ESTIMAÇÃO DE MODELOS

ARMA MULTIVARIADOS

MARIA APARECIDA GOUVÊA

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO: ESTATÍSTICA

ORIENTADOR: Prof. Dr. PEDRO ALBERTO MORETTIN

- São Paulo, outubro de 1982 -

AGRADECIMENTOS

Quero registrar meus agradecimentos ao Professor Doutor Pedro Alberto Morettin pela amizade e incentivo a mim transmitidos na orientação deste trabalho. Agradeço também à Maria Cristina Bonesio pelo auxílio prestado na pesquisa do material bibliográfico; à Edmarí Garro Teixeira por sua inestimável ajuda na elaboração da bibliografia; ao Sr. João Baptista Esteves de Oliveira por seu excelente trabalho de datilografia.

Finalmente, agradeço aos meus pais pelo apoio e estímulo que me deram durante todo o tempo dedicado a este trabalho.

Í N D I C E

CAP. 1 - INTRODUÇÃO	1
1.1. Preliminares.	1
1.2. Causalidade e "feedback".	2
1.3. Modelos Multivariados	3
1.4. Objetivo do Trabalho	5
CAP. 2 - ESTIMAÇÃO EM MODELOS LINEARES UNIVARIADOS.	8
2.1. Introdução.	8
2.2. A Função Exata de Verossimilhança para um Processo AR(p) segundo Box e Jenkins	10
2.3. A Função de Verossimilhança Exata para um Processo MA(q)	16
2.3.1. Metodologia de Box e Jenkins	16
2.3.2. Metodologia de Durbin.	25
2.3.3. Considerações sobre as Condições Impostas sobre os Parâmetros de um Processo MA(q)	30
2.4. A função de Verossimilhança Exata para um Processo ARMA(p,q)	34
2.4.1. Metodologia de Box e Jenkins e Newbold	34
2.4.2. Metodologia de Durbin.	47
2.5. Variâncias e Covariâncias das Estimativas de Máxima Verossimilhança, segundo Box e Jenkins.	53
CAP. 3 - ESTIMAÇÃO EM MODELOS ARMA MULTIVARIADOS.	56
3.1. Introdução.	56
3.2. Metodologia de Nicholls e Hall.	56
3.3. Determinação da Matriz de Covariância dos Valores Pré-períodos, Segundo Nicholls e Hall	63
3.4. Metodologia de Wilson	69
3.5. Estimação em Modelos ARMA Multivariados não Invertí- veis.	83
3.6. Comparação entre as Metodologias Empregadas durante a Estimação de Modelos ARMA Multivariados	103
CAP. 4 - ESTIMAÇÃO EM MODELOS MA E AR MULTIVARIADOS	106
4.1. Introdução.	106
4.2. Metodologia de Osborn	106
4.3. Metodologia de Wilson	110
4.4. Metodologia de Phadke e Kedem	111
4.5. Comparação dos Métodos Propostos por Phadke e Kedem	117
4.6. Metodologia de Hillmer e Tiao	118

4.7. Comparação da Função de Verossimilhança Exata com a Condicional para um Processo MA(q) Invertível	120
4.8. Obtenção das Estimativas de Máxima Verossimilhança Segundo Osborn.	125
4.9. Comparação da Função de Verossimilhança Exata com a Condicional para um Processo MA(q) não Invertível, Segundo Hillmer e Tiao.	128
4.10. Metodologia de Hillmer e Tiao na Estimação do Modelo AR(p) Multivariado.	132
4.11. Comentários Finais.	135
CAP. 5 - TÉCNICAS DE IDENTIFICAÇÃO EM SÉRIES MULTIVARIADAS.	137
5.1. Introdução.	137
5.2. Metodologia de Tiao e Box	137
5.3. Metodologia de Pierce	148
5.4. Metodologia de Granger e Newbold.	152
5.5. Metodologia de Chan e Wallis.	155
5.6. Determinação da Ordem de um Processo AR Multivariado.	157
CAP. 6 - APLICAÇÕES	160
6.1. Introdução.	160
6.2. Exemplo 1 (Phadke e Kedem, 1978).	160
6.3. Exemplo 2 (Hillmer e Tiao, 1979).	162
6.4. Exemplo 3 (Tiao e Box, 1981).	164
6.5. Exemplo 4 (Pierce, 1980).	172
6.6. Exemplo 5 (Granger e Newbold, 1977)	177
BIBLIOGRAFIA.	183

CAPÍTULO 1

INTRODUÇÃO

1.1 - PRELIMINARES

O campo de estudo de séries temporais multivariadas encontra-se em crescente desenvolvimento em consequência de sua larga aplicação nos mais variados ramos da engenharia, ciências físicas, biológicas, econômicas e sociais.

Na literatura existem muitos livros e monografias abordando séries temporais multivariadas, mas no contexto das aplicações predominam os artigos que retratam a análise de séries temporais univariadas, o que é compreensível pois a metodologia de análises conjuntas de séries temporais é mais recente e acarreta uma carga computacional mais substancial, dificultando, assim, a sua utilização.

A importância do emprego de modelos multivariados de séries temporais reside no fato de podermos extrair simultaneamente informações a respeito do comportamento de cada série no decorrer do tempo e relações de dependência entre as séries. Assim, a contribuição que a evolução em um período de tempo oferece para o entendimento do desenvolvimento de cada série é acrescida da relação conjunta entre duas ou mais séries.

1.2 - CAUSALIDADE E "FEEDBACK"

Convém estabelecermos diferenciações entre as possíveis relações de causalidade que podemos detectar entre pares de séries temporais. Tais relações serão empregadas mais diretamente na seção 5.3.

Segundo Granger & Newbold (1977), podemos relacionar duas variáveis X_t e Y_t em termos de previsibilidade, sendo ambas integradas em um determinado conjunto universo de informações Ω_t , no instante t .

Se $P(X_{t+1}/\Omega_t) = P(X_{t+1}/\Omega_t - Y_t)$, então não há causalidade unidirecional de Y_t para X_t (Y_t não causa X_t).

Haverá causalidade instantânea entre Y_{t+1} e X_{t+1} se

$$P(X_{t+1}/\Omega_{t+1} - X_{t+1}) \neq P(X_{t+1}/\Omega_{t+1} - X_{t+1} - Y_{t+1}).$$

Ocorrerá "feedback" se for verificada causalidade entre Y_t e X_t e também entre X_t e Y_t .

Podemos ver também essas manifestações de causalidade sob o aspecto dos erros quadráticos médios dos previsores a um passo à frente da variável X_t .

$$Y_t \text{ causa } X_t \text{ equivale a } \sigma^2(X_t/\bar{\Omega}) < \sigma^2(X_t/\bar{\Omega} - \bar{Y}_t)$$

$$Y_t \text{ causa } X_t \text{ instantaneamente equivale a } \sigma^2(X_t/\bar{\Omega}, \bar{\bar{Y}}_t) < \sigma^2(X_t/\bar{\Omega})$$

onde $\{\Omega_t, t=0, \pm 1, \dots\}$ é o conjunto de informações disponíveis, incluindo $\{(X_t, Y_t)\}$;

$$\bar{\Omega}_t = \{\Omega_s : s < t\}; \quad \bar{\bar{\Omega}}_t = \{\Omega_s : s \leq t\}.$$

sendo análogas as definições de $\bar{X}_t$, $\bar{Y}_t$, $\bar{\bar{X}}_t$, $\bar{\bar{Y}}_t$.

Intuitivamente, Y_t causa X_t (com respeito a um dado conjunto de informações que inclui também o passado de X_t e Y_t) se o valor presente de X_t pode ser previsto melhor usando o passado de Y_t do que se não o usássemos.

1.3 - MODELOS MULTIVARIADOS

Inicialmente, em cada série temporal envolvida no estudo de modelação multivariada é feita uma apreciação com respeito à sua estacionariedade. Se esta não se verificar, será feita alguma transformação da série em questão a fim de se obter uma série estacionária.

No caso multivariado, os modelos são delineados de forma semelhante ao caso univariado mas recebendo um tratamento matricial.

Consideremos K séries temporais observadas e seja N o número de observações realizadas para cada série. Temos, então, a construção dos seguintes vetores:

$$\begin{aligned} W' &= (W'_1, \dots, W'_N) \\ W'_1 &= (W'_{11}, \dots, W'_{K1}) \\ W'_2 &= (W'_{12}, \dots, W'_{K2}) \\ &\vdots \\ W'_N &= (W'_{1N}, \dots, W'_{KN}) \end{aligned}$$

Assim $W_t = (W_{1t}, \dots, W_{Kt})'$, $t = 1, \dots, N$ representa as séries observadas e seja

$$\omega_t = (\omega_{1t}, \dots, \omega_{Kt})' = T * W_t$$

o correspondente vetor das séries transformadas para se obter estacionariedade.

Podemos, sob certas condições (ver Box & Jenkins, 1970), representar ω_t como um modelo de médias móveis, como segue:

$$\omega_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B) a_t,$$

onde $\{\psi_j\}$ é uma seqüência de matrizes de dimensão $K \times K$; $\{a_t\}$ é uma seqüência de vetores ruído branco, independentes e identicamente distribuídos com média 0 e covariância Σ e

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$$

é um polinômio matricial em B , o qual representa o operador retroativo tal que $B^j a_t = a_{t-j}$. Além disso, $\psi(B)$ é convergente para $|B| \leq 1$.

De modo análogo, sob certas condições, existe também a representação auto-regressiva esboçada por:

$$\sum_{j=0}^{\infty} \pi_j \omega_{t-j} = \pi(B) \omega_t = a_t$$

tal que para $|B| \leq 1$, $\pi^{-1}(B) = \psi(B)$ e se $|\pi(B)| = 0$ então $|B| > 1$.

Na análise conjunta das séries destacam-se alguns modelos e para exemplificação consideremos $K = 2$.

Sejam x_t e y_t séries estacionárias, conforme as transformações:

$$x_t = T_x * X_t \text{ e } y_t = T_y * Y_t.$$

O modelo bivariado pode ser esboçado na forma de médias móveis como:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \psi_{11}(B) & \psi_{12}(B) \\ \psi_{21}(B) & \psi_{22}(B) \end{pmatrix} \begin{pmatrix} a_t \\ b_t \end{pmatrix} \quad (1.1)$$

ou na forma auto-regressiva:

$$\begin{pmatrix} \pi_{11}(B) & \pi_{12}(B) \\ \pi_{21}(B) & \pi_{22}(B) \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a_t \\ b_t \end{pmatrix}. \quad (1.2)$$

As formulações (1.1) e (1.2) retratam a relação de "feed-back" entre as séries na qual uma série depende do presente e do passado da outra série valendo a recíproca.

Se a causalidade for unidirecional no sentido de X_t para Y_t , então $\psi_{12}(B)$ é considerado nulo e (1.1) é simplificado para:

$$\begin{aligned} x_t &= \psi_{11}(B) a_t \\ y_t &= \psi_{21}(B) a_t + \psi_{22}(B) b_t \\ y_t &= V(B) x_t + \psi_{22}(B) b_t \end{aligned} \quad (1.3)$$

onde $V(B) = \psi_{21}(B) \psi_{11}^{-1}(B)$.

A equação (1.3) é conhecida como modelo de regressão dinâmica, no qual a série y_t é delineada em função do presente e passado de x_t e de um termo de erros autocorrelacionados.

1.4 - OBJETIVO DO TRABALHO

O objetivo básico do presente trabalho consiste na des-

crição de diferentes metodologias direcionadas para a estimação dos parâmetros de modelos lineares multivariados. Quanto aos modelos multiplicativos ou sazonais que descrevem o comportamento periódico intra e inter séries, não será efetuado um estudo correspondente devido à escassez de artigos e pesquisas divulgadas dando este enfoque, apesar do importante papel desempenhado pelos modelos sazonais dentro da classe geral de modelos temporais multivariados.

Não abordaremos, neste trabalho, questões relativas à representação em espaço de estados de modelos ARMA (representação Markoviana) e a conseqüente utilização do Filtro de Kalman. Para detalhes ver Priestley (1981), Cap. 9.

O problema da identificação de modelos ARMA multivariados será tratado brevemente no Capítulo 5.

No tocante à parte aplicada a ser exposta no último capítulo, será feita uma apreciação de pesquisas já realizadas e publicadas na literatura, em virtude de ainda não dispormos de um programa para a resolução dos cálculos das expressões deduzidas no estágio de identificação e estimação. Tiao et al. (1979) desenvolveram o "Wisconsin Multiple Time Series Program Package" (WMTS-1), ao qual não foi possível termos acesso no decorrer deste trabalho. Outro programa de interesse na área é o "TIMSAC-78" desenvolvido por Akaike e outros (1979).

A organização deste trabalho é constituída de seis capítulos. O segundo (o próximo) retrata a estimação em modelos lineares univariados. O Capítulo 3 apresenta as técnicas desenvolvidas para a estimação em modelos ARMA multivariados, dos quais os

modelos ARMA univariados incluídos no segundo capítulo são um caso particular. O Capítulo 4 descreve técnicas de estimação dos processos puros multivariados MA e AR, casos particulares do processo misto multivariado ARMA. O Capítulo 5 apresenta diferentes procedimentos no estágio de identificação em séries multivariadas. O Capítulo 6 fornece ilustrações da teoria abordada nos capítulos que o antecedem.

CAPÍTULO 2

ESTIMAÇÃO EM MODELOS LINEARES UNIVARIADOS

2.1 - INTRODUÇÃO

Box & Jenkins (1970) ao derivarem o método de elaboração de um modelo teórico para análise e previsão de séries temporais seguiram as seguintes etapas:

- i) postular uma classe geral de modelos;
- ii) identificar um modelo tentativo;
- iii) estimar os seus parâmetros;
- iv) diagnosticar se o mesmo é adequado ou não e usá-lo para previsão ou controle se o seu diagnóstico for favorável ou tentar novas alternativas na identificação, se o modelo for constatado não adequado.

A etapa referente à estimação será discutida neste capítulo.

Primeiramente, suponhamos que temos uma amostra

$$Z = (Z_1, \dots, Z_N)' \text{ de } Z_t.$$

Esta variável aleatória tem $p(Z/\xi)$ como distribuição de probabilidade conhecida, dependendo de um conjunto de parâmetros desconhecidos ξ .

Ao ser observada Z vamos encarar esta densidade como uma

função de ξ . Assim, vamos verificar os possíveis valores paramétricos que poderiam ter dado origem ao conjunto de observações Z obtido. Temos, então, que Z tornou-se fixado enquanto ξ passou a ser variável.

Denotemos $L(\xi/Z)$ como sendo a função de verossimilhança associada ao conjunto de dados.

Consideremos $N = n+d$ observações de uma determinada série temporal, representadas da forma:

$$Z_{-d+1}, Z_{-d+2}, \dots, Z_0, Z_1, Z_2, \dots, Z_n.$$

Supondo que esta série decorre de um modelo ARIMA de ordem (p, d, q) , obteremos uma nova série $\omega = (\omega_1, \dots, \omega_n)'$ após d diferenças adequadas para tornarem o modelo estacionário, onde $n = N-d$.

Um processo ARMA de ordem (p, q) é definido por:

$$\tilde{\omega}_t - \phi_1 \tilde{\omega}_{t-1} - \dots - \phi_p \tilde{\omega}_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (2.1)$$

$$\tilde{\omega}_t = \nabla^d Z_t = \omega_t - \mu_\omega, \quad E(\tilde{\omega}_t) = \mu_\omega, \quad \xi = (\phi, \theta, \sigma_a^2),$$

onde ∇ é o operador diferença e $\nabla = 1-B$, com $\phi = (\phi_1, \dots, \phi_p)$, $\theta = (\theta_1, \dots, \theta_q)$ e onde $\{a_t\}$ constitui uma sequência de variáveis aleatórias normais e independentes com média zero e variância σ_a^2 . O processo pode ser escrito mais sinteticamente como:

$$\phi(B) \tilde{\omega}_t = \theta(B) a_t,$$

sendo $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ e B é o ope-

rador retroativo $B^k \bar{\omega}_t = \bar{\omega}_{t-K}$.

Para garantirmos estacionariedade e invertibilidade devemos ter que as raízes de $\phi(B) = 0$ e $\theta(B) = 0$ devem estar fora do círculo unitário respectivamente.

Quando $d > 0$ supomos em geral que $\mu_\omega = 0$. Entretanto, se esta suposição for discutível teremos duas alternativas: estimar μ_ω através de $\bar{\omega} = \frac{1}{n} \sum_{t=1}^n \omega_t$ ou considerar μ_ω como um parâmetro adicional a ser estimado juntamente com ϕ , θ e σ_a^2 .

Nosso objetivo básico é a obtenção dos valores de ξ que maximizam $L(\xi/Z)$ ou $\ell(\xi/Z) = \ln L(\xi/Z)$, os quais serão os estimadores de máxima verossimilhança.

Para isso é necessário obtermos a função exata ou aproximada de verossimilhança para o processo aventado.

A seguir apresentaremos a metodologia de obtenção da função de verossimilhança para cada um dos possíveis modelos: AR(p), MA(q) e ARMA(p,q).

2.2 - A FUNÇÃO EXATA DE VEROSSIMILHANÇA PARA UM PROCESSO AR(p) SEGUNDO BOX & JENKINS

Um processo estacionário AR(p) é definido por:

$$\bar{\omega}_t - \phi_1 \bar{\omega}_{t-1} - \dots - \phi_p \bar{\omega}_{t-p} = a_t \quad (2.2)$$

Para garantirmos estacionariedade, as raízes de $\phi(B) = 0$ devem estar fora do círculo unitário.

Vamos supor $\mu_\omega = 0$, apesar de podermos admitir o contrário conforme a argumentação já explanada. Portanto $\bar{\omega}_t = \omega_t - \mu_\omega = \omega_t$.

Não podemos substituir imediatamente $\omega_1, \dots, \omega_n$ em (2.2) para calcularmos os a_t 's devido à dificuldade em inicializarmos a equação de diferenças. Vamos supor, então, que p valores $\omega_1, \dots, \omega_p$ são fornecidos antes do início da série $\omega_1, \dots, \omega_n$. Assim os valores $a_{p+1}, \dots, a_n$ poderão ser calculados de (2.2) condicionais aos valores dados.

Consideremos a_t e conseqüentemente ω_t normalmente distribuídos com a densidade conjunta de ω_t sendo dada por:

$$p(\omega/\phi, \sigma_a) = (2\pi\sigma_a^2)^{-n/2} \cdot |M_n^{(p,0)}|^{1/2} \cdot \exp\left\{-\frac{\omega' \cdot M_n^{(p,0)} \cdot \omega}{2\sigma_a^2}\right\} \quad (2.3)$$

onde $[M_n^{(p,0)}]^{-1} \cdot \sigma_a^2$ é a matriz $n \times n$ de covariância de ω_t para um processo AR(p).

Sabemos que

$$p(\omega/\phi, \sigma_a) = p(\omega_{p+1}, \dots, \omega_n / \omega_1, \dots, \omega_p, \phi, \sigma_a) \cdot p(\omega_1, \dots, \omega_p / \phi, \sigma_a)$$

pois

$$p(\omega/\phi, \sigma_a) = p(\omega_1, \dots, \omega_p, \omega_{p+1}, \dots, \omega_n)$$

Como $a_t \sim N(0, \sigma_a^2)$, teremos a distribuição

$$p(a_{p+1}, \dots, a_n) = (2\pi\sigma_a^2)^{-(n-p)/2} \cdot \exp\left\{-\frac{1}{2\sigma_a^2} \sum_{t=p+1}^n a_t^2\right\} \quad (2.4)$$

Há $n-p$ equações da forma:

$$\begin{aligned} a_{p+1} &= \omega_{p+1} - \phi_1 \omega_p - \dots - \phi_p \omega_1 \\ &\vdots \\ a_n &= \omega_n - \phi_1 \omega_{n-1} - \dots - \phi_p \omega_{n-p} \end{aligned}$$

A transformação acima tem Jacobiano unitário e temos dessa forma a distribuição

$$\begin{aligned} p(\omega_{p+1}, \dots, \omega_n / \omega_1, \dots, \omega_p, \phi, \sigma_a^2) &= (2\pi\sigma_a^2)^{-(n-p)/2} \cdot \\ &\cdot \exp\left\{ -\frac{1}{2\sigma_a^2} \sum_{t=p+1}^n (\omega_t - \dots - \phi_p \omega_{t-p})^2 \right\}. \end{aligned}$$

De (2.3) vem:

$$\begin{aligned} p(\omega_1, \dots, \omega_p / \phi, \sigma_a^2) &= (2\pi\sigma_a^2)^{-p/2} \cdot |M_p^{(p,0)}|^{1/2} \cdot \\ &\cdot \exp\left\{ -\frac{1}{2\sigma_a^2} (\omega_1, \dots, \omega_p)' \cdot M_p^{(p,0)} \cdot (\omega_1, \dots, \omega_p) \right\}. \end{aligned}$$

Portanto, a função exata de verossimilhança será dada por:

$$p(\omega / \phi, \sigma_a^2) = (2\pi\sigma_a^2)^{-n/2} \cdot |M_p^{(p,0)}|^{1/2} \cdot \exp\left\{ -\frac{S(\phi)}{2\sigma_a^2} \right\} \quad (2.5)$$

onde

$$S(\phi) = \sum_{i=1}^p \sum_{j=1}^p m_{ij}^{(p)} \omega_i \omega_j + \sum_{t=p+1}^n (\omega_t - \phi_1 \omega_{t-1} - \dots - \phi_p \omega_{t-p})^2 \quad (2.6)$$

e

$$M_p^{(p,0)} = \{m_{ij}^{(p)}\} = \{\gamma_{|i-j|}\}^{-1} \cdot \sigma_a^2 = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \dots & \gamma_0 \end{bmatrix}^{-1} \cdot \sigma_a^2, \quad (2.7)$$

sendo $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ as autocovariâncias teóricas do processo e $|M_p^{(p,0)}| = |M_n^{(p,0)}|$.

A função $S(\phi)$ é uma forma quadrática tanto em ω_t como nos parâmetros ϕ_i .

Podemos escrever

$$\omega' M_n^{(p)} \omega = \phi_u' D \phi_u$$

onde $\phi_u' = (1, \phi_1, \phi_2, \dots, \phi_p)$

$$D_{(p+1) \times (p+1)} = \begin{bmatrix} D_{11} & -D_{12} & -D_{13} & \dots & -D_{1,p+1} \\ -D_{12} & D_{22} & D_{23} & \dots & D_{2,p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ -D_{1,p+1} & D_{2,p+1} & D_{3,p+1} & \dots & D_{p+1,p+1} \end{bmatrix} \quad (2.8)$$

$$D_{ij} = D_{ji} = \omega_i \omega_j + \omega_{i+1} \omega_{j+1} + \dots + \omega_{n+1-j} \omega_{n+1-i}. \quad (2.9)$$

Portanto,

$$S(\phi) = \phi_u' D \phi_u. \quad (2.10)$$

Nosso objetivo é obter os valores de $\xi = (\phi, \sigma_a^2)$ que maximizam $L(\xi/Z)$ ou $\ln L(\xi/Z)$, os quais serão denominados de estimadores de máxima verossimilhança e onde $L(\xi/Z) = p(\omega/\phi, \sigma_a^2)$.

$$\ln L(\xi/Z) = \ell(\phi, \sigma_a^2/\omega) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_a^2 + \frac{1}{2} \ln |M_p^{(p,0)}| - \frac{S(\phi)}{2\sigma_a^2} \quad (2.11)$$

Vamos derivar a expressão (2.11) em relação a σ_a e a cada ϕ_j , igualando a zero as equações encontradas.

$$\frac{\partial \ell}{\partial \sigma_a} = -\frac{n}{\sigma_a} + \frac{S(\phi)}{\sigma_a^3} \quad (2.12)$$

$$\frac{\partial \ell}{\partial \phi_j} = \frac{\partial \left\{ \frac{1}{2} \ln |M_p^{(p,0)}| \right\}}{\partial \phi_j} + \sigma_a^{-2} \left\{ D_{1,j+1} - \phi_1 D_{2,j+1} - \dots - \phi_p D_{p+1,j+1} \right\} \quad (2.13)$$

$j = 1, 2, \dots, p.$

Temos por (2.12) que:

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi})}{n} \quad (2.14)$$

Para obtermos $\hat{\phi}_j$, $j = 1, 2, \dots, p$, faremos a seguir algumas considerações.

1ª) Na expressão (2.5), o termo $|M_p^{(p,0)}|^{\frac{1}{2}}$ é de magnitude desprezível comparada à de $S(\phi)$. Podemos, então, omiti-lo e, desta maneira, evitar cálculos complicados na sua derivação, chegando à aproximação:

$$\ell(\phi, \sigma_a^2 / \omega) \approx -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_a^2 - \frac{S(\phi)}{2\sigma_a^2} \quad (2.15)$$

A minimização de $S(\phi)$ conduz diretamente a $\hat{\phi}_j$, então estimativas de mínimos quadrados de ϕ_j .

Assim, resolvemos um sistema linear de equações, partindo de (2.13)

$$\left. \begin{aligned} D_{1,2} &= \hat{\phi}_1 D_{2,2} + \hat{\phi}_2 D_{2,3} + \dots + \hat{\phi}_p D_{2,p+1} \\ D_{1,3} &= \hat{\phi}_1 D_{2,3} + \hat{\phi}_2 D_{3,3} + \dots + \hat{\phi}_p D_{3,p+1} \\ D_{1,p+1} &= \hat{\phi}_1 D_{2,p+1} + \hat{\phi}_2 D_{3,p+1} + \dots + \hat{\phi}_p D_{p+1,p+1} \end{aligned} \right\} \quad (2.16)$$

ou seja, $d = D_p \hat{\phi}$ tal que $\hat{\phi} = D_p^{-1} d$.

2ª) Algumas vezes é preferível considerar os valores assintoticamente equivalentes $\hat{\phi}_j$ obtidos através das equações lineares de Yule-Walker, cujos coeficientes são as autocorrelações estimadas. Estas equações decorrem da relação:

$$\gamma_k = E(\omega_t \omega_{t-k}) = E(\phi_1 \omega_{t-1} \omega_{t-k} + \phi_2 \omega_{t-2} \omega_{t-k} + \dots + \phi_p \omega_{t-p} \omega_{t-k} + a_t \omega_{t-k})$$

$$\Rightarrow \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}, \quad k > 0 \quad (2.17)$$

Seja $\rho_k = \frac{\gamma_k}{\gamma_0} = \rho_{-k}$ a correlação teórica

$$\Rightarrow \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k > 0 \quad (2.18)$$

Fazendo k variar de 1 a p , teremos

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

Estimando estas equações, vem:

$$\left. \begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1 + \dots + \hat{\phi}_p r_{p-1} \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2 + \dots + \hat{\phi}_p r_{p-2} \\ &\vdots \\ r_p &= \hat{\phi}_1 r_{p-1} + \hat{\phi}_2 r_{p-2} + \dots + \hat{\phi}_p \end{aligned} \right\} \quad (2.19)$$

onde

$$\hat{\rho}_k = r_k = \frac{c_k}{c_0}, \quad \hat{\gamma}_k = c_k = \frac{1}{n} \sum_{t=1}^{n-k} (\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega}).$$

Em notação matricial, temos: $r = R\hat{\phi} \rightarrow \hat{\phi} = R^{-1} \cdot r$.

2.3 - A FUNÇÃO DE VEROSSIMILHANÇA EXATA PARA

UM PROCESSO MA(q)

2.3.1 - Metodologia de Box & Jenkins

Um processo MA(q) é definido por

$$\tilde{\omega}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (2.20)$$

onde $\tilde{\omega}_t = \omega_t - \mu_\omega$.

Faremos as suposições: $\mu_\omega = 0$ e a distribuição de a_t e, portanto, de ω_t é normal. Vamos considerar, também, que o modelo é invertível, ou seja, as raízes de $\theta(B) = 0$ estão fora do círculo unitário.

A densidade conjunta de ω_t é dada por:

$$p(\omega/\theta, \sigma_a) = (2\pi\sigma_a^2)^{-n/2} \cdot |M_n^{(0,q)}|^{1/2} \cdot \exp\left\{-\frac{\omega' \cdot M_n^{(0,q)} \cdot \omega}{2\sigma_a^2}\right\} \quad (2.21)$$

onde $M_n^{(0,q)}$ é a matriz $n \times n$ de covariância de ω_t para um processo MA(q) e $\omega' = (\omega_1, \dots, \omega_n)$.

Há $n+q$ equações da forma:

$$\begin{aligned} a_{1-q} &= a_{1-q} \\ &\vdots \\ a_{-1} &= a_{-1} \\ a_0 &= a_0 \\ a_1 &= \omega_1 + \theta_1 a_0 + \theta_2 a_{-1} + \dots + \theta_q a_{1-q} \\ &\vdots \\ a_n &= \omega_n + \theta_1 a_{n-1} + \theta_2 a_{n-2} + \dots + \theta_q a_{n-q} \end{aligned} \quad (2.22)$$

Substituindo sucessivamente a_1 em a_2 , a_2 em a_3 , etc. teremos a_t em função dos resíduos iniciais $a_{1-q}, \dots, a_0$ e das observações $\omega_1, \dots, \omega_n$. Sejam $a' = (a_{1-q}, \dots, a_n)$,

$$a'_* = (a_{1-q}, \dots, a_0).$$

Então

$$a = L\omega + Xa'_*, \quad (2.23)$$

sendo L e X matrizes $(n+q) \times n$ e $(n+q) \times q$, respectivamente. Estas matrizes são funções dos elementos de θ , então considerados fixos.

O (i,j) -ésimo elemento de L é dado por

$$\left. \begin{aligned} l_{i,j} &= 0 & i=1, \dots, q & \quad j=1, \dots, n \\ l_{ij} &= 0 & j > i-q & \quad i=q+1, \dots, q+n \\ l_{ij} &= 1 & j=1-q & \quad i=q+1, \dots, q+n \\ l_{ij} &= \theta_1 l_{i-1,j} + \theta_2 l_{i-2,j} + \dots + \theta_q l_{i-q,j} & \begin{cases} j < i-q \\ i=q+1, \dots, q+n \end{cases} \end{aligned} \right\} \quad (2.24)$$

O (i,j) -ésimo elemento de X é da forma

$$\left. \begin{aligned} x_{ii} &= 1 & i=1, \dots, q \\ x_{ij} &= 0 & i \neq j & \quad i, j=1, \dots, q \\ x_{ij} &= \theta_1 x_{i-1,j} + \theta_2 x_{i-2,j} + \dots + \theta_q x_{i-q,j} & \begin{cases} i=q+1, \dots, q+n \\ j=1, \dots, q \end{cases} \end{aligned} \right\}$$

Além disso $l_{i1} = x_{i-1,q}$, $i=2, \dots, n+q$

$$l_{11} = 0.$$

Assim os elementos não nulos da j -ésima coluna de L são obti-

dos da última coluna de X :

$$x_{ij} = x_{i-j,q}, \quad j \leq i-q, \quad i = q+1, \dots, q+n. \quad (2.25)$$

Como $a_t \sim N(0, \sigma_a^2)$, teremos a distribuição

$$p(a_{1-q}, \dots, a_n) = (2\pi\sigma_a^2)^{-(n+q)/2} \cdot \exp\left\{-\frac{1}{2\sigma_a^2} \sum_{t=1-q}^n a_t^2\right\} \quad (2.26)$$

A transformação (2.23) tem Jacobiano unitário e nos conduz à distribuição conjunta:

$$p(\omega, a_*/\theta, \sigma_a) = (2\pi\sigma_a^2)^{-(n+q)/2} \cdot \exp\left\{-\frac{1}{2\sigma_a^2} S(\theta, a_*)\right\}$$

onde

$$S(\theta, a_*) = \sum_{t=1-q}^n a_t^2 = a'a = (L\omega + Xa_*)'(L\omega + Xa_*)$$

Aplicando mínimos quadrados em (2.23), temos a estimativa de máxima verossimilhança $\hat{a}_*$ de a_* , dada por:

$$\hat{a}_* = -(X'X)^{-1} \cdot X' \cdot L \cdot \omega. \quad (2.27)$$

Portanto, $\hat{a}_*$ é o vetor que minimiza $S(\theta, a_*)$.

$S(\theta, a_*)$ pode ser decomposta em duas parcelas, a saber:

$$\begin{aligned} S(\theta, a_*) &= (L\omega + Xa_*)' \cdot (L\omega + Xa_*) \\ &= \underbrace{[L\omega + X\hat{a}_*]}'_A \underbrace{[X(a_* - \hat{a}_*)]}_B \cdot \underbrace{[L\omega + X\hat{a}_*]}_A \underbrace{[X(a_* - \hat{a}_*)]}_B \\ &= (A' + B') \cdot (A + B) \\ &= A'A + B'B \end{aligned}$$

$$S(\theta, a_*) = \underbrace{(L\omega + X\hat{a}_*)' \cdot (L\omega + X\hat{a}_*)}_{S(\theta)} + (a_* - \hat{a}_*)' \cdot X' \cdot X (a_* - \hat{a}_*)$$

Os produtos $A'B$ e $B'A$ são nulos pois A e B são ortogonais.

De fato,

$$B'A = (a_* - \hat{a}_*)' \cdot X' \cdot L\omega + (a_* - \hat{a}_*)' \cdot X' \cdot X\hat{a}_*$$

De (2.27) vem:

$$B'A = (a_* - \hat{a}_*)' \cdot X' \cdot L\omega - (a_* - \hat{a}_*)' \cdot X' \cdot L\omega = 0,$$

o mesmo para $A'B$.

$S(\theta)$ é função das observações ω mas não dos resíduos iniciais a_* , o mesmo acontecendo com $\hat{a}_*$.

Sabemos que

$$p(\omega, a_*/\theta, \sigma_a) = p(\omega/\theta, \sigma_a) \cdot p(a_*/\omega, \theta, \sigma_a).$$

Portanto

$$p(a_*/\omega, \theta, \sigma_a) = (2\pi\sigma_a^2)^{-q/2} \cdot |X'X|^{1/2} \cdot \exp\left\{-\frac{(a_* - \hat{a}_*)' \cdot X'X \cdot (a_* - \hat{a}_*)}{2\sigma_a^2}\right\} \quad (2.28)$$

e a função de verossimilhança incondicional exata é dada por

$$p(\omega/\theta, \sigma_a) = (2\pi\sigma_a^2)^{-n/2} \cdot |X'X|^{-1/2} \cdot \exp\left\{-\frac{S(\theta)}{2\sigma_a^2}\right\} \quad (2.29)$$

De (2.21), temos

$$|X'X|^{-1} = |M_n^{(0,q)}|, \quad S(\theta) = \omega' \cdot M_n^{(0,q)} \cdot \omega$$

Temos, ainda:

$$\hat{a}_* = [a_*/\omega, \theta] = [a_*],$$

onde $[a_*/\omega, \theta] = E(a_*/\omega, \theta)$, $[a] = L\omega + X[a_*]$

$$S(\theta) = \sum_{t=1-q}^n [a_t]^2 \quad (2.30)$$

Nosso objetivo é obter os valores de $\xi = (\theta, \sigma_a^2)$ que maximizam $L(\xi/Z) = p(\omega/\theta, \sigma_a)$ ou, de modo equivalente, $\ln L(\xi/Z) = \ell(\theta, \sigma_a/\omega)$.

$$\begin{aligned} \ell(\theta, \sigma_a/\omega) &= \ell[p(\omega/\theta, \sigma_a)] = \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_a^2 - \frac{1}{2} \ln |X'X| - \frac{S(\theta)}{2\sigma_a^2} \end{aligned} \quad (2.31)$$

Derivando (2.31) em relação a σ_a^2 e igualando a zero, obtemos o estimador de máxima verossimilhança de σ_a^2 :

$$\hat{\sigma}_a^2 = \frac{S(\hat{\theta})}{n} \quad (2.32)$$

Desta maneira, o logaritmo concentrado da função de verossimilhança é da forma:

$$\ell(\theta/\omega) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \frac{S(\theta)}{n} - \frac{1}{2} \ln |X'X| - \frac{n}{2} \quad (2.33)$$

O termo $|X'X|$ é desprezível comparado com $S(\theta)$, exceto quando n é pequeno. Podemos assim omiti-lo e a maximização de $\ell(\theta/\omega)$ é equivalente à minimização de $S(\theta)$.

O cálculo de $S(\theta)$ em (2.30) depende dos valores de

$$[a_t] = \underbrace{([a_{1-q}], \dots, [a_0], [a_1], \dots, [a_n])}_{[a_*]}'$$

Na estimação dos resíduos iniciais há diferentes metodo-

logias. Åström & Bohlin (1966) consideram $\hat{a}_* = 0$; Phillips, cujo método é detalhado por Trivedi (1970), estima as componentes de $\hat{a}_*$ como parâmetros de perturbação.

Para obter $[a_*] = \hat{a}_*$ podemos aplicar a técnica de "previsão retroativa" proposta por Box & Jenkins a ser descrita na seção 2.4.1 ou, então, utilizar a relação (2.27) como segue:

- i) a construção de X , seguindo as etapas descritas em (2.24) possibilita o cálculo de $\hat{a}_*$, por (2.27) e, ainda, (2.25);
- ii) os termos $[a_1], \dots, [a_n]$ podem ser calculados recursivamente, tomando esperanças condicionais em (2.22), pois:

$$[a_t] = [\omega_t] + \theta_1[a_{t-1}] + \dots + \theta_q[a_{t-q}]$$

$$[a_t] = \omega_t + \theta_1[a_{t-1}] + \dots + \theta_q[a_{t-q}]$$

A soma dos quadrados $S(\theta)$ é, então, calculada, assim como $\ell(\theta/\omega)$.

Embora o processo de médias móveis seja largamente empregado em análise de séries temporais, esforços na estimação de seus parâmetros têm sido infrutíferos, devido à impraticabilidade na resolução das equações de máxima verossimilhança.

Pesaran (1973) mostrou que no processo MA(1) é possível obter a maximização de (2.31) pois a matriz $X'X$, através de transformação ortogonal, torna-se uma matriz diagonal. Porém, esta propriedade não se verifica para processos de ordem superior.

As estimativas de máxima verossimilhança são próximas das estimativas de mínimos quadrados que minimizam a expressão:

$$S(\theta) = \sum_{t=1-q}^n [a_t]^2$$

Se $[a_t]$ é função linear de $\beta = (\theta)$ o cálculo da minimização é simplificado.

Como já vimos, no caso de um processo AR(p), obtemos um sistema linear de equações pois:

$$[a_t] = \phi(B)[\omega_t] = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)[\omega_t]$$

$$\frac{\partial}{\partial \phi_i} [a_t] = \underbrace{\phi(B) \frac{\partial}{\partial \phi_i} [\omega_t] - [\omega_{t-i}]}_{\text{função linear de } \phi_i}$$

Seja, agora, um processo MA(q):

$$[a_t] = \theta^{-1}(B)[\omega_t]$$

$$\frac{\partial}{\partial \theta_j} [a_t] = \underbrace{\theta^{-1}(B) \frac{\partial}{\partial \theta_j} [\omega_t] - [\omega_{t-j}] \theta^{-2}(B)}_{\text{função não linear em } \theta_j}$$

É possível linearizar o modelo através da expansão de $[a_t]$ em uma série de Taylor em torno de $[a_{t,0}]$ correspondente a um conjunto inicial de parâmetros $\beta'_0 = (\beta_{1,0}, \beta_{2,0}, \dots, \beta_{k,0})$.

Temos, assim:

$$[a_t] = [a_{t,0}] - \sum_{i=1}^k (\beta_i - \beta_{i,0}) x_{i,t} \quad (2.34)$$

sendo

$$[a_{t,0}] = [a_t/\omega, \beta_0] \quad \text{e} \quad x_{i,t} = - \left. \frac{\partial [a_t]}{\partial \beta_i} \right|_{\beta=\beta_0}$$

$$k = p+q, \quad t = -q+1, \dots, n.$$

Podemos expressar as $(n+q)$ equações em (2.34) na forma:

$$\begin{array}{ccccc} [a_0] & = & X(\beta-\beta_0) & + & [a] \\ (n+q) \times 1 & & \downarrow & & k \times 1 \quad (n+q) \times 1 \\ & & (n+q) \times k & & \end{array}$$

onde X é uma matriz $\{x_{i,t}\}_{(n+q) \times k}$.

Podemos obter os ajustes $(\beta-\beta_0)$ que minimizam $S(\theta) = [a'] [a]$ através de mínimos quadrados lineares, isto é, por regressão de $[a_0]$ sobre X .

Entretanto, um único ajuste não fornecerá de imediato valores de quadrados mínimos, uma vez que os $[a_t]$ não são lineares em θ . É preciso, então, usar os valores ajustados como novos palpites e repetir o processo sucessivamente até que haja convergência.

As derivadas $x_{i,t}$ podem ser obtidas numericamente ou através de um cálculo direto.

a) estimativas numéricas das derivadas

Para um determinado modelo, os valores

$$[a_t/\omega, \beta_{1,0}, \beta_{2,0}, \dots, \beta_{k,0}], \quad t = 1-q, \dots, n$$

são calculados recursivamente.

Repete-se o cálculo para $[a_t/\omega, \beta_{1,0} + \delta_1, \dots, \beta_{k,0}]$, depois para $[a_t/\omega, \beta_{1,0}, \beta_{2,0} + \delta_2, \dots, \beta_{k,0}]$, etc.

$$x_{i,t} = \frac{1}{\delta_i} \left\{ [a_t/\omega, \beta_{1,0}, \dots, \beta_{i,0}, \dots, \beta_{k,0}] \cdot - \right. \\ \left. - [a_t/\omega, \beta_{1,0}, \dots, \beta_{i,0} + \delta_i, \dots, \beta_{k,0}] \right\}. \quad (2.35)$$

O valor $X = X_{\hat{\beta}}$ obtido na iteração final pode ser usado no cálculo de variâncias, covariâncias e intervalos de confiança aproximados. Desta maneira, a matriz aproximada de variâncias e covariâncias dos $\hat{\beta}$'s será dada por $(X'_{\hat{\beta}} X_{\hat{\beta}})^{-1} \cdot \sigma_a^2$ e $\hat{\sigma}_a^2 = \frac{S(\hat{\beta})}{n}$.

Para obtermos um primeiro ajuste para β usamos:

$$\beta - \beta_0 = \frac{\sum_{t=0}^n [a_{t,0}] x_t}{\sum_{t=0}^n x_t^2}$$

b) cálculo direto das derivadas

Consideremos, para ilustração, o processo MA(1)

$$\omega_t = \begin{cases} (1-\theta F) e_t \\ (1-\theta B) a_t \end{cases} \Rightarrow \begin{cases} e_t = \omega_t + \theta e_{t+1} \\ a_t = \omega_t + \theta a_{t-1} \end{cases}$$

As duas formulações podem ser usadas alternadamente, uma fornecendo valores iniciais para se começar uma recursão sobre a outra.

Seja

$$a_t^{(\theta)} = \frac{\partial [a_t]}{\partial \theta}, \quad \omega_t^{(\theta)} = \frac{\partial [\omega_t]}{\partial \theta}, \quad e_t^{(\theta)} = \frac{\partial [e_t]}{\partial \theta}$$

Teremos as equações:

$$\begin{cases} e_t^{(\theta)} = \omega_t^{(\theta)} + \theta e_{t+1}^{(\theta)} + [e_{t+1}] \end{cases} \quad (2.36)$$

$$\begin{cases} a_t^{(\theta)} = \omega_t^{(\theta)} + \theta a_{t-1}^{(\theta)} + [a_{t-1}] \end{cases} \quad (2.37)$$

sendo que:

$$\left. \begin{aligned} [\omega_t] = \omega_t, \omega_t^{(\theta)} &= 0 \quad t = 1, \dots, n \\ [e_{-j}] &= 0, \quad j = 0, 1, \dots, n \end{aligned} \right\} \quad (2.38)$$

Considerando $e_{n+1}^{(\theta)} = 0$ em (2.36) iniciamos uma recursão retroativa e o uso de (2.38) torna possível obtermos $\omega_{-j}^{(\theta)}$, $j = 0, 1, \dots, q-1$.

Usando o fato de que $a_{-q}^{(\theta)}$, $a_{-q-1}^{(\theta)}$, ... podem ser supostos nulos, utilizamos (2.37) para obter recursivamente as derivadas $a_t^{(\theta)}$.

2.3.2 - Metodologia de Durbin

Durbin (1959) foi um dos primeiros a contribuir para o estudo da estimação de parâmetros de um processo $MA(q)$. A partir de suas pesquisas foram tentados caminhos alternativos para a maximização da função de verossimilhança devido à impossibilidade de uma otimização não linear na época.

Walker (1961) também se preocupou em procurar outras técnicas para a estimação de parâmetros de um processo $MA(q)$.

Surgiram dificuldades mesmo em modelos mais simples como o processo $MA(1)$:

$$\omega_t = a_t - \theta a_{t-1}.$$

Como vimos na seção anterior, a derivação de $[a_t]$ em função do parâmetro θ dá origem a uma função não linear em θ .

Outra alternativa é usar o método dos momentos e igualar

o coeficiente teórico de autocorrelação, $\rho_1 = \frac{-\theta}{1+\theta^2}$, ao correspondente valor amostral $r_1 = \frac{c_1}{c_0}$, onde

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega})$$

mas esta tentativa não foi bem sucedida devido à ineficiência de tal estimador.

De fato, no caso em que $\theta = -\frac{1}{2}$, Whittle (1953) mostrou que a variância do estimador sugerido tem um fator multiplicativo igual a 3,8 em relação à estimativa de máxima verossimilhança.

O método de Durbin, a seguir, fundamenta-se na representação do processo de médias móveis em termos de um esquema auto-regressivo de ordem alta que fornece meios de obtenção das estimativas dos parâmetros de médias móveis.

Destacam-se também Wold (1938) e Whittle (1951, 1953) por trabalhos anteriormente prestados com relação ao mesmo problema.

Inicialmente, consideremos o processo MA(1):

$$\begin{aligned}\omega_t &= a_t - \theta a_{t-1} = (1 - \theta B) a_t \\ a_t &= (1 - \theta B)^{-1} \omega_t = \frac{1}{1 - \theta B} \omega_t = \left\{ \sum_{j=0}^{\infty} \theta^j B^j \right\} \omega_t \\ a_t &= \omega_t + \theta B \omega_t + \theta^2 B^2 \omega_t + \theta^3 B^3 \omega_t + \dots \\ a_t &= \omega_t + \theta \omega_{t-1} + \theta^2 \omega_{t-2} + \theta^3 \omega_{t-3} + \dots\end{aligned}\tag{2.39}$$

A expressão (2.39) é a representação auto-regressiva de ordem infinita de um MA(1).

$$a_t = \omega_t + \theta \omega_{t-1} + \theta^2 \omega_{t-2} + \dots + \theta^k \omega_{t-k} + \theta^{k+1} \omega_{t-k-1} + \theta^{k+2} \omega_{t-k-2} + \dots$$

$$a_t = \omega_t + \theta(\omega_{t-1} + \theta\omega_{t-2} + \dots + \theta^{k-1}\omega_{t-k}) + \theta^{k+1}(\omega_{t-k-1} + \theta\omega_{t-k-2} + \dots)$$

$$a_t = \omega_t + \theta \sum_{j=1}^k \theta^{j-1} \omega_{t-j} + \theta^{k+1} \cdot a_{t-k-1}.$$

O termo $\theta^{k+1} \cdot a_{t-k-1}$ cuja variância é $\theta^{2k+2} \cdot \sigma_a^2$ tende a zero à medida que o valor de k aumenta em consequência do fato: $|\theta| < 1$.

Então podemos omitir as parcelas de (2.39) a partir da $(k+1)$ -ésima e teremos uma representação auto-regressiva finita de ordem k :

$$a_t = \omega_t + \theta \sum_{j=1}^k \theta^{j-1} \omega_{t-j} \quad (2.40)$$

cuja precisão é condicionada à nossa escolha do valor de k com magnitude tão grande quanto desejarmos. Entretanto seu valor será sempre pequeno ao ser comparado com n , o tamanho da amostra.

Em (2.40) seja $\alpha_i = \theta^i$ para evitarmos uma notação envolvendo potências.

Pela minimização de

$$\sum_{t=k+1}^n (\omega_t + \theta\omega_{t-1} + \dots + \theta^k \omega_{t-k})^2 = \sum_{t=k+1}^n (\omega_t + \alpha_1 \omega_{t-1} + \dots + \alpha_k \omega_{t-k})^2$$

vamos obter os estimadores de mínimos quadrados $\hat{\alpha}_1, \dots, \hat{\alpha}_k$.

Mann & Wald (1943) provaram que $\hat{\alpha}_1, \dots, \hat{\alpha}_k$ são assintoticamente normais com médias $\alpha_1, \dots, \alpha_k$ e a matriz de covariância V_k^{-1}/n , onde $\sigma_a^2 \cdot V_k$ é a matriz de covariância de $\omega_{t-1}, \dots, \omega_{t-k}$.

$$V_k = \begin{bmatrix} 1+\theta^2 & -\theta & 0 & \dots & 0 \\ -\theta & 1+\theta^2 & -\theta & \dots & \cdot \\ 0 & -\theta & 1+\theta^2 & \dots & \cdot \\ \vdots & & & \ddots & -\theta \\ 0 & \dots\dots\dots & -\theta & 1+\theta^2 \end{bmatrix}$$

Os coeficientes estimados $\hat{\alpha}_1, \dots, \hat{\alpha}_k$ têm como distribuição assintótica

$$p(\hat{\alpha}_1, \dots, \hat{\alpha}_k) = \frac{n^{1/2} \cdot |V_k|^{1/2}}{(2\pi)^{k/2}} \cdot \exp \left\{ -\frac{n}{2} \left[(1+\theta^2) \sum_{i=1}^k (\hat{\alpha}_i - \alpha_i)^2 - 2\theta \sum_{i=1}^{k-1} (\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_{i+1} - \alpha_{i+1}) \right] \right\} \quad (2.41)$$

Os coeficientes de auto-regressão $\alpha_1, \dots, \alpha_k$ satisfazem o sistema de equações:

$$\begin{aligned} c_1 &= \alpha_1 c_0 + \alpha_2 c_1 + \dots + \alpha_k c_{k-1} \\ c_2 &= \alpha_1 c_1 + \alpha_2 c_0 + \dots + \alpha_k c_{k-2} \\ &\vdots \\ c_k &= \alpha_1 c_{k-1} + \dots + \alpha_k c_0 \end{aligned}$$

onde c_r são as autocovariâncias amostrais.

$$c_0 = (1+\theta^2)\sigma_a^2, \quad c_1 = -\theta\sigma_a^2, \quad c_r = 0 \text{ para } r > 1.$$

Teremos, então:

$$-\theta = (1+\theta^2)\alpha_1 - \theta\alpha_2$$

$$0 = -\theta\alpha_{r-1} + (1+\theta^2)\alpha_r - \theta\alpha_{r+1}, \quad (r=2, \dots, k-1)$$

$$r=k \Rightarrow 0 = -\theta\alpha_{k-1} + (1+\theta^2)\alpha_k$$

Multiplicando estas equações por $-2\hat{\alpha}_i + \alpha_i$ ($i=1, \dots, k$), obtemos após algumas manipulações algébricas uma expressão quadrática, digamos Q , expoente da base exponencial em (2.41):

$$\begin{aligned} Q &= (1+\theta^2) \sum_{i=1}^k (\hat{\alpha}_i - \alpha_i)^2 - 2\theta \sum_{i=1}^{k-1} (\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_{i+1} - \alpha_{i+1}) \\ &= (1+\theta^2) \sum_{i=1}^k \hat{\alpha}_i^2 - 2\theta \sum_{i=1}^{k-1} \hat{\alpha}_i \hat{\alpha}_{i+1} - 2\theta \hat{\alpha}_1 + \theta \alpha_1 \end{aligned}$$

Para um valor alto de k , α_1 igual a θ e considerando $\hat{\alpha}_0 = 1$, temos:

$$Q = (1+\theta^2) \sum_{i=0}^k \hat{\alpha}_i^2 - 2\theta \sum_{i=0}^{k-1} \hat{\alpha}_i \hat{\alpha}_{i+1} - 1$$

Em (2.41), $|V_k| = (1-\theta^{2k+2})/(1-\theta^2) \approx 1/(1-\theta^2)$ quando k é suficientemente grande. Mesmo assim, a expressão $1/(1-\theta^2)$ é desprezível em relação a Q e, desta forma, vamos considerar a função de verossimilhança em função de Q basicamente. Então, a maximização da função de verossimilhança equivale à minimização de Q . Derivando Q em relação a θ e igualando a zero o resultado, teremos:

$$\hat{\theta} = \frac{\sum_{i=0}^{k-1} \hat{\alpha}_i \hat{\alpha}_{i+1}}{\sum_{i=0}^k \hat{\alpha}_i^2} \quad (2.42)$$

cuja eficiência é tão próxima da unidade quanto quisermos, à medida que formos tomando valores cada vez mais altos para k .

No caso mais geral:

$$\omega_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad t=1, \dots, n,$$

a mesma aproximação auto-regressiva fornece os estimadores $\hat{\theta}_1, \dots, \hat{\theta}_q$ obtidos a partir das equações lineares

$$\sum_{j=1}^q A_{|r-j|} \hat{\theta}_j = A_r, \quad r=1, \dots, q$$

onde

$$A_r = \sum_{i=0}^{k-r} \hat{\alpha}_i \hat{\alpha}_{i+r}.$$

Ver Durbin (1959).

2.3.3 - Considerações sobre as condições impostas sobre os parâmetros de um processo MA(q)

Costuma-se impor a condição de invertibilidade para um processo MA(q) que restringe os parâmetros a satisfazerem certas relações obtidas com a condição de que $\theta(B) = 0$ tenha raízes fora do círculo unitário. De fato, o cálculo das estimativas de máxima verossimilhança que minimizam (2.33) com respeito a θ não pode ser considerado um problema irrestrito de otimização.

Isto decorre do fato de que quando não impomos nenhuma condição aos elementos de θ pode haver 2^q pontos nos quais a função de verossimilhança assume seu valor mínimo, ou seja, há um problema de identificação.

Podem também ocorrer diferenças apreciáveis entre as estimativas obtidas pelo procedimento não linear de mínimos quadrados (Box & Jenkins, 1970) e as estimativas exatas de máxi-

ma verossimilhança se os parâmetros de médias móveis estiverem próximos ou na fronteira da região de invertibilidade.

Kang (1975) ao examinar as propriedades da função da soma de quadrados em (2.33) verificou que a mesma é decrescente quando a fronteira da região de invertibilidade é atravessada, sendo que o mínimo seria obtido em $|\alpha_i| = 1$ e a função tenderia ao seu mínimo valor possível, zero, quando $|\alpha_i|$ tendesse a infinito.

Esse resultado é indesejável pois além do modelo não ser invertível ainda não haveria unicidade na parametrização já que seria alcançado pela função de verossimilhança um mesmo valor ao ser aplicada às raízes α_i^{-1} e às suas inversas α_i .

Os termos α_i e α_i^{-1} são provenientes da representação do polinômio de médias móveis como segue:

$$\theta(B) = \prod_{i=1}^q (1 - \alpha_i B), \quad (2.43)$$

cujas raízes são dadas por α_i^{-1} . Invertibilidade Implica $|\alpha_i| < 1, \forall i$.

Embora este problema pareça contornável se atribuirmos zero a todos os coeficientes iniciais durante a estimação, geralmente quando uma ou mais raízes estão mesmo que moderadamente próximas da fronteira de admissibilidade, há risco da função perder o seu mínimo dentro da região aceitável, segundo Osborn (1976).

A técnica sugerida por Box & Jenkins (1970) de se estimar os valores iniciais dos coeficientes pelas autocorrela-

ções amostrais também não garante que as raízes permanecerão dentro da região de invertibilidade.

O mais seguro, de acordo com Osborn (1976) é usarmos como valores iniciais estimativas preliminares dentro da região de invertibilidade obtidas no estágio de identificação, testar as raízes no fim de cada iteração e alterar o valor de alguma raiz que tenha ultrapassado a fronteira desta região para um outro valor dentro da mesma. Nicholls (1972) é adepto deste procedimento.

Outra alternativa é adotar o método de assegurar invertibilidade proposto por Osborn (1976).

Para garantirmos invertibilidade vamos considerar os α_i divididos em pares que podem ser complexos conjugados ou ambos reais, sendo os componentes de cada par as raízes da expressão quadrática

$$\lambda^2 + a_i \lambda + b_i = 0 \quad (2.44)$$

Se $|b_i| < 1$ e $|a_i| < 1 + b_i$, ambas as raízes de (2.44) estarão dentro do círculo unitário e para isso basta definirmos a_i e b_i como:

$$b_i = \frac{\delta_i}{1 + |\delta_i|}, \quad a_i = \frac{\gamma_i(1 + b_i)}{1 + |\gamma_i|} \quad (2.45)$$

onde δ_i e γ_i são irrestritos.

Com isso, teremos as raízes de (2.44) satisfazendo a condição de invertibilidade.

Se o processo MA(q) for de ordem ímpar, restará uma raiz forçosamente real após serem tomados os pares. Esta raiz po-

de ser especificada como

$$d = - \frac{\beta}{1+|\beta|}$$

Usamos o sinal negativo devido ao fato de que d pode ser interpretada como a raiz não nula de (2.44) quando $b_i=0$, pois

$$a_i = \frac{\gamma_i(1+0)}{1+|\gamma_i|} = \frac{\gamma_i}{1+|\gamma_i|}$$

Em (2.44), temos:

$$\lambda^2 + \frac{\gamma_i}{1+|\gamma_i|} \cdot \lambda = 0$$

cujas possíveis raízes são:

$$\lambda_1 = 0$$

$$\lambda_2 = \frac{-\gamma_i}{1+|\gamma_i|}$$

Com esta parametrização estimamos as raízes α_i restritas a pertencerem ao círculo unitário e obtemos os coeficientes de médias móveis a partir destas estimativas, utilizando a relação (2.43). Desta maneira é possível o cálculo da função em (2.33).

Se formos estimar um processo MA de ordem maior do que um, podemos iniciar a estimação na ordem proposta atribuindo zero aos valores paramétricos iniciais. Por outro lado, também podemos aplicar o método "step-wise", começando com um processo MA(1) cuja ordem será acrescida de um a cada passo. Fa-

zendo uma correspondência dos parâmetros irrestritos como componentes de pares como (γ_i, δ_i) , sendo β o parâmetro restante se q for ímpar, usaremos as estimativas do passo anterior como valores iniciais para os primeiros q parâmetros quando, então, a ordem passará de q para $(q+1)$. Atribuindo zero ao valor inicial do parâmetro adicional, os primeiros q elementos de θ são as estimativas dos coeficientes do último passo e o $(q+1)$ -ésimo coeficiente é zero inicialmente. Se q é ímpar, o valor estimado de β passa a ser o valor inicial do último γ_i para o processo $(q+1)$.

Entretanto, algumas raízes podem arbitrariamente se aproximar do círculo unitário se os parâmetros γ_i , δ_i e β são irrestritos. Convém, então, especificar algum valor absoluto máximo que estes parâmetros possam eventualmente assumir, ou testar convergência em função das raízes do processo a qual é condicionada à essa reparametrização.

2.4 - A FUNÇÃO DE VEROSSIMILHANÇA EXATA PARA UM PROCESSO ARMA(p,q)

2.4.1 - Metodologia de Box & Jenkins e Newbold

Como já vimos na seção 2.1, um processo ARMA(p,q) expressa-se na forma

$$\tilde{\omega}_t = \phi_1 \tilde{\omega}_{t-1} + \dots + \phi_p \tilde{\omega}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.46)$$

Vamos considerar para simplificação $\mu_\omega = 0$ e, portanto,
 $\tilde{\omega}_t = \omega_t - \mu_\omega = \omega_t$.

Box & Jenkins (1970), ao considerarem o problema de es-

timação para o processo misto ARMA, fazem uma extensão a partir do resultado encontrado para modelos MA, escrevendo um processo ARMA como um MA de ordem infinita, isto é,

$$\omega_t = (1 + \psi_1 B + \psi_2 B^2 + \dots) a_t, \quad (2.47)$$

ou

$$\omega_t = \psi(B) a_t \text{ onde } \psi(B) = \frac{\theta(B)}{\phi(B)}$$

Fazem, a seguir, uma aproximação de (2.47) para um processo MA finito de ordem Q: $\omega_t = (1 + \psi_1 B + \psi_2 B^2 + \dots + \psi_Q B^Q) a_t$ ou seja,

$$\omega_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \approx \sum_{j=0}^Q \psi_j a_{t-j}.$$

A função de verossimilhança é expressada como

$$L(\xi/Z) = L(\phi, \theta, \sigma_a^2/Z) = (2\pi\sigma_a^2)^{-n/2} \cdot |M_n^{(p,q)}|^{1/2} \cdot \exp\left\{-\frac{S(\phi, \theta)}{2\sigma_a^2}\right\} \quad (2.48)$$

onde

$$S(\phi, \theta) = \sum_{t=-\infty}^n [a_t/Z, \phi, \theta]^2$$

Box & Jenkins propõem que na prática $[a_t/Z, \phi, \theta]$ sejam calculados recursivamente a partir de algum ponto $t = 1-Q$, Q assumindo valor suficientemente alto sendo desprezíveis os valores de $[a_t]$ anteriores a este ponto.

Tal estratégia é bem sucedida em geral. Contudo, há três situações que poderiam ocorrer, individual ou simultaneamente, e que gerariam problemas. São elas:

- 1) as raízes do polinômio $1 - \phi_1 B - \dots - \phi_p B^p = 0$ podem estar próximas da fronteira do círculo unitário;
- 2) as raízes de $1 - \theta_1 B - \dots - \theta_q B^q = 0$ podem estar no círculo unitário ou próximas dele;
- 3) a série de dados pode ser relativamente pequena.

Em tais situações Newbold (1974) sugere derivar diretamente a função de verossimilhança exata para um processo ARMA fazendo uma analogia com a técnica empregada por Box e Jenkins no caso do processo MA.

No caso mais simples, ARMA (1,1), Tiao & Ali (1971) conseguiram uma determinação exata dos elementos componentes da inversa da matriz de covariâncias amostrais bem como de seu determinante.

Entretanto no caso geral, ARMA(p,q), Shaman (1973) determinou aproximadamente os elementos da inversa da matriz de covariâncias amostrais mas não apresentou o determinante da mesma, o qual requer uma grande carga de cálculos computacionais. Recentemente, Dent (1977), Ali (1977) e Ansley (1979) detiveram-se no estudo da função de verossimilhança para o modelo geral. O método de Ansley destaca-se como um dos mais eficientes dentre os demais divulgados.

Vejamos o procedimento de Newbold para o caso geral. Usando o modelo (2.46) podemos escrever as equações:

$$\begin{aligned} a_{1-q} &= a_{1-q} \\ &\vdots \\ a_{-1} &= a_{-1} \end{aligned}$$

$$a_0 = a_0$$

$$\omega_{1-p} = \omega_{1-p}$$

$$\vdots$$

$$\omega_0 = \omega_0$$

$$a_1 = \omega_1 - \phi_1 \omega_0 - \dots - \phi_p \omega_{1-p} + \theta_1 a_0 + \dots + \theta_q a_{1-q}$$

$$\vdots$$

$$a_n = \omega_n - \phi_1 \omega_{n-1} - \dots - \phi_p \omega_{n-p} + \theta_1 a_{n-1} + \dots + \theta_q a_{n-q}.$$

Seja

$$e' = (e_{1-p-q}, \dots, e_{-1}, e_0, e_1, \dots, e_n)$$

$$e' = (e'_*, e'_n)$$

onde $e'_* = (e_{1-p-q}, \dots, e_0)$; $e'_n = (e_1, \dots, e_n)$ tal que:

$$e_{1-p-j} = a_{1-j} \quad (j=1, \dots, q) \Rightarrow \begin{cases} e_{1-p-q} = a_{1-q} \\ \vdots \\ e_{-p} = a_0 \end{cases}$$

$$e_{1-j} = \omega_{1-j} \quad (j=1, \dots, p) \Rightarrow \begin{cases} e_{-p+1} = \omega_{-p+1} \\ \vdots \\ e_0 = \omega_0 \end{cases}$$

$$e_t = a_t \quad (t=1, \dots, n) \Rightarrow \begin{cases} e_1 = a_1 \\ \vdots \\ e_n = a_n \end{cases}$$

Seja $\omega' = (\omega_1, \dots, \omega_n)$.

Assim

$$e = L\omega + Xe_* \quad (2.49)$$

sendo L e X matrizes $(n+p+q) \times n$ e $(n+p+q) \times (p+q)$ respectivamente. Estas matrizes são funções apenas dos coeficientes $\phi_1, \dots, \phi_p$ e $\theta_1, \dots, \theta_q$.

A equação (2.49) equivale à expressão

$$\begin{pmatrix} e_* \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ L_n \end{pmatrix} \cdot \omega + \begin{pmatrix} I \\ \vdots \\ X_n \end{pmatrix} e_* \quad (2.50)$$

sendo L_n e X_n matrizes $n \times n$ e $n \times (p+q)$ respectivamente envolvendo apenas os coeficientes $\phi_1, \dots, \phi_p$ e $\theta_1, \dots, \theta_q$.

Passemos, agora, à determinação das matrizes L_n e X_n .

Para simplificação extraímos de (2.50) a expressão

$$e_n = L_n \cdot \omega + X_n \cdot e_*.$$

A matriz L_n consiste do produto de matrizes $L_n = F^{-1} \cdot H$, onde F é uma matriz triangular inferior de dimensão $n \times n$, cujos elementos consistem do valor igual à unidade na diagonal principal, $-\theta_j$ ($j = 1, \dots, q$) na j -ésima diagonal a partir da principal e zeros nas outras posições; H é uma matriz triangular inferior de dimensão $n \times n$, com o valor igual à unidade na diagonal principal, $-\phi_i$ ($i = 1, \dots, p$) na i -ésima diagonal a partir da principal e zeros nas outras posições. A matriz X_n é consequência do produto $X_n = -F^{-1} \cdot G$, sendo G uma matriz de dimensão $n \times (p+q)$ definida por:

$$G = \begin{bmatrix} \tau & Y \\ 0 & 0 \end{bmatrix}$$

onde τ tem dimensão $q \times q$ com $-\theta_{i+j-1}$ na (i, j) -ésima posição ($i, j = 1, \dots, q$), com $\theta_m = 0$ se $m > q$; a matriz Y tem dimensão $p \times p$ com

$\phi_{k+\ell-1}$ na (k,ℓ) -ésima posição ($k,\ell = 1, \dots, p$) e $\phi_s = 0$ se $s > p$.

Para exemplificação, sejam $n = 3$, $p = 2$, $q = 2$.

Teremos, então:

$$F = \begin{bmatrix} 1 & 0 & 0 \\ -\theta_1 & 1 & 0 \\ -\theta_2 & -\theta_1 & 1 \end{bmatrix} \Rightarrow F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \theta_1 & 1 & 0 \\ \theta_2 + \theta_1^2 & \theta_1 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 \\ -\phi_1 & 1 & 0 \\ -\phi_2 & -\phi_1 & 1 \end{bmatrix}$$

$$L_n = F^{-1}H = \begin{bmatrix} 1 & 0 & 0 \\ \theta_1 - \phi_1 & 1 & 0 \\ \theta_2 + \theta_1^2 - \theta_1\phi_1 - \phi_2 & \theta_1 - \phi_1 & 1 \end{bmatrix}$$

$$G = \begin{bmatrix} -\theta_1 & -\theta_2 & \phi_1 & \phi_2 \\ -\theta_2 & 0 & \phi_2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$X_n = -F^{-1}G = \begin{bmatrix} \theta_1 & \theta_2 & -\phi_1 & -\phi_2 \\ \theta_1^2 + \theta_2 & \theta_1\theta_2 & -\theta_1\phi_1 - \phi_2 & -\theta_1\phi_2 \\ 2\theta_1\theta_2 + \theta_1^3 & \theta_2^2 + \theta_2\theta_1^2 & -\phi_1\theta_2 - \phi_1\theta_1^2 - \phi_2\theta_1 & -\phi_2\theta_2 - \phi_2\theta_1^2 \end{bmatrix}$$

Seja $E(e_*e'_*) = \sigma_a^2 \Omega$ e consideremos uma matriz quadrada não singular T com a propriedade:

$$T\Omega T' = I \quad (2.51)$$

Multiplicando (2.50) pela matriz

$$\begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix}$$

temos:

$$\begin{pmatrix} u_* \\ \dots \\ u_n \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ L_n \end{pmatrix} \omega + \begin{pmatrix} I \\ \dots \\ X_n^T - I \end{pmatrix} u_* \quad (2.52)$$

onde $u_* = Te_*$ e $u_n = e_n$.

Comparando (2.52) com (2.49), podemos escrever (2.52) como:

$$u = L\omega + Zu_* \quad (2.53)$$

sendo Z uma matriz $(n+p+q) \times (p+q)$ que é função apenas de

$$\phi_1, \dots, \phi_p \text{ e } \theta_1, \dots, \theta_q.$$

Temos ainda que u_* independe de u_n uma vez que seus elementos são funções de a_t , $t \leq 0$. Usando (2.51),

$$E(u_* u_*') = \sigma_a^2 T \Omega T' = \sigma_a^2 I.$$

Logo $u \sim N(0, \sigma_a^2)$ e temos a distribuição

$$p(u/\sigma_a) = (2\pi\sigma_a^2)^{-(n+p+q)/2} \cdot \exp\left\{\frac{-1}{2\sigma_a^2} u'u\right\}$$

A transformação (2.53) tem Jacobiano unitário e nos conduz à distribuição conjunta:

$$p(\omega, u_*/\phi, \theta, \sigma_a) = (2\pi\sigma_a^2)^{-(n+p+q)/2} \cdot \exp\left\{\frac{-1}{2\sigma_a^2} S(\phi, \theta, u_*)\right\} \quad (2.54)$$

onde

$$S(\phi, \theta, u_*) = u'u = (L\omega + Zu_*)'(L\omega + Zu_*).$$

Aplicando mínimos quadrados em (2.53) temos a estimativa de máxima verossimilhança $\hat{u}_*$ de u_* , dada por:

$$\hat{u}_* = -(Z'Z)^{-1} \cdot Z' \cdot L\omega \quad (2.55)$$

Podemos decompor $S(\phi, \theta, u_*)$ em duas parcelas:

$$S(\phi, \theta, u_*) = S(\phi, \theta) + (u_* - \hat{u}_*)' \cdot Z' \cdot Z \cdot (u_* - \hat{u}_*)$$

onde

$$S(\phi, \theta) = (L\omega + Z\hat{u}_*)' \cdot (L\omega + Z\hat{u}_*) \quad (2.56)$$

pois:

$$\begin{aligned} S(\phi, \theta, u_*) &= (L\omega + Zu_*)' \cdot (L\omega + Zu_*) \\ &= [\underbrace{L\omega + Z\hat{u}_*}_A + \underbrace{Z(u_* - \hat{u}_*)}_B]' \cdot [L\omega + Z\hat{u}_* + Z(u_* - \hat{u}_*)] \\ &= (A' + B')(A + B) \\ &= A'A + A'B + B'A + B'B \\ &= A'A + B'B \end{aligned}$$

Os produtos $A'B$ e $B'A$ são nulos uma vez que A e B são ortogonais. Por exemplo,

$$B'A = (u_* - \hat{u}_*)' \cdot Z' \cdot L\omega + (u_* - \hat{u}_*)' \cdot Z' \cdot Z \cdot \hat{u}_*$$

De (2.55) vem:

$$B'A = (u_* - \hat{u}_*)' \cdot Z' \cdot L\omega - (u_* - \hat{u}_*)' \cdot Z' \cdot L\omega$$

$$\Rightarrow B'A = 0$$

$S(\phi, \theta)$ é função das observações ω mas não dos resíduos iniciais u_* , o mesmo ocorrendo com $\hat{u}_*$.

Como (2.54) pode se expressar na forma:

$$p(\omega, u_*/\phi, \theta, \sigma_a) = p(\omega/\phi, \theta, \sigma_a) \cdot p(u_*/\omega, \phi, \theta, \sigma_a)$$

obtemos:

$$p(u_*/\omega, \phi, \theta, \sigma_a) = (2\pi\sigma_a^2)^{-(p+q)/2} \cdot |Z'Z|^{1/2} \cdot \exp\left\{\frac{-(u_* - \hat{u}_*)'Z'Z(u_* - \hat{u}_*)}{2\sigma_a^2}\right\}$$

e a função de verossimilhança incondicional exata é dada por

$$p(\omega/\phi, \theta, \sigma_a) = (2\pi\sigma_a^2)^{-n/2} \cdot |Z'Z|^{-1/2} \cdot \exp\left\{\frac{-1}{2\sigma_a^2} S(\phi, \theta)\right\} \quad (2.57)$$

onde $S(\phi, \theta)$ é dada por (2.56) e (2.55).

Também podemos escrever $S(\phi, \theta)$ na forma:

$$S(\phi, \theta) = \sum_{t=1-p-q}^n [u_t]^2 \quad (2.58)$$

onde

$$[u_t] = E(u_t/\omega, \phi, \theta) \quad \text{e} \quad [u_*] = \hat{u}_*.$$

Assim, obtivemos uma soma de quadrados com número finito de parcelas.

Vejamos, agora, o procedimento adotado por Box & Jenkins no caso em que usamos a expressão

$$S(\phi, \theta) = \sum_{t=-\infty}^n [a_t/Z, \phi, \theta]^2$$

quando não houver risco de ocorrer qualquer uma das três circunstâncias desfavoráveis que já apontamos.

Os valores de $[a_t]$ são calculados recursivamente, tomando-se esperanças condicionais em (2.46).

Um cálculo preliminar apresenta os valores $[\omega_{-j}]$, $j=0,1,2,\dots$, necessários para inicializarmos a recursão.

Sejam as equações:

$$\begin{cases} \phi(B) \omega_t = \theta(B) a_t \end{cases} \quad (2.59)$$

$$\begin{cases} \phi(F) \omega_t = \theta(F) e_t \end{cases} \quad (2.60)$$

A partir de (2.60) obtemos previsões retroativas $[\omega_{-j}/\phi, \theta, \omega]$ sendo que, em geral, $[\omega_t] = 0$, $t < -Q$. A partir de (2.59) geramos valores de a_t .

Considerando o sistema:

$$\begin{cases} \phi(B) [\omega_t] = \theta(B) [a_t] \end{cases} \quad (2.61)$$

$$\begin{cases} \phi(F) [\omega_t] = \theta(F) [e_t] \end{cases} \quad (2.62)$$

e as suposições:

$$\begin{cases} [e_{-j}/\phi, \theta, \omega] = 0, & j=0,1,2,\dots, \\ [a_{-j}/\phi, \theta, \omega] = 0, & j > Q-1 \end{cases}$$

podemos constatar a rapidez de convergência deste processo, aplicando um segundo ciclo de iterações juntamente com as suposições:

$$\begin{cases} [a_{n+i}/\phi, \theta, \omega] = 0, & i=1,2,\dots, \\ [e_{n+i}/\phi, \theta, \omega] = 0, & i > Q' \end{cases}$$

onde Q' é tal que $[\omega_{n+i}] \approx 0$, $i > Q'$.

Isolando $[a_t]$ em função de $[\omega_t]$ e dos valores defasados de ambos, vamos conseguir gerar ω_j , $n < j < Q'$. Estes valores de ω_j , substituídos numa equação análoga em que $[e_t]$ é função de $[\omega_t]$ e dos valores defasados de ambos, irão fornecer novas previsões retroativas $[\omega_0], [\omega_{-1}], \dots$. Poderemos obter $S(\phi, \theta/\omega)$ como resultante de:

$$\sum_{t=1}^{Q'} [e_t]^2.$$

Convém enfatizar que este procedimento é adequado desde que haja condições favoráveis e que o ciclo iterativo seja aplicado a defasagens bastante remotas.

Como o processo ARMA é o caso geral do qual são casos particulares os processos AR e MA, podemos também aplicar a estes últimos o esquema de "previsão para trás".

Consideremos o logaritmo da função de verossimilhança:

$$\begin{aligned} \ell(\phi, \theta, \sigma_a^2/\omega) &= \ell[p(\omega/\phi, \theta, \sigma_a^2)] = -\frac{n}{2} \ln(2\pi) - \\ &- \frac{n}{2} \ln \sigma_a^2 - \frac{1}{2} \ln |Z'Z| - \frac{S(\phi, \theta)}{2\sigma_a^2}. \end{aligned} \quad (2.63)$$

Derivando (2.63) em relação a σ_a^2 e igualando a zero a expressão encontrada, obtemos o estimador de máxima verossimilhança de σ_a^2 :

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n} \quad (2.64)$$

Temos assim o logaritmo concentrado da função de veros-

similhança:

$$\ell(\phi, \theta/\omega) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \frac{S(\phi, \theta)}{n} - \frac{1}{2} \ln |Z'Z| - \frac{n}{2} \quad (2.65)$$

O termo $|Z'Z|$ é insignificante quando comparado com $S(\phi, \theta)$, sendo importante somente quando n é pequeno. Então, podemos desconsiderá-lo e a maximização de $\ell(\phi, \theta/\omega)$ vai ser equivalente à minimização de $S(\phi, \theta)$.

Sabemos que

$$[u_t] = \theta^{-1}(B) \cdot \phi(B) \cdot [\omega_t]$$

$$\frac{\partial}{\partial \phi_i} [u_t] = \underbrace{\theta^{-1}(B) \phi(B) \frac{\partial}{\partial \phi_i} [\omega_t]}_{\text{função linear de } \phi_i} - \theta^{-1}(B) [\omega_{t-i}]$$

mas

$$\frac{\partial}{\partial \theta_j} [u_t] = \underbrace{\theta^{-1}(B) \phi(B) \frac{\partial}{\partial \theta_j} [\omega_t]}_{\text{função não linear em } \theta_j} - \phi(B) [\omega_{t-j}] \theta^{-2}(B)$$

Da mesma forma que foi apresentado na seção 2.3.1 um método de linearização através da expansão de $[a_t]$ em uma série de Taylor, o mesmo procedimento é válido para o modelo ARMA, desempenhando $[u_t]$ o mesmo papel de $[a_t]$.

Os estimadores obtidos pelo desenvolvimento deste procedimento são assintoticamente reconhecidos como estimadores de máxima verossimilhança, sendo, de fato, estimadores de mínimos quadrados.

Exemplifiquemos, agora, o cálculo direto das derivadas, presentes na expansão em uma série de Taylor no caso do processo ARMA.

Seja o modelo ARMA(1,1)

$$(1-\phi B)\omega_t = \begin{cases} (1-\theta F)e_t \\ (1-\theta B)a_t \end{cases} \Rightarrow \begin{cases} e_t = \omega_t - \phi\omega_{t+1} + \theta e_{t+1} \\ a_t = \omega_t - \phi\omega_{t-1} + \theta a_{t-1} \end{cases}$$

Vamos definir

$$a_t^{(\theta)} = \frac{\partial}{\partial \theta}[a_t], \quad a_t^{(\phi)} = \frac{\partial}{\partial \phi}[a_t], \quad e_t^{(\theta)} = \frac{\partial}{\partial \theta}[e_t],$$

$$e_t^{(\phi)} = \frac{\partial}{\partial \phi}[e_t], \quad \omega_t^{(\theta)} = \frac{\partial}{\partial \theta}[\omega_t], \quad \omega_t^{(\phi)} = \frac{\partial}{\partial \phi}[\omega_t].$$

Usando alternadamente as duas equações desse sistema vamos obter valores iniciais de uma para possibilitar a recursão sobre a outra.

Temos, agora, os sistemas:

$$\begin{cases} e_t^{(\phi)} = \omega_t^{(\phi)} - \phi\omega_{t+1}^{(\phi)} + \theta e_{t+1}^{(\phi)} - [\omega_{t+1}] \end{cases} \quad (2.66)$$

$$\begin{cases} a_t^{(\phi)} = \omega_t^{(\phi)} - \phi\omega_{t-1}^{(\phi)} + \theta a_{t-1}^{(\phi)} - [\omega_{t-1}] \end{cases} \quad (2.67)$$

$$\begin{cases} e_t^{(\theta)} = \omega_t^{(\theta)} - \phi\omega_{t+1}^{(\theta)} + \theta e_{t+1}^{(\theta)} + [e_{t+1}] \end{cases} \quad (2.68)$$

$$\begin{cases} a_t^{(\theta)} = \omega_t^{(\theta)} - \phi\omega_{t-1}^{(\theta)} + \theta a_{t-1}^{(\theta)} + [a_{t-1}] \end{cases} \quad (2.69)$$

sendo que:

$$[\omega_t] = \omega_t, \quad \omega_t^{(\phi)} = \omega_t^{(\theta)} = 0, \quad t=1, \dots, n \quad (2.70)$$

$$[e_{-j}] = 0, \quad j=0, 1, \dots, n$$

Considerando $e_{n+1}^{(\phi)} = 0$ em (2.66) iniciamos uma recursão re-
troativa e o uso de (2.70) possibilita o cálculo de $\omega_{-j}^{(\phi)}$, $j=0,$
 $1, \dots, q-1$.

Desde que $a_{-q}^{(\phi)}$, $a_{-q-1}^{(\phi)}, \dots$ podem ser supostos nulos, uti-
lizamos (2.67) para obter recursivamente as derivadas $a_t^{(\phi)}$. Com
procedimento análogo obtemos $a_t^{(\theta)}$.

2.4.2 - Metodologia de Durbin

Anteriormente, Walker (1950) e Quenouille (1958) dedica-
ram-se à tarefa de ajustar um modelo ARMA. Porém não se pro-
puseram a pesquisar técnicas eficientes na estimação dos pa-
râmetros. Durbin (1960), por sua vez, lançou dois métodos de
estimação a serem descritos nesta seção.

O primeiro método é não iterativo e não apresenta esti-
madores eficientes assintoticamente.

Para simplificação, consideremos o modelo ARMA(1,1):

$$\omega_t - \phi \omega_{t-1} = a_t - \theta a_{t-1}, \quad t = 1, \dots, n \quad (2.71)$$

ou

$$\omega_t + \gamma \omega_{t-1} = a_t + \delta a_{t-1},$$

onde $\gamma = -\phi$, $\delta = -\theta$.

Esse modelo pode ser representado em termos de um esque-
ma auto-regressivo de ordem alta, a saber:

$$(1 + \gamma B) \omega_t = (1 + \delta B) a_t$$

$$a_t = (1 + \gamma B)(1 + \delta B)^{-1} \omega_t$$

$$a_t = \omega_t + (\gamma - \delta)\omega_{t-1} - \delta(\gamma - \delta)\omega_{t-2} + \delta^2(\gamma - \delta)\omega_{t-3} - \\ - \delta^3(\gamma - \delta)\omega_{t-4} + \dots + \delta^{k-1}(\gamma - \delta)\omega_{t-k}. \quad (2.72)$$

Sejam $m_1, \dots, m_k$ os coeficientes de um modelo auto-regressivo ajustado de ordem alta, k , e seja e_t a série residual correspondente. Neste método vamos substituir a_{t-1} por e_{t-1} em (2.71).

Temos, ainda, a partir de (2.71) que:

$$e_t = \omega_t + \gamma\omega_{t-1} - \delta e_{t-1}.$$

A partir da minimização de $\sum_{t=1}^n (\omega_t + \gamma\omega_{t-1} - \delta e_{t-1})^2$ obtaremos os estimadores de mínimos quadrados $\hat{\gamma}$ e $\hat{\delta}$.

Desenvolvendo essa soma de quadrados, fazendo sua derivação em relação a γ e a δ e igualando a zero as expressões encontradas, vem:

$$\begin{cases} \gamma \sum_{t=1}^n \omega_{t-1}^2 - \delta \sum_{t=1}^n \omega_{t-1} e_{t-1} = - \sum_{t=1}^n \omega_t \omega_{t-1} \\ -\gamma \sum_{t=1}^n \omega_{t-1} e_{t-1} + \delta \sum_{t=1}^n e_{t-1}^2 = \sum_{t=1}^n \omega_t e_{t-1}. \end{cases} \quad (2.73)$$

A solução do sistema de equações (2.73) é assintoticamente equivalente às expressões

$$\hat{\gamma} = - \frac{m_1 r_2 + m_2 r_3 + \dots + m_k r_{k+1}}{m_1 r_1 + m_2 r_2 + \dots + m_k r_k} \quad (2.74)$$

$$\hat{\delta} = \hat{\gamma} + \frac{r_1 + m_1 r_2 + \dots + m_k r_{k+1}}{1 + m_1 r_1 + \dots + m_k r_k} \quad (2.75)$$

onde $r_k = \hat{\rho}_k$, isto é, r_k consiste na autocorrelação amostral de defasagem k e a sequência: $m_1, \dots, m_k$ é obtida a partir do ajuste de um modelo auto-regressivo de ordem k retratado em (2.72).

Esse método aplica-se também a processos mistos de ordem superior e pode fornecer estimativas que servirão de valores iniciais para o segundo método, que é iterativo.

No segundo método os parâmetros auto-regressivos e de médias móveis serão estimados alternadamente. Na primeira etapa deste método o procedimento é estimar δ para um dado valor de γ .

Consideremos $u_1, \dots, u_k$ os verdadeiros valores paramétricos de um modelo auto-regressivo ajustado de ordem alta, k .

Mann & Wald (1943) provaram que $m_1, \dots, m_k$, já definidos no primeiro método, são assintoticamente normais com densidade

$$p(m_1, \dots, m_k) = \text{constante} \times \exp \left[\frac{-n}{2\sigma^2} \sum_{i,j=1}^k (m_i - u_i)(m_j - u_j) \cdot E(\omega_{t-i} \omega_{t-j}) \right] \quad (2.76)$$

Supondo conhecido o valor de γ , é feita uma transformação dos valores antigos $m_1, \dots, m_k$ e $u_1, \dots, u_k$ em novos valores $\ell_1, \dots, \ell_k$ e $\lambda_1, \dots, \lambda_k$, juntamente com a inserção do valor de γ , como segue:

$$\begin{aligned} m_k &= \ell_k + \gamma \ell_{k-1} \\ u_k &= \lambda_k + \gamma \lambda_{k-1} \end{aligned} \quad (2.77)$$

Substituindo (2.77) na forma quadrática expoente da base exponencial em (2.76), é obtida a expressão:

$$Q = \varepsilon \left\{ \sum_{i=1}^{\infty} (\ell_i - \lambda_i) z_{t-i} \right\}^2 \quad (2.78)$$

onde

$$z_{t-i} = \omega_{t-i} + \gamma \omega_{t-i-1} = a_{t-i} + \delta a_{t-i-1}.$$

Aqui, ε indica a operação de tomar as esperanças dos valores de ω_t , os quais serão variáveis, ao passo que m_i serão considerados constantes fixadas.

Para maiores detalhes, ver Durbin (1960).

Em (2.78), $z_{t-1}, z_{t-2}, \dots$ correspondem a um modelo de médias móveis em que $\lambda_1, \lambda_2, \dots$ são os coeficientes auto-regressivos verdadeiros, enquanto que $\ell_1, \ell_2, \dots$ correspondem a coeficientes auto-regressivos ajustados aos dados referentes ao modelo

$$z_t = a_t + \delta a_{t-1}, \quad t=1, \dots, n.$$

Em (2.42) obtivemos $\hat{\theta}$, estimador do parâmetro θ , em função de $\hat{\alpha}_i$, coeficientes auto-regressivos ajustados. Analogamente, obtemos agora o estimador eficiente do parâmetro $\delta = -\theta$:

$$\hat{\delta} = - \frac{\sum_{i=0}^{\infty} \ell_i \ell_{i+1}}{\sum_{i=0}^{\infty} \ell_i^2}, \text{ sendo } \ell_0 = 1. \quad (2.79)$$

Na prática é suficiente termos essas somatórias finitas com limite superior igual a k .

O estimador $\hat{\delta}$ foi obtido após a derivação de Q em relação a δ , que surgiu na expressão (2.78) após a substituição: $\lambda_{k+r} = (-\delta)^r \cdot \lambda_k$, obtido através de manipulações algébricas.

Quando k é suficientemente grande, o determinante da matriz de covariâncias dos estimadores dos coeficientes auto-regressivos é desprezível em relação a Q . Esse determinante surge no fator constante de (2.76) e, por esse motivo, não foi levado em consideração na maximização da função de verossimilhança, tendo sido considerada somente a forma quadrática Q na obtenção de $\hat{\delta}$.

Uma expressão equivalente a (2.79) é dada por

$$\hat{\delta} = - \frac{\sum_{i=0}^{k-1} \ell_i \ell_{i+1} - \gamma \ell_k^2 / (1-\gamma^2)}{\sum_{i=0}^{k-1} \ell_i^2 + \ell_k^2 / (1-\gamma^2)} \quad (2.80)$$

Vamos, agora, estimar γ para um dado valor de δ .

Sejam $\varepsilon_1, \dots, \varepsilon_n$ definidos pela relação

$$\omega_t = \varepsilon_t + \delta \varepsilon_{t-1}, \quad t=1, \dots, n$$

sendo ε_0 escolhido arbitrariamente.

Como os ε 's satisfazem o modelo auto-regressivo

$$\varepsilon_t + \gamma \varepsilon_{t-1} = a_t,$$

o estimador eficiente de $\gamma = -\phi$ é dado por:

$$\hat{\gamma} = - \frac{\frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon_{t+1}}{\frac{1}{n} \sum_{t=1}^n \varepsilon_t^2} \quad (2.81)$$

Também encontramos outra expressão para $\hat{\gamma}$, dada por:

$$\hat{\gamma} = \frac{-\delta s_0 + (1+\delta^2)(s_1 - \delta s_2 + \delta^2 s_3 - \dots)}{s_0 - 2\delta s_1 + 2\delta^2 s_1 - \dots} \quad (2.82)$$

onde

$$s_r = \frac{1}{n} \sum_{t=1}^n \omega_t \omega_{t+r}.$$

Obtemos, assim, um método iterativo de estimar γ e δ , aplicando alternadamente as expressões (2.79) e (2.81). As expressões (2.74) e (2.75) podem ser usadas para fornecerem estimativas iniciais para o segundo método.

No caso do modelo geral, ARMA(p,q), basta fazer uma extensão dos resultados aqui obtidos.

Os parâmetros $\delta_1, \dots, \delta_q$ podem ser estimados a partir do resultado encontrado na seção 2.3.2, isto é:

$$\sum_{j=1}^q A_{|r-j|} \hat{\delta}_j = -A_r, \quad r=1, \dots, q$$

onde

$$A_r = \sum_{i=0}^{\infty} \ell_i \ell_{i+r}, \text{ sendo } \ell_0 = 1.$$

Na prática podemos truncar a somatória que define A_r nas vizinhanças de $i=k$.

As estimativas $\hat{\gamma}_1, \dots, \hat{\gamma}_p$ são obtidas a partir do sistema de equações:

$$\begin{aligned}
 q_0 \hat{Y}_1 + q_1 \hat{Y}_2 + \dots + q_{p-1} \hat{Y}_p &= -q_1 \\
 q_1 \hat{Y}_1 + q_0 \hat{Y}_2 + \dots + q_{p-2} \hat{Y}_p &= -q_2 \\
 &\vdots \\
 q_{p-1} \hat{Y}_1 + \dots + q_0 \hat{Y}_p &= -q_p
 \end{aligned}
 \tag{2.83}$$

supondo $\delta_1, \dots, \delta_q$ conhecidos e $q_r = \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon_{t+r}$.

2.5 - VARIÂNCIAS E COVARIÂNCIAS DAS ESTIMATIVAS DE MÁXIMA VEROSSIMILHANÇA, SEGUNDO BOX & JENKINS

Vamos nesta seção nos referir ao processo ARMA, lembrando que a teoria aqui desenvolvida também se aplica aos processos AR e MA.

Suponhamos que o tamanho amostral é moderadamente grande.

Já definimos anteriormente

$$\beta = (\phi, \theta), \quad \xi = (\beta, \sigma_a) = (\phi, \theta, \sigma_a).$$

Sobre uma região apropriada do espaço paramétrico, o logaritmo da função de verossimilhança de um modelo ARMA será aproximadamente quadrático em β , tal que:

$$\ell(\xi) \approx \ell(\hat{\beta}, \sigma_a) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \ell_{ij} (\beta_i - \hat{\beta}_i) (\beta_j - \hat{\beta}_j)$$

sendo que as derivadas

$$\ell_{ij} = \frac{\delta^2 \ell(\beta, \sigma_a)}{\delta \beta_i \delta \beta_j}$$

são constantes.

A função $\ell(\beta, \sigma_a)$ será quadrática em β se $S(\beta)$ o for, porque quando n é grande podemos omitir o termo $|Z'Z|$ em (2.63).

De modo equivalente, $\ell(\beta, \sigma_a)$ será quadrática em β se os $[a_t/\beta, \omega]$ forem aproximadamente lineares em β . Logo estamos partindo da suposição de que os $[a_t]$ eram originalmente lineares ou foram submetidos a uma linearização através da expansão em série de Taylor.

A matriz de variâncias e covariâncias será dada por:

$$V(\hat{\beta}) \approx \left\{ -E(\ell_{ij}) \right\}^{-1}$$

onde $-E(\ell_{ij}) = I(\beta)$ = matriz de informação para os parâmetros β com dimensão $K \times K$, onde $K = p + q$, sendo p e q as ordens dos processos AR e MA respectivamente.

A expressão (2.63) nos fornece

$$\ell_{ij} \approx \frac{-S_{ij}}{2\sigma_a^2} \text{ onde } S_{ij} = \frac{\delta^2 S(\beta/\omega)}{\delta\beta_i \delta\beta_j}.$$

Quando tivermos amostras grandes, poderemos aproximar $E(S_{ij})$ pelos respectivos valores observados.

Assim:

$$V(\hat{\beta}) \approx 2\sigma_a^2 \left\{ E(S'_{ij}) \right\}^{-1} \approx 2\sigma_a^2 \left\{ S_{ij} \right\}^{-1}.$$

Na seção 2.4.1 vimos que

$$\hat{\sigma}_a^2 = \frac{S(\hat{\beta})}{n}, \quad (2.84)$$

sendo $\hat{\sigma}_a^2$ e $\hat{\beta}$ não correlacionados em amostras grandes.

Temos também que

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) \approx 2\hat{\sigma}_a^2 \{S_{ij}\}^{-1}.$$

Para um determinado σ_a , fazendo a suposição de que $\ell(\beta, \sigma_a)$ é aproximadamente quadrática em β na vizinhança de $\hat{\beta}$, obtemos uma região de confiança para β , cujo coeficiente de confiança é $1 - \epsilon$ e é expressada como:

$$- \sum_{i,j} E(\ell_{ij})(\beta_i - \hat{\beta}_i)(\beta_j - \hat{\beta}_j) < \chi_\epsilon^2(K).$$

Usando (2.84) e o fato:

$$S(\beta) - S(\hat{\beta}) = \frac{1}{2} \sum_{i,j} S_{ij}(\beta_i - \hat{\beta}_i)(\beta_j - \hat{\beta}_j)$$

chegamos à fronteira da região de confiança:

$$S(\beta) = S(\hat{\beta}) \left[1 + \frac{\chi_\epsilon^2(K)}{n} \right].$$

As expressões para variâncias e covariâncias juntamente com esta região de confiança constituem um instrumento bastante eficaz e prático para a obtenção gráfica das estimativas paramétricas de máxima verossimilhança sendo o cálculo iterativo da estimação não linear uma outra alternativa com o mesmo objetivo.

CAPÍTULO 3

ESTIMAÇÃO EM MODELOS ARMA MULTIVARIADOS

3.1 - INTRODUÇÃO

Passamos, agora, à exposição de diferentes métodos objetivando a obtenção de estimativas paramétricas em modelos ARMA multivariados, extensão dos modelos ARMA univariados abordados no Capítulo 2.

3.2 - METODOLOGIA DE NICHOLLS E HALL

Nicholls e Hall (1979) desenvolveram uma técnica de derivação da função de verossimilhança exata de modelos mistos ARMA multivariados. Basearam-se numa extensão do procedimento seguido por Newbold (1974) no caso do modelo misto ARMA univariado. Nesta seção apresentaremos o seu método.

Primeiramente, vamos considerar $\{\omega_t\}$ como sendo um vetor de séries temporais de dimensão k , seguindo a estrutura de um modelo ARMA multivariado e estacionário. Tal processo expressa-se na forma:

$$\phi_p(B)\tilde{\omega}_t = \theta_q(B)a_t \quad (3.1)$$

onde

$$\phi_p(B) = (I_k - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

$$\theta_q(B) = (I_k - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

B: operador retroativo tal que $B\tilde{\omega}_t = \tilde{\omega}_{t-1}$

ϕ_i ($i=1, \dots, p$) e θ_j ($j=1, \dots, q$): matrizes paramétricas desconhecidas com dimensão $k \times k$

a'_t 's ($t=0, \pm 1, \pm 2, \dots$): vetores de dimensão $k \times 1$ independentes e identicamente distribuídos como $N(0, \Phi)$, onde Φ é uma matriz finita de dispersões.

Vamos supor que as raízes dos determinantes polinomiais em B, $|\phi_p(B)| = 0$, $|\theta_q(B)| = 0$, estão fora do círculo unitário, o que vai assegurar respectivamente a estacionariedade e invertibilidade do modelo (3.1).

Iremos também supor que $\mu_\omega = 0$ e, portanto,

$$\tilde{\omega}_t = \omega_t - \mu_\omega = \omega_t.$$

Podemos reescrever este modelo como:

$$\omega_t = \sum_{i=1}^p \phi_i \omega_{t-i} + a_t - \sum_{i=1}^q \theta_i a_{t-i}. \quad (3.2)$$

Assim, para uma amostra de N observações ω_t ($t=1, \dots, N$) teremos como objetivo básico a estimação dos parâmetros desconhecidos ϕ_i ($i=1, \dots, p$), θ_j ($j=1, \dots, q$) e a matriz de covariância Φ . Acrescentem-se a esses parâmetros os valores "pré-períodos" também desconhecidos a_j ($j=0, -1, \dots, -q+1$) e ω_j ($j=0, -1, \dots, -p+1$), os quais serão estimados neste procedimento.

Façamos uma analogia com o sistema de equações desenvol-

vido na seção 2.4 com adaptações de cada parâmetro ϕ_i , θ_j à estrutura de matrizes que assumem agora, lembrando que os a_t 's e ω_t 's também receberão tratamento vetorial.

Temos, então:

$$a_{1-q} = a_{1-q}$$

$$\vdots$$

$$a_{-1} = a_{-1}$$

$$a_0 = a_0$$

$$\omega_{1-p} = \omega_{1-p}$$

$$\vdots$$

$$\omega_0 = \omega_0$$

$$\omega_1 = \phi_1 \omega_0 + \dots + \phi_p \omega_{1-p} + a_1 - \theta_1 a_0 - \dots - \theta_q a_{1-q}$$

$$\vdots$$

$$\omega_N = \phi_1 \omega_{N-1} + \dots + \phi_p \omega_{N-p} + a_N - \theta_1 a_{N-1} - \dots - \theta_q a_{N-q}$$

Consideremos:

$$a' = (a'_1, \dots, a'_N)$$

$$\omega' = (\omega'_1, \dots, \omega'_N)$$

valores pré-períodos:

$$a'_* = (a'_0, \dots, a'_{1-q}, \omega'_0, \dots, \omega'_{1-p}).$$

O modelo (3.2) pode ser expressado na forma:

$$H\omega = Fa + Ga_* \quad (3.3)$$

onde H é uma matriz triangular inferior de dimensão $Nk \times Nk$, consistindo de I_k 's na diagonal principal, $-\phi_i$ ($i=1, \dots, p$) na i -

ésima diagonal a partir da principal e zeros nas posições restantes; analogamente, F é uma matriz triangular inferior de dimensão $Nk \times Nk$ consistindo de I_k 's na diagonal principal, $-\theta_j$ ($j=1, \dots, q$) na j -ésima diagonal a partir da principal e zeros nas demais posições; G é uma matriz de dimensão $Nk \times (p+q)k$ definida como

$$G = \begin{bmatrix} \tau & Y \\ 0 & 0 \end{bmatrix}$$

onde a matriz τ é de dimensão $qk \times qk$ com $-\theta_{i+j-1}$ no (i,j) -ésimo bloco ($i,j=1, \dots, q$), sendo $\theta_m = 0$ para $m > q$ e a matriz Y cuja dimensão é $pk \times pk$, contém ϕ_{k+l-1} no (k,l) -ésimo bloco ($k,l=1, \dots, p$), sendo $\phi_s = 0$ para $s > p$.

Para exemplificação, consideremos $N=6$, $p=2$, $q=3$, $k=1$.

Assim, temos:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -\phi_1 & 1 & 0 & 0 & 0 & 0 \\ -\phi_2 & -\phi_1 & 1 & 0 & 0 & 0 \\ 0 & -\phi_2 & -\phi_1 & 1 & 0 & 0 \\ 0 & 0 & -\phi_2 & -\phi_1 & 1 & 0 \\ 0 & 0 & 0 & -\phi_2 & -\phi_1 & 1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -\theta_1 & 1 & 0 & 0 & 0 & 0 \\ -\theta_2 & -\theta_1 & 1 & 0 & 0 & 0 \\ -\theta_3 & -\theta_2 & -\theta_1 & 1 & 0 & 0 \\ 0 & -\theta_3 & -\theta_2 & -\theta_1 & 1 & 0 \\ 0 & 0 & -\theta_3 & -\theta_2 & -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} +$$

$$+ \begin{bmatrix} -\theta_1 & -\theta_2 & -\theta_3 & \phi_1 & \phi_2 \\ -\theta_2 & -\theta_3 & 0 & \phi_2 & 0 \\ -\theta_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_{-1} \\ a_{-2} \\ \omega_0 \\ \omega_{-1} \end{bmatrix}$$

Se definirmos $L = F^{-1}H$ e $X = -F^{-1}G$ com dimensões $Nk \times Nk$ e $Nk \times (p+q)k$, respectivamente, e utilizarmos (3.3), teremos:

$$\begin{bmatrix} a_* \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix} \omega + \begin{bmatrix} I_{(p+q)k} \\ X \end{bmatrix} a_*. \quad (3.4)$$

Consideremos, agora, duas matrizes quadradas não singulares T e R , com dimensões $(p+q)k \times (p+q)k$ e $k \times k$, respectivamente, tais que:

se $\text{cov}(a_*) = \Omega$, então $T\Omega T' = I_{(p+q)k}$ e $R\Omega R' = I_k$.

Pré-multiplicando ambos os lados de (3.4) pela matriz

$$\begin{bmatrix} T & 0 \\ 0 & I_N \otimes R \end{bmatrix},$$

teremos:

$$\begin{bmatrix} Ta_* \\ (I_N \otimes R)a \end{bmatrix} = \begin{bmatrix} 0 \\ (I_N \otimes R)L \end{bmatrix} \omega + \begin{bmatrix} I_{(p+q)k} \\ (I_N \otimes R)XT^{-1} \end{bmatrix} Ta_*. \quad (3.5)$$

Dessa maneira, o vetor transformado pertencente ao primeiro membro de (3.5) tem matriz de covariância identidade de dimensão $(N+p+q)k \times (N+p+q)k$.

Podemos reescrever (3.5) na forma:

$$u = \psi\omega + \Lambda e_* \quad (3.6)$$

onde u tem dimensão $(N+p+q)k \times 1$, pois u tem duas partições a saber:

Ta_* e $(I_N \otimes R)a$ com as respectivas dimensões $(p+q)k \times 1$ e $Nk \times 1$;

analogamente ψ tem dimensão $(N+p+q)k \times Nk$ devido às partições 0 e $(I_N \otimes R)L$ com dimensões $(p+q)k \times Nk$ e $Nk \times Nk$ respectivamente; da mesma forma Λ tem dimensão $(N+p+q)k \times (p+q)k$, tendo as partições $I_{(p+q)k}$ e $(I_N \otimes R)XT^{-1}$ com as respectivas dimensões: $(p+q)k \times (p+q)k$ e $Nk \times (p+q)k$.

A matriz de covariância de u é dada por

$$\begin{bmatrix} TE(a_* a'_*)T' & TE(a_* a') (I_N \otimes R') \\ (I_N \otimes R)E(aa'_*)T' & (I_N \otimes R)E(aa') (I_N \otimes R') \end{bmatrix} = I_{(N+p+q)k}$$

Essa igualdade decorre da independência entre a_* e a , uma vez que os elementos de a_* são função de a_t , $t \leq 0$.

Assim, $u \sim N(0, I)$ e temos a distribuição

$$p(u) = L(u) = (2\pi)^{-1/2 (N+p+q)k} \cdot \exp\left\{-\frac{1}{2} u'u\right\}$$

onde $L(u)$ é a função de verossimilhança de u .

A transformação (3.6) tem por Jacobiano a expressão

$$|(I_N \otimes R)L| = |I_N \otimes R| \cdot |L| = |I_N|^k \cdot |R|^N \cdot |L|.$$

De acordo com a definição de L sua estrutura consiste numa matriz diagonal com os elementos da diagonal principal iguais à unidade e valores nulos acima desta diagonal.

Dessa maneira temos, então, uma matriz cujo determinante é igual à unidade e isto também ocorre com a matriz identidade.

Portanto,

$$|(I_N \otimes R)L| = |R|^N = |I|^{-N/2}.$$

Logo, a distribuição conjunta obtida $\tilde{e}$ expressada na forma:

$$p(\omega, e_*) = L(\omega, e_*) = (2\pi)^{-1/2(N+p+q)k} \cdot |\mathbb{X}|^{-N/2} \cdot \exp\left\{-\frac{1}{2}S(\omega, e_*)\right\} \quad (3.7)$$

onde

$$S(\omega, e_*) = (\psi\omega + \Lambda e_*)' (\psi\omega + \Lambda e_*)$$

$\tilde{e}$ função de $\mathbb{X}$, ϕ_i ($i=1, \dots, p$) e θ_j ($j=1, \dots, q$).

Aplicando mínimos quadrados em (3.6) temos a estimativa de máxima verossimilhança $\hat{e}_*$ de e_* dada por:

$$\hat{e}_* = -(\Lambda' \Lambda)^{-1} \Lambda' \psi \omega. \quad (3.8)$$

Podemos decompor $S(\omega, e_*)$ em duas parcelas:

$$S(\omega, e_*) = S(\omega, \hat{e}_*) + (e_* - \hat{e}_*)' \Lambda' \Lambda (e_* - \hat{e}_*)$$

onde

$$S(\omega, \hat{e}_*) = (\psi\omega + \Lambda \hat{e}_*)' (\psi\omega + \Lambda \hat{e}_*), \quad (3.9)$$

pois

$$\begin{aligned} S(\omega, e_*) &= (\psi\omega + \Lambda e_*)' (\psi\omega + \Lambda e_*) \\ &= \underbrace{[\psi\omega + \Lambda \hat{e}_*]}_A + \underbrace{\Lambda(e_* - \hat{e}_*)}_B \quad \underbrace{[\psi\omega + \Lambda \hat{e}_* + \Lambda(e_* - \hat{e}_*)]}_B \\ &= (A' + B') (A + B) = A'A + A'B + B'A + B'B \\ &= A'A + B'B. \end{aligned}$$

Como A e B são ortogonais os produtos $A'B$ e $B'A$ são nu-

los. Por exemplo,

$$B'A = (e_* - \hat{e}_*)' \Lambda' \psi \omega + (e_* - \hat{e}_*)' \Lambda' \Lambda \hat{e}_*.$$

De (3.8) vem:

$$B'A = (e_* - \hat{e}_*)' \Lambda' \psi \omega - (e_* - \hat{e}_*)' \Lambda' \psi \omega,$$

logo $B'A = 0$, o mesmo para $A'B$.

Conseqüentemente, (3.7) torna-se

$$L(\omega, e_*) = (2\pi)^{-1/2 Nk} \cdot |\Phi|^{-N/2} \cdot |\Lambda' \Lambda|^{-1/2} \cdot \exp\left\{-\frac{1}{2} S(\omega, \hat{e}_*)\right\} \cdot (2\pi)^{-1/2 (p+q)k} \cdot |\Lambda' \Lambda|^{1/2} \cdot \exp\left\{-\frac{1}{2} (e_* - \hat{e}_*)' \Lambda' \Lambda (e_* - \hat{e}_*)\right\} = L(\omega) L(e_*/\omega).$$

Assim,

$$L(\omega) = (2\pi)^{-1/2 Nk} \cdot |\Phi|^{-N/2} \cdot |\Lambda' \Lambda|^{-1/2} \cdot \exp\left\{-\frac{1}{2} (\psi\omega + \Lambda \hat{e}_*)' (\psi\omega + \Lambda \hat{e}_*)\right\} \quad (3.10)$$

Essa é a função de verossimilhança incondicional exata, a qual é função dos parâmetros desconhecidos do modelo (3.2). Podemos maximizar $L(\omega)$ usando um algoritmo não linear, tal como o de Newton-Raphson, a fim de obter as estimativas de máxima verossimilhança exata dos parâmetros.

3.3 - DETERMINAÇÃO DA MATRIZ DE COVARIÂNCIA DOS VALORES

PRÉ-PERÍODOS, SEGUNDO NICHOLLS E HALL

Nesta seção vamos expressar a matriz de covariância dos valores pré-períodos a_* em função de Φ , ϕ_i e θ_j . Estabelecida essa relação basta usarmos estimativas iniciais consistentes de Φ , ϕ_i e θ_j para determinarmos, então, a matriz de covariância, Ω , e também a matriz T .

Detalhes referentes à obtenção de estimativas iniciais consistentes fazem parte do Apêndice de Nicholls (1976).

Definimos a matriz Ω como:

$$\Omega = E(a_* a_*') = \begin{bmatrix} A & B \\ B' & C \end{bmatrix}$$

onde $a_*' = (a_0', \dots, a_{1-q}', \omega_0', \dots, \omega_{1-p}')$, Ω tem dimensão $(p+q)k \times (p+q)k$, A consiste numa matriz diagonal de dimensão $qk \times qk$ $A = \text{diag}(\lambda, \dots, \lambda)$, B é uma matriz de ordem $qk \times pk$ definida para $i = 0, \dots, 1-q$; $j = 0, \dots, 1-p$; tal que $B = \{E(a_i \omega_j')\} = \{\Gamma_{a\omega}(i-j)\}$, C é uma matriz de ordem $pk \times pk$ definida para $i, j = 0, \dots, 1-p$, tal que $C = \{E(\omega_i \omega_j')\} = \{\Gamma_{\omega}(i-j)\}$.

Vamos agora achar as transpostas em (3.2) e inserir à esquerda o fator a_{t-j} .

Tomando esperanças, teremos as seguintes possibilidades:

i) $m = 0$ onde $m = t-j-l$, sendo l o índice de ω .

$$\omega_{t-j}' = \sum_{i=1}^p \omega_{t-j-i}' \phi_i' + a_{t-j}' - \sum_{i=1}^q a_{t-j-i}' \theta_i'$$

$$a_{t-j} \omega_{t-j}' = \sum_{i=1}^p a_{t-j} \omega_{t-j-i}' \phi_i' + a_{t-j} a_{t-j}' - \sum_{i=1}^q a_{t-j} a_{t-j-i}' \theta_i'$$

$$\begin{aligned} E(a_{t-j} \omega_{t-j}') &= \sum_{i=1}^p E(a_{t-j} \omega_{t-j-i}') \phi_i' + E(a_{t-j} a_{t-j}') - \\ &\quad - \sum_{i=1}^q E(a_{t-j} a_{t-j-i}') \theta_i' \end{aligned}$$

$$\Gamma_{a\omega}(0) = 0 + \lambda - 0.$$

$$\text{Logo, } \Gamma_{a\omega}(0) = \lambda.$$

ii) $m > 0$, $t-j > \ell$.

Para ilustração suponhamos que $\ell = t-j-1$. Então

$$\begin{aligned} E(a_{t-j} \omega'_{t-j-1}) &= \sum_{i=1}^p E(a_{t-j} \omega'_{t-j-1-i}) \phi'_i + E(a_{t-j} a'_{t-j-1}) - \\ &- \sum_{i=1}^q E(a_{t-j} a'_{t-j-1-i}) \theta'_i \end{aligned}$$

$$\Gamma_{a\omega}(1) = 0 + 0 - 0$$

$$\text{Logo, } \Gamma_{a\omega}(m) = 0, \quad m > 0.$$

iii) $m < 0$, $t-j < \ell$.

Seja, por exemplo, $\ell = t-j+3$. Então:

$$\begin{aligned} E(a_{t-j} \omega'_{t-j+3}) &= \sum_{i=1}^p E(a_{t-j} \omega'_{t-j+3-i}) \phi'_i + E(a_{t-j} a'_{t-j+3}) - \\ &- \sum_{i=1}^q E(a_{t-j} a'_{t-j+3-i}) \theta'_i \end{aligned}$$

$$\begin{aligned} \Gamma_{a\omega}(-3) &= \Gamma_{a\omega}(-2) \phi'_1 + \Gamma_{a\omega}(-1) \phi'_2 + \Gamma_{a\omega}(0) \phi'_3 + \Gamma_{a\omega}(1) \phi'_4 + \\ &+ \Gamma_{a\omega}(2) \cdot \phi'_5 + \dots + \Gamma_{a\omega}(p-3) \cdot \phi'_p + 0 - \cancel{4} \cdot \theta'_3. \end{aligned}$$

Usando (ii) temos:

$$\Gamma_{a\omega}(-3) = \sum_{i=1}^3 \Gamma_{a\omega}(i-3) \phi'_i - \cancel{4} \theta'_3$$

Logo,

$$\Gamma_{a\omega}(m) = \sum_{i=1}^{-m} \Gamma_{a\omega}(i+m) \phi'_i - \cancel{4} \theta'_{-m}, \quad m < 0 \quad (3.11)$$

$-m = 1, \dots, q$, onde se $p < q$ então $\phi_i = 0$ para $i = p+1, \dots, q$.

Podemos reescrever estes resultados como:

$$[\Gamma_{aw}(0) \begin{vmatrix} \vdots \\ \Gamma_{aw}(-1) \\ \vdots \end{vmatrix} \dots \begin{vmatrix} \vdots \\ \Gamma_{aw}(-q) \\ \vdots \end{vmatrix}] \Delta' = \sharp \theta' \quad (3.12)$$

onde a matriz Δ tem dimensão $(q+1)k \times (q+1)k$ e é da forma:

$$\Delta = \begin{bmatrix} -I_k & 0 & 0 & \dots & 0 \\ \phi_1 & -I_k & 0 & \dots & 0 \\ \phi_2 & \phi_1 & -I_k & \dots & 0 \\ \phi_3 & \phi_2 & \phi_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_q & \phi_{q-1} & \phi_{q-2} & \dots & \phi_1 & -I_k \end{bmatrix}$$

$$\text{e } \theta' = (-I_k \begin{vmatrix} \vdots \\ \theta'_1 \\ \vdots \end{vmatrix} \dots \begin{vmatrix} \vdots \\ \theta'_q \\ \vdots \end{vmatrix}).$$

Sejam

$$\delta = \text{vet}[\Gamma_{aw}(0) \begin{vmatrix} \vdots \\ \Gamma_{aw}(-1) \\ \vdots \end{vmatrix} \dots \begin{vmatrix} \vdots \\ \Gamma_{aw}(-q) \\ \vdots \end{vmatrix}] \quad \text{e} \quad \sigma = \text{vet}(\sharp).$$

Consideremos a relação

$$\text{vet}(ABC) = (C' \Theta A) \text{vet}(B) \quad (3.13)$$

para colocar (3.12) em forma vetorial, chegando aos resultados:

$$\text{vet} \left\{ I_k [\Gamma_{aw}(0) \begin{vmatrix} \vdots \\ \Gamma_{aw}(-1) \\ \vdots \end{vmatrix} \dots \begin{vmatrix} \vdots \\ \Gamma_{aw}(-q) \\ \vdots \end{vmatrix}] \Delta' \right\} = \text{vet} \left\{ I_k \sharp \theta' \right\}$$

$$(\Delta \Theta I_k) \delta = (\Theta \Theta I_k) \sigma$$

Portanto

$$\delta = (\Delta \Theta I_k)^{-1} (\Theta \Theta I_k) \sigma = (\Delta^{-1} \Theta \Theta I_k) \sigma \quad (3.14)$$

Desta maneira, com uma seleção de elementos corretos para δ

obtemos a matriz B pois esta é função de $\Gamma_{a\omega}(i-j)$.

Vamos agora inserir à direita em (3.2) o fator ω'_{t-j} . Tomando esperanças e considerando o resultado de (3.11), obtemos:

$$\sum_{i=0}^p \phi_i \omega_{t-i} \omega'_{t-j} = \sum_{i=j}^q \theta_i a_{t-i} \omega'_{t-j}$$

$$\sum_{i=0}^p \phi_i \Gamma_{\omega}(j-i) = \sum_{i=j}^q \theta_i \Gamma_{a\omega}(j-i), \quad j=0,1,\dots,p \quad (3.15)$$

onde se $p > q$ então o segundo membro de (3.15) será nulo para $j = q+1, \dots, p$.

Sejam $\delta(j) = \text{vet}\{\Gamma_{a\omega}(j)\}$, $\gamma(j) = \text{vet}\{\Gamma_{\omega}(j)\}$.

Podemos colocar (3.15) em forma vetorial utilizando (3.13), isto é,

$$\sum_{i=0}^p \text{vet}(\phi_i \Gamma_{\omega}(j-i) I'_k) = \sum_{i=j}^q \text{vet}(\theta_i \Gamma_{a\omega}(j-i) I'_k)$$

$$\sum_{i=0}^p (I_k \otimes \phi_i) \gamma(j-i) = \sum_{i=j}^q (I_k \otimes \theta_i) \delta(j-i), \quad j=0, \dots, p \quad (3.16)$$

A expressão (3.16) é função de

$$\gamma(-p), \gamma(-p+1), \dots, \gamma(0), \dots, \gamma(p-1), \gamma(p).$$

Vamos agora definir uma matriz de permutação P com dimensão $K^2 \times K^2$ tal que se D é uma matriz $k \times k$, então

$$P \text{ vet}(D) = \text{vet}(D'). \quad (3.17)$$

Considerando (3.17) e o fato de que para $s > 0$,

$$\Gamma_{\omega}(-j) = \left\{ E(\omega_{-(s+1)j} \omega'_{-sj}) \right\}$$

$$\Gamma_{\omega}(-j) = \left\{ E(\omega_{-sj} \omega'_{-(s+1)j}) \right\}$$

$$\Gamma_{\omega}(-j) = \Gamma'_{\omega}(j),$$

teremos:

$$P \text{ vet} \left\{ \Gamma_{\omega}(j) \right\} = \text{vet} \left\{ \Gamma'_{\omega}(j) \right\} = \text{vet} \left\{ \Gamma_{\omega}(-j) \right\}.$$

Portanto, $P\gamma(j) = \gamma(-j)$.

Desta maneira é possível escrevermos (3.16) somente em função de $\gamma(0), \gamma(1), \dots, \gamma(p)$, para $j = 0, 1, \dots, p$.

Assim,

$$\sum_{i=0}^j (I_k \otimes \phi_i) \gamma(j-i) + \sum_{i=j+1}^p (I_k \otimes \phi_i) P\gamma(i-j) = \sum_{i=j}^q (I_k \otimes \theta_i) \delta(j-i).$$

Temos, então, $(p+1)k^2$ equações lineares constituídas de $(p+1)k^2$ incógnitas: $\gamma(0), \gamma(1), \dots, \gamma(p)$.

Este sistema de equações pode ser escrito na forma:

$$\gamma = \Phi^{-1} \Theta \delta \quad (3.18)$$

onde

$$\gamma' = [\gamma(0)', \gamma(1)', \dots, \gamma(p)']$$

e as matrizes Φ e Θ têm dimensões $(p+1)k^2 \times (p+1)k^2$ e $(p+1)k^2 \times (q+1)k^2$, respectivamente, sendo expressadas por:

$$\Phi = \begin{bmatrix} -I_{k^2} & \phi_1^+ & \phi_2^+ & \dots & \phi_{p-1}^+ & \phi_p^+ \\ \phi_1^+ & -I_{k^2} + \phi_2^+ & \phi_3^+ & \dots & \phi_p^+ & 0 \\ \phi_2^+ & \phi_1^+ + \phi_3^+ & -I_{k^2} + \phi_4^+ & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_{p-1}^+ & \phi_{p-2}^+ + \phi_p^+ & \phi_{p-3}^+ & \dots & -I_{k^2} & 0 \\ \phi_p^+ & \phi_{p-1}^+ & \phi_{p-2}^+ & \dots & \phi_1^+ & -I_{k^2} \end{bmatrix}$$

$$\Theta = \begin{bmatrix} -I_{k^2} & \theta_1^+ & \dots & \theta_{q-1}^+ & \theta_q^+ \\ \theta_1^+ & \theta_2^+ & \dots & \theta_q^+ & 0 \\ \theta_2^+ & \theta_3^+ & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \theta_p^+ & \theta_{p+1}^+ & \dots & \theta_q^+ & 0 \end{bmatrix}$$

onde $\phi_j^+ = (I_k \otimes \phi_j)$, $j = 1, \dots, p$; e $\theta_j^+ = (I_k \otimes \theta_j)$, $j = 1, \dots, q$.

De (3.14) e (3.18), vem:

$$\gamma = \Phi^{-1} \Theta (\Delta^{-1} \Theta \Theta I_k) \sigma. \quad (3.19)$$

Portanto, conseguimos expressar, através de (3.19), os elementos de C em termos de Φ , ϕ_i e θ_j como desejávamos.

3.4 - METODOLOGIA DE WILSON

Wilson (1973) desenvolveu um método iterativo a fim de estimar os parâmetros em modelos mistos ARMA multivariados.

Seu método é uma generalização do de Marquardt (1963) aplicado por Box & Jenkins (1970) a séries temporais univariadas.

Inicialmente é suposto que os inteiros p e q , graus dos polinômios $\phi_p(B)$ e $\theta_q(B)$ de (3.1) respectivamente, já estão corretamente especificados.

Teremos, então, como parâmetros desconhecidos as matrizes ϕ_i ($i = 1, \dots, p$) e θ_j ($j = 1, \dots, q$) juntamente com a matriz de dispersões Φ , todas com dimensão $k \times k$.

Vamos definir o vetor paramétrico

$$\beta = \left\{ \phi_{rs,i}, \theta_{rs,i} \right\} = (\beta_1, \beta_2, \dots, \beta_H)'.$$

Suponhamos que as condições de invertibilidade e estacionariedade, já explicitadas na seção 3.2, estão plenamente satisfeitas.

Neste novo procedimento de estimação, para algum conjunto proposto de parâmetros $\hat{\beta}$ é gerada a partir de ω_t uma correspondente série residual $\hat{a}_t = \hat{a}_t(\hat{\beta})$.

Esta etapa baseia-se na suposição de que o modelo está representando corretamente a série $\{\omega_t\}$.

Vamos distinguir os verdadeiros valores paramétricos e residuais dos propostos, utilizando na notação um ponto sobre os símbolos relativos aos valores propostos e quantidades derivadas destes.

Consideremos obtidas as observações: $\omega_1, \omega_2, \dots, \omega_N$.

Usando (3.2), segue que:

$$\hat{a}_t = \omega_t - \dot{\phi}_1 \omega_{t-1} - \dots - \dot{\phi}_p \omega_{t-p} + \dot{\theta}_1 \hat{a}_{t-1} + \dots + \dot{\theta}_q \hat{a}_{t-q} \quad (3.20)$$

A expressão (3.20) será usada recursivamente para $t = 1, 2, \dots, N$.

No segundo membro de (3.20) existem termos que são desconhecidos no início das iterações. Como os $\dot{a}_t$ são funções destes termos, os seus valores sofrerão influência pela escolha destes elementos iniciais. Vejamos um exemplo.

A expressão (3.20) pode ser reescrita de modo que $\dot{a}_t$ seja definido na forma do processo puro AR, ou seja:

$$\dot{a}_t = \omega_t - \sum_{j=1}^{\infty} \dot{\pi}_j \omega_{t-j} = \omega_t - \sum_{j=1}^{t-1} \dot{\pi}_j \omega_{t-j} - \sum_{j=t}^{\infty} \dot{\pi}_j \omega_{t-j} \quad (3.21)$$

Se escolhêssemos o valor zero para atribuímos aos termos $\omega_0, \dots, \omega_{1-p}, \dot{a}_0, \dots, \dot{a}_{1-q}$, teríamos

$$\begin{aligned} \dot{a}_t &= \omega_t - \sum_{j=1}^{t-1} \dot{\pi}_j \omega_{t-j} - \sum_{j=t}^{t+p-1} \dot{\pi}_j \omega_{t-j} - \sum_{j=t+p}^{\infty} \dot{\pi}_j \omega_{t-j} \\ \dot{a}_t &= \omega_t - \sum_{j=1}^{t-1} \dot{\pi}_j \omega_{t-j} - (\dot{\pi}_t \omega_0 + \dot{\pi}_{t+1} \omega_{-1} + \dots + \dot{\pi}_{t+p-1} \omega_{1-p}) - \sum_{j=t+p}^{\infty} \dot{\pi}_j \omega_{t-j} \end{aligned}$$

O valor de $\dot{a}_t$ seria a soma truncada

$$\omega_t - \sum_{j=1}^{t-1} \dot{\pi}_j \omega_{t-j}$$

e o erro em $\dot{a}_t$ seria a quantidade desconhecida $\sum_{j=t}^{\infty} \dot{\pi}_j \omega_{t-j}$.

Os termos $\dot{\pi}_j$ tornam-se desprezíveis à medida que j aumenta, isto é,

$$\dot{\pi}_j = O(\rho^j), \quad \rho < 1.$$

Portanto,

$$\sum_{j=t}^{\infty} \dot{\pi}_j \omega_{t-j} = O(\rho^t)$$

e essa expressão decresce para zero conforme t aumenta.

Há ainda outra alternativa para iniciarmos a recursão cujos erros têm menor magnitude embora sejam da mesma ordem. Consiste em considerarmos o início da recursão no instante $t=p+1$, tal que em (3.20) precisamos atribuir valor zero apenas aos termos $a_p, \dots, a_{p-q+1}$, pois

$$\dot{\theta}_1 \dot{a}_{t-1} + \dots + \dot{\theta}_q \dot{a}_{t-q}$$

$$t = p+1 \Rightarrow \dot{\theta}_1 \dot{a}_p + \dots + \dot{\theta}_q \dot{a}_{p-q+1}.$$

O método de previsão retroativa usado por Box & Jenkins para resolver o problema de obtenção do valor inicial no caso univariado torna-se difícil de ser aplicado em modelos multivariados.

Vamos agora introduzir uma notação relativa à derivação. Ao nos referirmos à derivada dos resíduos com respeito aos parâmetros β_h usaremos o símbolo $\dot{a}_t^h$, $h=1,2,\dots,H$, sendo que as segundas derivadas serão simbolizadas por $\dot{a}_t^{h\ell}$, $h,\ell=1,\dots,H$.

Essas derivadas são facilmente obtidas a partir dos dados ω_t e dos resíduos $\dot{a}_t$ após a derivação de (3.20) com respeito ao parâmetro de interesse.

A dependência dos resíduos $\dot{a}_t$ em relação a ω_t pode ser observada através da expressão que envolve $\dot{a}_t$ em termos de um processo puro AR, isto é,

$$\dot{a}_t = \omega_t - \sum_{j>0} \pi_j \omega_{t-j}.$$

Analogamente é possível observar a dependência das derivadas residuais em relação a ω_t , partindo de:

$$\dot{a}_t^h = - \sum_{j>0} \dot{\pi}_j^h \omega_{t-j} \quad (3.22)$$

$$\dot{a}_t^{h\ell} = - \sum_{j>0} \dot{\pi}_j^{h\ell} \omega_{t-j} \quad (3.23)$$

Sabemos que

$$E(a_s a_t') = \begin{cases} 1, & s = t \\ 0, & s \neq t. \end{cases} \quad (3.24)$$

Usando o fato de que a série ω_t pode ser expressada na forma do processo puro MA, vem:

$$\omega_t = a_t + \sum_{i>0} \psi_i \cdot a_{t-i}$$

$$\omega_{t-j} = a_{t-j} + \sum_{i>0} \psi_i \cdot a_{t-i-j}$$

Logo, concluímos que, para $j > 0$, $E(\omega_{t-j} \cdot a_t') = 0$.

Aplicando esse resultado a (3.22) e (3.23), segue que:

$$E(\dot{a}_t^h a_t') = 0 \quad \text{e} \quad E(\dot{a}_t^{h\ell} a_t') = 0,$$

uma vez que $\dot{a}_t^h$ e $\dot{a}_t^{h\ell}$ são combinações lineares de ω_{t-j} , $j > 0$.

De acordo com um teorema ergódico provado por Hannan (1970, pg. 210) obtemos os resultados quando $N \rightarrow \infty$:

$$\frac{1}{N} \sum_{t=1}^N (a_{r,t} a_{s,t}') \longrightarrow E(a_{r,t} a_{s,t}') = \delta_{rs},$$

$$\frac{1}{N} \sum_{t=1}^N (\dot{a}_{r,t}^h a_{s,t}) \longrightarrow E(\dot{a}_{r,t}^h a_{s,t}) = 0, \quad (3.25)$$

$$\frac{1}{N} \sum_{t=1}^N (\dot{a}_{r,t}^{hl} a_{s,t}) \longrightarrow E(\dot{a}_{r,t}^{hl} a_{s,t}) = 0,$$

$$\frac{1}{N} \sum_{t=1}^N (\dot{a}_{r,t}^h \dot{a}_{s,t}^l) \longrightarrow E(\dot{a}_{r,t}^h \dot{a}_{s,t}^l)$$

Com esse teorema temos estabelecido que as covariâncias cruzadas amostrais entre os elementos de uma série temporal multivariada gerada por um processo linear, convergem em média quadrática para os valores teóricos, desde que a série temporal geradora tenha momentos finitos de 4.^a ordem.

As condições impostas por esse teorema estão satisfeitas em nosso modelo, pois nas equações (3.25) o primeiro e segundo membros constituem respectivamente as covariâncias amostrais e teóricas entre elementos de série temporal multivariada, incluindo a série residual $\dot{a}_t$ juntamente com as séries de derivadas $\dot{a}_t^h$ e $\dot{a}_t^{hl}$.

Estas séries consistem num processo linear pois são geradas através da série ω_t de forma linear, sendo que a própria série ω_t é gerada linearmente pelos verdadeiros resíduos a_t . A aplicação desse teorema requer que a série a_t tenha momentos finitos de 4.^a ordem.

Cumpramos enfatizarmos que devido ao fato de que os erros em $\dot{a}_t$ (e suas derivadas), provenientes da atribuição do valor inicial, serem de ordem desprezível, $O(\rho^t)$, seu efeito sobre os valores amostrais nas equações (3.25) é de ordem $O(1/N)$ e, conseqüentemente, é dissipado quando $N \rightarrow \infty$.

Vamos, agora, determinar a função de verossimilhança dos parâmetros $\hat{\beta}$ e $\hat{\Sigma}$. Usando a hipótese de normalidade conjunta de $(a_{1,t}, \dots, a_{k,t})$, segue que:

$$p(\omega; \hat{\beta}, \hat{\Sigma}) = L(\omega; \hat{\beta}, \hat{\Sigma}) = (2\pi)^{-N/2} |\hat{\Sigma}|^{-N/2} \exp\left\{-\frac{1}{2} \sum_{t=1}^N \hat{a}_t' \hat{\Sigma}^{-1} \hat{a}_t\right\} \quad (3.26)$$

Não consideraremos os efeitos dos valores iniciais.

$$\ln L(\omega; \hat{\beta}, \hat{\Sigma}) = \ln(\hat{\beta}, \hat{\Sigma} / \omega) = -\frac{1}{2} N k \ln(2\pi) - \frac{1}{2} N F(\hat{\beta}, \hat{\Sigma}), \quad (3.27)$$

onde

$$F(\hat{\beta}, \hat{\Sigma}) = \ln |\hat{\Sigma}| + \frac{1}{N} \sum_{t=1}^N \hat{a}_t' \hat{\Sigma}^{-1} \hat{a}_t \quad (3.28)$$

ou, de modo equivalente,

$$\ln(\hat{\beta}, \hat{\Sigma} / \omega) = -\frac{1}{2} N k \ln(2\pi) - \frac{N}{2} \ln |\hat{\Sigma}| - \frac{1}{2} \sum_{t=1}^N \hat{a}_t' \hat{\Sigma}^{-1} \hat{a}_t \quad (3.29)$$

Através de (3.27) podemos concluir que a maximização de $\ln(\hat{\beta}, \hat{\Sigma} / \omega)$ implica a minimização de $F(\hat{\beta}, \hat{\Sigma})$ com respeito aos elementos de $\hat{\beta}$ e $\hat{\Sigma}$. Quando obtivermos o mínimo dessa função serão conhecidas as nossas estimativas de máxima verossimilhança, $\hat{\beta}$ e $\hat{\Sigma}$.

Embora esse procedimento de estimação seja fundamentado na suposição de normalidade, poderemos aplicá-lo mesmo quando essa suposição não se verificar. Nessa circunstância, as estimativas resultantes não serão estimativas de máxima verossimilhança. Adotando o critério de Whittle (1962), vamos denotá-las de estimativas Gaussianas.

Podemos obter o mínimo de $F(\dot{\beta}, \dot{\Phi})$, igualando suas derivadas a zero.

As derivadas de $F(\dot{\beta}, \dot{\Phi})$ em relação a $\dot{\beta}_h$ ($h=1,2,\dots,H$) são

$$\frac{2}{N} \sum_{t=1}^N \dot{a}_t^{h'} \dot{\Phi}^{-1} \dot{a}_t \quad (3.30)$$

pois

$$[F(\dot{\beta}, \dot{\Phi})]^h = [\ln |\dot{\Phi}|]^h + \frac{1}{N} \left[\sum_{t=1}^N \dot{a}_t \dot{\Phi}^{-1} \dot{a}_t \right]^h$$

$$[F(\dot{\beta}, \dot{\Phi})]^h = \frac{1}{N} \left[\sum_{t=1}^N (\dot{a}_t \dot{\Phi}^{-1} \dot{a}_t)^h \right]$$

$$[F(\dot{\beta}, \dot{\Phi})]^h = \frac{1}{N} \left[\sum_{t=1}^N (\dot{a}_t^{h'} \dot{\Phi}^{-1} \dot{a}_t + \dot{a}_t \dot{\Phi}^{-1} \dot{a}_t^h) \right]$$

$$[F(\dot{\beta}, \dot{\Phi})]^h = \frac{1}{N} \left[\sum_{t=1}^N (2 \dot{a}_t^{h'} \dot{\Phi}^{-1} \dot{a}_t) \right]$$

$$[F(\dot{\beta}, \dot{\Phi})]^h = \frac{2}{N} \sum_{t=1}^N \dot{a}_t^{h'} \dot{\Phi}^{-1} \dot{a}_t$$

e as derivadas de $F(\dot{\beta}, \dot{\Phi})$ em relação a $\dot{\Phi}_{r,s}^{-1}$ ($r,s=1,2,\dots,k$) são

$$-\dot{\Phi}_{r,s} + \frac{1}{N} \sum_{t=1}^N \dot{a}_{r,t} \dot{a}_{s,t} \quad (3.31)$$

pois

$$\frac{\partial}{\partial \dot{\Phi}_{r,s}^{-1}} [F(\dot{\beta}, \dot{\Phi})] = \frac{\partial}{\partial \dot{\Phi}_{r,s}^{-1}} \left\{ -\ln |\dot{\Phi}|^{-1} + \frac{1}{N} \sum_{t=1}^N \dot{a}_{r,t} \dot{\Phi}_{r,s}^{-1} \dot{a}_{s,t} \right\}$$

$$\frac{\partial}{\partial \dot{\Phi}_{r,s}^{-1}} [F(\dot{\beta}, \dot{\Phi})] = -\dot{\Phi}_{r,s} + \frac{1}{N} \sum_{t=1}^N \dot{a}_{r,t} \dot{a}_{s,t}$$

Para obtermos as estimativas de máxima verossimilhança basta encontrarmos as raízes simultâneas de (3.30) e (3.31). Suponhamos que haja somente uma solução para esse problema de estimação.

Em (3.31), se fixarmos $\dot{\beta}$, a raiz obtida será:

$$\dot{\Phi}_{r,s} = \frac{1}{N} \sum_{t=1}^N \dot{a}_{r,t} \dot{a}_{s,t}. \quad (3.32)$$

Essa é a estimativa condicional de $\dot{\Phi}$, que será denotada por $\tilde{\Phi}$.

Em (3.32) temos justamente a matriz de dispersão amostral de $\dot{a}_t$, a qual podemos expressar por:

$$\tilde{\Phi} = \tilde{\Phi}(\dot{\beta}) = \frac{1}{N} \sum_{t=1}^N \dot{a}_t \dot{a}_t'. \quad (3.33)$$

Podemos reescrever o segundo termo de (3.28), usando (3.33). Teremos, assim:

$$\frac{1}{N} \sum_{t=1}^N \dot{a}_t' \dot{\Phi}^{-1} \dot{a}_t = \sum_{r,s=1}^k \left(\frac{1}{N} \sum_{t=1}^N \dot{a}_{r,t} \dot{a}_{s,t} \right) \left\{ \dot{\Phi}^{-1} \right\}_{rs}.$$

A primeira igualdade decorre de (3.24).

Podemos simplificar ainda mais essa expressão, usando (3.33).

$$\frac{1}{N} \sum_{t=1}^N \dot{a}_t' \dot{\Phi}^{-1} \dot{a}_t = \sum_{r,s=1}^k \tilde{\Phi}_{rs} \left\{ \dot{\Phi}^{-1} \right\}_{rs} = \text{tr}(\tilde{\Phi} \dot{\Phi}^{-1}).$$

Dessa maneira conseguimos simplificar nossa expressão para $F(\dot{\beta}, \dot{\Phi})$, que se transforma em:

$$F(\dot{\beta}, \dot{\Phi}) = \ln |\dot{\Phi}| + \text{tr}(\dot{\Phi} \dot{\Phi}^{-1}). \quad (3.34)$$

Quando substituirmos $\dot{\Phi} = \tilde{\Phi}$ obtemos o mínimo de $F(\dot{\beta}, \dot{\Phi})$, ou seja:

$$\begin{aligned} \inf_{\dot{\Phi}} F(\dot{\beta}, \dot{\Phi}) &= \ln |\tilde{\Phi}| + \text{tr}(\tilde{\Phi} \tilde{\Phi}^{-1}) \\ \inf_{\dot{\Phi}} F(\dot{\beta}, \dot{\Phi}) &= \ln |\tilde{\Phi}| + \text{tr}(I_k) \\ \inf_{\dot{\Phi}} F(\dot{\beta}, \dot{\Phi}) &= \ln |\tilde{\Phi}| + k \end{aligned} \quad (3.35)$$

Devido ao fato de que

$$\inf_{\dot{\beta}, \dot{\Phi}} F(\dot{\beta}, \dot{\Phi}) = \inf_{\dot{\beta}} \left\{ \inf_{\dot{\Phi}} F(\dot{\beta}, \dot{\Phi}) \right\},$$

podemos obter estimativas de β através da minimização de $\{\ln |\tilde{\Phi}| + k\}$ em relação a $\tilde{\beta}$.

De (3.33), vem:

$$|\tilde{\Phi}| = \left| \frac{1}{N} \sum_{t=1}^N \dot{a}_t \dot{a}_t' \right| \quad (3.36)$$

Podemos usar (3.36) na estimação de β , sendo este procedimento recomendável somente se houver poucos parâmetros no modelo pois, caso contrário, a minimização envolverá cálculos muito extensos.

Em (3.30), se fixarmos $\dot{\Phi}$, a raiz obtida será:

$$\frac{2}{N} \sum_{t=1}^N \dot{a}_t \dot{h}' \dot{\Phi}^{-1} \dot{a}_t = 0$$

$$\frac{1}{N} \sum_{t=1}^N \ddot{a}_t^{h'} \ddot{a}_t^{-1} \dot{a}_t = 0, \quad h=1, 2, \dots, H. \quad (3.37)$$

É nosso objetivo, agora, encontrar a estimativa condicional de β .

Para linearização de $\dot{a}_t$ vamos expandi-lo em série de Taylor em torno de algum ponto escolhido $\dot{\beta}_0$, tal que:

$$\dot{a}_t(\dot{\beta}) = \dot{a}_t(\dot{\beta}_0) + \sum_{\ell=1}^H \ddot{a}_t^{\ell} \delta \dot{\beta} + o\|\delta \dot{\beta}\| \quad (3.38)$$

onde

$$\delta \dot{\beta} = \dot{\beta} - \dot{\beta}_0. \quad (3.39)$$

Após a substituição de (3.38) em (3.37), obtemos equações lineares para $\delta \dot{\beta}$, omitindo o termo $o\|\delta \dot{\beta}\|$, isto é:

$$\dot{A} \delta \dot{\beta} = -\dot{g}, \quad (3.40)$$

onde a matriz $\dot{A}$ de dimensão $h \times h$ consiste dos elementos

$$\dot{A}_{h\ell} = \frac{1}{N} \sum_{t=1}^N \ddot{a}_t^{h'} \ddot{a}_t^{-1} \ddot{a}_t^{\ell} \quad (3.41)$$

e $\dot{g}$ é metade do vetor da primeira derivada de $F(\dot{\beta}, \ddot{a})$ em relação a $\dot{\beta}$, ou seja:

$$\dot{g}_h = \frac{1}{N} \sum_{t=1}^N \ddot{a}_t^{h'} \ddot{a}_t^{-1} \dot{a}_t \quad (3.42)$$

Calculamos, então, $\dot{A}$ e $\dot{g}$ em $\dot{\beta}_0$.

Se em nosso modelo não houvesse termos de média móvel, isto é, $q=0$, teríamos $\dot{a}_t$ linear em $\dot{\beta}$ e, conseqüentemente, o ter-

mo $0 \|\delta\dot{\beta}\|$ seria nulo.

Os parâmetros corrigidos

$$\dot{\beta}_1 = \dot{\beta}_0 + \delta\dot{\beta}. \quad (3.43)$$

forneceriam a solução para (3.37).

Por outro lado, se $q \neq 0$, devemos substituir $\dot{\beta}_0$, nosso primeiro ponto escolhido, por $\dot{\beta}_1$ e repetir a linearização a fim de obter uma nova correção paramétrica. Este procedimento é repetido sucessivamente até que a seqüência $\dot{\beta}_0, \dot{\beta}_1, \dots$ convirja para a estimativa condicional $\tilde{\beta} = \tilde{\beta}(\dot{\beta})$. Para que a obtenção de convergência seja garantida, correções paramétricas sucessivas podem ser restringidas através de modificação das equações (3.40) em cada iteração, usando o método de Marquardt (1963) para problemas não lineares de mínimos quadrados.

Iniciamos, agora, o processo de estimação simultânea de β e β .

Aplicando os dois esquemas de estimação condicional alternadamente, ou seja,

$$\left. \begin{array}{l} \text{i) considerar } \dot{\beta}_n = \tilde{\beta}(\dot{\beta}_n) \\ \text{ii) considerar } \dot{\beta}_{n+1} = \tilde{\beta}(\dot{\beta}_n) \end{array} \right\} n = 0, 1, 2, \dots \quad (3.44)$$

os parâmetros $\dot{\beta}_n, \dot{\beta}_n$ irão forçosamente convergir para um mínimo global $\hat{\beta}, \hat{\beta}$, uma vez que ambos os passos i) e ii) consistem em uma minimização condicional.

Para elucidar esse processo de estimação consideremos o algoritmo que segue.

ALGORITMO - Dados valores iniciais de $\dot{\beta}$ e um parâmetro de restrição $\lambda > 0$

a) Considerar $\dot{\Phi} = \tilde{\Phi}(\beta) = \frac{1}{N} \sum_{t=1}^N \dot{a}_t' \dot{a}_t$

b) Obter $\dot{\Phi}^{-1}$

c) Construir a matriz $\dot{A}$, que consiste dos elementos

$$\dot{A}_{h\ell} = \frac{1}{N} \sum_{t=1}^N \dot{a}_t' \dot{\Phi}^{-1} \dot{a}_t$$

e o vetor $\dot{g}$ com elementos

$$\dot{g}_h = \frac{1}{N} \sum_{t=1}^N \dot{a}_t' \dot{\Phi}^{-1} \dot{a}_t$$

e as quantidades de escala $\alpha_h = \dot{A}_{hh}^{1/2}$

d) Considerar a matriz escalonada $\dot{B}$ cujos elementos são definidos por:

$$\dot{B}_{h\ell} = \frac{\dot{A}_{h\ell}}{\alpha_h \alpha_\ell}$$

e o vetor escalonado $\dot{\epsilon}$ cujos elementos são definidos por:

$$\dot{\epsilon}_h = \frac{\dot{g}_h}{\alpha_h}$$

e) Substituir os elementos diagonais de $\dot{B}$ por $1+\lambda$, resolver as equações $\dot{B}\omega = \dot{\epsilon}$ e avaliar um novo conjunto proposto de parâmetros

$$\beta_h = \dot{\beta}_h - (\omega_h / \alpha_h)$$

f) Considerar $\dot{\Phi} = \tilde{\Phi}(\beta)$ e testar se

$$\text{tr}(\dot{\beta}\dot{\beta}^{-1}) < k.$$

i) se esta condição for satisfeita, é preciso reduzir o parâmetro de restrição, λ , de um fator pré-estabelecido, fazer as substituições de $\dot{\beta}$ por β , $\dot{\lambda}$ por λ e retornar ao passo b).

ii) se esta condição não for satisfeita, é preciso aumentar o parâmetro de restrição, λ , de um fator pré-estabelecido e retornar ao passo e).

É conveniente testar, após o passo f)(i), se houve convergência com uma precisão pré-fixada, como também decidir, após o passo f)(ii), pela interrupção das iterações se o parâmetro λ tornar-se excessivamente elevado.

Marquardt (1963) fornece uma orientação para se escolher o valor inicial e o fator de variação para o parâmetro de restrição bem como sugere um critério sensato de convergência.

Nesta seção vimos o método de Wilson com o qual obtivemos estimativas que, segundo resultados dados por Whittle (1962, seção 8), têm as seguintes propriedades:

i) as estimativas $\hat{\beta}$ são consistentes e têm assintoticamente uma distribuição normal conjunta com média β e matriz de dispersão $\frac{1}{N}C^{-1}$, onde C é uma matriz com elementos

$$C_{h\ell} = -\frac{1}{2}E\left\{\frac{\partial^2 F}{\partial \beta_h \partial \beta_\ell}\right\}$$

calculados em β .

Esse resultado não necessita da suposição da série de tempo ω_t ser Gaussiana.

ii) as estimativas $\hat{\beta}$ são assintoticamente não correlacio-

nadas com $\hat{\beta}$, estimativas consistentes de β .

Acrescentemos a essas propriedades o fato de que, desde que N seja grande e o modelo ajustado seja adequado para explicar os dados, a metodologia de Wilson goza das propriedades de convergência quadrática.

Um aspecto prático também relevante é que esse método requer apenas a geração dos resíduos $\hat{a}_t$ e suas primeiras derivadas $\dot{\hat{a}}_t^h$ ($h = 1, 2, \dots, H$), totalizando $KN(1+H)$ quantidades escalares. O método tradicional utiliza também as segundas derivadas de $\hat{a}_t$ e, portanto, totaliza $\frac{1}{2}KNH(H+1)$ quantidades escalares.

Assim, através do esquema de Wilson há uma redução de cálculos computacionais, redução esta bastante apreciável principalmente quando, ao ajustarmos um modelo, depararmos com um grande número de parâmetros a serem estimados.

3.5 - ESTIMAÇÃO EM MODELOS ARMA MULTIVARIADOS

NÃO INVERTÍVEIS

Quando analisamos séries temporais, geralmente precisamos tomar diferenças dos dados originais a fim de obtermos estacionariedade. Entretanto, se houver um número excessivo de diferenças teremos, como consequência, modelos não invertíveis. Modelos dessa natureza podem também ocorrer durante análises de séries temporais multivariadas. Por exemplo, consideremos duas séries de tempo $\{\omega_{1t}\}$ e $\{\omega_{2t}\}$ seguindo, respectivamente, os modelos:

$$\omega_{1t} = \omega_{1(t-1)} + a_{1t} - \alpha a_{1(t-1)};$$

$$\omega_{2t} = \beta \omega_{1t} + a_{2t}.$$

Logo ambas as séries são não estacionárias individualmente, uma vez que ω_{1t} decorre de um processo ARMA(1,1) onde o parâmetro auto-regressivo, coeficiente de $\omega_{1(t-1)}$, é igual à unidade. Esse fato reflete a não estacionariedade das duas séries pois ω_{2t} é combinação linear de ω_{1t} e será, conseqüentemente, não estacionária também.

Seja

$$y_{it} = (1-B)\omega_{it}, \quad i = 1, 2. \quad (3.45)$$

São tomadas essas primeiras diferenças com o objetivo de se obter estacionariedade.

Assim, o modelo bivariado para as séries das primeiras diferenças pode ser expressado por:

$$y_t = a_t^* - \theta a_{t-1}^*, \quad (3.46)$$

onde

$$\begin{aligned} a_t^* &= (a_{1t}, \beta a_{1t} + a_{2t})' \\ \text{e } \theta &= \begin{bmatrix} \alpha & 0 \\ \beta(\alpha-1) & 1 \end{bmatrix} \end{aligned} \quad (3.47)$$

pois, substituindo (3.47) em (3.46), vem:

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ \beta a_{1t} + a_{2t} \end{bmatrix} - \begin{bmatrix} \alpha & 0 \\ \beta(\alpha-1) & 1 \end{bmatrix} \begin{bmatrix} a_{1(t-1)} \\ \beta a_{1(t-1)} + a_{2(t-1)} \end{bmatrix} =$$

$$y_t = \begin{bmatrix} a_{1t} \\ \beta a_{1t} + a_{2t} \end{bmatrix} - \begin{bmatrix} \alpha a_{1(t-1)} \\ \beta(\alpha-1)a_{1(t-1)} + \beta a_{1(t-1)} + a_{2(t-1)} \end{bmatrix} \quad (3.48)$$

$$y_t = \begin{bmatrix} a_{1t} \\ \beta a_{1t} + a_{2t} \end{bmatrix} - \begin{pmatrix} \alpha & 0 \\ \beta(\alpha-1) & 1 \end{pmatrix} \begin{bmatrix} a_{1(t-1)} \\ \beta a_{1(t-1)} + a_{2(t-1)} \end{bmatrix} \quad (3.49)$$

Usando (3.45), temos:

$$y_{1t} = (1-B)\omega_{1t} = \omega_{1t} - \omega_{1(t-1)} = a_{1t} - \alpha a_{1(t-1)}$$

$$y_{2t} = (1-B)\omega_{2t} = \omega_{2t} - \omega_{2(t-1)} = \beta\omega_{1t} + a_{2t} - \beta\omega_{1(t-1)} - a_{2(t-1)}$$

$$y_{2t} = \beta(\omega_{1t} - \omega_{1(t-1)}) + a_{2t} - a_{2(t-1)}$$

$$y_{2t} = \beta a_{1t} - \beta\alpha a_{1(t-1)} + a_{2t} - a_{2(t-1)}$$

Assim, temos:

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} - \alpha a_{1(t-1)} \\ \beta a_{1t} + a_{2t} - (\beta\alpha a_{1(t-1)} + a_{2(t-1)}) \end{bmatrix}$$

Essa expressão obtida coincide com (3.48) e, comparando (3.46) e (3.49), podemos encontrar

$$|\theta(B)| = |I - \theta_1 B| = \left| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \alpha B & 0 \\ \beta(\alpha-1)B & B \end{pmatrix} \right| =$$

$$= \begin{vmatrix} 1-\alpha B & 0 \\ -\beta(\alpha-1)B & 1-B \end{vmatrix} = (1-\alpha B)(1-B) \quad (3.50)$$

Como $|\theta(B)|$ gerou uma raiz unitária correspondente ao segundo fator em (3.50), concluímos que o modelo (3.46) é não invertível.

Passamos, agora, à determinação da função de verossimilhança exata para modelos ARMA MULTIVARIADOS não invertíveis, baseando-nos em estudo desenvolvido por Hillmer e Tiao (1979).

De acordo com nossa notação habitual, estamos tratando do modelo ARMA(p,q).

Consideremos, inicialmente, o caso $p = q$.

Suponhamos que dispomos de N observações $\omega' = (\omega'_1, \dots, \omega'_N)$.

Utilizaremos as seguintes matrizes:

1.^a) $D_{\lambda, N}$ com dimensão $Nk \times Nk$

$$D_{\lambda, N} = \begin{bmatrix} I & & & 0 \\ -\lambda_1 & & & \\ \vdots & & & \\ -\lambda_q & & & \\ 0 & -\lambda_q & \dots & -\lambda_1 & I \end{bmatrix}$$

2.^a) $C_{\lambda, q}$ com dimensão $qk \times qk$

$$C_{\lambda,q} = \begin{bmatrix} \lambda_q & \lambda_{q-1} & \cdots & \lambda_1 \\ & & \ddots & \vdots \\ & & & \lambda_{q-1} \\ 0 & & & \lambda_q \end{bmatrix}$$

3^a) $F_{\lambda,q}$ com dimensão $qk \times qk$

$$F_{\lambda,q} = \begin{bmatrix} I & & & 0 \\ & \ddots & & \\ -\lambda_1 & & \ddots & \\ \vdots & & & \\ -\lambda_{q-1} & \cdots & -\lambda_1 & I \end{bmatrix}$$

4^a) $\Delta_{N,q}$ com dimensão $Nk \times qk$

$$\Delta_{N,q} = \begin{bmatrix} I_{kq} \\ 0 \end{bmatrix}$$

As matrizes $\lambda_1, \dots, \lambda_q$ têm dimensão $k \times k$.

As matrizes apresentadas guardam entre si a relação

$$D_{\lambda,N+q} = \begin{bmatrix} F_{\lambda,q} & 0 \\ -\Delta_{N,q} C_{\lambda,q} & D_{\lambda,N} \end{bmatrix} \quad (3.51)$$

De fato, se tivéssemos por exemplo: $q = 2$, $N = 3$, $k = 1$, as matrizes assumiriam as seguintes estruturas:

$$F_{\lambda,2} = \begin{bmatrix} 1 & 0 \\ -\lambda_1 & 1 \end{bmatrix}, \quad C_{\lambda,2} = \begin{bmatrix} \lambda_2 & \lambda_1 \\ 0 & \lambda_2 \end{bmatrix}$$

$$\Delta_{3,2} = \begin{bmatrix} I_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$D_{\lambda,3} = \begin{bmatrix} 1 & 0 & 0 \\ -\lambda_1 & 1 & 0 \\ -\lambda_2 & -\lambda_1 & 1 \end{bmatrix}$$

$$\Delta_{3,2} \cdot C_{\lambda,2} = \begin{bmatrix} \lambda_2 & \lambda_1 \\ 0 & \lambda_2 \\ 0 & 0 \end{bmatrix}$$

$$D_{\lambda,3+2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\lambda_1 & 1 & 0 & 0 & 0 \\ -\lambda_2 & -\lambda_1 & 1 & 0 & 0 \\ 0 & -\lambda_2 & -\lambda_1 & 1 & 0 \\ 0 & 0 & -\lambda_2 & -\lambda_1 & 1 \end{bmatrix} = \begin{bmatrix} F_{\lambda,2} & 0 \\ -\Delta_{3,2} \cdot C_{\lambda,2} & D_{\lambda,3} \end{bmatrix}$$

O resultado encontrado é uma constatação da validade de (3.51).

O modelo ARMA(q,q) será, então, definido por:

$$D_{\phi,N} \omega = D_{\theta,N} a + \Delta_{N,q} v, \quad (3.52)$$

onde

$$v = C_{\phi,q} \omega_* - C_{\theta,q} a_*, \quad a' = (a'_1, \dots, a'_N)$$

$$\omega'_* = (\omega'_{-q+1}, \dots, \omega'_0), \quad a'_* = (a'_{-q+1}, \dots, a'_0)$$

sendo v e a independentes.

A expressão (3.52) é funcional. Por exemplo, sejam:

$p = q = 2$, $k = 1$ e $N = 3$. Teríamos, então:

$$v = \begin{bmatrix} \phi_2 & \phi_1 \\ 0 & \phi_2 \end{bmatrix} \begin{bmatrix} \omega_{-1} \\ \omega_0 \end{bmatrix} - \begin{bmatrix} \theta_2 & \theta_1 \\ 0 & \theta_2 \end{bmatrix} \begin{bmatrix} a_{-1} \\ a_0 \end{bmatrix} = \begin{bmatrix} \phi_2 \omega_{-1} + \phi_1 \omega_0 - \theta_2 a_{-1} - \theta_1 a_0 \\ \phi_2 \omega_0 - \theta_2 a_0 \end{bmatrix}$$

A partir de (3.52), vem:

$$\begin{bmatrix} 1 & 0 & 0 \\ -\phi_1 & 1 & 0 \\ -\phi_2 & -\phi_1 & 1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\theta_1 & 1 & 0 \\ -\theta_2 & -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \phi_2 \omega_{-1} + \phi_1 \omega_0 - \theta_2 a_{-1} - \theta_1 a_0 \\ \phi_2 \omega_0 - \theta_2 a_0 \end{bmatrix}$$

$$\begin{cases} \omega_1 = a_1 + \phi_2 \omega_{-1} + \phi_1 \omega_0 - \theta_2 a_{-1} - \theta_1 a_0 \\ -\phi_1 \omega_1 + \omega_2 = -\theta_1 a_1 + a_2 + \phi_2 \omega_0 - \theta_2 a_0 \\ -\phi_2 \omega_1 - \phi_1 \omega_2 + \omega_3 = -\theta_2 a_1 - \theta_1 a_2 + a_3 \end{cases}$$

$$\begin{cases} \omega_1 - \phi_1 \omega_0 - \phi_2 \omega_{-1} = a_1 - \theta_1 a_0 - \theta_2 a_{-1} \\ \omega_2 - \phi_1 \omega_1 - \phi_2 \omega_0 = a_2 - \theta_1 a_1 - \theta_2 a_0 \\ \omega_3 - \phi_1 \omega_2 - \phi_2 \omega_1 = a_3 - \theta_1 a_2 - \theta_2 a_1 \end{cases} \quad (3.53)$$

O sistema (3.53) verifica a relação (3.1) e podemos concluir que a notação implantada em (3.52) equivale à formulação em (3.1).

Seja

$$Y = D_{\phi, N} \omega, \quad \omega' = [\omega'_1, \omega'_2], \quad Y' = [Y'_1, Y'_2],$$

$$a' = [\alpha'_1, \alpha'_2],$$

onde ω_1 , Y_1 e α_1 consistem dos primeiros vetores de dimensão $qk \times 1$ de ω , Y e a , respectivamente.

Primeiramente, consideremos a distribuição de Y . A partir de (3.51) e (3.52), vem

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} F_{\theta,q} & 0 \\ -\Delta_{N-q,q} C_{\theta,q} & D_{\theta,N-q} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} v \\ 0 \end{bmatrix} \quad (3.54)$$

onde $Y_1 = F_{\phi,q} \omega_1$ e $Y_2 = D_{\phi,N-q} \omega_2 - \Delta_{N-q,q} C_{\phi,q} \omega_1$.

Sendo $p(Y)$ a distribuição de Y podemos expressá-la por $p(Y) = p(Y_2)p(Y_1/Y_2)$.

Utilizando as notações estabelecidas, expressaríamos um modelo MA(q) na forma:

$$z = D_{\theta,N} a - \Delta_{N,q} C_{\theta,q} a_* \quad (3.55)$$

pois se considerássemos, por exemplo, $q = 2$, $k = 1$ e $N = 3$, teríamos:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -\theta_1 & 1 & 0 \\ -\theta_2 & -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_2 & \theta_1 \\ 0 & \theta_2 \end{bmatrix} \begin{bmatrix} a_{-1} \\ a_0 \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\theta_1 & 1 & 0 \\ -\theta_2 & -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} - \begin{bmatrix} \theta_2 & \theta_1 \\ 0 & \theta_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{-1} \\ a_0 \end{bmatrix}$$

$$\begin{cases} z_1 = a_1^{-\theta} a_2^{-1-\theta} a_0 \\ z_2 = -\theta_1 a_1 + a_2^{-\theta} a_0 \\ z_3 = -\theta_2 a_1^{-\theta} a_2 + a_3 \end{cases}$$

$$\begin{cases} z_1 = a_1^{-\theta} a_0^{-\theta} a_2^{-1} \\ z_2 = a_2^{-\theta} a_1^{-\theta} a_0 \\ z_3 = a_3^{-\theta} a_2^{-\theta} a_1 \end{cases} \quad (3.56)$$

O sistema (3.56) verifica a relação apresentada por um processo MA(q), ou seja, $z_t = a_t^{-\theta} a_{t-1}^{-\theta} \dots a_{t-q}^{-\theta}$.

Portanto, a formulação adotada em (3.55) é válida.

Observando (3.55) e (3.54) segue que $p(Y_2)$ corresponde à densidade de N-q observações de um processo MA(q).

Então, torna-se necessária a determinação da função de verossimilhança exata de um processo puro MA(q). Tal processo, como já vimos, pode ser definido por:

$$z = D_{\theta, N} a - \Delta_{N, q} C_{\theta, q} a_* \quad (3.57)$$

Seja $b = Pa_*$ onde P é uma matriz não singular de dimensão $qk \times qk$ escolhida arbitrariamente tal que $|P| = 1$.

Podemos, então, escrever:

$$\begin{bmatrix} a_* \\ a \end{bmatrix} = \begin{bmatrix} P^{-1} & 0 \\ D_{\theta, N}^{-1} \Delta_{N, q} C_{\theta, q} P^{-1} & D_{\theta, N}^{-1} \end{bmatrix} \begin{bmatrix} b \\ z \end{bmatrix} = M-Xb \quad (3.58)$$

onde

$$M' = z' [0, D_{\theta, N}^{-1}] \text{ e } X' = -P^{-1} [I, (\Delta_{N, q} C_{\theta, q})' D_{\theta, N}^{-1}].$$

Aplicando a metodologia dos mínimos quadrados, chegamos a:

$$L(\theta, \Phi/z) \propto |\Phi|^{-1/2(N+q)} |X' \Phi_{N+q}^{-1} X|^{-1/2} \cdot \exp\left(-\frac{1}{2} [M' \Phi_{N+q}^{-1} M - \hat{b}' X' \Phi_{N+q}^{-1} X \hat{b}]\right) \quad (3.59)$$

onde

$$\hat{b} = (X' \Phi_{N+q}^{-1} X)^{-1} X' \Phi_{N+q}^{-1} M \quad (3.60)$$

e $\Phi_N = I_N \otimes \Phi$.

Podemos expressar (3.59) em uma outra forma equivalente:

$$L(\theta, \Phi/z) \propto |\Phi|^{-1/2(N+q)} |A_{N,q}|^{-1/2} \cdot \exp\left(-\frac{1}{2} [z' H_N z - \hat{a}_*' A_{N,q} \hat{a}_*]\right) \quad (3.61)$$

onde $H_N = D_{\theta,N}^{-1} \Phi_N^{-1} D_{\theta,N}^{-1}$,

$$A_{N,q} = \Phi_q^{-1} + (\Delta_{N,q} C_{\theta,q})' H_N (\Delta_{N,q} C_{\theta,q})$$

e

$$\hat{a}_* = -A_{N,q}^{-1} (\Delta_{N,q} C_{\theta,q})' H_N z.$$

O caso particular $q=1$ foi analisado por Box, Hillmer e Tiao (1976). Considerando um processo MA(q) para uma única série, isto é, $k=1$, o resultado encontrado em (3.59) ou (3.61) coincide com os obtidos por Tiao e Ali (1971), Newbold (1974), Ljung e Box (1976) e Dent (1977).

Há também uma expressão equivalente a (3.61) obtida por Osborn (1977).

Podemos utilizar a matriz $D_{\lambda,N}^{-1}$ na forma:

$$D_{\lambda, N}^{-1} = \begin{bmatrix} I & \dots & \dots & 0 \\ \pi_1(\lambda) & \ddots & & \\ \vdots & \ddots & & \\ \pi_{N-1}(\lambda) & \dots & \pi_1(\lambda) & I \end{bmatrix} \quad (3.62)$$

onde

$$\pi_j(\lambda) = \lambda_1 \pi_{j-1}(\lambda) + \dots + \lambda_q \pi_{j-q}(\lambda), \quad (3.63)$$

$j = 1, \dots, N-1$; sendo

$$\pi_0(\lambda) = I_k \text{ e, para } q \geq 2, \pi_{-1}(\lambda) = \dots = \pi_{-q+1}(\lambda) = 0.$$

Ao calcularmos a verossimilhança através de (3.59) para (θ, λ) dados, podemos seguir as etapas:

i) escrever $M' = (m'_{-q+1}, \dots, m'_N)$ e portanto,

$$m_{-q+1} = \dots = m_0 = 0 \text{ e } m_j = z_j + \theta_1 m_{j-1} + \dots + \theta_q m_{j-q},$$

$j = 1, \dots, N$.

ii) se $P = F_{\theta, q}$, então a matriz X é constituída dos primeiros q blocos colunas de $D_{\theta, N+q}^{-1}$ de tal maneira que os termos de X podem ser obtidos por θ , usando a relação (3.63).

Dessa maneira, os elementos de X e M podem ser determinados recursivamente.

Para a obtenção das estimativas de máxima verossimilhança, observemos que o expoente de (3.59) pode ser expressado por

$$M' \Phi_{N+q}^{-1} M - \hat{b}' X' \Phi_{N+q}^{-1} X \hat{b} = \sum_{t=1-q}^N \hat{a}'_t \Phi_t^{-1} \hat{a}_t \quad (3.64)$$

onde

$$\hat{a}_t = z_t + \theta_1 \hat{a}_{t-1} + \dots + \theta_q \hat{a}_{t-q}, \quad t=1, \dots, N$$

e

$$\hat{a}'_t = (\hat{a}'_{t-q+1}, \dots, \hat{a}'_0).$$

Sejam

$$\Phi^{-1} = H'H \text{ e } u_t = H\hat{a}_t = (u_{1t}, \dots, u_{kt})'.$$

Então,

$$\sum_{t=1-q}^N \hat{a}'_t \Phi^{-1} \hat{a}_t = \sum_{t=1-q}^N \sum_{j=1}^k u_{jt}^2.$$

Além disso, podemos verificar que $|\Phi|^q |X' \Phi_{N+q}^{-1} X| \geq 1$.

É possível, então, expressar a função de verossimilhança como

$$L(\theta, \Phi/z) \propto |\Phi|^{-1/2N} \cdot \exp\left[-\frac{1}{2} S(\theta, \Phi)\right] \quad (3.65)$$

onde

$$S(\theta, \Phi) = u_0^2 + \sum_{t=1-q}^N \sum_{j=1}^k u_{jt}^2$$

e

$$u_0^2 = \ln |\Phi|^q |X' \Phi_{N+q}^{-1} X|.$$

As estimativas de (Φ, θ) são calculadas através de iterações sucessivas, começando com uma estimativa preliminar $\theta^{(0)}$ de θ e $\Phi^{(0)}$ de Φ , onde $\Phi^{(0)} = \frac{1}{N} \sum_{t=1-q}^N \hat{a}_t \hat{a}'_t$. Um procedimento não

linear de mínimos quadrados é, então, aplicado, objetivando a minimização de $S(\theta^{(0)}, \phi^{(0)})$ e determinando um novo valor paramétrico $\theta^{(1)}$ que vai dar origem a $\phi^{(1)}$.

O procedimento é repetido sucessivamente até que seja atingida a convergência estabelecida de maneira que mudanças paramétricas consecutivas sejam desprezíveis. É seguido o mesmo procedimento para processos mistos ARMA(p,q).

Em virtude da presença do termos u_0^2 em $S(\theta, \phi)$, as estimativas paramétricas então obtidas não são as estimativas de máxima verossimilhança exata a menos que ϕ seja conhecido.

Procedimentos alternativos que maximizam diretamente a expressão (3.59) necessitam ser desenvolvidos.

Há dificuldades na obtenção das estimativas de máxima verossimilhança exata principalmente devido à necessidade de calcularmos $\hat{a}_*$ e u_0^2 .

Aplicando o resultado encontrado em (3.61), a função de verossimilhança de $\phi = (\phi_1, \dots, \phi_q)$, θ e ϕ assumirá a forma:

$$L_1(\theta, \phi, \phi/\omega) \propto |\phi|^{-1/2N} |A_{N-q,q}|^{-1/2} \cdot \exp\left(-\frac{1}{2}[Y_2' H_{N-q} Y_2 - \tilde{\alpha}_1' A_{N-q,q} \tilde{\alpha}_1]\right) \quad (3.66)$$

onde

$$\tilde{\alpha}_1 = -A_{N-q,q}^{-1} (\Delta_{N-q,q} C_{\theta,q})' H_{N-q} Y_2.$$

Por outro lado, a verossimilhança associada à distribuição condicional $p(Y_1/Y_2)$ é definida por:

$$L_2(\theta, \phi, \lambda/\omega) \propto |\Gamma + F_{\theta, q} A_{N-q, q}^{-1} F'_{\theta, q}|^{-1/2} \cdot \exp \left[-\frac{1}{2} Y'_{1.} (\Gamma + F_{\theta, q} A_{N-q, q}^{-1} F'_{\theta, q})^{-1} Y_{1.} \right] \quad (3.67)$$

onde $Y_{1.} = Y_1 - F_{\theta, q} \tilde{\alpha}_1$ e Γ é a matriz de covariância de v .

Temos, agora, a função completa de verossimilhança:

$$L(\theta, \phi, \lambda/\omega) \propto L_1(\theta, \phi, \lambda/\omega) L_2(\theta, \phi, \lambda/\omega). \quad (3.68)$$

Para um modelo misto ARMA(q, q) expressões explícitas para Γ são complicadas. Mostraremos, agora, que à medida que N cresce, $\ln L_1$ tende a prevalecer em relação a $\ln L_2$, independentemente da condição de $|\theta_q(B)|$ ter ou não ter raízes no círculo unitário. Assim, para N grande, podemos aproximar $L(\theta, \phi, \lambda/\omega)$ de $L_1(\theta, \phi, \lambda/\omega)$.

Vejamos alguns resultados preliminares.

1º) Sejam B_1 e B_2 duas matrizes positivas definidas simétricas de mesma dimensão. Usaremos a notação $B_1 \geq B_2$ para considerarmos $B_1 - B_2$ uma matriz positiva semi-definida. A relação $B_1 \geq B_2$ implica que $B_1^{-1} \leq B_2^{-1}$.

Consideremos a matriz $(F_{\theta, q} A_{N-q, q}^{-1} F'_{\theta, q})^{-1}$ em (3.67).

De acordo com a definição dada a $A_{N, q}$ segue que $A_{N-q, q} \geq \lambda_q^{-1}$, uma vez que, por analogia

$$A_{N-q, q} = \lambda_q^{-1} + (\Delta_{N-q, q} C_{\theta, q})' H_{N, q} (\Delta_{N-q, q} C_{\theta, q}).$$

Temos ainda, que Γ é uma matriz positiva definida. Aplicando sucessivamente os resultados citados neste primeiro ítem, temos:

$$\begin{aligned} \lambda_q^{-1} &\leq A_{N-q,q}, \quad \Gamma + F_{\theta,q} A_{N-q,q}^{-1} F'_{\theta,q} \geq \Gamma \\ \Rightarrow (\Gamma + F_{\theta,q} \lambda_q F'_{\theta,q})^{-1} &\leq (\Gamma + F_{\theta,q} A_{N-q,q}^{-1} F'_{\theta,q})^{-1} \leq \Gamma^{-1} \quad (3.69) \end{aligned}$$

Logo, os valores característicos de $(\Gamma + F_{\theta,q} A_{N-q,q}^{-1} F'_{\theta,q})^{-1}$ são limitados para qualquer valor de θ .

2º) Suponhamos que N seja um múltiplo de q , por exemplo, $m = N/q$. Isto resultará numa simplificação na obtenção dos resultados assintóticos.

Podemos escrever:

$$\begin{aligned} D_{\theta, N-q} &= \begin{bmatrix} F_{\theta,q} & & & 0 \\ -C_{\theta,q} & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -C_{\theta,q} & F_{\theta,q} \end{bmatrix} \\ D_{\theta, N-q}^{-1} &= \begin{bmatrix} I & & & 0 \\ G & \ddots & & \\ \vdots & \ddots & \ddots & \\ G^{m-2} & \dots & G & I \end{bmatrix} \begin{bmatrix} F_{\theta,q}^{-1} & & & 0 \\ & \ddots & & \\ 0 & & & F_{\theta,q}^{-1} \end{bmatrix} \quad (3.70) \end{aligned}$$

onde $G = F_{\theta,q}^{-1} C_{\theta,q}$.

Então:

$$A_{N-q,q} = \lambda_q^{-1} + \sum_{j=1}^{m-1} G^j \lambda_q^{-1} G^j = R_{(m-1)}. \quad (3.71)$$

Ainda temos que:

$$(\Delta_{N-q,q}^C \theta, q)' H_{N-q} = \\ = [G'R_{(m-2)} F_{\theta,q}^{-1}, G'^2 R_{(m-3)} F_{\theta,q}^{-1}, \dots, G'^{m-1} R_{(0)} F_{\theta,q}^{-1}],$$

onde $R_{(0)} = \frac{1}{q}^{-1}$.

Logo, se escrevermos $Y'_2 = (y'_2, \dots, y'_m)$ onde cada componente é um vetor de dimensão $1 \times qk$, poderemos escrever a quantidade $\tilde{\alpha}_1$ em (3.66) como segue:

$$\tilde{\alpha}_1 = -R_{(m-1)}^{-1} \sum_{j=1}^{m-1} G'^j R_{(m-1-j)} F_{\theta,q}^{-1} y_{j+1}. \quad (3.72)$$

3º) Seja $G = L^q$,

$$L = \begin{bmatrix} 0 & & I & & \\ & \ddots & & \ddots & \\ & & 0 & I & \\ \theta_q & \dots & \theta_2 & \theta_1 & \end{bmatrix}. \quad (3.73)$$

O polinômio característico de L é $|\lambda^q I - \lambda^{q-1} \theta_1 - \dots - \theta_q| = 0$.

Portanto, os valores característicos de L são os recíprocos das raízes de $|\theta_q(B)|$.

4º) Existe uma matriz não singular H tal que $L = H \Lambda H^{-1}$ onde Λ é a forma canônica de Jordan de L , isto é,

$$\Lambda = \begin{bmatrix} \lambda_1 & \delta_1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \delta_{qk-1} & \\ & & & \lambda_{qk} & \end{bmatrix} \quad (3.74)$$

sendo δ_i igual a 1 ou 0. Então, $G^m = H \Lambda^N H^{-1}$ onde Λ^N é uma ma-

triz triangular superior de dimensão $qk \times qk$ de tal maneira que

- a) o i -ésimo elemento diagonal é λ_i^N ;
- b) para $N > qk$, o módulo do valor máximo dos elementos da diagonal é limitado por

$$\left\{ \begin{matrix} N \\ qk-1 \end{matrix} \right\} |\lambda_*|^{N-qk+1}, \text{ onde } |\lambda_*| = \max_i |\lambda_i|.$$

Voltando à formulação em (3.67), consideremos

$$\begin{aligned} -\frac{2}{N} \ln L_2(\theta, \phi, \frac{1}{\omega}) &= -\frac{1}{N} \ln |\Gamma + F_{\theta, q} A_{N-q, q}^{-1} F'_{\theta, q}|^{-1} + \\ &+ \frac{1}{N} Y'_1 (\Gamma + F_{\theta, q} A_{N-q, q}^{-1} F'_{\theta, q})^{-1} Y_1. \end{aligned} \quad (3.75)$$

De acordo com (3.69), o primeiro termo do lado direito vai se aproximar de zero quando $N \rightarrow \infty$ para todo valor de θ .

Suponhamos que $|\theta_q(B)|$ tenha todas as suas raízes fora do círculo unitário. Através de (3.74), vemos que à medida que N cresce, G^m aproxima-se da matriz nula. Conseqüentemente, a matriz $R_{(m-1)}$ em (3.71) tenderá a uma matriz positiva definida R finita.

Assim, dado que os elementos de ω sejam finitos, os elementos do vetor $\tilde{\alpha}_1$ em (3.72) serão limitados. Este resultado juntamente com (3.69) implica que o segundo termo do lado direito de (3.75) vai tender a zero.

Voltemos, agora, a (3.66). Teremos:

$$\begin{aligned} -\frac{2}{N} \ln L_1(\theta, \phi, \frac{1}{\omega}) &= \ln |\frac{1}{\omega}| + \frac{1}{N} Y'_2 H_{N-q} Y_2 + \\ &+ \frac{1}{N} (\ln |A_{N-q, q}| - \tilde{\alpha}'_1 A_{N-q, q} \tilde{\alpha}_1). \end{aligned} \quad (3.76)$$

A medida que N cresce, o último termo de (3.76) aproxima-se de zero. Como $H_{N-q} = D_{\theta, N-q}^{-1} \Phi_{N-q}^{-1} D_{\theta, N-q}^{-1}$, temos a partir de (3.70) que

$$\frac{1}{N} Y_2' H_{N-q} Y_2 = \frac{1}{mq} \sum_{j=2}^m u_j' \Phi_q^{-1} u_j, \quad (3.77)$$

onde

$$u_j = \sum_{k=2}^j G^{k-2} F_{\theta, q}^{-1} Y_{j-(k-2)},$$

cujos elementos são limitados para todo valor de j . Daí decorre que $Y_2' H_{N-q} Y_2$ aumenta de valor de acordo com o aumento de N .

Portanto, é possível concluir que há predomínio da influência de L_1 sobre a de L_2 e, assim, a função de verossimilhança pode ser aproximada por

$$L(\theta, \phi, \Phi/\omega) \propto |\Phi|^{-N/2} \cdot \exp\left(-\frac{1}{2} Y_2' H_{N-q} Y_2\right). \quad (3.78)$$

Suponhamos, agora, que $|\theta_q(B)|$ tem r ($r \leq kq$) de suas raízes no círculo unitário.

Então, por (3.74), os elementos de G^m podem ser de ordem tão alta quanto N^{r-1} e, assim, os elementos de $A_{N-q, q}$ serão de ordem N pelo menos.

Ainda neste caso L_1 prevalece em relação a L_2 . De fato, primeiramente vamos escrever o expoente em (3.66) na forma:

$$Y_2' H_{N-q} Y_2 - \tilde{\alpha}_1' A_{N-q, q} \tilde{\alpha}_1 = Y_1' H_N Y_1 - Y_1' F_{\theta, q}^{-1} A_{N-q, q} F_{\theta, q}^{-1} Y_1 \quad (3.79)$$

Como, por (3.69), $\frac{1}{N} (\Gamma + F_{\theta, q} \Lambda_{N-q, q}^{-1} F_{\theta, q}')^{-1}$ aproxima-se da

matriz nula conforme N aumenta de valor, temos que

$$\frac{1}{N} Y_1' F_{\theta,q}^{-1} A_{N-q,q} F_{\theta,q}^{-1} Y_1$$

prevalecerá em relação ao segundo termo do lado direito de (3.75).

Concluimos, assim, que num modelo ARMA(q,q), independentemente de $|\theta_q(B)|$ ter ou não raízes no círculo unitário, para N grande, podemos dispensar a contribuição de $L_2(\theta, \phi, \lambda/\omega)$ e aproximar a função de verossimilhança de $L_1(\theta, \phi, \lambda/\omega)$.

Daremos início, agora, à extensão do resultado encontrado na determinação da função de verossimilhança de um processo ARMA(q,q) para o caso geral, ou seja, ARMA(p,q).

Primeiramente, seja $p > q$.

Substituindo q por p em (3.54) e mudando a notação de Y_1 para Y_{1p} e Y_2 para Y_{2p} , vem:

$$\begin{bmatrix} Y_{1p} \\ Y_{2p} \end{bmatrix} = \begin{bmatrix} F_{\theta,p} & 0 \\ -\Delta_{N-p,p} C_{\theta,p} & D_{\theta,N-p} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} v \\ 0 \end{bmatrix} \quad (3.80)$$

onde

$$Y_{1p} = F_{\phi,p} \omega_1 \quad \text{e} \quad Y_{2p} = D_{\phi,N-p} \omega_2 - \Delta_{N-p,p} C_{\phi,p} \omega_1.$$

Segue que $\theta_{q+1} = \dots = \theta_p = 0$ em $F_{\theta,p}$, $C_{\theta,p}$ e $D_{\theta,N-p}$. A função de verossimilhança deste processo também poderá ser aproximada por (3.66) após mudanças convenientes na notação. Como $-\Delta_{N-p,p} C_{\theta,p} = [0, -\Delta_{N-p,p} C_{\theta,q}]$, temos por (3.80) que:

$$Y_{2p} = D_{\theta, N-p} \alpha_2 - \Delta_{N-p, q} C_{\theta, q} \alpha_{1*} \quad (3.81)$$

onde $\alpha_{1*}' = (a'_{p-q+1}, \dots, a'_p)$.

Logo, a verossimilhança associada a $p(Y_2)$ em (3.66) pode ser alternadamente escrita como

$$L_{1*}(\theta, \phi, \lambda/\omega) \alpha |\lambda|^{-1/2(N-p+q)} \cdot |A_{N-p, q}|^{-1/2} \cdot \exp\left(-\frac{1}{2} [Y_{2p}' H_{N-p} Y_{2p} - \tilde{\alpha}_{1*}' A_{N-p, q} \tilde{\alpha}_{1*}]\right), \quad (3.82)$$

onde $\tilde{\alpha}_{1*} = -A_{N-p, q}^{-1} (\Delta_{N-p, q} C_{\theta, q})' H_{N-p} Y_{2p}$.

Há vantagem nesta nova formulação pois será invertida a matriz $A_{N-p, q}$ de dimensão $qk \times qk$ ao invés da matriz $A_{N-p, p}$ de dimensão $pk \times pk$ implícita em (3.66).

Veremos, agora, que a aproximação (3.82) também é válida quando $p < q$.

Suponhamos que o vetor Y é particionado em Y_1 e Y_2 como em (3.54), onde fica implícito que $\phi_{p+1}, \dots, \phi_q$ são considerados nulos em $D_{\phi, N}$ ao se fazer a transformação $Y = D_{\phi, N} \omega$.

Além disso, vamos decompor Y_1 em Y_{1p} e Y_{1*} .

Anteriormente obtivemos que a verossimilhança associada a $p(Y_2)$ prevalece em relação a de $p(Y_1 | Y_2)$. Analogamente a verossimilhança associada a $p(Y_2, Y_{1*})$ prevalecerá em relação a de $p(Y_{1p} | Y_2, Y_{1*})$. Como $Y_{2p}' = [Y_{1*}', Y_2']$ e a estrutura de (3.81) é válida também para $p < q$, segue que a função de verossimilhança pode ser aproximada por L_{1*} em (3.82).

Portanto, obtivemos uma aproximação para a função de ve-

verossimilhança do modelo misto ARMA(p,q), efetuando as seguintes etapas:

- a) aplicar a transformação $Y = D_{\phi, N} \omega$;
- b) ignorar os elementos Y_{1p} , e
- c) considerar Y_{2p} como as N-p observações vetoriais de um processo MA(q).

Cumpra enfatizar que na obtenção da aproximação (3.82) não fizemos suposição alguma a respeito da invertibilidade do polinômio de médias móveis $\theta_q(B)$.

3.6 - COMPARAÇÃO ENTRE AS METODOLOGIAS EMPREGADAS DURANTE

A ESTIMAÇÃO DE MODELOS ARMA MULTIVARIADOS

Primeiramente, compararemos as expressões relativas à função de verossimilhança obtidas por Nicholls e Hall em (3.10) com a encontrada por Wilson em (3.26).

Em (3.10) a parte exponencial é igual a:

$$-\frac{1}{2}(\psi\omega + \Lambda\hat{e}_*)'(\psi\omega + \Lambda\hat{e}_*) = -\frac{1}{2}\hat{u}'\hat{u} = -\frac{1}{2}\sum_{t=1-p-q}^N \hat{a}_t' \Phi^{-1} \hat{a}_t.$$

Em (3.26) os termos da parte exponencial relativos a a_t , t de 1-p-q até 0, são omitidos devido à atribuição do valor zero aos resíduos iniciais na metodologia de Wilson.

O efeito dos resíduos iniciais sobre u em (3.6) é determinado pela matriz Λ , a qual é constituída de duas partições: $I_{(p+q)k}$ e $(I_N \otimes R)XT^{-1}$.

No caso do modelo ARMA univariado determinou-se a maneira de se obter a matriz X_n na seção 2.4.1. Fazendo as devidas

adaptações para o caso multivariado, podemos delinear a matriz X , componente da matriz Λ , de modo análogo. No próximo capítulo, a expressão verificada em (4.26) também é aplicável na matriz X do processo ARMA multivariado.

Conseqüentemente, como estão sendo consideradas satisfeitas as condições de estacionariedade e invertibilidade, segue que $X_{t+q,j} \rightarrow 0$ conforme $t \rightarrow \infty$.

Assim, os resíduos iniciais desconsiderados por Wilson têm um efeito irrelevante assintoticamente e as formulações (3.10) e (3.26) neste item de comparação atingem um nível de proximidade ainda mais acentuado.

Novamente comparando (3.10) com (3.26), podemos observar nesta última expressão a presença do termo $|\Phi|^{-N/2}$ somente, no lugar da expressão $|\Phi|^{-N/2} |\Lambda' \Lambda|^{-1/2}$ apresentada em (3.10).

Em decorrência da propriedade da matriz X ter componentes que decrescem no decorrer do tempo, temos que a matriz Λ tenderá, no limite, a uma matriz finita correspondente à sua primeira partição, cuja influência é pouco significativa na expressão (3.10).

Portanto, as expressões (3.10) e (3.26) mantêm em sua estrutura um forte nível de aproximação.

No tocante à carga computacional necessária para a determinação das estimativas paramétricas a partir da função de verossimilhança, convém ponderar a metodologia de Nicholls e Hall em confronto com a técnica de Wilson.

Na seção 3.3 é apresentada uma expressão em função de Φ ,

ϕ_i e θ_j , através da qual obtêm-se as matrizes Ω e T . Consequentemente, fica determinada a matriz A necessária para o delineamento da função de verossimilhança exata. A partir dessas expressões poderíamos, então, através de algoritmos não lineares, tais como o de Newton-Raphson, chegar às estimativas paramétricas de modelos ARMA multivariados.

Na seção 3.4 é descrito o método seguido por Wilson na estimação dos parâmetros destes modelos. Também surge a necessidade de se contornar problemas não lineares e, para isso, Wilson recomenda o método de Marquardt (1963). Foi visto que, através do esquema de estimação proposto por Wilson, ocorre uma redução na quantidade de cálculos computacionais e esta propriedade deve ser avaliada com sensatez na decisão sobre o método a ser seguido, tendo em vista as vantagens que oferece.

Não será feita uma comparação entre os métodos de Nicholls e Hall e Wilson com as técnicas de Hillmer e Tiao, tendo em vista que estes últimos abordaram modelos ARMA multivariados não invertíveis apenas.

CAPÍTULO 4

ESTIMAÇÃO EM MODELOS MA E AR MULTIVARIADOS

4.1 - INTRODUÇÃO

O conteúdo deste capítulo refere-se ao enfoque das diversas técnicas empregadas com o fim de estimar os parâmetros em modelos MA e AR multivariados, generalização dos respectivos modelos univariados expostos no Capítulo 2.

4.2 - METODOLOGIA DE OSBORN

Osborn (1977) desenvolveu uma técnica para obter a função de verossimilhança exata de modelos MA multivariados, baseando-se numa extensão do método adotado por Box & Jenkins (1970) no caso univariado.

Nesta seção é apresentada sua metodologia, a qual encontra um procedimento paralelo seguido por Newbold (1974) na estimação de modelos ARMA univariados.

Um processo vetorial MA pode ser expressado na forma:

$$\tilde{\omega}_t = \theta_q(B) a_t \quad (4.1)$$

onde: $\theta_q(B) = (I_k - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$;

$\tilde{\omega}_t$ = vetor de dimensão $k \times 1$;

B = operador retroativo tal que $B\tilde{\omega}_t = \tilde{\omega}_{t-1}$;

θ_i ($i=1, \dots, q$) = matrizes paramétricas desconhecidas com dimensão $k \times k$;

a_t 's ($t=0, \pm 1, \pm 2, \dots$) = vetores de dimensão $k \times 1$ independentes e identicamente distribuídos como $N(0, \Phi)$, onde Φ é uma matriz finita de dispersões.

Suponhamos que $\mu_\omega = 0$ e, conseqüentemente, $\tilde{\omega}_t = \omega_t - \mu_\omega = \omega_t$.

Com N observações amostrais de ω_t , vamos proceder à estimação dos parâmetros desconhecidos θ_i ($i=1, \dots, q$) e a matriz de covariância Φ juntamente com os valores "pre-períodos" também desconhecidos a_j ($j=0, -1, \dots, -q+1$).

Definimos, agora, o vetor completo das perturbações:

$$a = (a'_{1-q}, \dots, a'_0, a'_1, \dots, a'_N)' \quad (4.2)$$

cuja dimensão é $k(N+q) \times 1$.

Como $a \sim N(0, \Omega)$, onde $\Omega = I \otimes \Phi$, segue que sua função de densidade expressa-se na forma:

$$p(a) = (2\pi)^{\frac{1}{2}(N+q)k} |\Omega|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2} a' \Omega^{-1} a\right\} \quad (4.3)$$

Aqui, a matriz identidade, I , tem dimensão $(N+q) \times (N+q)$. Há $N+q$ equações da forma:

$$\begin{aligned} a_{1-q} &= a_{1-q} \\ &\vdots \\ a_{-1} &= a_{-1} \\ a_0 &= a_0 \\ a_1 &= \omega_1 + \theta_1 a_0 + \theta_2 a_{-1} + \dots + \theta_q a_{1-q} \\ &\vdots \\ a_N &= \omega_N + \theta_1 a_{N-1} + \theta_2 a_{N-2} + \dots + \theta_q a_{N-q} \end{aligned} \quad (4.4)$$

Substituindo sucessivamente a_1 em a_2 , a_2 em a_3 , etc. teremos a_t em função dos resíduos iniciais $a_{1-q}, \dots, a_0$ e das observações $\omega_1, \dots, \omega_N$.

Sejam $a_* = (a'_{1-q}, \dots, a'_0)'$, cuja dimensão é $kq \times 1$, e $\omega = (\omega'_1, \dots, \omega'_N)'$, cuja dimensão é $kN \times 1$.

Usando (4.2), podemos expressar o sistema (4.4) na forma:

$$a = L\omega + Xa_*, \quad (4.5)$$

onde as matrizes L e X têm dimensão $k(N+q) \times kN$ e $k(N+q) \times kq$, respectivamente. Estas matrizes são funções dos elementos de θ_i , então considerados fixos.

A transformação (4.5) tem Jacobiano unitário, levando-nos à determinação da seguinte distribuição conjunta:

$$p(\omega, a_*) = L(\omega, a_*) = (2\pi)^{\frac{1}{2}(N+q)k} |\Omega|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(L\omega + Xa_*)' \Omega^{-1} (L\omega + Xa_*)\right\} \quad (4.6)$$

Ao aplicar mínimos quadrados em (4.5), obtemos a estimativa de máxima verossimilhança $\hat{a}_*$ de a_* , expressada por:

$$\hat{a}_* = -(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} L \omega. \quad (4.7)$$

Assim, $\hat{a}_*$ é o vetor que minimiza $a' \Omega^{-1} a$ em (4.6).

Podemos decompor a forma quadrática $S(\theta, a_*) = a' \Omega^{-1} a$ em duas parcelas:

$$S(\theta, a_*) = S(\theta, \hat{a}_*) + (a_* - \hat{a}_*)' X' \Omega^{-1} X (a_* - \hat{a}_*) \quad (4.8)$$

onde

$$S(\theta, \hat{a}_*) = (L\omega + X\hat{a}_*)' \Omega^{-1} (L\omega + X\hat{a}_*). \quad (4.9)$$

A decomposição (4.8) é válida pois:

$$\begin{aligned} S(\theta, a_*) &= (L\omega + Xa_*)' \Omega^{-1} (L\omega + Xa_*) \\ &= [\underbrace{L\omega + X\hat{a}_*}_A + \underbrace{X(a_* - \hat{a}_*)}_B]' \Omega^{-1} [\underbrace{L\omega + X\hat{a}_*}_A + \underbrace{X(a_* - \hat{a}_*)}_B] \\ &= (A' + B') \Omega^{-1} (A + B) \\ &= A' \Omega^{-1} A + B' \Omega^{-1} B. \end{aligned}$$

Portanto,

$$S(\theta, a_*) = (L\omega + X\hat{a}_*)' \Omega^{-1} (L\omega + X\hat{a}_*) + (a_* - \hat{a}_*)' X' \Omega^{-1} X (a_* - \hat{a}_*).$$

Os produtos cruzados $A' \Omega^{-1} B$ e $B' \Omega^{-1} A$ foram omitidos por serem iguais a zero. Vejamos um deles como exemplo

$$B' \Omega^{-1} A = (a_* - \hat{a}_*)' X' \Omega^{-1} (L\omega + X\hat{a}_*)$$

$$B' \Omega^{-1} A = (a_* - \hat{a}_*)' X' \Omega^{-1} L\omega + (a_* - \hat{a}_*)' X' \Omega^{-1} X\hat{a}_*.$$

De (4.7) vem:

$$B' \Omega^{-1} A = (a_* - \hat{a}_*)' X' \Omega^{-1} L\omega - (a_* - \hat{a}_*)' X' \Omega^{-1} L\omega = 0.$$

A forma quadrática $S(\theta, \hat{a}_*)$ em (4.9) é uma função das observações ω mas não dos resíduos iniciais a_* , o mesmo ocorrendo com $\hat{a}_*$.

Como $p(\omega, a_*) = p(\omega)p(a_*/\omega)$, utilizando (4.6) e (4.8) segue que:

$$p(\omega, a_*) = (2\pi)^{-1/2 Nk} |\Omega|^{-1/2} |X' \Omega^{-1} X|^{-1/2} \cdot \exp \left\{ -\frac{1}{2} (L\omega + X\hat{a}_*)' \Omega^{-1} \cdot \right. \\ \cdot (L\omega + X\hat{a}_*) \left. \right\} (2\pi)^{-1/2 qk} |X' \Omega^{-1} X|^{1/2} \cdot \exp \left\{ -\frac{1}{2} (a_* - \hat{a}_*)' X' \Omega^{-1} X \cdot \right. \\ \cdot (a_* - \hat{a}_*) \left. \right\}.$$

Portanto, a função de verossimilhança incondicional exata é dada por:

$$p(\omega) = L(\omega) = (2\pi)^{-1/2 Nk} |\Omega|^{-1/2} |X' \Omega^{-1} X|^{-1/2} \\ \cdot \exp \left\{ -\frac{1}{2} (L\omega + X\hat{a}_*)' \Omega^{-1} (L\omega + X\hat{a}_*) \right\} \quad (4.10)$$

4.3 - METODOLOGIA DE WILSON

De acordo com a seção 3.4 do Capítulo 3 é deduzida a função de máxima verossimilhança condicional para modelos ARMA multivariados. O método se baseia no fato de não serem considerados os efeitos dos valores iniciais. Novamente podemos seguir as técnicas empregadas por Wilson (1973), supondo, desta vez, que $\phi_p(B) = I$.

Além disso, vamos nos basear na hipótese de que as raízes dos determinantes polinomiais em B, $|\theta_q(B)| = 0$, estão fora do círculo unitário para garantir a invertibilidade do modelo (4.1).

Dessa maneira, segue que a função de verossimilhança condicional é dada por:

$$p(\omega) = L(\omega) = (2\pi)^{-1/2 Nk} |\Phi|^{-N/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^N a_t' \Phi^{-1} a_t \right\}. \quad (4.11)$$

4.4 - METODOLOGIA DE PHADKE E KEDEM

Phadke e Kedem (1978) propuseram três métodos a fim de obter a função de verossimilhança exata para modelos MA multivariados. Veremos como cada um deles emprega a estrutura da matriz de covariância. Antes, porém, observemos que nesta nova metodologia é imposta a condição de invertibilidade mencionada na seção 4.3.

Denotaremos por γ_s a matriz de dimensão $k \times k$ que consiste na covariância de defasagem s , ou seja, $\gamma_s = E\{\omega_t \omega_{t+s}'\}$.

Segue, então, que:

$$\gamma_s = \gamma_{-s}' = \begin{cases} \sum_{j=0}^{q-s} \theta_j \theta_{j+s}', & 0 \leq s \leq q \\ 0, & s > q. \end{cases}$$

Seja ω o vetor constituído de N observações consecutivas ω_t , $t = 1, \dots, N$, ou seja,

$$\omega' = \{\omega_{11}, \dots, \omega_{k1}, \omega_{12}, \dots, \omega_{k2}, \dots, \omega_{1N}, \dots, \omega_{kN}\}.$$

A matriz de covariância de ω , denotada por Γ , é uma matriz Toeplitz, simétrica, em blocos, na qual a primeira linha de blocos com dimensão $k \times k$ é representada por $\gamma_0, \gamma_1, \dots, \gamma_{N-1}$, isto é

$$\Gamma = \text{TSB}\{\gamma_0, \dots, \gamma_{N-1}\}.$$

Uma vez que $\gamma_s = 0$ para $s > q$, temos como consequência o fato de Γ ter uma estrutura "band" em blocos.

Consideremos a_t e conseqüentemente ω_t normalmente distribuídos tal que a função de verossimilhança pode ser expressa da por:

$$p(\omega) = L(\omega) \alpha |\Gamma|^{-1/2} \cdot \exp\left\{-\frac{\omega' \Gamma^{-1} \omega}{2}\right\}.$$

Portanto,

$$\ln p(\omega) \propto -\frac{1}{2}(\ln |\Gamma| + \omega' \Gamma^{-1} \omega). \quad (4.12)$$

Como já foi comentado, o procedimento usual para o cálculo das estimativas de máxima verossimilhança (durante o qual é montado um sistema de equações relativas às derivadas do logaritmo da função de verossimilhança igualadas a zero) não é aplicável. Isto decorre do fato dessas equações serem altamente não lineares e a sua solução analítica incorre em cálculos laboriosos.

Um possível artifício seria maximizar numericamente o logaritmo da função de verossimilhança. Entretanto, surgem dificuldades na obtenção de Γ^{-1} e seu determinante.

Todos esses obstáculos deram origem ao desenvolvimento de novas técnicas visando a proposição de métodos aproximados direcionados para o cálculo e maximização da função de verossimilhança.

Nesta seção veremos algoritmos eficientes com esse fim.

Primeiramente, vamos estabelecer duas denominações distintas no que concerne à função de verossimilhança, a saber: função de verossimilhança aproximada e exata.

A primeira está ligada à atribuição do valor nulo aos com-

ponentes da série do ruído branco correspondentes às posições cronológicas retroativas, isto é, $a_0, a_{-1}, \dots, a_{-q+1}$.

Feito isso, obtemos recursivamente

$$a_t = \omega_t + \sum_{j=1}^q \theta_j a_{t-j}, \quad t=1, \dots, N. \quad (4.13)$$

Impondo a condição $a_0 = a_{-1} = \dots = a_{-q+1} = 0$, o logaritmo da função de verossimilhança, $\ln L(\xi/\omega)$, é dado por:

$$\ln L(\xi/\omega) = \ln p(\omega) = -\frac{N}{2} \ln |\Phi| - \frac{1}{2} \sum_{t=1}^N a_t' \Phi^{-1} a_t.$$

Através da maximização de $\ln p(\omega)$, encontramos as estimativas de máxima verossimilhança aproximada.

Outra alternativa para determinarmos $\ln p(\omega)$ reside na aproximação da matriz de covariância de ω pela matriz $\Gamma_t = LDL'$, onde $L = (L_{jr})$ é uma matriz triangular inferior de dimensão $Nk \times Nk$, cujos componentes seguem a relação

$$L_{jr} = \begin{cases} \theta_{j-r}, & 0 \leq j-r \leq q, \quad j, r = 1, \dots, N \\ 0, & \text{caso contrário.} \end{cases} \quad (4.14)$$

e $D = (D_{jj})$ é uma matriz diagonal em blocos de dimensão $Nk \times Nk$ tal que

$$D_{jj} = \Phi, \quad j=1, \dots, N.$$

Trataremos, agora, do item relativo à função de verossimilhança exata.

Serão apresentados dois métodos a seguir.

O primeiro método é baseado na decomposição de Cholesky.

Dado que Γ é uma matriz positiva definida, simétrica, com dimensão $[2(q+1)k-1]$, sua decomposição de Cholesky é expressa por $\Gamma = LDL'$, onde L é uma matriz triangular inferior de dimensão $Nk \times Nk$, sendo a matriz diagonal "band" de dimensão $(q+1)k$ com as entradas da diagonal principal iguais à unidade e D é uma matriz diagonal de dimensão $Nk \times Nk$.

A partir de (4.12), segue que os dois componentes do logaritmo da função de verossimilhança podem ser determinados por:

$$\omega' \Gamma^{-1} \omega = (L^{-1} \omega)' D^{-1} (L^{-1} \omega) \quad (4.15)$$

$$\ln |\Gamma| = \ln |D| = \sum_{r=1}^{Nk} \ln D_{rr}.$$

Tendo um conjunto de valores paramétricos como um ponto de partida, para calcular o logaritmo da função de verossimilhança basta seguir os passos:

- i) calcular as covariâncias γ_s , $s = 0, \dots, q$;
- ii) calcular os fatores L e D através de um algoritmo de triangularização para matrizes "band";
- iii) calcular $L^{-1}y = u$ através de um algoritmo de substituição retroativa para matrizes "band";
- iv) calcular o produto interno ponderado $u'D^{-1}u$ e $\ln |D|$.

É possível reduzirmos o espaço de armazenagem dos dados e o tempo dispendido nos cálculos. Para obtenção de informações amplas ver Phadke e Kedem (1978, p. 514).

Vejamos, agora, o segundo método eficiente no cálculo do logaritmo da função de verossimilhança exata, denominado método no domínio do tempo.

Este processo explora o fato de que $\Gamma - \Gamma_t$ é uma matriz espaçada com posto baixo assumindo no máximo o valor qk .

É possível verificar também que

$$\Gamma = \Gamma_t + UGCG'U',$$

onde:

$C = (C_{js})$ é uma matriz diagonal em blocos de dimensão $qk \times qk$

e $C_{jj} = I_{qk}$, $j = 1, \dots, q$;

$G = (G_{js})$ é uma matriz de dimensão $qk \times qk$ tal que

$$G_{js} = \begin{cases} \theta_{j-s+q}, & s \geq j, j, s = 1, \dots, q; \\ 0, & \text{caso contrário.} \end{cases}$$

$U = (I_{qk}, 0, \dots, 0)'$ é uma matriz de dimensão $Nk \times Nk$.

Ambas as matrizes C e Γ_t são não singulares.

Podemos, então, encontrar a forma quadrática procurada para ω , ou seja:

$$\omega' \Gamma^{-1} \omega = \omega' L'^{-1} D^{-1} L^{-1} \omega - B' G (C^{-1} + G' R G)^{-1} G' B, \quad (4.16)$$

onde: $B = F' D^{-1} L^{-1} \omega$ é um vetor de dimensão $qk \times 1$,

$R = F D^{-1} F'$ é uma matriz simétrica de dimensão $qk \times qk$, e

$F = L^{-1} U$ é uma matriz de dimensão $Nk \times Nk$.

Com essa decomposição a forma quadrática desejada consiste de uma forma quadrática no domínio do tempo aproximada e um

termo de correção de posto baixo.

Há uma maneira conveniente de calcular $|\Gamma|$, a qual vem a seguir:

$$|\Gamma| = |\Gamma_t| |C| |C^{-1} + G'RG|. \quad (4.17)$$

O seguinte algoritmo tem a finalidade de calcular o logaritmo da função de verossimilhança, partindo de um vetor paramétrico com valor determinado.

Passos a serem executados:

i) calcular a_t ($t=1, \dots, N$) a partir de (4.13) supondo

$$a_0 = a_{-1} = \dots = a_{-q+1} = 0;$$

ii) calcular a forma quadrática aproximada expressada por

$$\sum_{t=1}^N a_t' \Phi^{-1} a_t;$$

iii) considerar a matriz F particionada em sub-matrizes F_{rs} de dimensão $k \times k$, definidas por:

$$F = (F_{rs}), \quad r=1, \dots, N; \quad s=1, \dots, q,$$

onde

$$F_{1,1} = I_k; \quad F_{r,1} = I_k \left\{ \sum_{j=1}^{\min(q,r)} \theta_j F_{r-j,1} \right\}, \quad r=2, \dots, N$$

$$F_{r,j} = \begin{cases} 0, & r < j, \quad j = 2, \dots, q \\ F_{r-j+1,1}, & r \geq j, \quad j = 2, \dots, q. \end{cases}$$

iv) calcular:

$$b_j = \sum_{r=1}^N F'_{r,j} \Phi^{-1} a_r, \quad j=1, \dots, q.$$

Os vetores b_j têm dimensão $k \times 1$.

Temos, então: $B = \{b'_1, \dots, b'_q\}'$;

v) considerar a matriz R particionada em sub-matrizes R_{js} de dimensão $k \times k$, definidas por:

$$R = (R_{js}), \quad j, s=1, \dots, q.$$

A matriz R é simétrica e suas entradas na parte referente ao triângulo superior são definidas por;

$$R_{js} = \sum_{r=1}^{N-s-1} F'_{r+s-j,1} \Phi^{-1} F_{r,1};$$

vi) aplicar (4.16), (4.17) e (4.12) para concluir o cálculo do logaritmo da função de verossimilhança.

4.5 - COMPARAÇÃO DOS MÉTODOS PROPOSTOS POR PHADKE E KEDEM

Como os métodos de cálculo da função de verossimilhança exata abordam a estrutura de Γ de diferentes maneiras, há diferença ao avaliarmos seus desempenhos.

Utilizaremos para comparação desses métodos uma formulação envolvendo o número de operações executadas em cada um.

Seja O = número de operações.

Assim, temos:

Método da decomposição de Cholesky

$$O \approx N \left\{ \frac{1}{2} q^2 k^3 + \frac{1}{2} q k^2 (k+4) + \frac{1}{8} k(k+3)(k+5) \right\}. \quad (4.18)$$

Método no domínio do tempo

$$O \approx N \left\{ q k^2 (2k+3) + \frac{1}{2} k(k^2+3) \right\}. \quad (4.19)$$

O termo operação está sendo aqui usado referindo-se à multiplicação ou divisão de dois números reais. Dessa maneira decorre que multiplicação de dois números complexos equivale a quatro operações.

Nas fórmulas (4.18) e (4.19) foram omitidos os termos independentes de N.

Observando (4.18) e (4.19), segue que:

- i) se $k = 1$ e $q \leq 4$ o método de Cholesky apresenta o menor 0;
- ii) se $k = 1$ e $q \geq 5$ o método no domínio do tempo apresenta o menor 0;
- iii) se $k = 2$ e $q \leq 3$ o método de Cholesky apresenta o menor 0;
- iv) se $k = 3$ ou 4 e $q \leq 2$ o método de Cholesky também conserva o menor 0.

Assim, temos um critério para escolha do método mais adequado de acordo com a especificação dos valores k e q , com os quais estivermos efetuando os cálculos para determinação da função de verossimilhança exata.

4.6 - METODOLOGIA DE HILLMER E TIAO

De acordo com a seção 3.5 do Capítulo 3 é deduzida a fun-

ção de máxima verossimilhança para modelos ARMA multivariados não invertíveis. No decorrer da dedução houve necessidade da determinação da função de máxima verossimilhança para modelos MA multivariados não invertíveis.

Usando a notação empregada por Hillmer e Tiao (1979) para processos MA(q), temos:

$$\omega = D_{\theta, N} a - \Delta_{N, q} C_{\theta, q} a_*$$

De acordo com (3.59) e (3.61), seguem duas maneiras equivalentes de expressar a função de verossimilhança exata:

$$\begin{aligned} L(\theta, \Phi/\omega) &= p(\omega) \alpha |\Phi|^{-1/2 (N+q)} \cdot |X' \Phi_{N+q}^{-1} X|^{-1/2} \cdot \\ &\cdot \exp \left\{ -\frac{1}{2} \left[M' \Phi_{N+q}^{-1} M - \hat{b}' X' \Phi_{N+q}^{-1} X \hat{b} \right] \right\}. \end{aligned} \quad (4.20)$$

ou

$$\begin{aligned} L(\theta, \Phi/\omega) &= p(\omega) \alpha |\Phi|^{-1/2 (N+q)} |A_{N, q}|^{-1/2} \cdot \\ &\cdot \exp \left\{ -\frac{1}{2} \left[\omega' H_N \omega - \hat{a}_*' A_{N, q} \hat{a}_* \right] \right\}. \end{aligned} \quad (4.21)$$

A formulação apresentada em (4.21) equivale ao resultado encontrado por Osborn (1977) e apresentado em (4.10).

Utilizando a notação usada por Hillmer e Tiao, o resultado obtido por Wilson e destacado em (4.11) assume a seguinte representação:

$$L(\theta, \Phi/\omega, a_* = 0) = p(\omega) \alpha |\Phi|^{-1/2 N} \cdot \exp \left(-\frac{1}{2} \omega' H_N \omega \right). \quad (4.22)$$

Novamente convém enfatizar que a técnica de Wilson des-

considera os efeitos dos valores iniciais. Portanto, os termos

$$\hat{a}_* A_{N,q} \hat{a}_* \text{ e } |\lambda|^{-1/2 q} |A_{N,q}|^{-1/2}$$

são omitidos.

Na próxima seção veremos uma comparação das propriedades da função de verossimilhança condicional e exata, feita por Osborn (1977).

4.7 - COMPARAÇÃO DA FUNÇÃO DE VEROSSIMILHANÇA EXATA COM A CONDICIONAL PARA UM PROCESSO MA(q) INVERTÍVEL

Compararemos, a seguir, as expressões (4.10) e (4.11) com o objetivo de considerar seus aspectos computacionais confrontados.

Contudo, primeiramente, trataremos a respeito do problema de identificação de processos MA multivariados.

Sabemos que no caso univariado para um processo MA(q) pode haver 2^q processos com as mesmas autocovariâncias. No caso multivariado, o número de fatores em $|\theta_q(B)|$ determina o número de maneiras independentes nas quais as covariâncias poderiam ter sido geradas (Quenouille, 1957, Cap. 3). Entretanto, somente um dos processos tem todas as raízes fora do círculo unitário, raízes essas provenientes da equação

$$|\theta_q(B)| = 0. \quad (4.23)$$

Por outro lado, a função de densidade conjunta (4.10) é

uma densidade normal multivariada, sendo, portanto, completamente especificada por seus dois primeiros momentos. Conseqüentemente, surge o problema de identificação relativo às covariâncias, o qual acarreta a possibilidade de haver 2^r processos com a mesma função de verossimilhança, onde r é o número de raízes em (4.23). Não há possibilidade de distinguir os processos com base nas informações amostrais.

Na etapa referente à obtenção da função de verossimilhança exata seguindo a técnica de Osborn não foi feita suposição a respeito das raízes do processo. Entretanto, na fase relativa à estimação será necessário impor algumas restrições visando a identificação dos parâmetros.

Por essa razão, Osborn passa a requerer que todas as raízes de (4.23) estejam fora do círculo unitário.

Convém destacar ainda que ao considerarmos (4.11) como uma aproximação de (4.10), é necessário impormos essa condição referente às raízes a fim de garantirmos a validade da aproximação.

Vamos, agora, examinar os termos envolvidos nas expressões (4.10) e (4.11).

Em (4.10) a parte exponencial é igual a:

$$\begin{aligned} -\frac{1}{2} (L\omega + X\hat{a}_*)' \Omega^{-1} (L\omega + X\hat{a}_*) &= -\frac{1}{2} \hat{a}' \Omega^{-1} \hat{a} = \\ &= -\frac{1}{2} \sum_{t=1-q}^N \hat{a}_t' \Phi^{-1} \hat{a}_t. \end{aligned} \quad (4.24)$$

Assim, reportando-nos para a expressão (4.11) e comparando a sua parte exponencial com a última igualdade de (4.24), é imediato o fato da função de verossimilhança condicional ser considerada uma aproximação da função de verossimilhança exata, ambas retratadas em (4.11) e (4.10), respectivamente.

Nessa aproximação dois níveis são envolvidos. O primeiro diz respeito à atribuição do valor zero aos resíduos iniciais. Por esse motivo, os termos correspondentes a a_t , t variando de $1-q$ até 0 , não fazem parte da somatória apresentada em (4.11).

O segundo nível de aproximação é o emprego de $|\Phi|^{-N/2}$ no lugar da expressão exata

$$\begin{aligned}
 |\Omega|^{-1/2} |X' \Omega^{-1} X|^{-1/2} &= |I \otimes \Phi|^{-1/2} |X' (I \otimes \Phi)^{-1} X|^{-1/2} \\
 &= \left[|I|^k \right]^{-1/2} \left[|\Phi|^{N+q} \right]^{-1/2} |X' (I \otimes \Phi)^{-1} X|^{-1/2} \\
 &= |\Phi|^{-1/2(N+q)} |X' (I \otimes \Phi)^{-1} X|^{-1/2} \\
 &= |\Phi|^{-1/2(N+q)} |X' (I^{-1} \otimes \Phi^{-1}) X|^{-1/2} \\
 &= |\Phi|^{-1/2(N+q)} |X' (I \otimes \Phi^{-1}) X|^{-1/2} \quad (4.25)
 \end{aligned}$$

Passaremos, agora, a examinar esses níveis de aproximação individualmente.

O efeito dos resíduos iniciais sobre a em (4.5) é determinado pela matriz X , na qual cada bloco coluna segue a relação recursiva

$$X_{t+q,j} = \theta_1 X_{t+q-1,j} + \theta_2 X_{t+q-2,j} + \dots + \theta_q X_{t,j}, \quad (4.26)$$

$t = 1, \dots, N, j = 1, \dots, q,$

com valores iniciais explicitados no sistema (4.4) e que seguem a relação

$$X_{ij} = \delta_{ij} I, \quad i, j = 1, \dots, q \quad (4.27)$$

onde δ_{ij} é o delta de Kronecker.

Podemos verificar a funcionalidade de (4.26) através de um exemplo. Sejam $q = 2$ e $N = 3$.

De (4.4), vem:

$$\begin{aligned} a_{-1} &= a_{-1} \\ a_0 &= a_0 \\ a_1 &= \omega_1 + \theta_1 a_0 + \theta_2 a_{-1} \\ a_2 &= \omega_2 + \theta_1 a_1 + \theta_2 a_0 \\ a_3 &= \omega_3 + \theta_1 a_2 + \theta_2 a_1 \end{aligned}$$

$$a_{-1} = a_{-1}$$

$$a_0 = a_0$$

$$a_1 = \omega_1 + \theta_1 a_0 + \theta_2 a_{-1}$$

$$a_2 = \omega_2 + \theta_1 \omega_1 + \theta_1^2 a_0 + \theta_1 \theta_2 a_{-1} + \theta_2 a_0$$

$$a_3 = \omega_3 + \theta_1 \omega_2 + \theta_1^2 \omega_1 + \theta_1^3 a_0 + \theta_1^2 \theta_2 a_{-1} + \theta_1 \theta_2 a_0 + \theta_2 \omega_1 + \theta_1 \theta_2 a_0 + \theta_2^2 a_{-1}$$

$$\begin{bmatrix} a_{-1} \\ a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ \theta_1 & 1 & 0 \\ \theta_1^2 + \theta_2 & \theta_1 & 1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \theta_2 & \theta_1 \\ \theta_1 \theta_2 & \theta_1^2 + \theta_2 \\ \theta_1^2 \theta_2 + \theta_2^2 & \theta_1^3 + 2\theta_1 \theta_2 \end{bmatrix} \begin{bmatrix} a_{-1} \\ a_0 \end{bmatrix}$$

$$X_{2+2,2} = X_{4,2} = \theta_1 X_{3,2} + \theta_2 X_{2,2}$$

$$X_{4,2} = \theta_1 \theta_1 + \theta_2 1$$

$$X_{4,2} = \theta_1^2 + \theta_2$$

$$X_{3+2,2} = X_{5,2} = \theta_1 X_{4,2} + \theta_2 X_{3,2}$$

$$X_{5,2} = \theta_1 (\theta_1^2 + \theta_2) + \theta_2 \theta_1$$

$$X_{5,2} = \theta_1^3 + \theta_1 \theta_2 + \theta_1 \theta_2$$

$$X_{5,2} = \theta_1^3 + 2\theta_1 \theta_2$$

Se todas as raízes de (4.23) estiverem fora do círculo unitário, então $X_{t+q,j} \rightarrow 0$ à medida que $t \rightarrow \infty$.

Conseqüentemente, os resíduos iniciais têm um efeito assintoticamente desprezível.

Consideremos a seguir a segunda aproximação enunciada.

O termo $X'(I \otimes \mathbb{A}^{-1})X$ implícito em (4.10) tem como (i,j) -ésimo bloco a expressão

$$\sum_{r=1}^{N+q} X'_{ri} \mathbb{A}^{-1} X_{rj}$$

Os q primeiros blocos linhas de X formam uma matriz identidade. Assim, segue que:

$$\begin{aligned} \sum_{r=1}^{N+q} X'_{ri} \mathbb{A}^{-1} X_{rj} &= \delta_{ij} \mathbb{A}^{-1} + X'_{q+1,i} \mathbb{A}^{-1} X_{q+1,j} + \\ &+ X'_{q+2,i} \mathbb{A}^{-1} X_{q+2,j} + \dots + X'_{q+N,i} \mathbb{A}^{-1} X_{q+N,j} \end{aligned} \quad (4.28)$$

Novamente, supondo satisfeita a condição imposta sobre as

raízes de (4.23), vem:

$$X'_{t+q,i} \Phi^{-1} X_{t+q,j} \rightarrow 0 \text{ quando } t \rightarrow \infty. \quad (4.29)$$

Portanto, a soma em (4.28) tende, no limite, a uma matriz finita.

Daí decorre que os termos de (4.28) representam uma contribuição pouco significativa para a função de verossimilhança quando N é grande e a aproximação (4.11), ao ser comparada com (4.10), é equivalente a esta apenas com a inclusão do fator $\delta_{ij} \Phi^{-1}$ extraído de (4.28).

Concluimos, então, que a aproximação das funções de verossimilhança condicional e exata é bastante nítida e as diferenças entre as mesmas são assintoticamente irrelevantes.

4.8 - OBTENÇÃO DAS ESTIMATIVAS DE MÁXIMA VEROSSIMILHANÇA, SEGUNDO OSBORN

O processo para determinação de estimativas de máxima verossimilhança exata de Φ e θ_i ($i=1, \dots, q$) é executado através da maximização da função de verossimilhança exata expressada em (4.10) com respeito a esses parâmetros.

Podemos esboçar essa formulação de outra maneira alternativa apresentada a seguir.

Aplicando (4.24) e (4.25) em (4.10), temos:

$$p(\omega) = L(\omega) = (2\pi)^{-1/2 Nk} |\Phi|^{-1/2 (N+q)} |X' (I \otimes \Phi^{-1}) X|^{-1/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1-q}^N \hat{a}'_t \Phi^{-1} \hat{a}_t \right\}$$

$$\begin{aligned} \ln p(\omega) = & -\frac{1}{2} N k \ln(2\pi) - \frac{1}{2} (N+q) \ln |\Phi| - \frac{1}{2} \ln |X' (I \otimes \Phi^{-1}) X| \\ & - \frac{1}{2} \sum_{t=1-q}^N \hat{a}_t' \Phi^{-1} \hat{a}_t . \end{aligned}$$

Omitindo o primeiro termo e multiplicando ambos os membros da igualdade por (-2) , vem:

$$\begin{aligned} \ln^* p(\omega) = (-2) \ln p(\omega) = & (N+q) \ln |\Phi| + \ln |X' (I \otimes \Phi^{-1}) X| + \\ & + \sum_{t=1-q}^N \hat{a}_t' \Phi^{-1} \hat{a}_t . \end{aligned} \quad (4.30)$$

Conseqüentemente, a maximização de $\ln p(\omega)$ implica a minimização de $\ln^* p(\omega)$, usando uma técnica de otimização não linear apropriada.

Podemos calcular (4.30) de uma maneira eficaz empregando os dados em duas etapas.

A primeira consiste na determinação de $\hat{a}_*$ por meio de (4.7). Embora as matrizes L e X tenham grandes dimensões, nenhuma precisa ser obtida integralmente.

A matriz L é usada somente no produto $L\omega$ em (4.7). Podemos observar a partir de (4.5) que $L\omega = a$ quando $a_* = 0$. Assim, $L\omega$ equivale ao conjunto de resíduos obtidos no procedimento condicional com valores iniciais iguais a zero. Portanto, a matriz L não precisa ser calculada diretamente. Analogamente, não é necessário armazenar a matriz X totalmente e, sim, apenas os q blocos linhas requisitados para obtermos $X_{t+q,j}$ ($j = 1, \dots, q$) por meio da equação de diferenças (4.26). Este tratamen-

to "step-wise" aplicado em X é apropriado também na etapa referente à construção dos produtos $X'\Omega^{-1}L\omega$ e $X'\Omega^{-1}X$ em (4.7) em virtude da forma diagonal em blocos da matriz Ω^{-1} .

Denotemos por $\hat{a}_{0t}$ o vetor dos resíduos correspondente ao tempo t , supondo $a_* = 0$. Nesse caso, $a = L\omega$, e, dessa maneira, o i -ésimo bloco de $X'\Omega^{-1}L\omega$ é expressado por:

$$\sum_{t=1}^N X'_{t+q,i} \Phi^{-1} \hat{a}_{0t}. \quad (4.31)$$

O (i,j) -ésimo bloco de $X'\Omega^{-1}X$ será dado por (4.28). Para cada valor de t ($t=1, \dots, N$) o esquema sugerido funciona conforme a seqüência: o vetor $\hat{a}_{0t}$ é calculado de acordo com (4.4) e a restrição $a_* = 0$ e $X_{t+q,j}$ ($j=1, \dots, q$) é obtido através de (4.26). A contribuição de t para $X'\Omega^{-1}L\omega$ e $X'\Omega^{-1}X$ pode, então, ser avaliada ao usarmos (4.31) e (4.28) respectivamente.

Na segunda etapa os resíduos iniciais são utilizados no cálculo de $\hat{a}_t$ e $\hat{a}'_t \Phi^{-1} \hat{a}_t$, $t = 1, \dots, N$.

Comparando os cálculos necessários para (4.10) e (4.11), podemos observar que há operações adicionais para a determinação das estimativas de máxima verossimilhança exata e a primeira etapa descrita no esquema computacional não é necessária no caso da determinação das estimativas de máxima verossimilhança condicional.

O fator pelo qual o número de operações é aumentado no caso da verossimilhança exata depende da ordem q do polinômio de médias móveis como também do número de variáveis, k , em

ω_t . Esse fator independe do tamanho amostral, N , se q e k forem pequenos quando comparados com N .

Concluimos, então, que o procedimento exato para a obtenção das estimativas de máxima verossimilhança de um modelo de médias móveis é justificável sob o ponto de vista teórico, mas em certas circunstâncias pode não ser computacionalmente prático.

4.9 - COMPARAÇÃO DA FUNÇÃO DE VEROSSIMILHANÇA

EXATA COM A CONDICIONAL PARA UM PROCESSO

MA(q) NÃO INVERTÍVEL, SEGUNDO HILLMER E TIAO

Compararemos nesta seção as expressões (4.21) e (4.22) com o objetivo de considerar as suas propriedades. Usaremos como fundamento de nossa análise a hipótese do modelo MA(q) ser não invertível, ou seja, $|\theta_q(B)|$ tem uma ou mais raízes no círculo unitário.

Dunsmuir e Hannan (1976) mostraram que as estimativas de máxima verossimilhança com respeito a (4.21) são consistentes para modelos invertíveis e não invertíveis. Reinsel (1976) verificou que as estimativas correspondentes a (4.22) também gozam da propriedade da consistência.

Consideremos, a título de ilustração, a situação particular na qual trataremos de um modelo univariado ($k=1$) MA(1) expressado por:

$$\omega_t = a_t - \theta_0 a_{t-1}, \quad (4.32)$$

onde θ_0 é o valor paramétrico verdadeiro. Discutiremos o caso $\theta_0 = 1$ e, para simplificação, vamos supor σ^2 conhecido e igual à unidade. O tratamento do caso $\theta_0 = -1$ seguiria de forma análoga.

De (4.22), vem:

$$-\frac{1}{N} E_{\omega/\theta_0=1} \left[\ln L(\theta, \omega, a_*=0) \right] = C + R_0(\theta) \quad (4.33)$$

onde

$$R_0(\theta) = \frac{1}{1+\theta} \left[1 + \frac{\theta}{N} \frac{(1-\theta^{2N})}{(1-\theta^2)} \right]. \quad (4.34)$$

De (4.21), vem:

$$-\frac{1}{N} E_{\omega/\theta_0=1} \left[\ln L(\theta, \omega) \right] = C' + R_1(\theta) \quad (4.35)$$

onde

$$R_1(\theta) = R_0(\theta) + \frac{1}{2N} \ln \frac{1-\theta^{2(N+1)}}{1-\theta^2} - \frac{\theta^2(1-\theta^{2N})(1-\theta^{2N+1}) + N\theta^{2N+1}(1+\theta)(1-\theta)^2}{N(1-\theta^2)(1+\theta)(1-\theta^{2(N+1)})}. \quad (4.36)$$

Em (4.33) nosso objetivo é a minimização do primeiro membro da equação uma vez que, inicialmente, direcionamos nosso estudo para a maximização do logaritmo da função de verossimilhança. O mesmo acontece em (4.35). As parcelas C e C' , presentes em (4.33) e (4.35) respectivamente, são constantes de maneira que visaremos a minimização de $R_0(\theta)$ e $R_1(\theta)$.

Podemos observar em (4.34) que:

a) $R_0(-1) = N$ e $R_0(0) = R_0(1) = 1$. De fato,

$$\begin{aligned} R_0(\theta) &= \frac{1}{1+\theta} \left[1 + \frac{\theta}{N} \frac{(1-\theta^{2N})}{(1-\theta^2)} \right] \\ &= \frac{1}{1+\theta} \left[\frac{N + \theta(\theta^{2N-2} + \theta^{2N-4} + \theta^{2N-6} + \theta^{2N-8} + \dots + \theta^{2N-2N})}{N} \right] \\ &= \frac{1}{1+\theta} \left[\frac{N + \theta^{2N-1} + \theta^{2N-3} + \theta^{2N-5} + \theta^{2N-7} + \dots + \theta}{N} \right] \\ &= \frac{\theta^{2N-2} - \theta^{2N-3} + 2\theta^{2N-4} - 2\theta^{2N-5} + 3\theta^{2N-6} - 3\theta^{2N-7} + \dots + (N-1)\theta^2 - (N-1)\theta + N}{N} \end{aligned}$$

ou seja,

$$\begin{aligned} R_0(\theta) &= \frac{\theta^{2N-2} + 2\theta^{2N-4} + 3\theta^{2N-6} + \dots + (N-1)\theta^2 + N}{N} \\ &\quad - \frac{\theta^{2N-3} + 2\theta^{2N-5} + 3\theta^{2N-7} + \dots + (N-1)\theta}{N} \end{aligned}$$

Logo,

$$\begin{aligned} R_0(-1) &= \frac{1 + 2 + 3 + \dots + (N-1) + N}{N} - \\ &\quad - \frac{(-1)[1 + 2 + 3 + \dots + (N-1)]}{N} \end{aligned}$$

$$R_0(-1) = \frac{N(N+1)}{2N} + \frac{(N-1)N}{2N} = N$$

$$R_0(0) = \frac{N}{N} = 1$$

$$\begin{aligned} R_0(1) &= \frac{1 + 2 + 3 + \dots + (N-1) + N}{N} \\ &\quad - \frac{1 + 2 + 3 + \dots + (N-1)}{N} \end{aligned}$$

$$R_0(1) = \frac{N(N+1)}{2N} - \frac{(N-1)N}{2N} = 1;$$

b) quando $N > 1$, $R'_0(\theta) > 0$ para $-1 < \theta \leq 0$.

Assim, a função $R_0(\theta)$ é crescente para θ pertencente ao intervalo $(-1; 0]$;

c) para N suficientemente grande, $R_0(\theta)$ é aproximadamente igual a $\frac{1}{1+\theta}$ se $-1 < \theta < 1$.

Portanto, a minimização de $R_0(\theta)$ é alcançada para algum valor de θ pertencente ao intervalo $(0; 1)$. Entretanto, para N suficientemente grande, esse valor de θ pode se aproximar de 1 tanto quanto o desejarmos. A função $R_1(\theta)$ correspondente à verossimilhança exata difere de $R_0(\theta)$ por dois termos apresentados em (4.36).

A seguir, temos através de uma tabela a relação de valores de θ que minimizam $R_0(\theta)$ em função de alguns valores sugeridos para N

N	10	50	100	200	500	1.000
θ	0,72	0,87	0,90	0,93	0,96	0,97

Como podemos notar, boas estimativas são obtidas na maximização de (4.22) se dispusermos de amostras de grande dimensão. Por outro lado, pode ser verificado que quando θ assume o valor igual à unidade, $R_1(\theta)$ é minimizado qualquer que seja o valor de N . Dessa maneira, não há viés no procedimento de obtenção de estimativas relativas à função de verossimilhança exata. Portanto, estimativas correspondentes à verossimilhan-

ça exata são superiores às correspondentes à verossimilhança aproximada. Essa afirmação poderá ser comprovada de maneira mais concreta no capítulo onde focalizarmos as aplicações de resultados alcançados em experimentos.

4.10 - METODOLOGIA DE HILLMER E TIAO NA ESTIMAÇÃO DO MODELO AR(p) MULTIVARIADO

Um processo vetorial AR expressa-se na forma:

$$\phi_p(B)\tilde{\omega}_t = a_t \quad (4.37)$$

onde

$$\phi_p(B) = (I_k - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

$\tilde{\omega}_t$: vetor de dimensão $k \times 1$;

B: operador retroativo tal que $B\tilde{\omega}_t = \tilde{\omega}_{t-1}$;

ϕ_i ($i=1, \dots, p$): matrizes paramétricas desconhecidas com dimensão $k \times k$;

a_t 's ($t=0, \pm 1, \pm 2, \dots$): vetores de dimensão $k \times 1$ independentes e identicamente distribuídos como $N(0, \Phi)$, sendo Φ uma matriz finita de dispersões.

Novamente consideraremos $\mu_\omega = 0$, ou seja, $\tilde{\omega}_t = \omega_t - \mu_\omega = \omega_t$.

Com relação à região de admissibilidade para o campo de variação dos parâmetros do modelo, devemos ter as raízes dos determinantes polinomiais em B, $|\phi_p(B)| = 0$, fora do círculo

unitário. Dessa maneira, está assegurada a estacionariedade do processo.

A respeito da invertibilidade de modelos auto-regressivos, sabemos que não é necessário impormos restrições sobre os valores paramétricos, pois esses processos são sempre invertíveis.

Foi feita por Hillmer & Tiao (1979) uma adaptação dos resultados encontrados na especificação da função de máxima verossimilhança do processo misto ARMA para o processo puro AR, o qual é um caso particular do primeiro.

Consideremos as expressões (3.66) e (3.67).

O fator $A_{N-q,q}$ presente em (3.66) é

$$A_{N-q,q} = \Phi_q^{-1} + (\Delta_{N-q,q} C_{\theta,q})' H_{N-q} (\Delta_{N-q,q} C_{\theta,q}). \quad (4.38)$$

As matrizes utilizadas em (4.38) já foram definidas na seção 3.5.

No processo AR o polinômio $\theta_q(B)$ é reduzido à matriz identidade. Logo, os demais coeficientes são nulos.

Como

$$C_{\theta,q} = \begin{bmatrix} \theta_q & \theta_{q-1} & \cdots & \theta_1 \\ & . & \ddots & \vdots \\ & & . & \theta_{q-1} \\ 0 & & & \theta_q \end{bmatrix}$$

teremos $C_{\theta,q}$ igual à matriz nula e, conseqüentemente, $A_{N-q,q} = \Phi_q^{-1}$.

Da mesma forma $\tilde{\alpha}_1 = 0$, pois

$$\tilde{\alpha}_1 = -A_{N-q,q}^{-1} (\Delta_{N-q,q} C_{\theta,q})' H_{N-q} Y_2.$$

Analogamente, $H_{N-q} = \Phi_{N-q}^{-1}$ e $F_{\theta,q} = I$, pois:

$$H_{N-q} = D_{\theta,N-q}^{-1} \Phi_{N-q}^{-1} D_{\theta,N-q}^{-1}$$

e

$$D_{\theta,N-q} = \begin{bmatrix} I & \dots & \dots & \dots & \dots & 0 \\ -\theta_1 & & & & & \vdots \\ \vdots & & & & & \vdots \\ -\theta_q & & & & & \vdots \\ & \ddots & & & & \vdots \\ 0 & & -\theta_q & \dots & -\theta_1 & I \end{bmatrix} = \begin{bmatrix} I & \dots & \dots & \dots & \dots & 0 \\ 0 & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & 0 & \dots & I \end{bmatrix}$$

$$F_{\theta,q} = \begin{bmatrix} I & \dots & \dots & \dots & \dots & 0 \\ -\theta_1 & & & & & \vdots \\ \vdots & & & & & \vdots \\ -\theta_{q-1} & \dots & -\theta_1 & & & I \end{bmatrix} = \begin{bmatrix} I & & & & & 0 \\ 0 & & & & & \vdots \\ 0 & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & \dots & 0 & 0 & I \end{bmatrix}$$

Portanto, de acordo com (3.66) e (3.67), a função de verossimilhança de $\phi = (\phi_1, \dots, \phi_p)$ e Φ assumirá a forma:

$$\begin{aligned} L_1(\phi, \Phi/\omega) &\propto |\Phi|^{-1/2 N} |\Phi_q^{-1}|^{-1/2} \cdot \exp\left(-\frac{1}{2} Y_2' \Phi_{N-q}^{-1} Y_2\right) = \\ &= |\Phi|^{-1/2 N} |\Phi|^{+q/2} \cdot \exp\left(-\frac{1}{2} Y_2' \Phi_{N-q}^{-1} Y_2\right) = \\ &= |\Phi|^{-1/2 (N-q)} \cdot \exp\left(-\frac{1}{2} Y_2' \Phi_{N-q}^{-1} Y_2\right) \end{aligned} \quad (4.39)$$

e a função de verossimilhança associada à distribuição con-

dicional $p(Y_1/Y_2)$ é dada por

$$L_2(\phi, \Sigma/\omega) \propto |\Gamma + (\Sigma_q^{-1})^{-1}|^{-1/2} \cdot \exp\left\{-\frac{1}{2} Y_1' [\Gamma + (\Sigma_q^{-1})^{-1}]^{-1} Y_1\right\}$$

$$L_2(\phi, \Sigma/\omega) \propto |\Gamma + \Sigma_q|^{-1/2} \cdot \exp\left\{-\frac{1}{2} Y_1' [\Gamma + \Sigma_q]^{-1} Y_1\right\} \quad (4.40)$$

sendo $Y = D_{\phi, N} \omega$, $Y' = [Y_1', Y_2']$.

Como podemos notar, L_2 está relacionada com a distribuição marginal de ω_1 e L_1 refere-se à distribuição condicional de ω_2 dado ω_1 . Se desconsiderarmos a participação de L_2 na determinação da função de verossimilhança, iremos nos restringir a empregar a distribuição condicional $p(\omega_2/\omega_1)$, dada em (4.39). Dessa maneira, a estimação dos parâmetros ϕ pode ser feita por mínimos quadrados e tais estimativas são assintoticamente consistentes e eficientes de acordo com trabalho desenvolvido por Anderson (1971).

4.11 - COMENTÁRIOS FINAIS

Através das seções 4.7 e 4.8 foram confrontados os métodos de Osborn e Wilson. Quanto à metodologia de Phadke e Kedem exposta na seção 4.4, foi possível abordar o problema da determinação da função de verossimilhança empregando a matriz de covariância de diferentes maneiras. Foi feita uma apreciação de cada uma delas na seção 4.5.

Na determinação da função de verossimilhança aproximada as expressões a que chegaram Phadke e Kedem e Wilson são análogas. Porém, Phadke e Kedem sugerem outra alternativa para a apresentação da matriz de covariância de ω .

Na seção 3.4 foram descritos algoritmos propostos por Wilson, objetivando a obtenção das estimativas paramétricas de modelos ARMA. Com as devidas adaptações, esses algoritmos também podem ser aplicados em modelos MA. Assim, Wilson fornece-nos uma teoria mais abrangente, que inclui não só a determinação da função de verossimilhança como também meios de estimar os parâmetros do modelo em questão.

Consideremos, agora, os métodos de Phadke e Kedem na obtenção da função de verossimilhança exata: o método de decomposição de Cholesky e o método no domínio do tempo. Na seção 4.5 vimos que a opção referente à escolha de cada procedimento é condicionada à especificação dos valores de k e q .

Comparando estes procedimentos com o de Osborn, um critério de opção entre eles depende da carga computacional acarretada em cada um como também das magnitudes dos erros quadráticos médios gerados na obtenção das estimativas paramétricas.

CAPÍTULO 5

TÉCNICAS DE IDENTIFICAÇÃO EM SÉRIES MULTIVARIADAS

5.1 - INTRODUÇÃO

Em virtude da importância da fase de identificação para o alcance de resultados satisfatórios em análises de séries temporais, este capítulo abordará brevemente alguns caminhos a serem seguidos no estágio de identificação, para se chegar a uma decisão a respeito dos modelos que mais fielmente descrevem o comportamento de séries temporais. O problema da identificação de modelos ARMA multivariados é fundamental e muito necessita ser feito nesta área.

5.2 - METODOLOGIA DE TIAO E BOX

Serão vistos nesta seção os estágios constituintes de processos de identificação de modelos com base em um conjunto de dados amostrais e estatísticas derivadas a partir desses dados, através das quais levamos avante a investigação dos modelos mais cotados como delineadores das séries temporais analisadas conjuntamente. Tais enfoques são provenientes de análises feitas por Tiao e Box (1981).

Primeiramente vamos focar as estatísticas que serão usadas como auxiliares não só para esboçar a forma de possíveis

modelos como também para fornecer estimativas aproximadas dos valores paramétricos.

Consideremos as matrizes de covariâncias cruzadas e correlações cruzadas.

Seja $\{\omega_t\}$ um vetor de séries temporais de dimensão k , seguindo a estrutura de um modelo ARMA multivariado e estacionário, tendo como média o vetor μ_ω . Denotaremos por $\Gamma(\ell)$ a matriz de covariâncias cruzadas de defasagem ℓ , expressada por:

$$\Gamma(\ell) = E(\omega_{t-\ell} \omega_t') = \{\gamma_{ij}(\ell)\}, \ell=0, \pm 1, \pm 2, \dots; i, j=1, \dots, k. \quad (5.1)$$

A matriz de correlações cruzadas correspondentes será denotada por $\rho(\ell) = \{\rho_{ij}(\ell)\}$.

O fato de termos um modelo ARMA estacionário implica a seguinte relação:

$$\Gamma(\ell) = \begin{cases} \sum_{j=\ell-r}^{\ell-1} \Gamma(j) \phi_{\ell-j}' - \sum_{j=0}^{r-\ell} \psi_j \theta_{j+\ell}', & \ell=0, \dots, r \\ \sum_{j=1}^r \Gamma(\ell-j) \phi_j', & \ell > r \end{cases} \quad (5.2)$$

sendo ψ_j 's obtidos da expressão

$$\psi(B) = \phi^{-1}(B) \theta(B) = (I + \psi_1 B + \dots)$$

e considerando

$$\theta_0 = -I,$$

$$r = \max(p, q)$$

$$\phi_{p+1} = \dots = \phi_r = 0, \text{ se } p < q,$$

$$\theta_{q+1} = \dots = \theta_r = 0, \text{ se } q < p.$$

No caso particular do processo MA(q), temos:

$$\Gamma(\ell) = \begin{cases} \sum_{j=0}^{q-\ell} \theta_j \theta'_{j+\ell}, & \ell=0, \dots, q \\ 0, & \ell > q. \end{cases} \quad (5.3)$$

Podemos concluir que quando $\ell > q$ as autocorrelações e as correlações cruzadas são nulas no caso do processo MA(q).

No caso do processo AR o comportamento dessas estatísticas é diferente sendo que, em geral, as autocorrelações e as correlações cruzadas decrescem gradualmente para zero à medida que $|\ell|$ aumenta.

Através do sistema de equações (5.2), podemos associar diretamente as matrizes de autocovariâncias $\Gamma(\ell)$'s com os coeficientes auto-regressivos matriciais como segue:

$$\begin{bmatrix} A(p, m) & b(p, m) \\ g'(p, m) & \Gamma(m) \end{bmatrix} \begin{bmatrix} \phi_{p-1} \\ \phi'_p \end{bmatrix} = \begin{bmatrix} c(p, m) \\ \Gamma(p+m) \end{bmatrix}, \quad m = q, q+1, \dots \quad (5.4)$$

onde

$$A(p, m) = \begin{bmatrix} \Gamma(m) & \Gamma(m-1) & \dots & \Gamma(m-p+2) \\ \Gamma(m+1) & \Gamma(m) & & \\ \vdots & \cdot & \ddots & \\ \Gamma(m+p-2) & \dots & \Gamma(m+1) & \Gamma(m) \end{bmatrix}$$

$$b(p,m) = \begin{bmatrix} \Gamma(m-p+1) \\ \vdots \\ \Gamma(m-1) \end{bmatrix}, \quad c(p,m) = \begin{bmatrix} \Gamma(m+1) \\ \vdots \\ \Gamma(m+p-1) \end{bmatrix}$$

$$g'(p,m) = [\Gamma(m+p-1) \dots \Gamma(m+1)]$$

$$\Phi'_{p-1} = [\phi_1, \dots, \phi_{p-1}]$$

Seja

$$D(\ell, m) = [d_{ij}(\ell, m)], \quad \ell=1, 2, \dots; \quad m=0, 1, \dots \quad (5.5)$$

cuja dimensão é $(k \times k)$, onde

$$d_{ij}(\ell, m) = \det \begin{bmatrix} A(\ell, m) & c_j(\ell, m) \\ g'_i(\ell, m) & \gamma_{ij}(\ell+m) \end{bmatrix}, \quad i, j=1, \dots, k$$

sendo

$c_j(\ell, m)$ a j -ésima coluna de $c(\ell, m)$,

$g'_i(\ell, m)$ a i -ésima linha de $g'(\ell, m)$,

$\gamma_{ij}(\ell+m)$ o (i, j) -ésimo elemento de $\Gamma(\ell+m)$.

Através de (5.4), podemos concluir que no caso de um processo ARMA(p,q), temos:

$$D(\ell, m) = 0 \quad \text{para } \ell > p \text{ e } m \geq q. \quad (5.6)$$

Esse resultado é uma extensão dos resultados obtidos por Gray, Kelley e McIntire (1978) correspondentes a processos ARMA univariados.

Se atribuirmos o valor zero aos índices m e q , a for-

mulação em (5.4) transforma-se em uma extensão no campo multivariado das equações de Yule-Walker referentes a modelos auto-regressivos univariados.

Passamos, agora, à definição da estatística $\underline{P}(\ell)$, função matricial das auto-regressões parciais, generalizando a função escalar das autocorrelações parciais.

No caso do processo AR(p), o comportamento da função $\underline{P}(\ell)$ é o seguinte:

$$\underline{P}(\ell) = \begin{cases} \phi_{\ell}, & \ell = p \\ 0, & \ell > p. \end{cases} \quad (5.7)$$

A função $\underline{P}(\ell)$ pode ser expressada em termos das matrizes de autocovariâncias a partir das relações em (5.4). Temos, assim

$$\underline{P}'(\ell) = \begin{cases} \Gamma^{-1}(0)\Gamma(1), & \ell = 1 \\ [\Gamma(0) - b'(\ell, 0)A^{-1}(\ell, 0)b(\ell, 0)]^{-1} & \\ [\Gamma(\ell) - b'(\ell, 0)A^{-1}(\ell, 0)c(\ell, 0)], & \ell > 1. \end{cases} \quad (5.8)$$

No estágio de identificação utilizamos as propriedades do comportamento teórico das estatísticas amostrais a fim de reduzirmos o número de sub-classes de modelos apropriados para análises subseqüentes.

Veremos, agora, as estatísticas amostrais.

As correlações cruzadas amostrais, $\hat{\rho}_{ij}(\ell)$, são dadas por

$$\hat{\rho}_{ij}(\ell) = \frac{\sum (\omega_{it} - \bar{\omega}_i)(\omega_{j,t+\ell} - \bar{\omega}_j)}{\left\{ \sum (\omega_{it} - \bar{\omega}_i)^2 \sum (\omega_{jt} - \bar{\omega}_j)^2 \right\}^{1/2}},$$

onde $\bar{\omega}_i$ é a média amostral da série que consiste na i -ésima componente de ω_t .

Devido aos resultados de (5.3), podemos utilizar $\hat{\rho}_{ij}(\ell)$ como indicadores da ordem dos polinômios através dos quais são expressados os modelos de médias móveis, uma vez que $\hat{\rho}_{ij}(\ell) = 0$ para $\ell > q$.

Supondo que um vetor de séries temporais seja estacionário, um esquema prático para uma visualização compreensível das magnitudes dos valores amostrais $\hat{\rho}_{ij}(\ell)$ é a listagem dos mesmos através de símbolos indicadores, abolindo, dessa forma, os valores numéricos. Convencionaremos, então, que o sinal "+" indicará um valor maior do que $2N^{-1/2}$, onde N é o número de observações de cada série temporal componente do vetor ω_t ; o sinal "-" indicará um valor menor do que $-2N^{-1/2}$ e um ponto "." simbolizará um valor entre $-2N^{-1/2}$ e $2N^{-1/2}$. Esse artifício é empregado em função do fato de que se uma série é ruído branco e N é grande, então os $\hat{\rho}_{ij}(\ell)$'s são normalmente distribuídos com média 0 e variância N^{-1} .

Com respeito às auto-regressões parciais amostrais, obtivemos em (5.7) que as mesmas são nulas para $\ell > p$. Assim, elas servem como indicadores da ordem dos polinômios auto-regressivos de um modelo AR. Para obter estimativas de $\underline{p}(\ell)$ e de seus desvios padrões é necessário ajustar sucessivamente modelos auto-regressivos de ordem alta para $\ell = 1, 2, \dots$ usando a

técnica de mínimos quadrados multivariados.

Analogamente ao artifício empregado no caso das correlações cruzadas utilizaremos em substituição aos valores numéricos uma listagem de símbolos dentre os quais o sinal "+" surgirá quando um coeficiente em $\hat{p}(\ell)$ for maior do que o dobro de seus desvios padrões estimados; o sinal "-" estará associado ao fato de um coeficiente ser menor do que o oposto do dobro dos desvios padrões estimados e um ponto "." indicará valores intermediários no intervalo tendo como limite inferior o oposto do dobro dos desvios padrões e como limite superior o dobro dos desvios padrões. Esse artifício utilizado é motivado pelo fato de que, segundo Anderson (1971), em um processo estacionário AR(p) as estimativas $\phi'_1, \dots, \phi'_p$ são normalmente distribuídas.

No caso de modelos auto-regressivos, temos ainda uma estatística conhecida como aproximação de Bartlett (1938), que orienta na decisão a respeito da ordem do polinômio auto-regressivo.

Vamos denotar essa estatística por $M(\ell)$ e expressá-la por:

$$M(\ell) = -(n - \frac{1}{2} - \ell k) \log_e U \quad (5.9)$$

onde

$$U = |S(\ell)| / |S(\ell-1)| \quad (5.10)$$

$$S(\ell) = \sum_{t=\ell+1}^N (\omega_t - \hat{\phi}_1 \omega_{t-1} - \dots - \hat{\phi}_\ell \omega_{t-\ell}) \times$$

$$\times (\omega_t - \hat{\phi}_1 \omega_{t-1} - \dots - \hat{\phi}_\ell \omega_{t-\ell})' \quad (5.11)$$

A função $S(\ell)$ é a matriz da soma de quadrados e produtos cruzados residuais gerados após o ajuste de um modelo $AR(\ell)$.

A função U é uma estatística da razão de verossimilhança correspondente ao seguinte teste de hipóteses:

$$\text{hipótese nula } (H_0): \phi_\ell = 0$$

contra a

$$\text{hipótese alternativa } (H_a): \phi_\ell \neq 0.$$

Retomando a expressão em (5.9) e supondo verdadeira a hipótese nula, a estatística $M(\ell)$ tem como distribuição assintótica uma variável χ^2 com k^2 graus de liberdade, onde $n=N-p-1$ é o número efetivo de observações, admitindo a inclusão de um termo constante no modelo. Assim, quando o valor calculado para $M(\ell)$ for maior do que o valor tabelado, consideraremos significativo o coeficiente auto-regressivo estimado; caso contrário, não aceitaremos a sua participação na formulação do modelo.

Outra informação adicional útil para indicar o nível de melhoria obtida no ajuste de modelo auto-regressivo, à medida que a ordem do polinômio auto-regressivo é aumentada, consiste nos elementos diagonais das matrizes de covariâncias residuais $\hat{\Sigma}$, correspondendo a modelos auto-regressivos ajustados sucessivamente. O ideal seria obtermos valores de pequena magnitude na diagonal dessas matrizes.

Um artifício também adotado corresponde à análise das matrizes de correlações cruzadas dos resíduos $\hat{a}_t$ de cada ajuste $AR(\ell)$, $\ell = 1, \dots, p$. Para facilitar a visualização do comportamento dessas correlações são usados os símbolos "+", "-" e ".", cujo significado já foi explicitado anteriormente. Este procedimento fundamenta-se no fato de que se a série for ruído branco e N for grande, então as correlações cruzadas residuais serão normalmente distribuídas com média 0 e variância N^{-1} .

Com respeito às autocorrelações residuais, recentemente Li e McLeod (1980) desenvolveram expressões para as suas distribuições assintóticas, chegando ao mesmo resultado encontrado no caso univariado, ou seja, as autocorrelações de ordem baixa têm variância menor do que $N^{-1/2}$.

No caso de processos mistos ARMA, as matrizes populacionais das correlações cruzadas e das auto-regressões parciais apresentam seus elementos decrescendo gradualmente para zero. Às vezes é possível tentar identificar a ordem de modelos mistos após o ajuste de modelos auto-regressivos, analisando o comportamento das correlações cruzadas residuais. Entretanto, essa análise pode ser enganosa.

Consideremos, para exemplificação, o modelo estacionário $ARMA(1,1)$:

$$(I - \phi B)\omega_t = (I - \theta B)a_t. \quad (5.12)$$

Se ajustarmos um modelo $AR(1)$ às séries componentes do

vetor $\{\omega_t\}$ obteremos uma estimativa $\hat{\phi}$ viesada. De fato, o estimador $\hat{\phi}$ assintoticamente converge em probabilidade para

$$\phi_0 = \Gamma'(1)\Gamma(0)^{-1}. \quad (5.13)$$

Portanto, os resíduos $\hat{a}_t$ seguem aproximadamente o modelo

$$\begin{aligned} \hat{a}_t &= (I - \hat{\phi}B)\omega_t \\ &= (I - \phi_0 B)\omega_t \end{aligned} \quad (5.14)$$

Substituindo (5.12) em (5.14), vem:

$$\hat{a}_t = (I - \phi_0 B)(I - \phi B)^{-1}(I - \theta B)a_t$$

ou, de forma equivalente:

$$(I - \phi B)\hat{a}_t = (I - \phi_0 B)(I - \theta B)a_t$$

$$(I - \phi B)\hat{a}_t = (I - (\theta + \phi_0)B + \theta\phi_0 B^2)a_t.$$

Supondo $k=1$, a série $\{\hat{a}_t\}$ tem a estrutura de um modelo ARMA(1,2) e, portanto, suas autocorrelações mantêm a relação

$$\rho_{\hat{a}}(j) = \phi \rho_{\hat{a}}(j-1), \quad j > 2 \quad (5.15)$$

e para $j=1,2$ as autocorrelações são funções de ϕ e θ .

A seguir, temos a Tabela 5.1, que contém valores assintóticos de $\rho_{\hat{a}}(1)$ e $\rho_{\hat{a}}(2)$ correspondentes a combinações de valores de ϕ e θ .

Através dessa tabela, podemos observar que se o verdadeiro valor de ϕ fosse alto em módulo, então elegeríamos um mo-

TABELA 5.1

θ	-0,95		-0,50		0,50		0,95	
$\rho_{\hat{a}}(j)$	$\rho_{\hat{a}}(1)$	$\rho_{\hat{a}}(2)$	$\rho_{\hat{a}}(1)$	$\rho_{\hat{a}}(2)$	$\rho_{\hat{a}}(1)$	$\rho_{\hat{a}}(2)$	$\rho_{\hat{a}}(1)$	$\rho_{\hat{a}}(2)$
ϕ								
-0,95	-	-	0,265	0,085	-0,381	-0,030	-0,481	-0,036
-0,50	0,049	-0,222	-	-	-0,223	-0,201	-0,321	-0,267
0,50	0,321	-0,267	0,223	-0,201	-	-	-0,049	-0,222
0,95	0,481	-0,036	0,381	-0,030	-0,265	0,085	-	-

delo MA(1) para a série $\{\hat{a}_t\}$, uma vez que apenas $\rho_{\hat{a}}(1)$ é significativamente diferente de zero e, assim, a identificação seria correta. Para valores moderados de ϕ seria escolhido um modelo MA(2) ou, possivelmente, um modelo de médias móveis de ordem superior a 2, acarretando uma super-parametrização. O problema de determinação da ordem de modelo ARMA univariado tem sido analisado por Gray, Kelley e McIntire (1978) e Beguin Gourieroux e Monfort (1980), os quais apresentaram alguns métodos. Tais métodos equivalem à estimação de $D(\ell, m)$ em (5.5) supondo $k=1$ e usando estimativas amostrais das autocovariâncias. A ordem dos polinômios auto-regressivos e de médias móveis é obtida através do resultado explicitado em (5.6). Atualmente, Tiao e Box estão desenvolvendo estudos sobre as propriedades amostrais de estimativas de funções adequadas de $D(\ell, m)$ no caso multivariado.

Todas as estatísticas apresentadas nesta seção fazem parte do estágio de identificação proposto por Tiao e Box. Convém enfatizar que todas elas funcionam como artifícios propostos para delineamentos preliminares de modelos tentativos e,

portanto, não devem ser encaradas como testes exatos de significância.

Posteriormente, é empreendido o estágio de estimação paramétrica, seguido dos diagnósticos de verificação da adequabilidade dos resultados anteriores e o processo termina com a previsão de valores futuros da série multivariada com base no modelo aprovado e adotado como verdadeiro.

5.3 - METODOLOGIA DE PIERCE

Antes de apresentarmos as técnicas sugeridas por Pierce (1980), introduziremos alguns conceitos essenciais para o perfeito entendimento das relações existentes entre duas ou mais séries temporais.

Esses conceitos referem-se aos padrões de causalidade entre variáveis analisadas conjuntamente, conforme visto no Capítulo 1.

A configuração desses padrões de causalidade emprega a função $\rho_{uv}(s)$, que consiste das correlações cruzadas de séries "branqueadas", onde $\{u_t\}$ e $\{v_t\}$ são séries de ruído branco em que foram transformadas duas séries temporais. Vamos estabelecer as possíveis relações existentes entre duas séries, lembrando que podemos dispor de um número maior de componentes do vetor de séries temporais mas que, no caso, estas seriam confrontadas duas a duas. Sejam $\{X_t\}$ e $\{Y_t\}$ duas séries associadas respectivamente às séries $\{u_t\}$ e $\{v_t\}$, cujo processo de transformação será visto posteriormente.

A seguir, temos a configuração de relações entre duas séries e o correspondente comportamento de $\rho_{uv}(s)$. (Adaptado de Pierce, 1980).

TABELA 5.2

Relação	Restrições sobre $\rho_{uv}(s)$
1. X_t causa Y_t	$\rho_{uv}(s) \neq 0$ para algum $s > 0$
2. Y_t causa X_t	$\rho_{uv}(s) \neq 0$ para algum $s < 0$
3. Causalidade instantânea	$\rho_{uv}(0) \neq 0$
4. "Feedback"	$\rho_{uv}(s) \neq 0$ para algum $s > 0$ e para algum $s < 0$
5. Y_t não causa X_t	$\rho_{uv}(s) = 0, \forall s < 0$
6. causalidade unidirecional de X_t para Y_t	$\rho_{uv}(s) \neq 0$ para algum $s > 0$ e $\rho_{uv}(s) = 0, \forall s < 0$
7. X_t e Y_t são relacionadas instantaneamente mas de nenhuma outra forma	$\rho_{uv}(s) = 0, \forall s \neq 0$, e $\rho_{uv}(0) \neq 0$
8. X_t e Y_t são independentes	$\rho_{uv}(s) = 0, \forall s.$

No item número 3 figura a causalidade instantânea, a qual é discutível uma vez que é esperado haver uma defasagem no tempo entre uma causa e o conseqüente efeito. Essa questão foi abordada por Granger e Newbold (1977), que concluíram ser possível ocorrer causalidade instantânea devido às condições limitadas em que os dados podem apresentar-se. Assim, se a verdadeira defasagem entre causa e efeito corresponder ao período de um dia mas se as nossas observações forem efetuadas mensalmente, a relação de causalidade instantânea poderá ocorrer

perfeitamente.

A relação de "feedback" referente ao ítem número 4 corresponde ao fato de X_t causar Y_t e também Y_t causar X_t .

Passamos, agora, à descrição do procedimento proposto por Pierce. No primeiro estágio são ajustados modelos de séries temporais univariadas para cada série de interesse.

Suponhamos, para exemplificação, que dispomos de duas séries X_t e Y_t que são transformadas em x_t e y_t após ser tomado um número conveniente de diferenças para torná-las estacionárias. Cada uma dessas séries é, então, delineada segundo possíveis modelos constituintes da classe geral de modelos ARMA univariados.

Assim, as séries x_t e y_t são transformadas, respectivamente, nas séries de ruído branco não correlacionadas u_t e v_t , como segue:

$$\phi_x(B) \theta_x^{-1}(B) x_t = u_t$$

$$\phi_y(B) \theta_y^{-1}(B) y_t = v_t.$$

No estágio seguinte, são avaliados os padrões de comportamento entre as variáveis, refletidos pelas correlações cruzadas amostrais:

$$\hat{r}_s = r_{\hat{u}\hat{v}}(s) = \frac{\sum (\hat{u}_{t-s} - \bar{\hat{u}}) (\hat{v}_t - \bar{\hat{v}})}{\left[\sum (\hat{u}_t - \bar{\hat{u}})^2 \sum (\hat{v}_t - \bar{\hat{v}})^2 \right]^{1/2}}$$

ou seja,

$$\hat{r}_s = r_{\hat{u}\hat{v}}(s) = \frac{\sum_{t-s} \hat{u}_t \hat{v}_t}{\left[\sum_{t-s} \hat{u}_t^2 \sum_{t-s} \hat{v}_t^2 \right]^{1/2}} \quad (5.16)$$

Haugh (1976) demonstrou que, sob a hipótese nula de que x_t e y_t não estão relacionadas em termos de causalidade, as correlações $\{r_{\hat{u}\hat{v}}(s)\}$ são assintoticamente independentes e normalmente distribuídas com médias zero e desvios padrões $(\sqrt{N})^{-1}$.

Definindo

$$R = N \sum_{i=1}^m \hat{r}_{s_i}^2 \quad (5.17)$$

onde $s_1, \dots, s_m$ são inteiros, temos que R , sob a hipótese nula de independência das séries, tem distribuição assintótica de χ^2 com m graus de liberdade.

As condições de causalidade da Tabela 5.2 podem ser examinadas de acordo com esse resultado. Assim, a relação de número 1 " X_t causa Y_t " pode ser afirmada ao nível de significância $\leq \alpha$ se

$$N \sum_{s=1}^m \hat{r}_s^2 > \chi_{\alpha}^2(m), \quad (5.18)$$

onde o segundo membro de (5.18) é o ponto α -% superior da distribuição $\chi^2(m)$. Analogamente, a relação de número 8 " X_t e Y_t são independentes" seria confirmada ao nível α se e somente se

$$N \sum_{s=-m}^m \hat{r}_s^2 < \chi_{\alpha}^2(2m+1), \quad (5.19)$$

onde m é de alta magnitude.

Porém, em testes tais como (5.18) há o risco de sub-estimarmos a significância se alguns valores de $\rho_{uv}(s)$, correspondentes a $\hat{r}_s$, não presentes na estatística do teste forem não nulos significativamente.

No estágio seguinte, são especificados modelos preliminares para o caso multivariado ou modelos de regressão dinâmica se a causalidade for unidirecional.

A seguir, os parâmetros do modelo sugerido são estimados e é verificada a adequabilidade do ajuste do modelo, sendo o processo encerrado na fase de previsão.

5.4 - METODOLOGIA DE GRANGER E NEWBOLD

A técnica empregada por Granger e Newbold (1977) também tem início com a construção de modelos ARMA univariados para as séries componentes do vetor de séries temporais e a análise das correlações cruzadas residuais formuladas em (5.16).

Podemos algumas vezes relacionar k séries $\{z_{1t}\}, \dots, \{z_{kt}\}$ através de modelos lineares de função de transferência

$$z_{it} = \sum_{j \neq i} \frac{\omega_{ij}(B)}{\delta_{ij}(B)} z_{jt} + \frac{\theta_i(B)}{\phi_i(B)} a_{it}, \quad i=1, \dots, k. \quad (5.20)$$

No caso bivariado, (5.20) seria reduzida para:

$$z_{1t} = v_1(B) z_{2t} + \psi_1(B) a_{1t} \quad (5.21)$$

$$z_{2t} = v_2(B) z_{1t} + \psi_2(B) a_{2t}.$$

onde

$$v_j(B) = \frac{\omega_j(B)}{\delta_j(B)}, \quad j=1,2; \quad \psi_j(B) = \frac{\theta_j(B)}{\phi_j(B)}, \quad j=1,2.$$

Na fase de identificação, o problema consiste em encontrar formas apropriadas para os polinômios em B em (5.21), após terem sido feitos ajustes univariados para cada série.

No caso de haver "feedback", um passo útil é a aplicação de (5.21) para a associação conjunta das séries residuais geradas na modelação univariada. Assim, temos:

Modelos univariados:

$$\phi_{Z_1}(B) \theta_{Z_1}^{-1}(B) Z_{1t} = \epsilon_{1t}$$

$$\phi_{Z_2}(B) \theta_{Z_2}^{-1}(B) Z_{2t} = \epsilon_{2t}$$

Modelo bivariado:

$$\epsilon_{1t} = V_1^*(B) \epsilon_{2t} + \psi_1^*(B) a_{1t}^*$$

$$\epsilon_{2t} = V_2^*(B) \epsilon_{1t} + \psi_2^*(B) a_{2t}^*.$$

No caso de causalidade unidirecional, não é estritamente necessário pré-branquear a série dependente a menos que haja suspeita de "feedback" no sistema, em cuja situação torna-se essencial a construção de modelos univariados para ambas as séries.

Se houver causalidade unidirecional no sentido de Z_{2t} pa-

ra Z_{1t} , teremos as equações:

$$Z_{1t} = V_1(B)Z_{2t} + \psi_1(B)a_{1t} \quad (5.22)$$

$$\phi_{Z_2}(B)\theta_{Z_2}^{-1}(B)Z_{2t} = \varepsilon_{2t}.$$

onde a_{1t} e ε_{2t} são séries de ruído branco não correlacionadas.

Analogamente à situação de "feedback", costuma-se construir para as séries residuais o modelo de causalidade unidirecional proposto.

Assim, vem:

$$\varepsilon_{1t} = V_1^*(B)\varepsilon_{2t} + \psi_1^*(B)a_{1t}^*$$

Tanto no caso de "feedback" como causalidade unidirecional, a fase de identificação é finalizada com a combinação dos modelos univariados com os modelos envolvendo as séries residuais, objetivando a construção do modelo bivariado para as séries originais.

Na prática, se houver um grande número de parâmetros a serem estimados, surgirão dificuldades em termos de cálculos computacionais. Para contornar esse problema, devemos procurar a simplificação de expressões polinomiais através de cancelamentos e efetuar a remoção de coeficientes cuja magnitude for desprezível ao ser comparada com os correspondentes desvios padrões.

Após a estimação paramétrica são efetuados testes de cons-

tatação da adequabilidade do modelo, tais como um exame das autocorrelações residuais do modelo ajustado e a comparação do valor observado e do tabelado para a estatística $Q = N \cdot \sum_{i=1}^m \hat{r}_i^2$ derivada por Box e Jenkins (1970), a qual tem assintoticamente a distribuição de χ^2 com $m - R$ graus de liberdade, onde R é a soma das ordens dos polinômios $\theta(B)$ e $\phi(B)$. Também pode ser verificada a inexistência de correlação entre resíduos de modelos ajustados como, por exemplo, em (5.22), onde podemos confrontar as séries $\{a_{1t}\}$ e $\{\epsilon_{2t}\}$.

Como os próprios autores Granger e Newbold já admitiram, sua metodologia é complexa e quase impraticável quando $k > 2$, além do fato dos parâmetros constituintes do modelo dos resíduos estarem sujeitos a complicadas restrições não lineares.

5.5 - METODOLOGIA DE CHAN E WALLIS

A técnica de Chan e Wallis (1978) tem início com a especificação de modelos univariados para cada uma das séries componentes do vetor de séries temporais. É também explorado o fato de que a formulação de um modelo ARMA multivariado

$$\phi(B)\omega_t = \theta(B)a_t \quad (5.23)$$

pode ser alterada para:

$$|\phi(B)|\omega_t = H(B)a_t \quad (5.24)$$

onde $H(B) = A(B)\theta(B)$, sendo $A(B)$ a matriz adjunta e $|\phi(B)|$ o determinante de $\phi(B)$.

Escrevendo a relação (5.24) como

$$D_{\ell}(B)\omega_t = H(B)a_t, \quad (5.25)$$

teremos uma forma alternativa de representar um modelo ARMA multivariado, na qual a matriz dos coeficientes auto-regressivos é diagonal e de grau ℓ ; são ajustados modelos da forma (5.25) sucessivamente para ℓ assumindo valores $r, r-1, \dots$ e r é um valor correspondente a uma ordem máxima especificada de maneira a ser obtido um valor apropriado para ℓ .

Podemos efetuar um teste da razão de verossimilhança para verificar se os elementos diagonais de $D_{\ell}(B)$ são idênticos como na forma (5.24). A partir dos ajustes para $H(B)$ e $|\phi(B)|$ ou $D_{\ell}(B)$, podemos estabelecer valores de p e q em (5.23) bem como os valores paramétricos em $\phi(B)$ e $\theta(B)$.

As relações apresentadas nessa seção têm sido estudadas em um contexto econômico por Zellner e Palm (1974), Prothero e Wallis (1976) e Wallis (1977); Jenkins (1975) também emprega esse tipo de modelação de séries temporais multivariadas.

A eficiência dessa metodologia é discutida por Tiao e Box (1981) em alguns aspectos, tais como:

1 - É possível que o grau do polinômio $H(B)$ em (5.24) seja maior do que o máximo grau de $\theta_{q_i}(B)$ para cada uma das séries. Podemos notar essa propriedade a partir de um exemplo obtido com a suposição de que $k=2$, os elementos de a_t são independentes e $H(B)$ tem a seguinte configuração:

$$H(B) = \begin{pmatrix} 1 & -h_1 B \\ -h_2 B & 1 \end{pmatrix}.$$

Conseqüentemente, temos $q_1 = q_2 = 0$, conduzindo falsamente à conclusão de que $H(B)$ seria de grau zero.

2 - Usando a modelação sugerida em (5.24), incorremos em um aumento no número máximo de parâmetros a serem estimados em contraposição ao modelo apresentado em (5.23). Além da matriz de covariância Φ , esse número corresponde a $k^2(p+q)$ em (5.23) e a $kp + [(k-1)p+q]k^2$ em (5.24), havendo, portanto, um aumento de $pk(k-1)^2$ parâmetros neste último modelo. Tal aumento seria ainda maior no caso de empregarmos a representação (5.25).

3 - Há uma certa obscuridade na técnica de determinação dos valores de p e q em (5.23) a partir de (5.24), uma vez que a correspondência entre os graus de $|\phi(B)|$ e $H(B)$ e os valores de p e q não é necessariamente biunívoca.

4 - Segundo pesquisas feitas por Tiao e Box, o método de Chan e Wallis converge lentamente e, portanto, sua aplicação seria admissível somente no estágio final do processo de estimação.

Para melhor compreensão do método exposto nesta seção, ver exemplificação em Chan e Wallis (1978).

5.6 - DETERMINAÇÃO DA ORDEM DE UM PROCESSO AR MULTIVARIADO

Dentre os modelos existentes para ajustes e projeções de

séries temporais, o processo auto-regressivo é o menos complexo. Por esse motivo, um procedimento específico para determinar a ordem polinomial de processos auto-regressivos foi desenvolvido.

Tal procedimento proposto por Hannan e Quinn (1979) determina a ordem dos polinômios auto-regressivos multivariados com base em uma extensão do mesmo procedimento para o caso univariado.

A formulação sugerida para a determinação do grau p do polinômio auto-regressivo consiste em determinar a ordem p que minimiza

$$\phi(p) = \ln(|\hat{\Phi}_p|) + N^{-1} \cdot 2 \cdot p \cdot c \cdot \ln \ln N, \quad c > 1 \quad (5.26)$$

onde $\hat{\Phi}_p$ é a estimativa da matriz de covariância residual obtida através das equações de Yule-Walker no ajuste de um processo auto-regressivo de ordem p , cuja amostra tem tamanho N .

No caso univariado há diferentes técnicas para obtenção do valor ideal para p . O procedimento AIC proposto por Akaike (1969, 1970) foi provado ser não consistente de acordo com estudo desenvolvido por Shibata (1976). O procedimento BIC de Akaike e outras técnicas sugeridas por Schwarz (1978) e Rissanen (1978) embora fortemente consistentes sub-estimam o valor de p no caso de pequenas amostras comparativamente com o procedimento de Hannan e Quinn (1979), que também é fortemente consistente.

Bhansali e Downham (1977) propuseram uma modificação do procedimento AIC, a qual não é fortemente consistente, ao passo que Akaike (1978,1979) empregou um argumento Bayesiano para conservar o procedimento AIC sem modificação.

Em (5.26), Hannan e Quinn (1979) provaram que a estimativa $\hat{p}$ obtida através da minimização de $\phi(p)$ é fortemente consistente e tende a p_0 , com $0 < p \leq P$, onde P é escolhido a priori como sendo maior do que o verdadeiro grau p_0 . Para a atribuição do valor de c é sugerido que seja $c = k^2$.

No caso multivariado o procedimento AIC pode super-estimar o verdadeiro valor de p .

Embora existam resultados referentes à estimação da ordem de modelos ARMA no caso univariado (ver Hannan, 1980), os análogos multivariados não foram ainda desenvolvidos.

CAPÍTULO 6

APLICAÇÕES

6.1 - INTRODUÇÃO

No decorrer dos capítulos anteriores foram apresentadas a teoria referente às técnicas empregadas na estimação em séries multivariadas, bem como algumas idéias referentes à identificação de modelos.

Para complementar este estudo, focalizaremos, agora, ilustrações dos diferentes métodos já abordados neste trabalho.

6.2 - EXEMPLO 1 (Phadke e Kedem, 1978)

A metodologia de Phadke e Kedem foi descrita na seção 4.4.

Foram realizadas algumas simulações com o objetivo de obter os erros quadráticos médios das estimativas referentes à maximização da função de verossimilhança aproximada e a exata. Com posse das magnitudes dos erros quadráticos médios, podemos estabelecer comparações entre os desempenhos alcançados em cada técnica.

Para isso, foram feitas simulações, ajustando-se um modelo bivariado MA(1), com os seguintes parâmetros:

$$\theta_0 = I,$$

$$\theta_1 = \begin{bmatrix} 0,98 & 1,96 \\ 0 & 0,98 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Na Tabela 6.1 a seguir, temos os valores dos erros quadráticos médios correspondentes às estimativas paramétricas resultantes da maximização das funções de verossimilhança aproximada e exata.

TABELA 6.1

Métodos Parâmetros	Aproximado	Exato
$\theta_{1,11}$	0,05580	0,03940
$\theta_{1,21}$	0,00917	0,01029
$\theta_{1,12}$	0,00893	0,00145
$\theta_{1,22}$	0,00696	0,00066
σ_1^2	0,05298	0,00251
σ_2^2	0,00801	0,00266

Observamos assim, que o método exato fornece erros quadráticos médios consideravelmente menores do que os obtidos pelo método aproximado.

Também foram pesquisados por Phadke e Kedem os casos em que todas as raízes de $\theta(B)$ estão fora do círculo unitário. A conclusão resultante é que os métodos aproximado e exato produzem erros quadráticos médios pró-

ximos quando as raízes de $\theta(B)$ estão fora do círculo unitário, mas quando elas estiverem próximos ou no próprio círculo unitário o método exato acarreta erros quadráticos médios muito menores do que os referentes ao método aproximado.

6.3 - EXEMPLO 2 (Hillmer e Tiao, 1979)

Nas seções 3.5 e 4.6 foi apresentada a metodologia de Hillmer e Tiao.

Para ilustração de suas técnicas foram geradas 100 realizações a partir de um modelo bivariado MA(1) para vários valores do tamanho amostral N.

Para cada realização foi utilizada uma rotina de regressão não linear para maximizar as funções de verossimilhança condicional e exata (4.22) e (4.21) respectivamente.

Uma síntese dos resultados é divulgada na Tabela 6.2.

Através da Tabela 6.2 podemos tecer os seguintes comentários:

1 - Pela função de verossimilhança exata obtemos estimativas bastante próximas dos verdadeiros valores paramétricos para todos os valores de N.

2 - Os erros quadráticos médios relativos ao método condicional decrescem à medida que N aumenta.

3 - Os erros quadráticos médios relativos ao método exato são consideravelmente menores do que os gerados pelo método condicional.

TABELA 6.2

N	Método	Médias das Estimativas paramétricas				Médias das Estimativas valores característicos		E.Q.M. $\times 100$			
		θ_{11}	θ_{12}	θ_{21}	θ_{22}			θ_{11}	θ_{12}	θ_{21}	θ_{22}
50	E	0,71	0,55	0,21	0,56	0,98	0,29	0,849	1,848	0,801	2,685
	C	0,69	0,51	0,17	0,50	0,90	0,29	1,720	2,699	1,762	3,744
100	E	0,72	0,56	0,21	0,57	1,00	0,29	0,290	0,848	0,324	0,953
	C	0,72	0,53	0,19	0,53	0,96	0,29	0,835	1,137	0,510	1,538
200	E	0,72	0,57	0,21	0,57	1,00	0,29	0,113	0,382	0,108	0,406
	C	0,70	0,54	0,19	0,54	0,95	0,29	0,185	0,493	0,201	0,665
400	E	0,72	0,56	0,21	0,57	1,00	0,29	0,054	0,191	0,057	0,207
	C	0,70	0,54	0,19	0,55	0,95	0,30	0,011	0,238	0,121	0,354
1.000	E	0,72	0,56	0,21	0,58	1,00	0,30	0,015	0,057	0,016	0,063
	C	0,71	0,55	0,20	0,56	0,98	0,29	0,034	0,093	0,041	0,122
Valores verdadeiros dos parâmetros		0,72	0,56	0,21	0,58	1,00	0,30				

E: verossimilhança exata

C: verossimilhança condicional.

4 - Os resultados encontrados nos conduzem à conclusão que quando o modelo não é invertível o desempenho do método exato é superior ao do método condicional.

6.4 - EXEMPLO 3 (Tiao e Box, 1981)

São utilizadas duas séries temporais já empregadas por Box e Jenkins (1970), os quais, em sua pesquisa, interessaram-se pela otimização da concentração de Gás Carbônico (CO_2) em uma fornalha, usando como entrada uma combinação de ar com metano em intervalos de 9 segundos para formar uma mistura de gases contendo CO_2 . A quantidade de ar permaneceu constante mas a razão do gás metano variou implicando uma variação da concentração de CO_2 . Seja Z_{1t} a razão de entrada do gás metano e Z_{2t} a concentração de CO_2 .

A priori, sabe-se que $\{Z_{2t}\}$ é uma variável dependente de $\{Z_{1t}\}$ e é proposto um modelo linear de função de transferência já delineado em (5.20).

Na fase de identificação, foram formulados os modelos abaixo, de acordo com a teoria de modelos de função de transferência exposta por Box e Jenkins (1970), cap. 11.

$$(1-1,97B+1,37B^2-0,34B^3)Z_{1t} = a_{1t}, \quad \hat{\sigma}_{a_1}^2 = 0,0353 \quad (6.1)$$

$$Z_{2t} = \frac{\omega(B)}{\delta(B)}B^b Z_{1t} + \frac{a_{2t}}{\phi(B)}, \quad \hat{\sigma}_{a_2}^2 = 0,0561 \quad (6.2)$$

onde:

$$\omega(B) = -(0,53+0,37B+0,51B^2),$$

$$\delta(B) = 1 - 0,57B$$

$\phi(B) = 1 - 1,53B + 0,63B^2$, sendo $\{a_{1t}\}$ e $\{a_{2t}\}$ séries independentes.

Veremos, agora, os resultados a que chegaríamos procedendo de acordo com Tiao e Box (1981).

Consideremos a Tabela 6.3

TABELA 6.3 a) Autocorrelações e correlações cruzadas dos dados originais das séries $\{Z_{1t}\}$ e $\{Z_{2t}\}$

	Z_{1t}	Z_{2t}
Z_{1t}	+++++++++	-----
Z_{2t}	-----	+++++++++

TABELA 6.3 b) Estatística $M(\ell)$ para a auto-regressão parcial

ℓ	1	2	3	4	5	6	7	8	9	10	11
$M(\ell)$	1650	665	31,7	22,5	5,6	12,9	1,8	8,0	3,5	0	2,0

TABELA 6.3 c) Correlações cruzadas dos resíduos $\hat{a}_t$ após o ajuste do modelo AR(6)

	$\hat{a}_{1t}$	$\hat{a}_{2t}$
$\hat{a}_{1t}$	-	
$\hat{a}_{2t}$		

Em 6.3a) e 6.3c) foram empregados símbolos indicadores da magnitude das correlações cujo significado encontra-se na seção 5.2.

Através de 6.3a) eliminamos a hipótese de termos um modelo de médias móveis, pois o comportamento das correlações nessa tabela diverge do estabelecido em (5.3), uma vez que neste exemplo as autocorrelações persistem em apresentar altas magnitudes.

De acordo com 6.3b), os valores da estatística $M(l)$ revelam serem significantes os coeficientes auto-regressivos estimados até a defasagem de ordem 6, após a qual os valores dessa estatística passam a ser inferiores a seu valor tabelado correspondente a um χ^2 com 4 graus de liberdade e igual a 9,49.

Conseqüentemente foi ajustado um modelo AR(6) aos dados das duas séries e em 6.3c) o comportamento das correlações cruzadas residuais confirma o acerto da escolha desse modelo.

No estágio seguinte, foi feita a estimação de verossimilhança condicional para o modelo AR(6), cujos fundamentos encontram-se na seção 3.4. Este procedimento de estimação foi apresentado por Wilson (1973) e as propriedades das estimativas de máxima verossimilhança foram ponderadas por Nicholls (1976,1977) e Anderson (1980). Vejamos o motivo da utilização deste método.

Para um modelo geral ARMA(p,q) podemos verificar que a correspondente função de verossimilhança exata tem grande pro-

ximidade com a obtida através da representação do processo ARMA em termos de um processo MA(q). Sendo assim, os algoritmos de estimação de um modelo MA(q) foram programados no computador para o estágio de estimação. Como no modelo MA(q) o método exato não converge com rapidez, foi empregado o método condicional nas primeiras iterações e o exato na final.

Assim, para o modelo AR(6) expressado por:

$$(I - \phi_1 B - \dots - \phi_6 B^6) Z_t = a_t \quad (6.3)$$

apresentamos, a seguir, as estimativas paramétricas acompanhadas dos respectivos erros padrões colocados entre parênteses.

São também fornecidos a matriz de covariâncias dos resíduos e a sua correlação.

$$\hat{\phi}_1 = \begin{bmatrix} 1,93 & -0,05 \\ (0,06) & (0,05) \\ 0,06 & 1,55 \\ (0,08) & (0,06) \end{bmatrix} \quad \hat{\phi}_2 = \begin{bmatrix} -1,20 & 0,10 \\ (0,13) & (0,08) \\ -0,14 & -0,59 \\ (0,16) & (0,11) \end{bmatrix}$$

$$\hat{\phi}_3 = \begin{bmatrix} 0,17 & -0,08 \\ (0,15) & (0,09) \\ -0,44 & -0,17 \\ (0,19) & (0,11) \end{bmatrix} \quad \hat{\phi}_4 = \begin{bmatrix} -0,16 & 0,03 \\ (0,15) & (0,09) \\ 0,15 & 0,13 \\ (0,19) & (0,11) \end{bmatrix}$$

$$\hat{\phi}_5 = \begin{bmatrix} 0,38 & -0,04 \\ (0,14) & (0,08) \\ -0,12 & 0,06 \\ (0,18) & (0,10) \end{bmatrix} \quad \hat{\phi}_6 = \begin{bmatrix} -0,22 & 0,03 \\ (0,08) & (0,03) \\ 0,25 & -0,04 \\ (0,11) & (0,04) \end{bmatrix}$$

$$\hat{\Phi} = \begin{bmatrix} 0,0345 & \\ -0,0023 & 0,0566 \end{bmatrix}, \quad \hat{\rho}(a_1, a_2) = 0,045. \quad (6.4)$$

Podemos observar que $\hat{\phi}_{12}^{(\ell)}$ é de pequena magnitude em comparação com seu erro padrão para todos os valores das defasagens $\ell = 1, \dots, 6$. Confirmamos, assim, a relação de causalidade unidirecional existente entre Z_{1t} e Z_{2t} e eliminamos a possibilidade da relação de "feedback" entre as séries. Como $\hat{\phi}_{21}^{(\ell)}$ é de pequena magnitude para $\ell = 1, 2$ e os resíduos $\hat{a}_{1t}$ e $\hat{a}_{2t}$ são não correlacionados, segue que há um retardamento de 3 períodos entre as séries. Coincidentemente, as variâncias encontradas em (6.4) para a_{1t} e a_{2t} aproximam-se bastante das obtidas em (6.1) e (6.2).

Atribuindo o valor zero aos coeficientes cujas estimativas são desprezíveis quando comparadas com seus erros padrões, obtemos um modelo simplificado exposto a seguir, juntamente com os resultados da fase de estimação.

$$\hat{\Phi}_1 = \begin{bmatrix} 1,98 & \cdot \\ (0,06) & \\ \cdot & 1,53 \\ & (0,06) \end{bmatrix}, \quad \hat{\Phi}_2 = \begin{bmatrix} -1,38 & \cdot \\ (0,10) & \\ \cdot & -0,58 \\ & (0,11) \end{bmatrix}$$

$$\hat{\Phi}_3 = \begin{bmatrix} 0,35 & \cdot \\ (0,06) & \\ -0,53 & -0,14 \\ (0,07) & (0,10) \end{bmatrix}, \quad \hat{\Phi}_4 = \begin{bmatrix} \cdot & \cdot \\ 0,11 & 0,12 \\ (0,16) & (0,04) \end{bmatrix}$$

$$\hat{\phi}_5 = \begin{bmatrix} \cdot & \cdot \\ -0,04 & \cdot \\ (0,17) & \cdot \end{bmatrix}, \quad \hat{\phi}_6 = \begin{bmatrix} \cdot & \cdot \\ 0,21 & \cdot \\ (0,11) & \cdot \end{bmatrix}$$

$$\hat{\Phi} = \begin{bmatrix} 0,0359 & -0,0029 \\ \cdot & 0,0561 \end{bmatrix}, \quad \hat{\rho}(a_1, a_2) = 0 \quad (6.5)$$

De acordo com $\hat{\Phi}$ e $\hat{\rho}$ podemos concluir que o modelo é apropriado.

Portanto, temos a formulação final para o processo AR(6), expressada por:

$$\begin{bmatrix} \phi_{11}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{bmatrix} \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \quad (6.6)$$

onde

$$\begin{bmatrix} \phi_{11}(B) \\ \phi_{21}(B) & \phi_{22}(B) \end{bmatrix} = I - \hat{\phi}_1 B - \dots - \hat{\phi}_6 B^6.$$

Temos, então:

$$\begin{aligned} \phi_{11}(B) &= 1 - 1,98B + 1,38B^2 - 0,35B^3 \\ \phi_{21}(B) &= 0,53B^3 - 0,11B^4 - 0,21B^6 \\ \phi_{22}(B) &= 1 - 1,53B + 0,58B^2 + 0,14B^3 - 0,12B^4 \end{aligned} \quad (6.7)$$

Considerando a_{1t} e a_{2t} não correlacionados, o modelo para a variável independente $\phi_{11}(B)Z_{1t} = a_{1t}$ com $\text{var}(a_{1t}) = 0,0359$ é basicamente o mesmo apresentado em (6.1).

Para a variável dependente, Z_{2t} , temos a partir de (6.6)

$$\begin{aligned}\phi_{21}(B)Z_{1t} + \phi_{22}(B)Z_{2t} &= a_{2t} \\ Z_{2t} &= -\frac{\phi_{21}(B)}{\phi_{22}(B)}Z_{1t} + \frac{1}{\phi_{22}(B)}a_{2t}\end{aligned}\quad (6.8)$$

com $\text{Var}(a_{2t}) = 0,0561$.

A parcela correspondente ao modelo do ruído branco em (6.8) pouco difere de sua correspondente em (6.2).

Aparentemente o primeiro termo das expressões (6.2) e (6.8) divergem. Entretanto, podemos reescrever a primeira parcela do segundo membro de (6.8) da forma:

$$-\phi_{21}(B)\phi_{22}^{-1}(B) = \sum_{j=0}^{\infty} v_j^* B^j. \quad (6.9)$$

Analogamente, a primeira parcela do segundo membro de (6.2) pode ser expressada por:

$$\omega(B)B^b\delta^{-1}(B) = \sum_{j=0}^{\infty} v_j B^j. \quad (6.10)$$

Os termos v_j e v_j^* são conhecidos como funções de resposta de impulso em modelos de função de transferência. Ver Box e Jenkins (1970), Cap. 11.

Podemos observar a semelhança dos valores assumidos pelas

expressões (6.9) e (6.10) através de comparação das funções de resposta de impulso apresentadas na Tabela 6.4.

TABELA 6.4 - Funções de resposta de impulso para os dados da fornalha de gases

	0	1	2	3	4	5	6	7	8	9	10	11	12
v_j	•	•	•	-0,53	-0,67	-0,89	-0,51	-0,29	-0,17	-0,09	-0,05	-0,03	-0,02
v_j^*	•	•	•	-0,53	-0,70	-0,77	-0,48	-0,26	-0,09	-0,01	0,01	0,00	-0,01

É interessante que examinemos as alterações que ocorrem nos modelos auto-regressivos ajustados, à medida que a ordem do polinômio auto-regressivo sofre um acréscimo.

Para isso, consideremos a Tabela 6.5.

TABELA 6.5 - Resultados correspondentes a ajustes auto-regressivos sucessivos AR(p)

p	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	$\hat{\sigma}^2$
1	$\begin{bmatrix} + & + \\ - & + \end{bmatrix}$						$\begin{bmatrix} 0,102 & 0,090 \\ & 0,346 \end{bmatrix}$
2	$\begin{bmatrix} + & - \\ + & + \end{bmatrix}$	$\begin{bmatrix} - & + \\ - & - \end{bmatrix}$					$\begin{bmatrix} 0,037 & -0,004 \\ & 0,069 \end{bmatrix}$
3	$\begin{bmatrix} + & \cdot \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ - & - \end{bmatrix}$	$\begin{bmatrix} + & \cdot \\ \cdot & + \end{bmatrix}$				$\begin{bmatrix} 0,036 & -0,002 \\ & 0,063 \end{bmatrix}$
4	$\begin{bmatrix} + & \cdot \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & - \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ - & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ + & + \end{bmatrix}$			$\begin{bmatrix} 0,036 & -0,003 \\ & 0,059 \end{bmatrix}$
5	$\begin{bmatrix} + & \cdot \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & - \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ - & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$		$\begin{bmatrix} 0,035 & -0,003 \\ & 0,058 \end{bmatrix}$
6	$\begin{bmatrix} + & \cdot \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & - \end{bmatrix}$	$\begin{bmatrix} + & \cdot \\ - & - \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ + & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ - & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} 0,035 & -0,002 \\ & 0,057 \end{bmatrix}$

Podemos tirar as seguintes conclusões:

1) Se fosse adotado um processo AR(1) ou AR(2), chegaríamos a admitir relação de "feedback" entre as variáveis pois incluiríamos o coeficiente $\phi_{12}(B)$ no modelo, o que incorreria em uma formulação errônea para o modelo.

2) A relação de causalidade unidirecional é bem retratada quando a ordem do polinômio auto-regressivo é aumentada para 3.

3) A defasagem $b = 3$ em (6.2) fica explícita quando a ordem p é aumentada para 4, pois a posição (2,1) torna-se significativa a partir da estimativa de ϕ_3 .

6.5 - EXEMPLO 4 (Pierce, 1980)

Consideremos as séries: taxa de desemprego (TD) e índice de produção industrial (IPI), sendo os correspondentes valores obtidos mensalmente, no decorrer do período de Dezembro de 1968 a Junho de 1976 nos EUA (dados do Federal Reserve Board and Bureau of Labor Statistics).

Após convenientes transformações, as séries tornam-se estacionárias e serão denotadas, respectivamente, por y_t e x_t .

Cada série é, então, ajustada separadamente em função do tempo e é delineada de acordo com um modelo ARMA univariado.

Foram obtidos, dessa maneira, os seguintes resultados:

$$(1 - 0,70B)x_t = \hat{u}_t, \quad \hat{\sigma}_u^2 = 0,066 \cdot 10^{-3} \quad (6.11)$$

$$(1-0,42B)y_t = (1+0,26B^4)\hat{v}_t, \quad \hat{\sigma}_v^2 = 0,827 \cdot 10^{-3}. \quad (6.12)$$

Foram calculadas as estatísticas de Box e Pierce (1970) definidas por:

$$Q_u = N \sum_{k=1}^k r_k^2(\hat{u}) \quad \text{e} \quad Q_v = N \sum_{k=1}^k r_k^2(\hat{v})$$

onde k é tomado com magnitude suficientemente grande de maneira que os pesos ψ_j para um modelo ARMA univariado expressado por:

$$\omega_t = \phi_\omega^{-1}(B)\theta_\omega(B)a_t = \psi(B)a_t \quad (6.13)$$

tornem-se desprezíveis após $j = k$.

Se o modelo ajustado é adequado, Q tem aproximadamente uma distribuição $\chi^2(k-p-q)$.

Obtivemos $Q_x = 13,2$ com 11 graus de liberdade e $Q_y = 8,2$ com 10 graus de liberdade ambas as estatísticas baseadas nas 12 primeiras autocorrelações residuais. Comparadas com os correspondentes valores tabelados, estas estatísticas comprovam a adequabilidade dos modelos sugeridos.

A seguir, é fornecida uma tabela com as correlações cruzadas amostrais apresentadas em (5.16).

Vejamos as condições de causalidade entre as séries.

Podemos observar a existência de uma relação instantânea entre x_t e y_t retratada em $r_0 = -0,24$ e de uma implicação de x_t para y_t em $r_1 = -0,35$.

TABELA 6.6 - Correlações cruzadas amostrais para as séries IPI e TD

Valores de s	$\hat{r}_s$										
-11 a -20	0,12	-0,03	-0,13	0,13	0,10	0,03	-0,05	0,04	0,08	-0,04	
-1 a -10	-0,05	-0,07	0,02	-0,09	0,04	0,00	0,06	-0,03	0,05	0,06	
0 a 10	-0,24	-0,35	-0,08	-0,11	0,02	-0,16	0,11	0,07	-0,02	-0,05	-0,19
11 a 20		-0,02	0,09	0,05	0,07	0,15	0,03	0,02	-0,03	-0,06	-0,01

Não podemos presumir a relação de "feedback" entre x_t e y_t , pois o ítem de número 4 da Tabela 5.2 não se manifesta na Tabela 6.6.

Segue a construção de um modelo bivariado envolvendo x_t e y_t . Neste processo, primeiramente esboçamos um modelo para as séries u_t e v_t , o qual ao ser combinado com os modelos univariados (6.11) e (6.12), fornecerá subsídios para a especificação de um modelo para as séries x_t e y_t . (Ver Parzen, 1969; Haugh e Box, 1977).

Foi, então, derivado o seguinte modelo de função de transferência:

$$v_t = (\beta_0 + \beta_1 B)u_t + (1 - \theta_1 B)a_t = \omega_t + e_t. \quad (6.14)$$

A estrutura de médias móveis estipulada para e_t é consequente da propriedade de v_t ser ruído branco e portanto

$$E(\omega_t \omega_{t-s}) = -E(e_t e_{t-s}) \quad (6.15)$$

para todo valor de s.

De (6.14), podemos encontrar uma expressão relacionando

as correlações cruzadas com os coeficientes β_0 e β_1 .

Temos, então:

$$\begin{aligned}\rho_s &= E(v_t u_{t-s}) \cdot \frac{1}{\sigma_u \sigma_v} \\ \rho_0 &= E(v_t u_t) \cdot \frac{1}{\sigma_u \sigma_v} = E(\beta_0 u_t^2 + \beta_1 u_{t-1} u_t) \cdot \frac{1}{\sigma_u \sigma_v} \\ \rho_0 &= \beta_0 \sigma_u^2 \cdot \frac{1}{\sigma_u \sigma_v}\end{aligned}$$

e, portanto,

$$\rho_0 = \beta_0 \cdot \frac{\sigma_u}{\sigma_v} \quad (6.16)$$

$$\begin{aligned}\rho_1 &= E(v_t u_{t-1}) \cdot \frac{1}{\sigma_u \sigma_v} = E(\beta_0 u_t u_{t-1} + \beta_1 u_{t-1}^2) \cdot \frac{1}{\sigma_u \sigma_v} \\ \rho_1 &= \beta_1 \sigma_u^2 \cdot \frac{1}{\sigma_u \sigma_v} \\ \rho_1 &= \beta_1 \cdot \frac{\sigma_u}{\sigma_v} \quad (6.17)\end{aligned}$$

Através de (6.11), (6.12), (6.16) e (6.17), obtemos as estimativas iniciais

$$\hat{\beta}_0 = -0,85 \text{ e } \hat{\beta}_1 = -1,24.$$

Resolvendo em (6.15) as equações para $s=0$ e $s=1$, podemos estimar $\text{Var}(a_t)$ e θ_1 mas, embora consistentes, tais estimativas não são eficientes.

Utilizando (6.11) e (6.12) e reescrevendo x_t e y_t nas formas auto-regressivas:

$$u_t = F(B)x_t \text{ e } v_t = G(B)y_t$$

podemos substituir em (6.14) as expressões assim obtidas para u_t e v_t e converter a formulação delineada em função de y_t , chegando ao seguinte resultado:

$$\begin{aligned} G(B)y_t &= (\beta_0 + \beta_1 B)F(B)x_t + (1 - \theta_1 B)a_t \\ y_t &= (\beta_0 + \beta_1 B) \frac{F(B)}{G(B)} x_t + \frac{(1 - \theta_1 B)}{G(B)} a_t \end{aligned} \quad (6.18)$$

Assim, temos estimativas para os ajustes iniciais de y_t em função de x_t . No próximo estágio, são feitas estimativas paramétricas.

Obtém-se a equação:

$$y_t = (-0,8-1,2B) \frac{(1+0,26B^4)}{(1-0,42B)} (1-0,7B)x_t + \frac{(1+0,26B^4)}{(1-0,42B)} (1-\theta_1 B)a_t. \quad (6.19)$$

A formulação final é especificada por:

$$y_t = (-1,12-0,87B-0,21B^4-0,74B^5)x_t + a_t \quad (6.20)$$

$$\hat{\sigma}_a^2 = 0,609 \cdot 10^{-3}.$$

Supondo que os modelos estimados (6.20), (6.11) e (6.12) são verdadeiros, podemos determinar um índice comparativo dos erros quadráticos acarretados pelos modelos univariados emul-tivariados.

Assim, temos:

$$R_*^2 = \frac{\hat{\sigma}_v^2 - \hat{\sigma}_a^2}{\hat{\sigma}_v^2} = \frac{0,827 - 0,609}{0,827} = 0,26.$$

Portanto, o ajuste da série TD recebe uma redução em sua variância quando é delineado em termos dos valores passados de TD e dos valores presentes e passados de IPI, comparativamente ao ajuste de TD sem a incorporação das informações relativas a IPI.

6.6 - EXEMPLO 5 (Granger e Newbold, 1977)

Foram feitas 144 observações mensais também relativas a número de desempregados (ND) e índice de produção industrial (IPI) nos EUA. Novamente, denotaremos essas séries por y_t e x_t respectivamente.

Para essas séries foram ajustados modelos ARMA univariados. São eles:

$$\begin{matrix} (1+0,06B-0,21B^2-0,18B^3-0,19B^4) & (1-B)y_t & = & \hat{v}_t \\ (0,08) & (0,08) & (0,08) & (0,08) \end{matrix} \quad (6.21)$$

$$\hat{\sigma}_v^2 = 13.617.$$

$$\begin{matrix} (1-0,26B) & (1-B)x_t & = & 0,21 + \hat{u}_t, & \hat{\sigma}_u^2 = 0,49 \\ (0,08) & & (0,06) & \end{matrix} \quad (6.22)$$

Os valores entre parênteses correspondem aos desvios padrões das estimativas.

Em seguida, foram analisadas as correlações cruzadas residuais apresentadas na Tabela 6.7.

Após terem sido tomadas diferenças nas séries para torná-las estacionárias, passamos a ter o número de observações,

TABELA 6.7 - Correlações cruzadas residuais para as séries IPI e ND

Valores de s	$\hat{r}_s$							
-8 a -15	-0,07	-0,08	0,02	0,05	0,08	-0,10	0,05	-0,01
-1 a -7		-0,08	-0,05	-0,07	-0,09	0,04	-0,08	-0,11
0 a 7	-0,16	-0,22	-0,22	-0,06	-0,13	-0,02	-0,04	0,05
8 a 15	0,01	-0,06	-0,03	-0,04	0,13	0,01	0,12	0,07

N, igual a 139. Como já vimos, $\hat{r}_s$ são aproximadamente normalmente distribuídas com média zero e desvios padrões $(\sqrt{N})^{-1}$. Assim, neste nosso exemplo, $(\sqrt{N})^{-1} \approx 0,085$.

Pela Tabela 6.7, vemos que $\hat{r}_s$ têm magnitude desprezível para $s < 0$ em comparação com seus desvios padrões. Podemos, então, concluir que não há relação de "feedback" entre as séries pois, de acordo com o item 4 da Tabela 5.2, deve haver pelo menos um valor $\hat{r}_s \neq 0$ tanto para $s > 0$ como $s < 0$.

Mesmo não havendo "feedback", decidiu-se fazer uma associação conjunta das séries residuais obtidas a partir dos modelos univariados.

Segue que:

$$\hat{v}_t = \frac{\omega_1(B)}{\delta_1(B)} \hat{u}_t + \frac{\theta_1(B)}{\delta_1(B)} \eta_{1,t}. \quad (6.23)$$

Para simplificar, as séries residuais foram padronizadas com a divisão dos resíduos pelos respectivos desvios padrões. Assim, (6.23) consiste em um modelo para as séries residuais

padronizadas.

Escrevendo

$$\frac{\omega_1(B)}{\delta_1(B)} = V_1(B) = V_{1,0} + V_{1,1}B + V_{1,2}B^2 + \dots$$

e tomando a esperança de (6.23) após multiplicarmos ambos os seus membros por $\hat{u}_{t-s}$, $s \geq 0$, temos que

$$E(\hat{v}_t \hat{u}_{t-s}) = V_{1,s}, \quad s = 0, 1, 2, \dots$$

De fato,

$$\hat{v}_t \hat{u}_{t-s} = (V_1(B) \hat{u}_t) \hat{u}_{t-s} + \left(\frac{\theta_1(B)}{\delta_1(B)} \eta_{1,t} \right) \hat{u}_{t-s}$$

$$E(\hat{v}_t \hat{u}_{t-s}) = V_{1,s} E(\hat{u}_t^2) + \frac{\theta_1(B)}{\delta_1(B)} E(\eta_{1,t} \hat{u}_{t-s})$$

$$E(\hat{v}_t \hat{u}_{t-s}) = V_{1,s},$$

uma vez que $\hat{u}_t$ e $\eta_{1,t}$ são não correlacionados e $\hat{u}_t$ têm variância unitária. Assim, os valores de $\hat{r}_s$, $s > 0$ fornecidos pela Tabela 6.7 são estimativas de $V_{1,s}$. Pelas magnitudes de $\hat{r}_s$, $s > 0$ é proposta a equação:

$$V_1(B) = \omega_{1,0} + \omega_{1,1}B + \omega_{1,2}B^2.$$

$$\text{Analogamente, } \theta_1(B) = 1 - \theta_{1,1}B - \theta_{1,2}B^2.$$

Para maiores detalhes, ver Granger e Newbold (1976, Cap. 7).

Através da minimização de $\sum \eta_{1,t}^2$, usando um algoritmo não

linear de regressão, é estimado um modelo para as séries ND e IPI, a saber:

$$\hat{v}_t = \begin{matrix} (-0,16-0,26B-0,25B^2) \\ (0,08) (0,08) (0,08) \end{matrix} \hat{u}_t + \begin{matrix} (1-0,21B-0,19B^2) \\ (0,09) (0,09) \end{matrix} \eta_{1,t} \quad (6.24)$$

O próximo passo foi a combinação de (6.21) e (6.22) com (6.24).

Após remoções de termos desprezíveis quando comparados com seus desvios padrões e reajustes para cada equação resultante em decorrência dessas remoções, foi obtido o modelo:

$$y_t = \begin{matrix} (13,2) (14,2) (17,0) \\ (-28,9-47,2B-36,8B^2) \\ (1+0,39B-0,53B^2-0,16B^3) \\ (0,18) (0,12) (0,12) \end{matrix} x_t + \begin{matrix} 1 \\ (1+0,24B) \\ (0,08) \end{matrix} \eta_{1,t}^* \quad (6.25)$$

$$\hat{\sigma}_{\eta_1^*}^2 = 10.591.$$

Para confirmar a adequabilidade do modelo, foram examinadas as autocorrelações residuais apresentadas na Tabela 6.8.

TABELA 6.8 - Autocorrelações residuais provenientes da equação (6.25)

Valores de τ	$r_\tau(\hat{\eta}_1^*)$									
1 a 10	-0,03	-0,05	0,00	0,08	0,05	0,03	-0,13	0,07	-0,04	-0,03
11 a 20	-0,02	-0,23	0,01	0,01	0,09	-0,04	-0,08	-0,09	0,20	-0,07

Os resultados são satisfatórios com exceção dos valores correspondentes a $\tau = 12$ e $\tau = 19$ que são inesperados e provavelmente decorrem de alguma imprecisão na fase de ajuste sazo-

nal a que foram submetidos os dados originais das séries para torná-las estacionárias. A estatística Q também avaliada apresentou como valor observado 22,25, o qual comparado com o valor extraído da tabela de χ^2 para 19 graus de liberdade, 30,14, não contradiz a hipótese da série residual ser ruído branco.

Foram também examinadas as correlações cruzadas entre os resíduos gerados de (6.25) e (6.22). Essas correlações são apresentadas na Tabela 6.9.

TABELA 6.9 - Correlações cruzadas entre os resíduos provenientes de (6.25) e (6.22)

Valores de τ	$r(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau})$									
-10 a -19	-0,04	0,02	0,03	-0,10	-0,01	-0,01	0,03	-0,06	-0,10	0,01
-1 a -9		-0,06	-0,03	-0,04	-0,10	0,04	-0,09	-0,10	-0,05	-0,05
0 a 9	-0,01	0,00	0,04	0,00	-0,05	-0,04	-0,03	0,04	0,01	-0,07
10 a 19	-0,01	0,06	0,14	0,00	0,12	0,10	0,04	0,14	-0,04	0,06

Sob a hipótese nula, Haugh (1972) mostrou que essas correlações cruzadas são normalmente distribuídas com média zero e desvios padrões $(\sqrt{N})^{-1}$.

A Tabela 6.9 evidencia o fato dessas séries residuais serem não correlacionadas.

Haugh também derivou as estatísticas:

$$Q_1 = N \sum_{\tau=0}^{M_1} r^2(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau}), \quad Q_2 = N \sum_{\tau=-M_2}^{M_2} r^2(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau})$$

$$Q_3 = N \sum_{\tau=-M_3}^{-1} r^2(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau})$$

as quais, sob a hipótese nula, têm aproximadamente distribuição χ^2 com $M_1 - U$, $2M_2 - U$ e M_3 graus de liberdade, respectivamente, onde U é a soma das ordens dos polinômios $\omega_1(B)$ e $\delta_1(B)$ implícitos em $V_1(B)$ em (5.22). Neste exemplo, o valor de U é extraído de (6.25) e é igual a 5.

Em relação aos resultados destacados na Tabela 6.9, temos

$$Q_1 = 139 \sum_{\tau=0}^{19} r^2(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau}) = 12,15$$

$$Q_2 = 139 \sum_{\tau=-12}^{12} r^2(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau}) = 11,16$$

$$Q_3 = 139 \sum_{\tau=-19}^{-1} r^2(\hat{\eta}_{1,t}^*, \hat{u}_{t-\tau}) = 9,52.$$

Os resultados das estatísticas Q_1 , Q_2 e Q_3 assim obtidos, quando comparados com os respectivos valores da Tabela de χ^2 com 14, 19 e 19 graus de liberdade iguais a 23,68 e 30,14, reforçam a nossa decisão de não rejeitarmos o modelo (6.25).

Portanto, a formulação conjunta exibida por (6.25) e (6.22) será usada para retratar a relação existente entre número de desempregados e índice de produção industrial.

BIBLIOGRAFIA

- AKAIKE, H. 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21:243-247.
- AKAIKE, H. 1970. Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, 22:203-217.
- AKAIKE, H. 1978. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, 30(1):9-14.
- AKAIKE, H. 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, 66(2):237-242.
- AKAIKE, H. at alii 1979. T1MSAC-78. Tokyo, Institute of Statistical Mathematics. (Computer Science Monographs, 11)
- ALI, M.M. 1977. Analysis of autoregressive-moving average models: estimation and prediction. *Biometrika*, 64(3):535-545.
- ANDERSON, T.W. 1971. *The statistical analysis of time series*. New York, John Wiley. 704p. (Wiley Series in Probability and Mathematical Statistics)

- ANDERSON, T.W. 1980. Maximum likelihood estimation for vector autoregressive moving average models. In: BRILLINGER D.R. & TIAO, G.C., eds. *Directions in time series: proceedings of the Meeting on Time Series Analysis*, Ames, 1978. S.L.p., Institute of Mathematical Statistics. p.49-59.
- ANSLEY, C.F. 1979. An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, 66(1):59-65.
- ASTROM, K.J. & T. BOHLIN, T. 1966. Numerical identification of linear dynamic systems from normal operating records. In: HAMMOND, P.H., ed. *Theory of self-adaptive control systems*. New York, Plenum Press.
- BARTLETT, M.S. 1938. Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, 34(1):33-40.
- BEGUIN, J.M.; GOURTIEROUX, C.; MONFORT, A. 1980. Identification of a mixed autoregressive-moving average process: the Corner method. In: ANDERSON, O.D., ed. *Time series: proceedings of the International Time Series Meeting*, Nottingham University, 1979. Amsterdam, North-Holland. p.423-426.
- PHANSALI, R.J. & DOWNHAM, D.Y. 1977. Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion. *Biometrika*, 64(3):547-551.
- BOX, G.E.P. & JENKINS, G.M. 1970. *Time series analysis: forecasting and control*. San Francisco, Holden-Day. 553p. (Holden-Day Series in Time Series Analysis)

- BOX, G.E.P. & PIERCE, D.A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509-1526.
- BOX, G.E.P.; HILLMER, S.C.; TIAO, G.C. 1976. Analysis and modelling of seasonal time series. In: CONFERENCE ON SEASONAL TIME SERIES, Washington. *Proceedings*. S.L.p. National Bureau of Economic Research.
- CHAN, W.Y.T. & WALLIS, K.F. 1978. Multiple time series modelling: another look at the mink-muskrat interaction. *Applied Statistics, Series C*, 27(2):168-175.
- DENT, W.T. 1977. Computation of the exact likelihood function for an ARIMA process. *Journal of Statistical, Computation and Simulation*, 5:193-206.
- DUNSMUIR, W. & HANNAN, E.J. 1976. Vector linear time series models. *Advances in Applied Probability*, 8(2):339-364.
- DURBIN, J. 1959. Efficient estimation of parameters in moving-average models. *Biometrika*, 46(3/4):306-316.
- DURBIN, J. 1960. The fitting of time-series models. *Revue de L'institut International de Statistique*, 28(3):233-243.
- GRANGER, C.W.J. & NEWBOLD, P. 1977. Multiple series modelling and forecasting. In _____. *Forecasting economic time series*. New York, Academic Press. p.215-267. (Economic Theory and Mathematical Economics)
- GRAY, H.L.; KELLEY, G.D.; MCINTIRE, D.D. 1978. A new approach to ARMA modelling. *Communications in Statistics, Simulation and Computation*, B7(1):1-77.

HANNAN, E.J. 1970. *Multiple time series*. New York, John Wiley. 536p. (Wiley Series in Probability and Mathematical Statistics)

HANNAN, E.J. 1980. The estimation of the order of an ARMA process. *Annals of Statistics*, 10:1071-1080.

HANNAN, E.J. & QUINN, B.G. 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41(2):190-195.

HAUGH, L.D. 1972. *Identification of time series interrelationships with special reference to dynamic regression models*. Madison. Dissertação (Doutoramento) - University of Wisconsin.

HAUGH, L.D. 1976. Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association*, 71(354):378-385.

HAUGH, L.D. & BOX, G.E.P. 1977. Identification of dynamic regression (distributed lag) models connecting two time series. *Journal of the American Statistical Association*, 72(357):121-130.

HILLMER, S.C. & TIAO, G.C. 1979. Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association*, 74(367): 652-660.

JENKINS, G.M. 1975. The interaction between the muskrat and mink cycles in North Canada. In: CORSTEN, L.C. & POSTELNICU T., eds. *International Biometric Conference, 8.: proceedings*, Bucharest, Academiei Republicii Socialiste Romania. p.55-71.

- KANG, K.M. 1975. A comparison of estimators for moving average processes. Canberra, Australian Bureau of Statistics. (Unpublished paper)
- LI, W.L. & MCLEOD, A.I. 1980. Distribution of the residual autocorrelations in multivariate ARMA time series models. London, University of Western Ontario. (TR-80-03)
- LJUNG, G.M. & BOX, G.E.P. 1976. Studies in the modelling of discrete time series-maximum likelihood estimation in the autoregressive moving average model. Madison, University of Wisconsin. (Department of Statistics. Technical Report, 476)
- MANN, H.B. & WALD, A. 1943. On the statistical treatment of linear stochastic difference equations. *Econometrica*, 11 (3/4):173-220.
- MARQUARDT, D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431-441.
- NEWBOLD, P. 1974. The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika*, 61(3):423-426.
- NICHOLLS, D.F. 1972. On Hannan's estimation of ARMA models. *Australian Journal of Statistics*, 14:262-269.
- NICHOLLS, D.F. 1976. The efficient estimation of vector linear time series models. *Biometrika*, 63(2):381-390.
- NICHOLLS, D.F. 1977. A comparison of estimation methods for vector linear time series models. *Biometrika*, 64(1):85-90.
- NICHOLLS, D.F. & HALL, A.D. 1979. The exact likelihood function of multivariate autoregressive-moving average models, *Biometrika*, 66(2):259-264.

- OSBORN, D.R. 1976. Maximum likelihood estimation of moving average processes. *Annals of Economics and Social Measurement*, 5(1):75-87.
- OSBORN, D.R. 1977. Exact and approximate maximum likelihood estimators for vector moving average processes. *Journal of the Royal Statistical Society, Series B*, 39(1):114-118.
- PARZEN, E. 1969. Multiple time series modelling. In: KRISHNAIAH, P.R., ed. *Multivariate analysis: proceedings of the Second International Symposium on Multivariate Analysis*, Dayton, 1968. New York, Academic Press. p.389-409.
- PESARAN, M.H. 1973. Exact maximum likelihood estimation of a regression equation with first-order moving-average error. *Review of Economic Studies*, 40(4):529-535.
- PHADKE, M.S. & KEDEM, G. 1978. Computation of the exact likelihood function of multivariate moving average models. *Biometrika*, 65(3):511-519.
- PIERCE, D.A. 1980. *Multivariate time series*. S.L.p., s.c.p. Trab. apres. à 5^a Escola Latino Americana de Matematica, Mar del Plata. (Notas de aula)
- PRIESTLEY, M.B. 1981. *Spectral analysis and time series*. London, Academic Press. 2v.
- PROTHERO, D.L. & WALLIS, K.F. 1976. Modelling macroeconomic time series. *Journal of the Royal Statistical Society, Series A*, 139(4):468-500.
- QUENOUILLE, M.H. 1957. *The analysis of multiple time-series*. New York, Hafner. 105p. (Griffin's Statistical Monographs and Courses, 1)

- QUENOUILLE, M.H. 1958. Discrete autoregressive schemes with varying time-intervals. *Metrika*, 1:21-27.
- REINSEL, G. 1976. A note on approximations for the inverse of the covariance matrix of a stationary vector process. Madison, University of Wisconsin. (Department of Statistics. Technical Report, 470)
- RISSANEN, J. 1978. Modelling by shortest data description. *Automatica*, 14(5):465-471.
- SCHWARZ, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2)461-464.
- SCHWARZ, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2)461-464.
- SHAMAN, P. 1973. On the inverse of the covariance matrix for an autoregressive-moving average process. *Biometrika*, 60(1):193-196.
- SHIBATA, R. 1976. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63(1):117-126.
- TIAO, G.C. & ALI, M.M. 1971. Analysis of correlated random effects: linear model with two random components. *Biometrika*, 58(1):37-51.
- TIAO, G.C. et alii. 1979. *Program: a preliminary guide*. Madison, University of Wisconsin. (Department of Statistics. Wisconsin Multiple Time Series, WMTS-1)
- TIAO, G.C. & BOX, G.E.P. 1981. Modelling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802-816.

- TRIVEDI, P.K. 1970. Inventory behaviour in V.K. manufacturing 1956-67. *Review of Economic Studies*, 37(4):517-536.
- WALKER, A.M. 1950. Note on a generalization of the large-sample goodness-of-fit test for linear autoregressive schemes. *Journal of the Royal Statistical Society, Series B*, 12:102-107.
- WALKER, A.M. 1961. Large sample estimation of parameters for moving-average models. *Biometrika*, 48(3/4):343-357.
- WALLIS, K.F. 1977. Multiple time series analysis and the final form of econometric models, *Econometrica*, 45(6): 1481-1497.
- WHITTIER, P. 1951. *Hypothesis testing in time series analysis*. Uppsala. 121p. Tese - Uppsala University.
- WHITTIE, P. 1953. Estimation and information in stationary time series. *Arkiv för Matematik*, 2(23):423-434.
- WHITTIE, P. 1962. Gaussian estimation in stationary time series. *Bulletin de L'Institut International de Statistique*, 39:105-129.
- WILSON, G.T. 1973. The estimation of parameters in multivariate time series models. *Journal of the Royal Statistical Society, Series B*, 35(1):76-85.
- WOLD, H. 1938. *A study in the analysis of stationary time series*. Uppsala, Almqvist and Wiksell. 236p.
- ZELLNER, A. & PALM, F. 1974. Time series analysis and simultaneous equation econometric models. *Journal of Econometrics*, 2:17-54.