos da teoria da informação foram estudados por Hannan (1980). [8]

II.4.3 - A Função de Autocorrelação Extendida (FACE)

A FAC de uma série temporal é simples e útil na identificação de um modelo ARIMA para a série se $\phi(B) = 1$ (modelo MA puro). Se $\phi(B) \neq 1$ (modelo mixto) o problema se reduziria ao anterior se pudessemos obter $\hat{\phi}(B)$: estimador consistente de $\phi(B)$, pois:

$$w_t = \phi(B) x_t = MA(q)$$

Tsay e Tiao (1982) [9], utilizaram-se deste fato e propuseram estimadores consistentes para os parâmetros autorregressivos de um modelo ARIMA, definindo em consequência a função de autocorrelação extendida e demonstrando sua utilização como ferramenta de identificação.

Os estimadores consistentes para $\phi(B)$ são obtidos através da regressão iterativa:

$$x_t = \sum_{L=1}^k \hat{\phi}_{L(k)}^{(j)} x_{t-L} + \sum_{L=1}^j e_{L,t-i}^{(j-i)} + e_{k,t}^{(j)}$$

onde: $e_{k,t-i}^{(j-i)}$ = resíduo obtido da $(j-i)$ -ésima regressão iterativa dos k primeiros termos autorregressivos, defasado de i passos.

A estimação é feita por mínimos quadrados ordinários.

A FACE de ordem k ($R_j(k)$), será:

$$R_j(k) = \text{FAC}_j [W_t(j) = x_t - \sum_{i=1}^k \phi_i^{(j)} x_{t-i}]$$

Obtemos, portanto, uma matriz em que cada linha k , corresponde a regressão com k elementos autorregressivos, e cada coluna j corresponde a j -ésima iteração. (Figura II.5)

A FACE tem o mesmo comportamento de corte brusco que a FAC dos MA puros, gerando a seguinte regra de identificação:

Procurar na tabela face o triângulo de zeros, com fronteiras nas linhas. $k = C_1 \geq 0$ e $j - k = C_2 \geq 0$ e identificar $P = C_1$ e $Q = C_2$.

		q			
AR	MA	0	1	2	3
p	0	$r_1(0)$	$r_2(0)$	$r_3(0)$	$r_4(0) \dots$
	1	$r_1(1)$	$r_2(1)$	$r_3(1)$	$r_4(1) \dots$
	2	$r_1(2)$	$r_2(2)$	$r_3(2)$	$r_4(2) \dots$
	3	$r_1(3)$	$r_2(3)$	$r_3(3)$	$r_4(3) \dots$
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.

FIGURA II.5 - QUADRO DA FACE ESTIMADA

III - MÉTODOS AUTOMÁTICOS

III.1 - INTRODUÇÃO

Muitos autores tem discutido as vantagens e desvantagens do uso das técnicas de Box & Jenkins para previsão. Por exemplo:

WHEELWRIGHT & MAKRIDAKIS [10] - "Em resumo, a metodologia B&J é uma das mais poderosas e precisas técnicas de previsão conhecidas hoje, mas é também custosa e complexa em comparação com outras técnicas como regressão, decomposição ou filtro adaptativo."

REID [11] - "O método de B&J é claramente a melhor técnica entre os estudos empíricos e existem boas razões teóricas para que assim seja...sua principal desvantagem é não ser uma técnica totalmente automática."

Apesar de toda evidência teórica e empírica das vantagens da metodologia B&J, esta tem produzido pouco impacto na massa de pessoas que trabalham com previsão. A razão principal parece estar na falta de pessoas capazes de produzir e interpretar os modelos além do custo relativamente alto do processo de modelagem. Estas considerações levam naturalmente à busca de um substituto automático para o processo de modelagem.

No restante do capítulo serão apresentadas algumas estratégias de implantação de um sistema B&J automático que serão comentadas no final.

III.2 - Um algoritmo de seleção automática de modelos combinando um método de reconhecimento de padrão com um procedimento de reexame e liberação guiado por um critério.

Este algoritmo foi apresentado por Hille Woodworth (1980) [12] e resultou no pacote SIFT da Unilever Research, de Liverpool.

1ª Parte - Procedimento de Seleção de Modelos

Estágio 1 - Determine o grau de diferenciação - selecione-o minimizando a variância usando enumeração de todas as combinações possíveis, isto é, resolva:

$$\text{MIN } \left\{ \sigma_w(d, D) \mid W_t = \delta^d \nabla^D X_t, \begin{array}{ll} 0 \leq d, D \leq 1 & \text{para séries sazonais} \\ 0 \leq d \leq 2 & \text{para séries não sazonais} \end{array} \right\}$$

Para a série diferenciada:

Estágio 2 - Selecione modelos MA puros - teste corte na FAC e/ou decaimento exponencial ou senoidal na FACP.

Estágio 3 - Selecione modelos AR puros - teste corte na FACP e/ou decaimento exponencial ou senoidal na FAC.

Estágio 4 - Selecione modelos ARMA - procure um modelo MA na FAC até o lag 4, que seja seguido por um comportamento puramente AR, decaimento senoidal ou exponencial. A ordem máxima selecionada para o termo MA de uma estrutura ARMA será menor ou igual à selecionada para o modelo MA puro.

Estágio 5-7 - Seleção de estruturas sazonais - repita os estágios 2 - 4 para os lags S, 2S, 3S,...MS $M < N/S$.

Estágio 8 - Forme um conjunto final de estruturas de modelos igual a todas combinações possíveis de estruturas sazonais com não sazonais - apesar do número máximo de modelos selecionado estar limitado a 25 na prática o número médio selecionado está em torno de 5.

29 Parte - Procedimento de Liberação de Modelos

Originalmente o sistema usava o critério FPE de Akaike para escolher o melhor modelo, dado por:

$$FPE = \left\{ \frac{N + P}{N - P} \right\} \sigma_a^2(P)$$

N = Número de pontos na série

P = Número de parâmetros no modelo

$\sigma_a^2(P)$ = Variância da série de resíduos para P parâmetros.

que apresenta a desvantagem de que para um $N \rightarrow P$, a razão $(N+P)/(N-P)$ é pouco afetada por uma variação de P , tornando difícil uma escolha consistente para a ordem do modelo. O critério foi substituído posteriormente pelo BIC de Akaike, que é uma medida da entropia esperada do modelo, dado por: [15]

$$BIC = N \ln \sigma_a^2(P) + P + P \ln N + \ln (\sigma_a^2(0) - \sigma_a^2(P)) / P$$

$\sigma_a^2(P)$, N, P os mesmos do FPE

$\sigma_a^2(0)$ = variância da série diferenciada

O fluxograma do procedimento de liberação é apresentado a seguir, onde os modelos considerados são colocados em uma lista chamada LC e os liberados em outra chamada LR.

Figura III.1 - Fluxograma do SIFT

O SIFT foi comparado com a literatura obtendo o seguinte resultado:

Número de séries analisadas = 17

Número de vezes em que a variância do modelo
obtido com o SIFT $\approx$ variância do modelo da
literatura = 11

Variância do SIFT < variância da literatura = 4

Variância do SIFT > variância da literatura = 2

Número médio de parâmetros nos modelos SIFT = 1.94

Número médio de parâmetros nos modelos literatura = 1.89

- III.3 - Uma abordagem simplificada do método de Box & Jenkins para a análise e previsão de séries temporais unidimensionais, Abadie & Travers (1980). [13]

Esta não é um método automático propriamente dito, mas uma proposta de algoritmo de identificação simplificado que pode ser automatizado facilmente. Consiste em estimar diversos modelos, escolhendo o seguinte a partir dos parâmetros estimados na iteração atual e no valor de $\hat{S} = \sum \hat{e}_t$.

Para um modelo não sazonal é proposto o seguinte algoritmo:

1 - Diferenciação

1.1 - Faça $d = 0$

1.2 - Estime os parâmetros do modelo $(1, d, 1)$

1.3 - Se $\hat{\phi} \sim 1$ faça $d = d + 1$ e volte para 1.2 (1)

1.4 - Se $\hat{\theta} \sim 1$ faça $d = d - 1$ e volte para 1.2

2 - Sobreidentificação na direção sugerida pelo modelo escolhido anteriormente

2.1 - Se $\hat{\theta} \sim 0$ estime $(2, d, 0)$

2.2 - Se $\hat{\phi} \sim 0$ estime $(0, d, 2)$

2.3 - Se $\hat{\phi} \sim \hat{\theta}$ estime $(2, d, 0)$ e $(0, d, 2)$ mantendo o que apresente menor $\hat{S}$

2.4 - Se $\hat{\theta} \neq 0, \hat{\phi} \neq 0, \hat{\phi} \neq \hat{\theta}$ estime $(2, d, 2)$

3 - Fatore o polinômio, se possível

(1) Para automatizar o método seria necessário cuidado para evitar um enlaçamento infinito.

4 - Reduza ou aumente o modelo progressivamente, levando em consideração:

- a) A precisão das estimativas para decidir se um parâmetro não é significativamente diferente de zero ou se dois fatores são iguais.
- b) O valor de $\hat{S}$ e o princípio da parcimônia para o número de parâmetros.

Para um modelo sazonal, o procedimento é semelhante iniciando com um modelo $(1, 0, 1) \times (1, 0, 1)_s$.

III.4 - ARIMAID - Kang (1981) [14]

Arimaid é o nome do pacote de identificação automática desenvolvido pelo Dr. Kang da Transco Companies.

Este pacote faz uso extensivo do critério AIC de Akaike [15], que assume duas formas diferentes neste contexto.

1 - AIC1 para determinar o grau de diferenciação.

$$\begin{aligned} \text{AIC1 (d, D)} &= N \ln \left\{ \begin{array}{l} \text{estimativa de máxima verossimilhança} \\ \text{da variância dos resíduos} \end{array} \right\} + \\ &+ 2\delta \frac{N}{N - d - D*S} \end{aligned}$$

N = número de termos da série

S = período sazonal

$\delta = 0$ se não há diferenciação do modelo

$\delta = 1$ se há diferenciação

2 - AIC2 para determinar a ordem do modelo

$$\begin{aligned} \text{AIC2} &= N \ln \left\{ \begin{array}{l} \text{estimativa de máxima verossimilhança da} \\ \text{variância dos resíduos} \end{array} \right\} + \\ &2 \left\{ \begin{array}{l} \text{número de parâmetros} \end{array} \right\} \end{aligned}$$

Parte 1 - Determinação do Grau de Diferenciação

1) Comece com $d = 0$

2) Calcule a FAC e AIC1 da série

3) Se AIC1 foi reduzida e a FAC decair vagarosamente faça $d = d + 1$

4) Se $d < 2$ volte para 2

- 5) Se a série for não sazonal passe para a parte 2.
- 6) Comece com $D = 0$
- 7) Calcule a ACF e AIC1 da série
- 8) Se AIC1 for reduzida e a ACF decair vagarosamente faça $D = D + 1$
- 9) Se $D < 2$ volte para 7

Parte 2 - Seja $w_t = \delta^d V_s^D (X_t - \mu)$ a série estacionária de comprimento $N - d - Ds$, obtida da diferenciação da série original.

Se as raízes da equação $(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0$, estiverem fora do círculo unitário, o que é a chamada condição de inversibilidade, então o processo ARMA (p, q) :

$\phi_p(B)w_t = \theta_q(B)e_t$, pode ser escrito na forma:

$$\phi(B)w_t = (1 - \sum_{i=1}^{\infty} \theta_i B^i) w_t = e_t$$

chamada de forma inversa, onde θ_i , tende para zero, para i suficientemente grande. Como θ_i tende para zero, o processo pode ser bem representado por um processo AR(k) finito de alta ordem:

$$\phi_k(B)w_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_k B^k) w_t = e_t$$

O Arimaid usa o AIC para selecionar o modelo inverso AR(k) que melhor se ajuste à série, e o estima.

Em seguida, os parâmetros do modelo ARMA (p, q) são obtidos da seguinte relação, onde $p + q \leq k$.

$$1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_k B^k = \frac{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p}{1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q}$$

Os parâmetros ϕ e θ obtidos, são usados como primeira aproximação para o procedimento de estimação propriamente dito, porém, para os modelos MA puros esta aproximação não é boa e o pacote as corrige através da FACI.

Para modelos sazonais, o mesmo procedimento é seguido, porém em dois passos. As componentes não sazonais são identificadas e estimadas no primeiro passo a partir da série W_t . No segundo passo, as componentes são então combinadas na forma multiplicativa e os parâmetros não significantes são eliminados para formar o modelo ARIMA $(p, d, q) * (P, D, Q)$ s recomendado.

O Arimaid foi testado com 46 séries, anteriormente identificadas manualmente ou simuladas. Para 45 destas séries, o Arimaid fez uma identificação estatisticamente melhor ou igual à identificação manual.

III.5 - COMENTÁRIOS

No quadro III.1, são apresentados os pacotes para modelagem Box & Jenkins automática mais conhecidos. Todos eles tem em comum a características de utilizarem-se do critério de Akaike como auxiliar na escolha do modelo, tornando tênue ou eliminando a separação das fases de identificação e estimação. Os métodos consistem de uma iteração entre essas fases, em que os modelos identificados em um estágio são levados para a estimação e conforme seu desempenho com relação aos critérios, são escolhidos ou abandonados, o que inicia a iteração seguinte.

NOME DO PACOTE	AUTOR	ANO
PROJECT [18]	MEHRA	1977
ASEP [18]	LENZ	1977
SIFT	HILL & WOODWORTH	1978
ARIMAID	KANG	1981

Quadro III.1

São duas as críticas principais a este tipo de abordagem:

- 1) Utilização indiscriminada dos critérios de Akaike.

Ray (1982) [16] apresentou um estudo de simulação em que foram analisadas quinze realizações do seguinte modelo ARMA (1, 2).

$$X_t - 0.5 X_{t-1} = e_t + 0.1 e_{t-1} - 0.5 e_{t-2}$$

e o AIC foi usado para seleccionar o modelo correto entre oito candidatos. O Quadro III.2 apresenta os resultados deste estudo.

MÓDELO ESCOLHIDO	NÚMERO DE VEZES
(0, 1)	6
(0, 2)	5
(1, 2)	1
(2, 1)	2
(2, 2)	1

Quadro 111.2

ou seja, em apenas uma das vezes o critério seleccionou o modelo correto, e em mais cinco o modelo correto poderia ser obtido pela simples sobrefixação do modelo seleccionado.

- 2) Número excessivo de estimações requeridas, tornando a utilização custosa, deixando conseqüentemente de satisfazer uma das motivações para a busca de algoritmos automáticos.

IV - O MÉTODO PROPOSTO

IV.1 - INTRODUÇÃO

Os comentários da seção III.5 indicam que a discussão sobre a automatização da metodologia Box & Jenkins está longe de um encerramento e a proposta deste capítulo é a de adicionar mais um elemento e este assunto.

Do que foi apresentado anteriormente, infere-se que esta contribuição pode ser feita de duas formas.

- a) Através da busca de novas funções de identificação
- b) Através da criação de um algoritmo de identificação passível de implantação em um sistema de computação, utilizando as funções já conhecidas, que não apresente os problemas detectados.

Foi escolhida a segunda forma de abordagem, e o algoritmo resultante será apresentado a seguir.

IV.2 - A FUNÇÃO DE IDENTIFICAÇÃO ESCOLHIDA

O primeiro passo na criação de um algoritmo consiste na escolha da função de identificação. Esta escolha foi feita a partir da análise das funções apresentadas no capítulo II.

IV.2.1 - FAC - FACP

O par FAC-FACP permite uma identificação fácil para modelos AR puros, fornecendo inclusive uma boa primeira aproximação para os parâmetros. Para modelos MA a identificação também é fácil, porém as funções não fornecem uma primeira aproximação consistente para os parâmetros.

Fácil no parágrafo anterior significa que o padrão a ser reconhecido, cortes bruscos, não oferece dificuldades de implantação em um computador, existindo fórmulas de cálculo do limite de significância do valor encontrado empiricamente.

Esta característica - facilidade - não é encontrada na identificação de modelos mistos ARMA, já que o padrão passa a ser o de decaimentos exponenciais ou senoidais.

IV.2.2 - FAC-FACI

As mesmas considerações do Ítem IV.2.1 valem para o par FAC-FACI, com a única vantagem para este par, o fato dos valores da FACI servirem como primeira aproximação consistente para os parâmetros MA.

IV.2.3 - Os Critérios de Akaike

Foram analisados na seção III.5

IV.2.4 - FACE

Em termos de padrão, a FACE permite identificação fácil fornecendo primeiras aproximação consistentes para os parâmetros AR tanto dos modelos puros como dos mistos.

Souza e Squadri (1983) [17], estudaram a consistência da identificação através da FACE, à luz do trabalho de Ray [16], obtendo os resultados apresentados, na Quadro IV.1.

MODELO	SELEÇÃO PARA FACE
ARMA (0, 1)	2
ARMA (0, 2)	9
ARMA (0, 3)	1
ARMA (1, 2)	3

Quadro IV.1

A FACE apresenta uma clara vantagem sobre o AIC, pois em três casos o modelo selecionado foi o correto, enquanto que em outros nove casos bastaria uma sobrefixação do parâmetro AR para a obtenção do modelo correto. Além disto, o custo envolvido no cálculo da FACE é muito menor que o do AIC.

Consequentemente, nossa ferramenta de identificação será a FACE.

IV.3 - AS "FIGURAS" DA FACE

Teoricamente, a matriz FACE de um modelo ARMA (p, q) deve apresentar um triângulo de zeros com vértice no ponto (p, q) , uma aresta na horizontal e outra sobre a linha $k = p - q$ onde $k = \text{constante}$, conforme a figura IV.1, onde "x" significa valores diferentes de zero e "0" valores iguais a zero.

	0	1	2.....q-1	q	q+1.....
0	X	X	X	X	.
1	X	X	X	X	.
2	X	X	X	X	.
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
p-1	X	X	X	X	X X X X X
p	X	X	X	X	0 0 0 0 0
p+1	X	X	X	X	X 0 0 0 0
.	.	.	.	.	. X 0 0 0
.	.	.	.	.	. . X 0 0
.	.	.	.	.	. . . X 0
.	.	.	.	.	 X

FIGURA IV.1 - O TRIÂNGULO

Porém, na prática além do triângulo, encontramos também o trapézio e o pentágono irregular, exemplificados nas figuras IV.2 e IV.3 para o modelo ARMA (1, 2).

	0	1	2	3	4
0	X	X	X	X	X
1	X	X	0	0	0
2	X	X	0	0	0
3	X	X	X	0	0
4	X	X	X	X	0

Figura IV. 2-O TRAPÉZIO

	0	1	2	3	4
0	X	X	X	X	X
1	X	X	0	0	0
2	X	0	0	0	0
3	X	0	0	0	0
4	X	X	0	0	0
5	X	X	X	0	0
6	X	X	X	X	0

Figura IV.3 - O PENTÁGONO

Estas figuras são também ideais mas prestam-se para apresentar casos possíveis e até frequentes na prática e que devem ser adequadamente tratados. No caso do trapézio, além do modelo do vértice superior esquerdo, a sobrefixação por mais um parâmetro AR apresenta-se também como candidata a ser o modelo correto (este foi o caso dos nove modelos identificados como ARMA (0, 2) no trabalho de Souza e Squadri). No caso do pentágono temos como candidato outro (ou outros) modelo(s) tão parcimonioso(s) quanto o do vértice superior esquerdo.

IV.4 - O ALGORÍTIMO

Estamos interessados em um algoritmo para a modelagem B&J não sazonal automática utilizando a FACE. O universo de pesquisa será limitado a:

$0 < p + d < 6$ - geralmente, utiliza-se $0 < p < 4$ e $0 < d < 2$, porém com a utilização da FACE, torna-se indistinto o que é p e o que é d .

$0 < q < 4$ - como em todas as propostas

Desta forma, a maior matriz FACE que precisaremos calcular, será uma 7×5 como na figura IV.3

Passo 1 - Calcule a 1ª linha da matriz FACE (ordem zero, correspondendo a FAC da série original)

Passo 2 - Localize a 1ª posição zerada da linha. Esta será a candidata. ($q = k$)

Passo 3 - Calcule a próxima linha, até a posição $k - j$, se esta for a J-ÉSIMA linha seguinte à detecção de k (mesmo grau de parcimônia)

Passo 4 - Localize a 1ª posição zerada da linha

- . Caso o modelo seja mais parcimonioso que o atual (ou atuais) candidato(s), substitua o(s) candidato(s) pelo novo calculado.
- . Caso o grau de parcimônia seja igual, inclua este modelo na lista de candidatos.

Passo 5 - Se $k - j > 1$ volte ao passo 3

Passo 6 - Fatore os polinômios AR dos candidatos. Os fatores próximos de um (no máximo 2) são transformados em fatores de diferenciação.

Passo 7 - Leve o(s) candidato(s) à estimação. Escolha o de menor

$$\hat{S} = \sum \hat{e}_t^2.$$

Passo 8 - Sobrefixe o modelo escolhido, com o modelo $p' = p + 1$, estime e calcule $\hat{S}$. Escolha o de menor $\hat{S}$.

Passo 9 - Efetue as previsões com o modelo escolhido.

Os passos 1 a 5 do algoritmo consistem no cálculo da FACE para o fenômeno estudado. O passo 4 especificamente, aplica o conceito de parcimônia ao desprezar zeros de ordem superior e cobre o caso do pentágono apresentado na figura IV.3.

O passo 6 transforma fatores autorregressivos próximos de um em fatores de diferenciação porque a FACE não os distingue.

O passo 7 introduz o critério de menor somatório do quadrado dos resíduos para a escolha entre candidatos de mesmo grau de parcimônia.

O passo 8 cobre o caso do trapézio apresentado na figura IV.2.

IV.5 - O PROBLEMA DA SAZONALIDADE

A algoritmo apresentado na seção anterior revela-se adequado e econômico para séries não sazonais. Apesar disto, a tentativa de ampliá-lo para englobar a sazonalidade, pela ampliação dos limites de cálculo de p e q (a face não diferencia p e P ou q e Q) não é conveniente porque o custo subiria muito. Por exemplo, para um modelo $AR(1) \times AR(1)_{12}$ ou similar, seriam preciso calcular até a face AR de ordem 13 até haver o decaimento, o que implicaria em resolver inúmeros sistemas de equação de ordem alta.

Portanto, é necessário proceder a dessazonalização da série (através do método de Holt-Winters ou qualquer outro filtro sazonal) antes de utilizar o algoritmo proposto para a modelagem não sazonal.

V - APLICAÇÃO

Para testar o algoritmo foi feito um programa em Fortran que o executa até o passo 5, ou seja, os passos que não envolvem estimação. O algoritmo não foi implantado totalmente porque a rotina de estimação disponível ultrapassa o porte do computador utilizado. Este, um micro Schumec com 64k de memória foi usado como parte dos objetivos do trabalho, implantação do algoritmo em uma máquina de pequeno porte. (A listagem do programa encontra-se no Apêndice).

O programa recebe como entrada a série temporal a ser analisada e informa quais os candidatos, e respectivos valores do polinômio autorregressivo calculados.

Foram testadas inicialmente as séries A e C de Box & Jenkins, anteriormente modeladas por outros autores como ARMA (1, 1) e ARIMA (1, 1, 0) respectivamente [9]. O programa apresentou como candidatos, o modelo ARMA (1, 1) para a série A (coincidindo com o estudo anterior) e o modelo ARIMA (0, 1, 0) para a série C, o que nos conduzirá eventualmente ao mesmo modelo obtido anteriormente através do passo 8 do algoritmo.

Em seguida, o programa foi testado através da análise de 40 séries simuladas, 10 de cada um dos seguintes modelos: AR(2), MA(2), ARMA(2,1) e ARMA(1,2). As séries utilizadas foram obtidas através de um programa gerador construído especialmente para este fim.

O programa gerador usou a função embutida no computador, de geração de números aleatórios com distribuição uniforme no intervalo (0,1), que foi testada convenientemente para obter uma série de comprimento maior ou igual a 200. Esta série foi transformada, por um algoritmo que aplica o teorema da função inversa, para se tornar um ruído branco [19]. O ruído branco obtido foi filtrado pelo modelo em estudo e a série final resultante foi analisada pelo programa de identificação automática. Este procedimento foi repetido dez vezes com diferentes séries iniciais para cada um dos quatro modelos estudados.

Como esperado, o modelo MA(2) foi corretamente identificado nas dez tentativas. O resultado obtido para as outras séries são apresentadas nos Quadros V.1, V.2 e V.3.

MODELOS OBTIDOS	FREQUÊNCIA
AR(2)	10
ARMA(1,1)	5

QUADRO V.1 - MODELO AR (2)

MODELOS OBTIDOS	FREQUÊNCIA
ARMA (2, 1)	4
ARMA (1, 1)	1
ARMA (1, 2)	5
ARMA (2, 2)	1

QUADRO V.2 - MODELO ARMA (2, 1)

MODELOS OBTIDOS	FREQUÊNCIA
ARMA (1, 2)	2
ARMA (0, 2)	6
ARMA (1, 3)	1
ARMA (2, 0)	2

QUADRO V.3 - MODELO ARMA (1, 2)

Nota-se que em todas as tentativas, o modelo AR(2) foi o escolhido, porém em 50% das vezes do modelo ARMA(1, 1) também foi selecionado.

Para o modelo ARMA (2, 1), em 4 tentativas foi escolhido o correto em uma delas juntamente com o modelo ARMA (1, 2). Em uma o modelo ARMA (1, 1), foi escolhido, o que resultará no modelo correto através da sobrefixação do passo 8. Em outras quatro o modelo selecionado foi o ARMA (1, 2), que na sobrefixação resultará em um modelo ARMA (2, 2), com um parâmetro MA insignificante, que foi o caso da última experiência.

O modelo correto ARMA (1, 2) foi selecionado em duas tentativas, e em outras seis o MA (2) que resulta no correto por sobrefixação foi o escolhido. Em uma tentativa o modelo escolhido tinha excesso de parâmetros ARMA (1, 3) e em uma o modelo selecionado AR (2) não resultaria no correto de forma nenhuma.

Ou seja, das 42 séries testadas, em 41 casos o algoritmo completo seria capaz de selecionar o modelo correto, ou sobrefixado (6 casos).

VI - CONCLUSÕES

A metodologia Box & Jenkins é uma das melhores ferramentas disponíveis para a previsão de séries temporais. Seu principal inconveniente é ter um processo de modelagem penoso e custoso.

Para sobrepujar esta dificuldade tem-se buscado a automatização computacional do método. As principais tentativas realizadas utilizaram-se dos critérios de Akaike como função de identificação. Estes critérios apresentam as desvantagens de necessitarem do estágio de estimação que é custoso e de não serem muito consistentes.

Procurou-se então um algoritmo alternativo com o uso de outra função, tendo sido apresentada uma proposta de utilização da FACE no capítulo IV. O algoritmo proposto foi testado com duas séries da literatura e com quarenta séries simuladas, tendo sido eficaz em 41 das tentativas (apresentando sobrefixação em seis casos). Foi ressaltado que a extensão do algoritmo para englobar séries sazonais não é adequada por causa do custo envolvido, devendo-se buscar uma alternativa dessazonalizante antes da aplicação do algoritmo.

Como linhas futuras de trabalho, serão buscadas e sugere-se a quem se interesse: (1) alternativas de análise de sazonalidade e (2) viabilização da implantação do algoritmo completo, incluindo a estimação, em microcomputadores.

APÊNDICE

LISTAGEM DO PROGRAMA

```

TYPE FACE.FOR
  DIMENSION FACE(6,6),Z(500),W(500),A(10,10),B(10),FI(10)
  DIMENSION LINHA(5),COLUNA(6),FIAUX(5,6)
  LOGICAL V
  COMMON /ERRO/E(6,500)
  BART(N,K,J)=2*((N-K-J)*(-0.5))

C
C      VARIÁVEIS
C  FACE - MATRIZ DA AUTOCORRELACAO EXTENDIDA
C  Z    - SERIE DE ENTRADA
C  W    - SERIE TRANSFORMADA
C  A    - MATRIZ PARA CALCULO DAS REGRESSOES
C  B    - VETOR RESULTADO DO SISTEMA DE EQUACOES
C  FI   - VETOR DE PARAMETROS DO SISTEMA DE EQUACOES
C  E    - MATRIZ DOS ERROS
C  N    - TAMANHO DA SERIE DE ENTRADA
C  K    - INDICE DAS LINHAS DA MATRIZ FACE
C  J    - INDICE DAS COLUNAS
C  L    - DEFASAGEM TEMPORAL
C  FI, IJ, IAUX, IAUX2, IAUX3 - VARIÁVEIS AUXILIARES PARA INDICES
C
  V=.FALSE.
  WRITE(1,900)
900  FORMAT(' NUMERO DE DADOS N = ')
  READ(1,1000)N
1000 FORMAT(I3)
  CALL OPEN(6,'DADOS.DAT',0)
  READ(6,1100)(Z(I),I=1,N)
1100 FORMAT(15(1X,F4.1))
  ENDFILE 6
C
C  CALCULAR A LINHA 0 DA MATRIZ FACE
C
  CALL MEDVAR(Z,1,N,AMEDIA,VARIAN)
  DO 10 J=1,5
    FACE(1,J) = ACF(Z,J,J+1,N,AMEDIA,VARIAN)
    IF (ABS(FACE(1,J))<.LT.BART(N,0,J)) GOTO 200
10  CONTINUE
200  IF (J.GT.5) GOTO 210
    NCAND=1
    LINHA(1)=1
    COLUNA(1)=J
    V=.TRUE.
    IJ=1
    AUXLIN=1
    AUXCOL=J

```


CALCULAR AS COLUNAS DA LINHA K

DO 20 K=2,7

REGRESSAO DE ORDEN K

CALL MONTAR(Z,N,K,A,B)

CALL GAUSS(A,B,K,FI)

CALCULAR AS COLUNAS DA LINHA K

IF(.NOT.U) AUXCOL=5+K+AUXLIN

IAUX=IAUXCOL-(K-AUXLIN)

DO 30 J=1,IAUX

CALCULAR OS RESIDUOS

IAUX2 = K+1-1

DO 40 I=IAUX2,N

TEMP = FI(1)

DO 50 L=2,K

II = I-L+1

TEMP = TEMP + FI(L) * Z(II)

IAUX3=J-1

IF(IAUX3.LT.1) GOTO 40

DO 60 L=1,IAUX3

II = K+L+1

JJ = I-(J-L)

TEMP = TEMP + FI(II) * E(L,JJ)

E(J,I) = Z(I) - TEMP

REGRESSAO DE ORDEN J

CALL MONTA2(Z,N,K,J,A,B)

CALL GAUSS(A,B,K+J,FI)

CALCULO DA SERIE TRANSFORMADA

DO 70 I=K,N

TEMP = Z(I) - FI(1)

DO 80 L=2,K

II = I-L+1

TEMP = TEMP - FI(L) * Z(II)

W(I) = TEMP

CALL MEDVAR(W,K,A,MEEDIA,VARIAN)

FACE(K,J) = ACF(W,J-J+K-R,MEEDIA,VARIAN)

IF(ABS(FACE(K,J)).LT.BART(N,K-1,J)) GOTO 500

CONTINUE

TESTE PARA INCLUSAO DE NOVO CANDIDATO

IF(ABS(FACE(K,J)).GE.BART(N,K-1,J)) GOTO 320

IF(U) GOTO 340

IF(J.GT.5) GOTO 20

V=.TRUE.

NCAND=1

LINHA(1)=K

COLUMA(1)=1

AUXLIN=1

AUXCOL=K

DO 300 III=1,N

1

ROTINA DE MONTAGEM DA MATRIZ DE REGRESSÃO DE ORDEN J

DIMENSION Z(500),A(10,10),B(10)

DO 10 I=1,K

 B(I) = 0.0

 DO 10 J=1,K

 A(I,J) = 0.0

A(1,1) = N-K+1

DO 20 I = K,N

 B(1) = B(1) + Z(I)

 DO 20 J=2,K

 II = I-J+1

 A(1,J) = A(1,J) + Z(II)

 B(J) = B(J) + Z(I) * Z(II)

 DO 20 L=J,K

 LL = I-L+1

 A(L,J) = A(L,J) + Z(LL)*Z(II)

IAUX=K-1

DO 30 J=1,IAUX

 DO 30 L=2,K

 A(L,J) = A(J,L)

RETURN

END

SUBROUTINE MONTAR(Z,N,K,J,A,B)

ROTINA DE MONTAGEM DA MATRIZ DE REGRESSÃO DE ORDEN J

DIMENSION Z(500),A(10,10),B(10)

COMMON /ERRO/E(6,500)

IAUX=K+J

IAUX2=K+1

DO 10 II=1,IAUX

 B(II)=0.0

 DO 10 JJ=1,IAUX

 A(II,JJ)=0.0

A(1,1)=N-K-J+1

DO 20 II=IAUX,N

 B(1)=B(1) + Z(II)

 DO 30 JJ=2,K

 IIJJ = II-JJ+1

 A(1,JJ) = A(1,JJ) + Z(IIJJ)

 B(JJ) = B(JJ) + Z(II) * Z(IIJJ)

 DO 30 LL=JJ,K

 IILL = II-LL+1

 A(JJ,LL) = A(JJ,LL) + Z(IILL) * Z(IIJJ)

DO 40 JJ = 1,J

 KJJ = K + JJ

 IIJJ = II + JJ - J - 1

 A(1,KJJ) = A(1,KJJ) + E(JJ,IIJJ)

 B(KJJ) = B(KJJ) + Z(II) * E(JJ,IIJJ)

 DO 40 LL = 2,K

 IILL = II - LL + 1

 A(LL,KJJ) = A(LL,KJJ) + Z(IILL) * E(JJ,IIJJ)

DO 20 JJ = IAUX2,IAUX

 JJJ = JJ - B

 IIJJ = II + JJ - K - J - 1

 DO 20 LL=JJ,IAUX

 LLK = LL - K

 IILL = II + LL - J - K - 1

 A(JJ,LL)=A(JJ,LL) + E(JJJ,IIJJ) * E(LLK,IILL)

```

DO 50 II=2,IAUX
  III=II-1
  DO 30 JJ=1,III
    A(II,JJ)=A(JJ,II)
  RETURN
END
SUBROUTINE GAUSS(A,B,K,FI)

  ROTINA DE RESOLUCAO DE UM SISTEMA DE EQUACOES PELO METODO DE GAUSS
  DIMENSION A(10,10),B(10),FI(10)
  H=K+1
  DO 10 I=1,K
    A(I,H)=B(I)
  DO 20 I=1,K
    PIVOT=A(I,I)
    DO 30 J=1,H
      A(I,J) = A(I,J) / PIVOT
    IAUX=I+1
    DO 20 J=IAUX,K
      PIVOT = A(J,I)
      DO 40 II=I,H
        A(J,II) = A(J,II) - PIVOT * A(I,II)
      A(J,I) = 0.0
    DO 50 I=2,K
      II=K+2-I
      IAUX2=II-1
      DO 50 J=1,IAUX2
        A(J,H) = A(J,H) - (A(J,II) * A(II,H))
        A(J,II) = 0
      DO 60 I=1,K
        FI(I) = A(I,H)
  RETURN
END

```

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BOX, G.E.P & JENKINS, G.M. (1970) - Time Series Analysis, Forecasting and Control - Holden Day.
- [2] SOUZA, R. C. (1981) - Metodologia Box & Jenkins para Previsão de Séries Temporais Univariada - Publicação Interna - PUC.
- [3] MAKRIDAKIS, S & WHEELWRIGHT (1978) - Forecasting - Methods And Applications - John Wiley & Sons.
- [4] PUKKILA, T (1980) - On Identification of ARMA (p, q) Models.
- [5] HIPPEL, K.W., MCLEOD, A.I. & LENNOX, W.C. (1977) - Advances In Box & Jenkins Modelling 1. Model Construction-Water Resources Research, 13 pp 567-575.
- [6] PEREIRA, B.B. & SOUZA, R.C. (1981) - Desenvolvimentos Recentes na Metodologia Box & Jenkins - Anais do Iº Encontro Sobre Previsão Quantitativa - PUC/RJ.
- [7] AKAIKE, H. (1974) - A New Look at The Statistical Model Identification - IEEE Transactions on Automatic Control AC- 15 pp 716-723
- [8] HANNAN, E. J. (1963) - The Estimation of Seasonal Variation in Economic Time Series - Journal of The American Statistical Association, V. 58, pp. 31-44.
- [9] TSAY, R.S. & TIAO, G.C. (1982) - Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function For Stationary and Non Stationary ARMA Models - Tech. Report 683 - Dept. of Statistics, Univ. of Wisconsin, Madison.

- [10] WHEELWRIGHT, S. & MAKRIDAKIS, S. (1975) - Forecasting Methods For Management - John Wiley & Sons.
- [11] REID, D.J. (1969) - A Comparative Study of Time Séries Prediction Techniques on Economic Data - PH.D Thesis, Dept. of Mathematics, Univ. of Nothngham.
- [12] HILL, G.W. & WOODWORTH, D. (1980) - Automatic Box-Jenkins Forecasting J. OPL RES.SOC. Vol 31 pp. 413-422.
- [13] ABADIE, J. & TRAVERS, D. (1980) - Une Approche Simplifiée de la Méthode de Box Et Jenkins pour L'Analyse et la Prévision des Séries Temporelles Unidimensionnelles - R.A.I. R.O. Vol 14, Nov. 1980 p. 355 à 380
- [14] KANG, C.A. (1981) - Automatic Identification of Univariate ARIMA models- -Fifth International Time Series Meeting, Houston, Texas.
- [15] AKAIKE, H. (1971) - Autoregressive Model Fitting for Control - Annals of Institue of Statistical Mathematics, Vol 23 pp 163-180.
- [16] RAY, W.D. (1982) - J. OPL. RES. SOC. Vol 33 p. 767-768.
- [17] SOUZA, R.C. & SQUADRI, A.L.B. (1983) - Identificação, Automática de Modelos ARIMA (p, d, q) Através da Função de Autocorre^lação extendida - Teste de consistência - Artigo Submetido à SOBRAPO para Participação no Simpósio Brasileiro de Pesquisa Operacional de 1983.
- [18] LENZ, H.J. (1978) - Strategies of Implementation of a Fully Automatic Box & Jenkins Forecasting Technique - Anais do III Symposium on Operations Research - Mannheim
- [19] SHANNON, R.E. (1975) - Systems Simulation: The Art and Science - Prentice Hall, Inc.

Dissertação apresentada no Departamento de Engenharia Elétrica da PUC/RJ,
fazendo parte da banca os seguintes professores.

Reinaldo Castro Souza
DEE (Orientador) PUC/RJ

Christine Ann Assis
PUC/RJ

Marcos Azevedo da Silveira
DEE - PUC/RJ

Visto e permitida a impressão

Rio de Janeiro, 18/08/83

Prof. Nelson Velho de Castro Faria
Coordenador dos Programas de Pós-Graduação
do Centro Técnico Científico da PUC/RJ