PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

TESTE DE RAZÃO DE VEROSSIMILHANÇA GENERALIZADO
NA IDENTIFICAÇÃO DE SÉRIES TEMPORAIS

por

José Mauro Pedro Fortes

Tese submetida como requisito parcial para
a obtenção do grau de

MESTRE EM CIÊNCIAS
DE
ENGENHARIA ELÉTRICA

José Paulo de Almeida e Albuquerque
José Paulo de Almeida e Albuquerque
Orientador

Rio de Janeiro, RJ, Julho de 1976

A Mônica

A meus pais

AGRADECIMENTOS

- Ao Professor José Paulo de Almeida e Albuquerque pela dedicada orientação
- À Srta. Maria da Graça L. Muniz pelo trabalho de datilografia
- Ao Sr. Luiz Claudio Fernandes de Oliveira pelo trabalho de impressão
- Ao Centro de Estudos em Telecomunicações da Pontifícia Universidade Católica do Rio de Janeiro
- A todos que de algum modo contribuíram para a realização deste trabalho.

SUMÁRIO

Muito frequentemente, as técnicas utilizadas na identificação de processos estocásticos conduzem a mais de um modelo passível de ser utilizado na caracterização do processo. O problema de escolher entre estes modelos é formulado como um problema de teste de hipóteses, e o teste de razão de verossimilhança é a ele aplicado. Considera-se então a situação particular onde se quer descrever processos de parâmetro discreto (séries temporais) através de modelos ARIMA (Autoregressive Integrated Moving Average). O teste de razão de verossimilhança associado ao problema é então deduzido e implementado através do algoritmo de Kalman-Bucy. Comparações com um outro teste usualmente empregado na escolha de modelos para séries temporais mostram a superioridade do teste de razão de verossimilhança.

SUMMARY

Very often random process identification techniques lead to several prospective models to characterize the process. The problem of choosing among these models is cast as a hypothesis testing problem, to which a likelihood ratio test is applied. For the special situation in which a choice between two autoregressive integrated moving average models is to be made, likelihood ratio is derived and afterwards implemented through the Kalman-Bucy algorithm. Comparisons with another procedure usually connected to time series model choices show likelihood ratio tests are definitely superior.

ÍNDICE

INTRODUÇÃO	1
CAPÍTULO 1 - O PROBLEMA DE IDENTIFICAÇÃO DE SÉRIES TEMPO RAIS	3
CAPÍTULO 2 - DEDUÇÃO DO TESTE DE RAZÃO DE VEROSSIMILHAN ÇA	5
2.1 - Teste de Razão de Verossimilhança	5
2.2 - O Modelo ARIMA (p,d,q)	12
2.3 - Teste Binário de Razão de Verossimilhança para Mo delos ARIMA	23
CAPÍTULO 3 - IMPLEMENTAÇÃO DO TESTE DE RAZÃO DE VEROSSI MILHANÇA	34
3.1 - Implementações de ℓ_{C_1} e ℓ_{C_2}	35
3.2 - Implementação de $\ell_B(\underline{R})$	36
3.3 - Implementação de $\ell_{R_1}(\underline{R})$	37
3.4 - Implementações de $\ell_{R_2}(\underline{R})$ e $\ell_{R_3}(\underline{R})$	47
3.5 - Implementação de $\ell(\underline{R})$	48
CAPÍTULO 4 - SIMULAÇÕES	50
4.1 - Parâmetros Conhecidos	51
4.2 - Parâmetros Desconhecidos	65
4.3 - Análise dos Resultados	72
CAPÍTULO 5 - CONCLUSÕES	77
APÊNDICE A - FILTRO LINEAR ÓTIMO	79
APÊNDICE B - AUTOVALORES E AUTOFUNÇÕES	88
APÊNDICE C - EQUAÇÃO DE DEFINIÇÃO DE UM FILTRO ÚTIL	95
APÊNDICE D - FILTRO DE KALMAN-BUCY	99
REFERÊNCIAS	104

INTRODUÇÃO

O presente trabalho diz respeito ao problema de identificação de séries temporais. Mais especificamente é aqui examinada a fase deste problema em que se deseja escolher entre vários possíveis modelos, capazes de caracterizar o processo estocástico em consideração.

No capítulo 1 o problema de identificação de séries temporais é brevemente descrito em suas três fases: identificação, estimação de parâmetros e decisão entre modelos, seguida de validação do modelo escolhido. É também claramente estabelecido que o objetivo do trabalho se concentra na utilização do teste de razão de verossimilhança (TRV) em conexão com a terceira fase do problema de identificação.

No capítulo 2 o TRV é definido e sua expressão deduzida para o caso específico em que se quer decidir entre dois modelos ARIMA¹. Essa expressão envolve a determinação de autofunções e autovalores de matrizes cujas ordens são iguais ao número de observações da série. Desse modo sua implementação direta se torna numericamente impraticável.

No capítulo 3 e apêndices A, B, C e D, é mostrada a possibilidade de implementar o TRV, a partir do algoritmo de Kalman-Bucy, o que elimina as dificuldades numéricas anteriormente mencionadas.

O capítulo 4 apresenta comparações entre os desempenhos do TRV e de um teste baseado nas autocorre_lações dos resíduos da série ^{1,2}.

Finalmente, conclusões são apresentadas no capítulo 5.

CAPITULO 1

O PROBLEMA DE IDENTIFICAÇÃO DE SÉRIES TEMPORAIS

O problema de identificação de uma série temporal é em terminologia mais ampla o problema de identificação de um processo estocástico de parâmetro discreto. Historicamente, como esse problema surgiu em conexão com o estudo das variações de uma grandeza ao longo do tempo, tornou-se usual denominá-lo como o problema de identificação de séries temporais. Matematicamente trata-se de caracterizar um processo estocástico de parâmetro discreto, a partir do conhecimento de uma de suas funções-amostra.

Dentro do presente trabalho, os modelos considerados serão paramétricos¹, ou seja dependerão de um determinado número finito de parâmetros. Dessa forma, ao procedimento de identificação segue-se a estimação dos parâmetros correspondentes ao modelo particular resultante da fase de identificação (eventualmente dois ou mais modelos terão que ser considerados até à fase de estimação de parâmetros). Sempre que mais de um modelo está sendo considerado, uma vez estimados os parâmetros é preciso decidir entre os diversos modelos. Esse problema pode ser formulado como um problema de teste de hipóteses e, embora essa possibilidade seja amplamente reconhecida, os

procedimentos usualmente empregados na escolha de um dos modelos não exploram as técnicas apropriadas à problemas de teste hipóteses. Assim, por exemplo, o teste de razão de verossimilhança não é comumente utilizado em conexão com problemas de identificação de processos. Nesse particular, parece haver um desequilíbrio entre os graus de sofisticação utilizados na fase de estimação de parâmetros (estimativas de máxima verossimilhança obtidas através de métodos de gradiente^{1,3}) e na fase de escolha entre os diversos modelos em consideração^{1,2}.

Em resumo, o problema de identificação de um processo estocástico através de modelos paramétricos, envolve tres fases: identificação (da qual resultam modelos dependendo de parâmetros ainda desconhecidos), estimação de parâmetros (na qual os parâmetros desconhecidos são estimados) e decisão por um dos modelos considerados, seguida da validade do mesmo. Dentro dessa formulação, o presente trabalho se concentra na terceira fase e mais especificamente no problema de escolha entre diversos modelos, para o que se propõe a utilização do teste de razão de verossimilhança. A definição desse teste, sua implementação em conexão com a escolha entre dois modelos ARIMA⁴ e sua comparação com outros procedimentos de decisão entre modelos são apresentadas nos capítulos que se seguem.

CAPÍTULO 2

DEDUÇÃO DO TESTE DE RAZÃO DE VEROSSIMILHANÇA

O objetivo do presente capítulo é formular o problema de decidir entre dois modelos ARIMA*^{1,4,5} como um problema de teste de hipóteses, de tal modo que o teste de razão de verossimilhança possa ser aplicado ao mesmo. Para isso o teste de razão de verossimilhança é sumarizado na seção 2.1. Na seção 2.2 considera-se um modelo ARIMA escrito sob diversas formas, visando especialmente a representação desse modelo sob uma forma conveniente para a dedução do teste de razão de verossimilhança associado ao problema em que se quer decidir entre dois modelos ARIMA. Esta dedução é apresentada na seção 2.3.

2.1 - Teste de Razão de Verossimilhança

Considere uma situação em que um processo $r(\cdot)$ é observado, e suponha que $r(\cdot)$ possa ser um qualquer entre K possíveis processos, de tal modo que a cada um desses processos se associa uma hipótese H_i , $i=1, \dots, K$. De acordo com a situação descrita pode ser definido um problema que consiste em observar o processo $r(\cdot)$ e escolher, segundo determinado critério, uma das K hipóteses possíveis como sendo a verdadeira.

* Autoregressive Integrated Moving Average

Um dos vários critérios usados para se solucionar este problema, é o teste de razão de verossimilhança, apresentado a seguir para o caso binário (2 hipóteses possíveis). Convém observar entretanto que a restrição feita ($K=2$) não altera a generalidade do desenvolvimento que se segue, o qual poderia ser facilmente estendido para $K>2$.

Considere então duas hipóteses possíveis H_0 e H_1 de modo que quando a hipótese H_0 é verdadeira o processo $r(.)$ observado é

$$\{ r(k) | H_0, \quad k \in \Pi \} \quad (2.1)$$

e quando H_1 é verdadeira, o processo $r(.)$ é

$$\{ r(k) | H_1, \quad k \in \Pi \} \quad (2.2)$$

onde Π é o espaço de definição do parâmetro k . Isto significa que dependendo da hipótese que seja a verdadeira, as observações $r(k)$ serão

$$\begin{cases} H_0 & ; \quad r(k) = r(k) | H_0, \quad k \in \Sigma \\ H_1 & ; \quad r(k) = r(k) | H_1, \quad k \in \Sigma \end{cases} \quad (2.3)$$

onde Σ é um subconjunto de Π , ou seja $\Sigma \subset \Pi$. Em particular será de interesse considerar aqui, processos de parâmetro

discreto com

$$\Pi = \{ \dots, -1, 0, 1, \dots \} \quad (2.4)$$

e

$$\Sigma = \{1, 2, \dots, N\} \quad (2.5)$$

A fim de se ter uma visualização melhor do problema, considere a representação das observações nas duas hipóteses em um espaço Euclidiano N-dimensional Z , conforme apresentado na figura 2.1. Associa-se então a ca da observação possível um determinado vetor $\underline{r}$ do espaço Z .

Figura 2.1 - Representação das observações associadas ao problema do teste de hipóteses binário em um espaço Euclidiano N-dimensional

Tem-se então como objetivo, determinar uma fronteira η , ótima segundo algum critério, que divida o espaço Z em duas regiões Z_0 e Z_1 .

Essas regiões são tais que se as observações pertencerem a Z_0 , decide-se pela hipótese H_0 , enquanto que se elas pertencerem a Z_1 , decide-se por H_1 .

Considere então um determinado custo associado a cada possível situação de decisão. Seja então C_{ij} o custo no caso de se decidir por H_i sendo H_j a hipótese verdadeira. Convém observar que nesse caso particular do problema binário, tem-se apenas quatro situações possíveis de decisão. Desse modo, é possível associar ao problema um custo C que é uma variável aleatória discreta, assumindo quatro valores: C_{00} , C_{01} , C_{10} e C_{11} .

Define-se como risco associado a um problema de teste de hipóteses, a quantidade R dada por

$$R \triangleq E[C] \quad (2.6)$$

O critério do teste de razão de verossimilhança consiste em se determinar a fronteira ótima η de modo a minimizar o risco, definido por (2.6). Serão aqui considerados como dados do problema: as probabilidades de cada uma das hipóteses ser a verdadeira (probabilidades a priori), $P(H_0)$ e $P(H_1)$ e os custos de cada situação possível de decisão, C_{00} , C_{01} , C_{10} e C_{11} .

A quantidade a se minimizar é então

$$R = E[C] = \sum_{j=0}^1 \sum_{i=0}^1 C_{ij} P[C=C_{ij}] \quad (2.7)$$

Representando-se por $P(H_i, H_j)$ a probabilidade de se decidir pela hipótese H_i e H_j ser a hipótese verdadeira, pode-se escrever (2.7) como

$$R = \sum_{j=0}^1 \sum_{i=0}^1 C_{ij} P(H_i, H_j) \quad (2.8)$$

ou ainda

$$R = \sum_{j=0}^1 \sum_{i=0}^1 C_{ij} P(H_j) P(H_i | H_j)$$

o que a partir da representação do problema no espaço Euclidiano apresentado na figura 2.1, se escreve

$$\begin{aligned} R = & C_{00} P(H_0) \int_{Z_0} p_{\underline{r}|H_0}(\underline{r}|H_0) d\underline{r} + C_{01} P(H_1) \int_{Z_0} p_{\underline{r}|H_1}(\underline{r}|H_1) d\underline{r} + \\ & + C_{10} P(H_0) \int_{Z_1} p_{\underline{r}|H_0}(\underline{r}|H_0) d\underline{r} + C_{11} P(H_1) \int_{Z_1} p_{\underline{r}|H_1}(\underline{r}|H_1) d\underline{r} \quad (2.9) \end{aligned}$$

onde $p_{\underline{r}|H_i}(\underline{r}|H_i)$ é a densidade de probabilidade do vetor de observações dado que a hipótese H_i é a verdadeira.

Observando ainda que

$$\int_Z p_{\underline{R}|H_0}(\underline{R}|H_0)d\underline{R} = \int_Z p_{\underline{R}|H_1}(\underline{R}|H_1)d\underline{R} = 1 \quad (2.10)$$

pode-se reduzir a equação (2.9) a

$$\begin{aligned} R = & C_{10} P(H_0) + C_{11} P(H_1) + \\ & + \int_{Z_0} \left\{ [C_{01} - C_{11}] P(H_1) p_{\underline{R}|H_1}(\underline{R}|H_1) - [C_{10} - C_{00}] P(H_0) p_{\underline{R}|H_0}(\underline{R}|H_0) \right\} d\underline{R} \end{aligned} \quad (2.11)$$

Por outro lado, tendo em vista o significado dos custos C_{ij} , é bastante razoável que se considere $C_{01} > C_{11}$ e $C_{10} > C_{00}$. Isto, levado em consideração juntamente com o fato de probabilidades e funções densidade de probabilidade serem sempre não negativas, permite afirmar que as duas parcelas do integrando de (2.11) são não negativas para todos os valores de $\underline{R}$.

Se o objetivo é minimizar o risco, isto pode ser feito definindo-se Z_0 como a região contendo todos os pontos $\underline{R}$ para os quais a segunda parcela do integrando de (2.11) é maior que a primeira. Obviamente os pontos $\underline{R}$ restantes, ou seja, aqueles para os quais o integrando de (2.11) é positivo devem pertencer à região Z_1 . Assim sendo, de modo a minimizar o risco, as regiões Z_0 e Z_1 devem ser definidas da seguinte forma:

$$Z_0 = \left\{ \underline{R} \mid \left[C_{01} - C_{11} \right] P(H_1) p_{\underline{R}|H_1}(\underline{R}|H_1) < \left[C_{10} - C_{00} \right] P(H_0) p_{\underline{R}|H_0}(\underline{R}|H_0) \right\} \quad (2.12)$$

e

$$Z_1 = \left\{ \underline{R} \mid \left[C_{01} - C_{11} \right] P(H_1) p_{\underline{R}|H_1}(\underline{R}|H_1) > \left[C_{10} - C_{00} \right] P(H_0) p_{\underline{R}|H_0}(\underline{R}|H_0) \right\} \quad (2.13)$$

Fica então bastante claro que a fronteira η será determinada por

$$\eta = \left\{ \underline{R} \mid \left[C_{01} - C_{11} \right] P(H_1) p_{\underline{R}|H_1}(\underline{R}|H_1) = \left[C_{10} - C_{00} \right] P(H_0) p_{\underline{R}|H_0}(\underline{R}|H_0) \right\} \quad (2.14)$$

Isto significa que os pontos $\underline{R}$ da fronteira η devem satisfazer a equação

$$\Lambda(\underline{R}) = \gamma \quad (2.15)$$

onde $\Lambda(\underline{R})$ é a razão de verossimilhança definida por

$$\Lambda(\underline{R}) = \frac{p_{\underline{R}|H_1}(\underline{R}|H_1)}{p_{\underline{R}|H_0}(\underline{R}|H_0)} \quad (2.16)$$

c

$$\gamma = \frac{[C_{10} - C_{00}]P(H_0)}{[C_{01} - C_{11}]P(H_1)} \quad (2.17)$$

De maneira análoga, as regiões Z_0 e Z_1 podem ser dadas por

$$Z_0 = \{\underline{R} | \Lambda(\underline{R}) < \gamma\} \quad (2.18)$$

e

$$Z_1 = \{\underline{R} | \Lambda(\underline{R}) > \gamma\} \quad (2.19)$$

Convém relembrar que $\underline{R}$ é obtido a partir da representação das observações no espaço Z . Desse modo, o teste de razão de verossimilhança consiste em se determinar a partir do vetor de observações $\underline{R}$ a quantidade $\Lambda(\underline{R})$ e compará-la com um limiar γ conhecido. Dependendo do resultado da comparação, faz-se então a opção por uma das duas hipóteses possíveis.

2.2 - O Modelo ARIMA (p,d,q)

Os modelos ARIMA (p,d,q), detalhadamente examinados na literatura^{1,4,5}, são modelos que expressam a maneira pela qual um processo de parâmetro discreto $z(\cdot)$

pode ser obtido pela passagem de um processo de ruído branco $a(\cdot)$ através de um sistema linear estável, causal e invariante no tempo. Verifica-se também ^{1,5} a existência de uma certa flexibilidade no sentido de que o processo $z(\cdot)$ possa ser não estacionário, desde que ele pertença a uma classe particular de processos não estacionários. Tal classe é constituída pelos processos $z(\cdot)$ para os quais existe um inteiro d positivo e finito de modo que o processo $w(\cdot)$, definido através da diferença de ordem d de $z(\cdot)$, seja estacionário no sentido amplo.

A presente seção tem por objetivo apresentar diversas formas de representação dos modelos ARIMA (p,d,q) que serão utilizadas mais adiante.

A equação usual de representação dos modelos ARIMA (p,d,q) é

$$\phi(B) \left[\nabla^d z(k) \right] = \theta_0 + \theta(B) a(k) \quad (2.20)$$

onde B é um operador tal que

$$B z(k) = z(k-1) \quad (2.21)$$

e $\phi(B)$ e $\theta(B)$ são polinômios dados por

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.22)$$

e

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.23)$$

Além disso, em (2.20), ∇ é o operador diferença definido por

$$\nabla = 1 - B \quad (2.24)$$

enquanto d é um número inteiro não negativo e θ_0 é uma constante real. O processo $a(\cdot)$ é caracterizado por

$$K_a(k, \ell) = \sigma_a^2 \delta(k, \ell) \quad , \quad \sigma_a^2 > 0 \quad (2.25)$$

e

$$E[a(k)] = 0 \quad (2.26)$$

Através de manipulações algébricas bastante simples, a equação (2.20) pode ainda ser escrita como

$$\theta(B) \left[\nabla^d z(k) - u \right] = \theta(B) a(k) \quad (2.27)$$

ou

$$\phi(B) \left[\nabla^d z(k) \right] = \theta(B) \left[a(k) + \xi \right] \quad (2.28)$$

onde os parâmetros μ e ξ se relacionam ao parâmetro θ_0 através de

$$\mu = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \dots - \phi_p} = \frac{(1 - \theta_1 - \theta_2 - \dots - \theta_q)\xi}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \quad (2.29)$$

Em algumas ocasiões é de interesse que se tenha o modelo ARIMA (p,d,q) representado em notação de estado. Para tanto considere um processo $v(\cdot)$ tal que

$$\nabla^d v(k) = \mu \quad (2.30)$$

Isto permite escrever (2.27) como

$$\phi(B) \nabla^d \left[w(k) \right] = \theta(B) a(k) \quad (2.31)$$

onde

$$w(k) = z(k) - v(k) \quad (2.32)$$

É possível ainda escrever (2.31) de forma mais compacta, resultando

$$\phi(B)w(k) = \theta(B)a(k) \quad (2.33)$$

onde

$$\phi(B) = \phi(B) \nabla^d = \phi(B) [1-B]^d \quad (2.34)$$

o que significa que $\phi(B)$ é um polinômio em B de grau $p'=p+d$, dado por

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^{p'} \quad (2.35)$$

É possível verificar que (2.33) pode ser escrita sob a forma

$$w(k) = \underline{H} \underline{x}(k) + a(k) \quad (2.36)$$

com

$$\underline{x}(k+1) = \underline{G} \underline{x}(k) + \underline{\Gamma} a(k) \quad (2.37)$$

onde $\underline{x}(k)$ é o vetor de estado de dimensão $n \times 1$ e $a(k)$ e $w(k)$ são escalares. As matrizes $\underline{H}$, $\underline{\Phi}$ e $\underline{\Gamma}$ possuem as dimensões apropriadas e são invariantes no tempo. A dimensão n do vetor de estado é igual ao maior dos números p' e q .

As matrizes $\underline{\Phi}$, $\underline{\Gamma}$ e $\underline{H}$ podem ser escritas em função dos coeficientes dos polinômios $\Phi(B)$ e $\theta(B)$. Se $p' > q$ tem-se

$$\underline{\Phi} = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ \phi_{p'-1} & 0 & 0 & \dots & 1 \\ \phi_{p'} & 0 & 0 & \dots & 0 \end{bmatrix}; \quad \underline{\Gamma} = \begin{bmatrix} \phi_1 - \theta_1 \\ \vdots \\ \phi_q - \theta_q \\ \phi_{q+1} \\ \vdots \\ \phi_{p'} \end{bmatrix} \text{ e } \underline{H} = [1 \ 0 \ \dots \ 0] \quad (2.38)$$

Por outro lado, se $p' < q$

$$\underline{\Phi} = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ \phi_{p'} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}; \quad \underline{\Gamma} = \begin{bmatrix} \phi_1 - \theta_1 \\ \vdots \\ \phi_{p'} - \theta_{p'} \\ \phi_{p'+1} \\ \vdots \\ \phi_q \end{bmatrix} \text{ e } \underline{H} = [1 \ 0 \ \dots \ 0] \quad (2.39)$$

Em alguns casos é ainda de interesse que se tenha o processo $w(.)$ que aparece em (2.33), escrito explícitamente em função do processo $a(.)$. Isto é feito através da equação

$$w(k) = \psi(B)a(k) \quad (2.40)$$

onde

$$\psi(B) = \frac{\theta(B)}{\phi(B)} \quad (2.41)$$

Como a razão entre dois polinômios de grau finito, $\psi(B)$ pode ser escrito como

$$\psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i \quad (2.42)$$

As equações (2.40) e (2.42) permitem que o processo $w(.)$ possa ser visto como a soma de um processo $s(.)$ com um ruído branco $a(.)$ correlatado com $s(.)$, ou seja

$$w(k) = s(k) + a(k) \quad (2.43)$$

onde

$$s(k) = \sum_{i=1}^{\infty} \psi_i a(k-i) \quad (2.44)$$

Observe-se que a partir de (2.44), (2.25) e (2.26) tem-se

$$K_{sa}(k, \ell) = K_{as}(\ell, k) = \begin{cases} \sigma_a^2 \psi_{k-\ell} & , \quad k > \ell \\ 0 & , \quad k \leq \ell \end{cases} \quad (2.45)$$

Levando em conta (2.43), verifica-se ainda que

$$K_{wa}(k, \ell) = K_{aw}(\ell, k) = \begin{cases} \sigma_a^2 \psi_{k-\ell} & , \quad k > \ell \\ \sigma_a^2 & , \quad k = \ell \\ 0 & , \quad k < \ell \end{cases} \quad (2.46)$$

Considere-se agora um problema de filtragem com o apresentado no apêndice A, onde as observações são dadas por (2.43) para $k=1, \dots, N$. A partir de (A.6) tem-se

$$\hat{s}(k|k-1) = \sum_{j=1}^{k-1} h(k, j:k-1) w(j) \quad (2.47)$$

Além disso, associado ao processo $w(\cdot)$, foi introduzido no apêndice A o processo de inovações $v(\cdot)$. A equação (A.15) pode também ser escrita como

$$v(k) = \tilde{s}(k|k-1) + a(k) \quad , \quad k=1, \dots, N \quad (2.48)$$

onde $\tilde{s}(k|k-1)$ representa o erro associado à predição, ou seja

$$\tilde{s}(k|k-1) = s(k) - \hat{s}(k|k-1) \quad , \quad k=1, \dots, N \quad (2.49)$$

É interessante que se determine agora a função covariância $K_v(k, \ell)$ associada ao processo de inovações. Tem-se então

$$\begin{aligned} K_v(k, \ell) &= E \left[v(k) v(\ell) \right] = \sigma_a^2 \delta(k, \ell) + K_{as}(k, \ell) + \\ &- \sum_{j=1}^{\ell-1} h(\ell, j: \ell-1) K_{aw}(k, j) + E \left[\tilde{s}(k|k-1) w(\ell) \right] + \\ &- \sum_{j=1}^{\ell-1} h(\ell, j: \ell-1) E \left[\tilde{s}(k|k-1) w(j) \right] \end{aligned} \quad (2.50)$$

Para $k > \ell$ tem-se

$$\sigma_a^2 \delta(k, \ell) = 0$$

A partir de (2.45) e (2.46) tem-se ainda

$$K_{as}(k, \ell) = 0 \quad , \quad k \geq \ell$$

e

$$K_{aw}(k, j) = 0 \quad , \quad k > j$$

que é satisfeita para $k > \ell$. Por outro lado, (A.28) permite escrever que

$$E\left[\tilde{s}(k|k-1)w(\ell)\right] = 0 \quad , \quad k > \ell$$

e

$$E\left[\tilde{s}(k|k-1)w(j)\right] = 0 \quad , \quad k > j$$

que é satisfeita para $k > \ell$. Isto significa que

$$K_v(k, \ell) = 0 \quad , \quad k > \ell \quad (2.51)$$

Para $k < \ell$, tem-se por simetria um resultado análogo, ou seja

$$K_v(k, \ell) = 0 \quad , \quad k < \ell \quad (2.52)$$

Finalmente, para $k=1$ tem-se

$$K_v(k,k) = E[v^2(k)] = \sigma_a^2 + E[\tilde{s}(k|k-1)]^2 + 2E[\tilde{s}(k|k-1)a(k)] \quad (2.53)$$

Como

$$E[\tilde{s}(k|k-1)a(k)] = K_{sa}(k,k) - \sum_{j=1}^{k-1} h(k,j:k-1)K_{wa}(j,k)$$

tem-se a partir de (2.45) e (2.46) que

$$E[\tilde{s}(k|k-1)a(k)] = 0$$

Isto significa que

$$K_v(k,k) = E[v^2(k)] = \sigma_a^2 + E\left\{\left[\tilde{s}(k|k-1)\right]^2\right\} \quad (2.54)$$

Em resumo, de (2.51), (2.52) e (2.54) tem-se finalmente

$$K_v(k,\ell) = \begin{cases} 0 & , \quad k > \ell \\ E[v^2(k)] = \sigma_a^2 + P_s(k|k-1) & , \quad k = \ell \\ 0 & , \quad k < \ell \end{cases} \quad (2.55)$$

onde

$$P_S(k|k-1) = E \left\{ \left[\tilde{s}(k|k-1) \right]^2 \right\} \quad (2.56)$$

Assim sendo o processo $v(.)$ é um processo de ruído branco, onde a variância de cada uma de suas componentes é dada por (2.54).

2.3 - Teste Binário de Razão de Verossimilhança para Modelos ARIMA

Na presente seção o problema de teste de hipóteses definido por (2.1) até (2.5) será particularizado para o caso em que os processos associados a cada uma das hipóteses são descritos por modelos ARIMA.

Admitindo-se que a hipótese H_0 está associada a um modelo ARIMA (p_0, d_0, q_0) e que a hipótese H_1 está associada a um modelo ARIMA (p_1, d_1, q_1) , com $d_1 \leq d_0$, tem-se a partir da equação (2.20) que as observações nas duas hipóteses se escrevem

$$\begin{cases} H_0: \phi^0(B) \left[\nabla^{d_0} z(k) \right] = \theta_0^0 + \theta^0(B) a_0(k) & , \quad k=1, \dots, N \\ H_1: \phi^1(B) \left[\nabla^{d_1} z(k) \right] = \theta_0^1 + \theta^1(B) a_1(k) & , \quad k=1, \dots, N \end{cases} \quad (2.57)$$

onde

$$E[a_0(k)] = E[a_1(k)] = 0 \quad , \quad k=1, \dots, N \quad (2.58)$$

$$K_{a_0}(k, \ell) = \sigma_{a_0}^2 \delta(k, \ell) \quad , \quad k=1, \dots, N ; \ell=1, \dots, N \quad (2.59)$$

e

$$K_{a_1}(k, \ell) = \sigma_{a_1}^2 \delta(k, \ell) \quad , \quad k=1, \dots, N ; \ell=1, \dots, N \quad (2.60)$$

onde $a_0(k)$ e $a_1(k)$ são linearmente dependentes, o que a partir de (2.59) e (2.60) permite escrever

$$a_0(k) = K a_1(k) \quad , \quad k=1, \dots, N \quad (2.61)$$

onde

$$K = \frac{\sigma_{a_0}}{\sigma_{a_1}} \quad (2.62)$$

A fim de se ter uma representação mais simples do problema, considere-se o modelo ARIMA, que aparece na segunda equação do conjunto (2.57), escrito sob a forma (2.31), ou seja

$$H_1: \phi^1(B) \nabla^{d_1} [u(k)] = \theta^1(B) a_1(k) \quad , \quad k=1, \dots, N \quad (2.63)$$

onde o processo $u(\cdot)$ seria dado por

$$u(k) = z(k) - v(k) \quad (2.64)$$

com $v(k)$ definido por (2.30) e (2.29). Utilizando estas duas equações, ao lado de (2.64), a primeira equação do conjunto (2.57) pode ser escrita como

$$H_0: \phi^0(B) \nabla^{d_0} [u(k)] = \theta_0^* + \theta^0(B) a_0(k) \quad , \quad k=1, \dots, N \quad (2.65)$$

onde

$$\theta_0^* = \theta_0^0 - \phi^0(B) \nabla^{d_0} v(k) \quad (2.66)$$

Considerando (2.30) e (2.29) e levando em conta que $d_0 \geq d_1$, uma expressão mais simples para θ_0^* seria dada por

$$\theta_0^* = \begin{cases} \theta_0^0 & , \quad d_0 > d_1 \\ \theta_0^0 - \frac{(1 - \epsilon_1^0 - \dots - \epsilon_{p_0}^0)}{(1 - \epsilon_1^1 - \dots - \epsilon_{p_1}^1)} \epsilon_0^1 & , \quad d_0 = d_1 \end{cases} \quad (2.67)$$

Observando agora que (2.65) pode ser escrita sob a forma (2.28), tem-se

$$H_0: \phi^0(B) \nabla^{d_0} [u(k)] = \theta^0(B) [a_0(k) + \xi] \quad , \quad k=1, \dots, N \quad (2.68)$$

com ξ calculado através de (2.29), ou seja

$$\xi = \frac{\theta_0^*}{1 - \theta_1^0 - \dots - \theta_{q_0}^0} \quad (2.69)$$

Utilizando (2.68), (2.63) e ainda (2.61) o conjunto (2.57) se escreve

$$\begin{cases} H_0: \phi^0(B) \nabla^{d_0} [u(k)] = \theta^0(B) [K a_1(k) + \xi] & , \quad k=1, \dots, N \\ H_1: \phi^1(B) \nabla^{d_1} [u(k)] = \theta^1(B) a_1(k) & , \quad k=1, \dots, N \end{cases} \quad (2.70)$$

o que pode também ser escrito como

$$\begin{cases} H_0: \frac{\phi^0(B) \nabla^{d_0}}{\theta^0(B)} [u(k)] = K a_1(k) + \xi & , \quad k=1, \dots, N \\ H_1: \frac{\phi^0(B) \nabla^{d_0}}{\theta^0(B)} [u(k)] = \frac{\phi^0(B) \nabla^{d_0} \theta^1(B)}{\theta^0(B) \phi^1(B) \nabla^{d_1}} a_1(k) & , \quad k=1, \dots, N \end{cases} \quad (2.71)$$

Fazendo finalmente

$$\frac{\phi^0(B) \nabla^{d_0}}{\theta^0(B)} \cdot [u(k)] = r(k) \quad , \quad k=1, \dots, N \quad (2.72)$$

e definindo ainda

$$\theta^*(B) = \phi^0(B) \nabla^{d_0} \theta^1(B) \quad (2.73)$$

e

$$\phi^*(B) = \theta^0(B) \phi^1(B) \nabla^{d_1} \quad (2.74)$$

o problema inicial de teste de hipóteses se reduz a

$$\left\{ \begin{array}{l} H_0 : r(k) = K a_1(k) + \xi \quad , \quad k=1, \dots, N \\ H_1 : r(k) = \frac{\theta^*(B)}{\phi^*(B)} a_1(k) = w(k) \quad , \quad k=1, \dots, N \end{array} \right. \quad (2.75)$$

o que significa que a partir das observações $r(k)$, obtidas a partir das observações $z(k)$, $k=1, \dots, N$, deseja-se escolher entre uma hipótese H_0 associada a um ruído branco de média ξ e uma hipótese H_1 associada a um processo $w(\cdot)$ descrito por um modelo ARIMA (p^*, d^*, q^*) com

$$p^* = q_0 + p_1 \quad (2.76)$$

$$q^* = p_0 + (d_0 - d_1) + q_1 \quad (2.77)$$

e

$$d^* = 0 \quad (2.78)$$

Considerando as condições de estacionariedade e invertibilidade do modelo ARIMA ^{1,2}, conclui-se a partir de (2.73) e (2.74) que se os modelos ARIMA(p_0, d_0, q_0) e ARIMA(p_1, d_1, q_1) são estacionários e invertíveis, o modelo ARIMA(p^*, d^*, q^*) é sempre estacionário. Quanto à sua invertibilidade, ela só ocorre quando $d_0 \leq d_1$, o que em presença da suposição inicial de que $d_1 \leq d_0$, se reduz a $d_0 = d_1$.

Considere-se agora a representação do problema definido por (2.75) no espaço das autofunções da função covariância $K_w(k, \ell)$ das observações do processo $w(\cdot)$, a fim de que nesta nova base, conforme mostrado no apêndice B, as coordenadas das observações sejam estatisticamente independentes. Considere então as autofunções $\{\phi_i(k)\}, k=1, \dots, N$ da função covariância $K_w(k, \ell)$. Essas autofunções devem satisfazer obviamente a (B.5), (B.9), (B.10) e (B.21). Nesta nova base, o problema em questão passa então a ser representado por

$$\begin{cases} H_0 : r^*(k|H_0) & , \quad k=1, \dots, N \\ H_1 : r^*(k|H_1) & , \quad k=1, \dots, N \end{cases} \quad (2.79)$$

onde

$$r^*(k|H_0) = \sum_{j=1}^N \left[K a_r(j) + \varepsilon \right] \phi_k(j) \quad , \quad j=1, \dots, N \quad (2.80)$$

e

$$r^*(k|H_1) = \sum_{j=1}^N w(j) \phi_k(j) \quad , \quad k=1, \dots, N' \quad (2.81)$$

Considerando o processo $w(\cdot)$ representado sob a forma (2.40) ao lado do fato de que

$$E[a(k)] = 0 \quad , \quad k=1, \dots, N \quad (2.82)$$

tem-se

$$E[r^*(k|H_1)] = 0 \quad , \quad k=1, \dots, N \quad (2.83)$$

Verifica-se ainda que

$$E[r^*(k|H_0)] = \sum_{j=1}^N \varepsilon \cdot \phi_k(j) = M(k) \quad , \quad k=1, \dots, N \quad (2.84)$$

$$K_{\underline{r}^*|H_0}(k, \ell) = K^2 \sigma_{a_1}^2 \delta(k, \ell) \quad , \quad k=1, \dots, N \quad ; \quad \ell=1, \dots, N \quad (2.85)$$

e

$$K_{\underline{r}^*|H_1}(k, \ell) = \lambda_k \delta(k, \ell) \quad , \quad k=1, \dots, N \quad ; \quad \ell=1, \dots, N \quad (2.86)$$

onde os $\{\lambda_k\}$ são os autovalores de $K_w(k, \ell)$.

Define-se agora um vetor $\underline{r}^*$ tal que

$$\underline{r}^*|_{H_0} = \begin{bmatrix} r^*(1|H_0) \\ r^*(2|H_0) \\ \vdots \\ r^*(N|H_0) \end{bmatrix} \quad ; \quad \underline{r}^*|_{H_1} = \begin{bmatrix} r^*(1|H_1) \\ r^*(2|H_1) \\ \vdots \\ r^*(N|H_1) \end{bmatrix} \quad (2.87)$$

Considerando (2.83), (2.84), (2.85), (2.86) e o fato dos processos serem Gaussianos em ambas as hipóteses, verifica-se facilmente que as funções densidade de probabilidade do vetor $\underline{r}^*$ em cada uma das hipóteses são dadas por

$$p_{\underline{r}^*|H_0}(\underline{r}|H_0) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi} K \sigma_{a_1}} e^{-\frac{[R(k)-M(k)]^2}{2K^2 \sigma_{a_1}^2}} \quad (2.88)$$

e

$$p_{\underline{R}}^* | H_1 (\underline{R} | \Pi_1) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi} \sqrt{\lambda_k}} e^{-\frac{R^2(k)}{2\lambda_k}} \quad (2.89)$$

A razão de verossimilhança, definida em (2.16), é então dada por

$$\Lambda(\underline{R}) = \frac{\prod_{k=1}^N \frac{1}{\sqrt{2\pi} \sqrt{\lambda_k}} e^{-\frac{R^2(k)}{2\lambda_k}}}{\prod_{k=1}^N \frac{1}{\sqrt{2\pi} K\sigma_{a_1}} e^{-\frac{[R(k)-M(k)]^2}{2K^2\sigma_{a_1}^2}}} \quad (2.90)$$

A razão de verossimilhança logaritmica, definida por

$$l(\underline{R}) = \ln[\Lambda(\underline{R})] \quad (2.91)$$

será então

$$\begin{aligned}
\ell(\underline{R}) = & \frac{1}{2} \sum_{k=1}^N \left[\frac{1}{K^2 \sigma_{a_1}^2} - \frac{1}{\lambda_k} \right] R^2(k) + \frac{1}{2} \sum_{k=1}^N \frac{M^2(k)}{K^2 \sigma_{a_1}^2} + \\
& + \sum_{k=1}^N \ln(K \sigma_{a_1}) - \sum_{k=1}^N \ln(\sqrt{\lambda_k}) - \sum_{k=1}^N \frac{R(k)M(k)}{K^2 \sigma_{a_1}^2} \quad (2.92)
\end{aligned}$$

A fim de facilitar a implementação do teste de razão de verossimilhança, a expressão (2.92), pode ainda ser escrita como

$$\ell(\underline{R}) = \ell_{C_1} + \ell_{C_2} + \ell_B(\underline{R}) + \ell_{R_1}(\underline{R}) + \ell_{R_2}(\underline{R}) + \ell_{R_3}(\underline{R}) \quad (2.93)$$

onde

$$\ell_{C_1} = \sum_{k=1}^N \ln(K \sigma_{a_1}) \quad (2.94)$$

$$\ell_{C_2} = \frac{1}{2} \sum_{k=1}^N \frac{M^2(k)}{K^2 \sigma_{a_1}^2} \quad (2.95)$$

$$\ell_B(\underline{R}) = -\frac{1}{2} \sum_{k=1}^N \ln(\lambda_k) \quad (2.96)$$

$$\ell_{R_1}(\underline{R}) = \frac{1}{2\sigma_{a_1}^2} \sum_{k=1}^N \left[\frac{\lambda_k - \sigma_{a_1}^2}{\lambda_k} \right] R^2(k) \quad (2.97)$$

$$\ell_{R_2}(\underline{R}) = \frac{1}{2} \sum_{k=1}^N \left[\frac{1}{K^2 \sigma_{a_1}^2} - \frac{1}{\sigma_{a_1}^2} \right] R^2(k) \quad (2.98)$$

e

$$\ell_{R_3}(\underline{R}) = - \sum_{k=1}^N \frac{R(k)M(k)}{K^2 \sigma_{a_1}^2} \quad (2.99)$$

CAPÍTULO 5

IMPLEMENTAÇÃO DO TESTE DE RAZÃO DE VEROSSIMILHANÇA

Reduzida a expressão da razão de verossimilhança logaritmica, apresentada nas equações (2.93) a (2.99), o próximo passo será determinar de que maneira cada uma das parcelas de (2.93) será obtida.

As quantidades ℓ_{C_1} , ℓ_{C_2} , $\ell_{R_2}(\underline{R})$ e $\ell_{R_3}(\underline{R})$ são facilmente obtidas a partir das observações, conforme será visto mais adiante. No entanto, implementar diretamente $\ell_B(\underline{R})$ e $\ell_{R_1}(\underline{R})$ envolveria dificuldades numéricas muito grandes, referentes à determinação de autovalores e autofunções da covariância das observações na hipótese H_1 . O procedimento adotado consiste então em obter $\ell_B(\underline{R})$ e $\ell_{R_1}(\underline{R})$, indiretamente, como sub-produtos da implementação do algoritmo de Kalman-Bucy associado a um problema de filtragem convenientemente definido. Essa possibilidade, que aparentemente foi destacada pela primeira vez por Scheppe⁶ tem sido largamente explorada na implementação de testes de razão de verossimilhança^{7,8,9}. A implementação detalhada para testar entre dois modelos ARIMA parece não ter sido feita antes e a dedução empregada para estabelecer, nesse caso, a conexão entre razão de verossimilhança e o problema de filtragem ótima é também original.

A implementação de cada uma das parcelas de (2.93) é apresentada a seguir.

3.1 - Implementações de ℓ_{c_1} e ℓ_{c_2}

De (2.94) tem-se imediatamente

$$\ell_{c_1} = \sum_{k=1}^N \ln [K \sigma_{a_1}] = \frac{N}{2} \ln [K^2 \sigma_{a_1}^2] \quad ,$$

o que a partir de (2.62) se escreve

$$\ell_{c_1} = \frac{N}{2} \ln [\sigma_{a_0}^2] \quad (3.1)$$

Por outro lado, considerando (2.84) e (2.62), tem-se que (2.95) pode ser escrita como

$$\ell_{c_2} = \frac{1}{2} \frac{1}{\sigma_{a_0}^2} \sum_{k=1}^N \sum_{i=1}^N \xi \phi_k(i) \sum_{j=1}^N \xi \phi_k(j) \quad (3.2)$$

o que, levando em conta (B.21) se escreve

$$\ell_{c_2} = \frac{1}{2 \sigma_{a_0}^2} \sum_{k=1}^N \xi^2 = \frac{N \xi^2}{2 \sigma_{a_0}^2} \quad (3.3)$$

3.2 - Implementação de $\ell_B(\underline{R})$

As equações (B.22), (B.15) e (B.16) permitem escrever

$$\det(\underline{K}_{\underline{w}}) = \det(\underline{L}) = \prod_{k=1}^N \lambda_k \quad (3.4)$$

Desse modo, (2.96) pode ser escrita como

$$\ell_B(\underline{R}) = -\frac{1}{2} \ln \left[\prod_{k=1}^N \lambda_k \right] = -\frac{1}{2} \ln [\det(\underline{K}_{\underline{w}})] \quad (3.5)$$

Considerando ainda (A.24), tem-se que

$$\ell_B(\underline{R}) = -\frac{1}{2} \ln [\det(\underline{K}_{\underline{v}})]$$

onde $v(\cdot)$ é o processo de inovações associado ao processo $w(\cdot)$ introduzido em (2.75). Conforme mostrado através de (2.55), $v(\cdot)$ é um processo de ruído branco. Isto permite escrever (3.5) como

$$\ell_B(\underline{R}) = -\frac{1}{2} \sum_{k=1}^N \ln \left\{ E \left[v^2(k) \right] \right\} \quad (3.6)$$

Para implementar esta última expressão observe-se que, em conexão com um processo $w(\cdot)$, caracterizado por um modelo ARIMA, foi definido através de (2.43) a (2.47) um problema de filtragem. A partir desse problema o processo de inovações $v(\cdot)$ pode ser expresso por (2.48) e (2.49) ou equivalente por (D.14). No apêndice D tem-se, por intermédio de (D.16) a maneira pela qual as quantidades

$$E[v^2(k)] = K_v(k,k) \quad , \quad k=1, \dots, N$$

podem ser obtidas, uma vez implementado o algoritmo de Kalman-Bucy associado ao problema de filtragem considerado

3.3 - Implementação de $\lambda_{R_1}(\underline{R})$

Considerando que

$$r^*(k) = \sum_{j=1}^N r(j) \phi_k(j)$$

tem-se

$$\lambda_{R_1}(\underline{R}) = \frac{1}{2\sigma_{a_1}^2} \sum_{j=1}^N \sum_{m=1}^N \bar{r}(m) \sum_{k=1}^N \frac{\lambda_k - \sigma^2}{\lambda_k} \phi_k(m) \phi_k(j) \bar{r}(j) \quad (3.7)$$

onde $\tilde{r}(k)$, $k=1, \dots, N$ é a realização das observações para a qual se deseja calcular $\lambda_{R_1}(\underline{R})$.

Para o caso onde se tem i observações ao invés de N , pode-se escrever

$$\lambda_{R_1}(\underline{R}; i) = \frac{1}{2\sigma_{a_1}^2} \sum_{j=1}^i \sum_{m=1}^i \tilde{r}(m) \sum_{k=1}^i \frac{\lambda_k(i) - \sigma_{a_1}^2}{\lambda_k(i)} \phi_k(m; i) \phi_k(j; i) \tilde{r}(j) \quad (3.8)$$

ou seja

$$\lambda_{R_1}(\underline{R}; i) = \frac{1}{2\sigma_{a_1}^2} \sum_{j=1}^i \sum_{m=1}^i \tilde{r}(m) h_1(m, j; i) \tilde{r}(j) \quad (3.9)$$

onde $h_1(m, j; i)$ é dado por

$$h_1(m, j; i) = \sum_{k=1}^i \frac{\lambda_k(i) - \sigma_{a_1}^2}{\lambda_k(i)} \phi_k(m; i) \phi_k(j; i) \quad (3.10)$$

Comparando (3.10) e (C.3), conclui-se a partir de (C.4) que $h_1(m, j; i)$ deve satisfazer ao conjunto de equações

$$\sigma_{a_1}^2 \delta(\lambda, m) + \sum_{j=1}^i h_1(m, j; i) K_w(\lambda, j) = K_w(\lambda, m), \quad \lambda=1, \dots, i; m=1, \dots, N; i=1, \dots, N \quad (3.11)$$

Considere agora $\nabla \ell_{R_1}(\underline{R}; i)$, definido por

$$\nabla \ell_{R_1}(\underline{R}; i) = \ell_{R_1}(\underline{R}; i+1) - \ell_{R_1}(\underline{R}; i) \quad (3.12)$$

de modo que $\ell_{R_1}(\underline{R})$ pode ser obtido a partir de

$$\ell_{R_1}(\underline{R}) = \ell_{R_1}(\underline{R}; N) = \sum_{i=0}^{N-1} \nabla \ell_{R_1}(\underline{R}; i) \quad (3.13)$$

com a condição

$$\ell_{R_1}(\underline{R}; 0) = 0 \quad (3.14)$$

Substituindo (3.9) em (3.12) tem-se

$$\begin{aligned} \nabla \ell_{R_1}(\underline{R}; i) = \frac{1}{2\sigma_a^2} & \left\{ \sum_{j=1}^{i+1} \sum_{m=1}^{i+1} \tilde{r}(m) h_1(m, j; i+1) \tilde{r}(j) + \right. \\ & \left. - \sum_{j=1}^i \sum_{m=1}^i \tilde{r}(m) h_1(m, j; i) \tilde{r}(j) \right\} \quad (3.15) \end{aligned}$$

que pode ainda ser escrito como

$$\begin{aligned}
\sigma_{a_1}^2 R_1(\underline{R}; \underline{i}) = & \frac{1}{2\sigma_{a_1}^2} \left\{ \sum_{j=1}^i \sum_{m=1}^{i+1} \tilde{r}(m) \tilde{h}_1(m, j; i) \tilde{r}(j) + \right. \\
& + \left[\sum_{m=1}^{i+1} \tilde{r}(m) h_1(i+1, m; i+1) \right] \tilde{r}(i+1) + \\
& \left. + \left[\sum_{j=1}^i \tilde{r}(j) h_1(i+1, j; i) \right] \tilde{r}(i+1) \right\}, \quad (3.16)
\end{aligned}$$

com

$$\tilde{v} h_1(m, j; i) = h_1(m, j; i+1) - h_1(m, j; i) \quad (3.17)$$

Neste ponto é necessário que sejam feitas algumas considerações, com relação à solução de (3.11). Observe-se que, para um dado número de observações i e fixado $m=m_0$, o problema da determinação do filtro fica reduzido à resolução de um sistema de i equações

$$\sigma_{a_1}^2 \delta(\ell, m_0) + \sum_{j=1}^i h_1(m_0, j; i) K_w(\ell, j) = K_w(\ell, m_0) \quad , \quad \ell=1, \dots, i \quad (3.18)$$

e i incógnitas

$$h_1(m_0, j; i) \quad , \quad j=1, \dots, i \quad (3.19)$$

Considere-se agora o problema da determinação do filtro

$$h(m, j:i) \quad , \quad j=1, \dots, i \quad ; \quad m \text{ inteiro}$$

que a partir das observações

$$w(k) = s(k) + a_1(k) \quad , \quad k=1, \dots, i$$

dã a estimativa $\hat{s}(m|i)$ de $s(m)$, minimizando o erro médio quadrático correspondente. Tal filtro deve satisfazer a um conjunto de equações análogo ao conjunto (A.14). Observe-se ainda que fixado $m=m_0$, o problema da determinação deste filtro fica reduzido à resolução de um sistema de i equações

$$K_{wa_1}(\ell, m_0) + \sum_{j=1}^i h(m_0, j:i) K_w(\ell, j) = K_w(\ell, m_0) \quad , \quad \ell=1, \dots, i \quad (3.20)$$

e i incógnitas

$$h(m_0, j:i) \quad , \quad j=1, \dots, i \quad (3.21)$$

Considerando a equação (2.46), verifica-se que se $m_0 \geq i$, o conjunto de equações (3.18) é idêntico ao conjunto (3.20). Isto significa que para $m_0 \geq i$ (predição ou filtragem) os coeficientes dos dois filtros são idênticos, ou seja

$$h_1(m_0, j:i) = h(m_0, j:i) \quad , \quad m_0 \geq i \quad ; \quad j=1, \dots, i$$

Isto permite escrever a equação (3.16) como

$$\begin{aligned} \nabla_{R_1}^2 (R:i) = & \frac{1}{2\sigma_{a_1}^2} \left\{ \sum_{j=1}^i \sum_{m=1}^{i+1} \tilde{r}(m) \nabla h_1(m, j:i) \tilde{r}(j) + \right. \\ & + \left[\sum_{m=1}^{i+1} h(i+1, m:i+1) \tilde{r}(m) \right] \tilde{r}(i+1) + \\ & \left. + \left[\sum_{j=1}^i h(i+1, j:i) \tilde{r}(j) \right] \tilde{r}(i+1) \right\} \end{aligned} \quad (3.22)$$

Em (3.22) os $h(i+1, m:i+1)$, $m=1, \dots, i+1$ são os coeficientes do filtro $h(\cdot, \cdot : \cdot)$ que dá a estimativa $\hat{s}(i+1|i+1)$ de $s(i+1)$ a partir das observações

$$w(k) = s(k) + a_1(k) \quad , \quad k=1, \dots, i+1$$

Analogamente, os $h(i+1, j:i)$, $j=1, \dots, i$ são os coeficientes do filtro $h(\cdot, \cdot|\cdot)$ que dá a estimativa $\hat{s}(i+1|i)$ de $s(i+1)$ a partir das observações

$$w(k) = s(k) + a_1(k) \quad , \quad k=1, \dots, i$$

Desse modo, tem-se

$$\begin{aligned} \nabla_{R_1}^2 (R:i) = \frac{1}{2\sigma_{a_1}^2} & \left\{ \sum_{j=1}^i \sum_{m=1}^{i+1} \tilde{r}(m) \nabla h_1(m, j:i) \tilde{r}(j) + \right. \\ & \left. + \left[s_{\tilde{r}}(i+1|i+1) + s_{\tilde{r}}(i+1|i) \right] \tilde{r}(i+1) \right\} \end{aligned} \quad (3.23)$$

onde $s_{\tilde{r}}(i+1|i+1)$ e $s_{\tilde{r}}(i+1|i)$ são as respostas do filtro $h(\cdot, \cdot|\cdot)$ às observações $\tilde{r}(\cdot)$.

Para determinar $\nabla h_1(m, j:i)$ aplica-se a (3.11) o operador ∇ relativo a i , obtendo-se

$$\sum_{j=1}^{i+1} h_1(m, j:i+1) K_w(\lambda, j) - \sum_{j=1}^i h_1(m, j:i) K_w(\lambda, j) = 0 \quad ,$$

$$\lambda=1, \dots, i \quad ; \quad m=1, \dots, N \quad ; \quad i=1, \dots, N-1 \quad (3.24)$$

ou ainda

$$\sum_{j=1}^i \left[h_1(m, j:i+1) - h_1(m, j:i) \right] K_w(\ell, j) + h_1(m, i+1:i+1) K_w(\ell, i+1) = 0 ,$$

$$\ell=1, \dots, i \quad ; \quad m=1, \dots, N \quad ; \quad i=1, \dots, N-1 \quad (3.25)$$

Calculando $K_w(\ell, i+1)$ a partir de (3.11) resulta

$$\begin{aligned} & \sum_{j=1}^i \nabla h_1(m, j:i) K_w(\ell, j) + h_1(m, i+1:i+1) \sigma_{a_1}^2 \delta(\ell, i+1) + \\ & + h_1(m, i+1:i+1) \sum_{j=1}^i h_1(i+1, j:i) K_w(\ell, j) = 0 \quad , \\ & \ell=1, \dots, i \quad ; \quad m=1, \dots, N \quad ; \quad i=1, \dots, N-1 \end{aligned} \quad (3.26)$$

ou seja

$$\begin{aligned} & \sum_{j=1}^i \left[\nabla h_1(m, j:i) + h_1(m, i+1:i+1) h_1(i+1, j:i) \right] K_w(\ell, j) = \\ & = - h_1(m, i+1:i+1) \sigma_{a_1}^2 \delta(\ell, i+1) \quad , \quad \ell=1, \dots, i \quad ; \quad m=1, \dots, N \quad ; \quad i=1, \dots, N-1 \end{aligned} \quad (3.27)$$

Levando em conta que i é menor ou igual a i , tem-se

$$\sum_{j=1}^i \left[\forall h_1(m, j:i) + h_1(m, i+1:i+1) h_1(i+1, j:i) \right] K_w(\ell, j) = 0 \quad ,$$

$$\ell=1, \dots, i \quad ; \quad m=1, \dots, N \quad ; \quad i=1, \dots, N-1 \quad (3.28)$$

Multiplicando ambos os membros de (3.28) por $f(\ell)$ definido como

$$f(\ell) = \forall h_1(m, \ell:i) + h_1(m, i+1:i+1) h_1(i+1, \ell:i) \quad (3.29)$$

e somando em relação a ℓ resulta

$$\sum_{\ell=1}^i \sum_{j=1}^i f(\ell) K_w(\ell, j) f(j) = 0 \quad (3.30)$$

Por outro lado, (A.24), (2.55) e (2.56) garantem que K_w é definida positiva, resultando que (3.30) só será satisfeita para

$$f(j) = 0 \quad , \quad j=1, \dots, i$$

ou seja

$$\forall h_1(m, j:i) + h_1(m, i+1:i+1) h_1(i+1, j:i) = 0 \quad , j=1, \dots, i \quad (3.31)$$

Tem-se então

$$\begin{aligned} \forall h_1(m, j:i) &= - h_1(m, i+1:i+1) h_1(i+1, j:i) = \\ &= - h_1(i+1, m:i+1) h_1(i+1, j:i) \end{aligned} \quad (3.32)$$

Substituindo agora (3.32) em (3.23) obtem-se

$$\begin{aligned} \nabla_{R_1}^2(\underline{R}:i) &= \frac{1}{2\sigma_{a_1}^2} \left\{ \sum_{j=1}^i \sum_{m=1}^{i+1} \tilde{r}(m) \left[-h_1(i+1, m:i+1) h_1(i+1, j:i) \right] \tilde{r}(j) + \right. \\ &\quad \left. + \left[s_{\tilde{r}}(i+1|i+1) + s_{\tilde{r}}(i+1|i) \right] \tilde{r}(i+1) \right\} \end{aligned} \quad (3.33)$$

Verifica-se facilmente que na equação (3.33) pode-se também substituir os coeficientes $h(\cdot, \cdot: \cdot)$ pelos coeficientes $h(\cdot, \cdot: \cdot)$ do filtro ótimo definido por (3.20). Isto permite escrever (3.33) como

$$\nabla_{R_1}^2(\underline{R}:i) = \frac{1}{2\sigma_{a_1}^2} \left\{ \tilde{r}(i+1) \left[s_{\tilde{r}}(i+1|i+1) + s_{\tilde{r}}(i+1|i) \right] - s_{\tilde{r}}(i+1|i+1) s_{\tilde{r}}(i+1|i) \right\} \quad (3.34)$$

Assim sendo, a obtenção de $\ell_{R_1}(\underline{R})$ é feita através das equações (3.13), (3.14) e (3.34), onde $s_{\tilde{r}}(i+1|i+1)$ e $s_{\tilde{r}}(i+1|i)$ são implementadas a partir do filtro de Kalman-Bucy associado ao problema de filtragem anteriormente definido.

3.4 - Implementações de $\ell_{R_2}(\underline{R})$ e $\ell_{R_3}(\underline{R})$

Considerando que

$$r^*(k) = \sum_{j=1}^N r(j) \phi_k(j) \quad (3.35)$$

tem-se a partir de (2.98), que

$$\ell_{R_2}(\underline{R}) = \frac{1}{2} \left[\frac{1}{K^2 \sigma_{a_1}^2} - \frac{1}{\sigma_{a_1}^2} \right] \sum_{j=1}^N \sum_{m=1}^N \tilde{r}(m) \tilde{r}(j) \sum_{k=1}^N \phi_k(j) \phi_k(m) \quad (3.36)$$

o que considerando (B.21) e (2.62) pode ser escrito como

$$\ell_{R_2}(\underline{R}) = \frac{1}{2} \frac{\sigma_{a_1}^2 - \sigma_{a_0}^2}{\sigma_{a_1}^2 \sigma_{a_0}^2} \sum_{j=1}^N \tilde{r}(j) \quad (3.37)$$

Por outro lado, a partir de (2.84), (3.35) e (2.99) $\ell_{R_3}(\underline{R})$ pode ser escrita como

$$\ell_{R_3}(\underline{R}) = - \frac{1}{K^2 \sigma_{a_1}^2} \sum_{j=1}^N \sum_{m=1}^N \bar{r}(m) \xi \sum_{k=1}^N \phi_k(j) \phi_k(m) \quad (3.38)$$

o que, levando em conta (B.21) e (2.62) dá finalmente

$$\ell_{R_3}(\underline{R}) = - \frac{\xi}{\sigma_{a_0}^2} \sum_{j=1}^N \bar{r}(j) \quad (3.39)$$

3.5 - Implementação de $\ell(\underline{R})$

Uma vez que nas seções anteriores cada uma das parcelas que aparecem em (2.93) foi implementada, tem-se que a implementação de $\ell(\underline{R})$ está completa. A figura 3.1 apresenta um diagrama em blocos do algoritmo usado para a obtenção de $\ell(\underline{R})$ a partir da realização disponível das observações e dos parâmetros que caracterizam os modelos $ARIMA(p_0, d_0, q_0)$ e $ARIMA(p_1, d_1, q_1)$ correspondentes a cada uma das hipóteses consideradas.

FIG. 3.1- DIAGRAMA EM BLOCOS DO ALGORITMO DE IMPLEMENTAÇÃO DE $\hat{l}(R)$

CAPÍTULO 4

SIMULAÇÕES

Com o objetivo de verificar o desempenho associado ao teste de razão de verossimilhança foram examinadas, através de simulações, algumas situações em que esse teste foi empregado, e nas quais utilizou-se como termo de comparação o desempenho associado a um outro teste ^{1,2,4,5} usualmente empregado. Este último teste utiliza como critério a escolha da hipótese para a qual uma função da soma dos quadrados das autocorrelações dos resíduos tem maior probabilidade de ser excedida em uma determinada distribuição χ^2 - quadrado.

Dois tipos de problemas foram analisados. No primeiro deles o desempenho dos dois testes foi comparado em conexão com o problema clássico de teste de hipóteses, no qual cada uma das hipóteses está completamente caracterizada a priori (parâmetros conhecidos). Neste problema a escolha desejável é óbvia pois se quer sempre escolher a hipótese verdadeira. Este problema é, no entanto, pouco significativo no que diz respeito à identificação de processos estocásticos, uma vez que nesse caso o problema de teste de hipóteses formulado para decidir entre dois modelos será sempre um problema envolvendo parâmetros desconhecidos. Assim, o segundo tipo de problemas analisado se refere ao caso no qual, em cada uma das hipóteses os parâmetros são desconhecidos. Além disso, se o objetivo é a determinação de um modelo para um processo estocástico, a escolha desejável não é tão óbvia como no

caso em que se tem parâmetros conhecidos. Com efeito, é perfeitamente concebível que um processo ARIMA (p_0, d_0, q_0) possa ser melhor representado por um modelo ARIMA (p_1, d_1, q_1) , simplesmente porque os erros cometidos na estimação dos parâmetros correspondentes ao modelo ARIMA (p_0, d_0, q_0) acabem conduzindo a um modelo pior do que o modelo ARIMA (p_1, d_1, q_1) estimado, mesmo levando em consideração os erros cometidos na estimação dos parâmetros deste último modelo. Assim sendo, em conexão com esse problema foi introduzido um outro critério capaz de determinar a escolha desejável. Esse critério será descrito posteriormente.

4.1 - Parâmetros Conhecidos

Nesse caso, a simulação consistiu em aplicar o teste, um determinado número de vezes, ao problema em que se quer decidir entre um modelo ARIMA (p_0, d_0, q_0) e um modelo ARIMA (p_1, d_1, q_1) , ambos com todos os parâmetros conhecidos.

Foi primeiramente gerado um determinado número N de funções amostra de um dos processos, no caso o ARIMA (p_0, d_0, q_0) , e verificado o número de vezes N_0 em que o teste de razão de verossimilhança optava pelo outro modelo. Observou-se então que à medida em que N crescia, a razão $\frac{N_0}{N}$ tendia para um valor constante. Esse valor foi então considerado como uma boa estimativa para a probabilidade de erro, dado que a hipótese H_0 era a verdadeira. Isto significa que para valores grandes de N considerou-se

$$Pe_{TRV} | H_0 \approx \frac{N_0}{N}$$

Por um procedimento análogo, pôde-se também determinar o número N_1 de vezes em que o teste de razão de verossimilhança optava pelo modelo ARIMA (p_0, d_0, q_0) , quando na verdade o modelo gerador das funções amostra era o ARIMA (p_1, d_1, q_1) . Supondo-se que nesse caso o número de funções amostra geradas era ainda N , o valor da probabilidade de erro dado que a hipótese verdadeira é H_1 foi estimado por

$$P_{e_{TRV}} | H_1 \cong \frac{N_1}{N}$$

Considerando-se que foram geradas N funções amostra de cada um dos modelos possíveis, as possibilidades de se ter uma função amostra gerada por um determinado modelo, $P(H_0)$ e $P(H_1)$, foram tomadas como sendo iguais a 0,5, o que foi aliás assumido na dedução do teste de razão de verossimilhança. Desse modo tem-se

$$P_{e_{TRV}} = \frac{1}{2} P_{e_{TRV}} | H_0 + \frac{1}{2} P_{e_{TRV}} | H_1 \cong \frac{N_0 + N_1}{2N}$$

A simulação descrita acima foi feita para alguns casos, onde considerou-se tanto modelos bastante próximos, como nos exemplos 4.1 e 4.2. quanto modelos não muito próximos como no exemplo 4.3.

Convém observar que, com o objetivo de se ter uma indicação relativa da qualidade do teste de razão de verossimilhança, a simulação apresenta também os resulta

dos obtidos para o teste feito a partir da distribuição χ^2 - quadrado de uma variável aleatória definida como função da soma dos quadrados das autocorrelações dos resíduos. O teste das autocorrelações dos resíduos (TAR) é descrito em mais detalhes na literatura^{1,4,5}.

Exemplo 4.1: Este exemplo consiste na simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos aplicados aos modelos

$$H_0 : \text{ARIMA } (2,0,1)$$

$$H_1 : \text{ARIMA } (1,0,0)$$

com o objetivo de se determinar uma estimativa para a probabilidade de erro associada a cada um desses testes. Os parâmetros de ambos os modelos são supostos conhecidos. Para o modelo ARIMA (2,0,1) tem-se

$$\phi_1^0 = 0,75$$

$$\phi_2^0 = -0,125$$

$$\theta_1^0 = 0,5025$$

$$\epsilon_0^0 = 0$$

e

$$\sigma_{a_0}^2 = 20.0$$

Para o modelo ARIMA (1,0,0) tem-se

$$\phi_1^1 = 0,25$$

$$\theta_0^1 = 0$$

e

$$\sigma_{a_1}^2 = 20,0$$

Cada função amostra $\tilde{e}$ constituída de 50 observações.

Convém observar que a escolha dos parâmetros foi feita de modo a se ter modelos bastante próximos, e com efeito tem-se nesse caso particular

$$\frac{\phi^0(B)}{\theta^0(B)} \equiv \phi^1(B)$$

Os resultados dessa simulação são apresentados nas figuras 4.1 e 4.2. Pode-se verificar através dos resultados da simulação, que nesse caso particular obteve-se

$$Pe_{TRV} \equiv 0,55$$

e

$$Pe_{TAR} \equiv 0,49$$

Figura 4.1 - Resultados da simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos, condicionados à ocorrência de cada uma das hipóteses H_0 e H_1 , relativos ao exemplo 4.1

Figura 4.2 - Resultados da simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos, relativos ao exemplo 4.1

Exemplo 4,2: Este exemplo consiste na simulação dos desem
penhos dos testes de razão de verossimilhança e das auto
correlações dos resíduos aplicados aos modelos

$$H_0 : \text{ARIMA } (1,0,2)$$

$$H_1 : \text{ARIMA } (0,0,1)$$

com o objetivo de se determinar uma estimativa para a pro
babilidade de erro associada a cada um desses testes. Os
 parâmetros de ambos os modelos são supostos conhecidos. Pa
ra o modelo ARIMA (1,0,2) tem-se

$$\phi_1^0 = 0,339$$

$$\theta_1^0 = 0,5833$$

$$\theta_2^0 = - 0,0833$$

$$\theta_0^0 = 0$$

e

$$\sigma_{a_0}^2 = 25,0$$

Para o modelo ARIMA (0,0,1) tem-se

$$\theta_1^1 = 0,25$$

$$\theta_0^1 = 0$$

e

$$\sigma_{a_1}^2 = 25,0$$

Cada função amostra é constituída de 40 observações

Convém observar ainda que a escolha dos parâmetros foi feita de modo a se ter modelos bastante próximos, e com efeito tem-se nesse caso particular

$$\phi^1(B) \approx \frac{\theta^0(B)}{\phi^0(B)}$$

Destaque-se ainda que os modelos aqui considerados estão mais próximos do que aqueles examinados no exemplo 4.1. Os resultados dessa simulação são apresentados nas figuras 4.3 e 4.4. Os resultados encontrados nesse caso particular foram

$$Pe_{TRV} \approx 0,42$$

e

$$Pe_{TAR} \approx 0,50$$

Figura 4.3 - Resultados da simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos, condicionados à ocorrência de cada uma das hipóteses H_0 e H_1 , relativos ao exemplo 4.2

Figura 4.4 - Resultados da simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos, relativos ao exemplo 4.2

Exemplo 4.3: Este exemplo consiste na simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos aplicados aos modelos

$$H_0 : \text{ARIMA}(2,0,1)$$

$$H_1 : \text{ARIMA}(1,0,0)$$

com o objetivo de se determinar uma estimativa para a probabilidade de erro associada a cada um desses testes. Os parâmetros de ambos os modelos são supostos conhecidos. Para o modelo ARIMA (2,0,1) tem-se

$$\phi_1^0 = 0,75$$

$$\phi_2^0 = - 0,22$$

$$\theta_1^0 = 0,5025$$

$$\theta_0^0 = 0$$

e

$$\sigma_{a_0}^2 = 15,0$$

Para o modelo ARIMA (1,0,0) tem-se

$$\phi_1^1 = 0,25$$

$$\theta_0^1 = 0$$

e

$$\sigma_{a_1}^2 = 20,0$$

Cada função amostra é constituída de 50 observações.

Convém observar que nesse exemplo, a escolha dos parâmetros foi feita de modo a se ter modelos próximos um ao outro, mas não tão próximos como nos exemplos 4.1 e 4.2. Com efeito, nesse caso particular, os resultados obtidos mostraram que a probabilidade de erro associada ao teste das autocorrelações dos resíduos diminuiu em relação aos valores obtidos nos exemplos 4.1 e 4.2. Tal diminuição foi entretanto muito pequena, comparada ao valor da variação da probabilidade de erro associada ao teste de razão de verossimilhança, que nesse caso reduziu-se praticamente a zero.

Os resultados dessa simulação são apresentados nas figuras 4.5 e 4.6. Pode-se verificar através desses resultados que nesse caso particular obteve-se

$$Pe_{TRV} \cong 0$$

e

$$Pe_{TAR} \cong 0,38$$

Figura 4.5 - Resultados da simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos, condicionados à ocorrência de cada uma das hipóteses H_0 e H_1 , relativos ao exemplo 4.3

Figura 4.6 - Resultados da simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos, relativos ao exemplo 4.3

4.2 - Parâmetros Desconhecidos

Tendo em vista que no caso real do problema de identificação de séries temporais, os parâmetros dos modelos não são conhecidos, foi feita uma simulação dos desempenhos dos testes de razão de verossimilhança e das autocorrelações dos resíduos para esse caso. Embora as ordens dos modelos em ambas as hipóteses tenham sido supostas conhecidas, seus parâmetros foram supostos desconhecidos.

Foi então gerado um determinado número de funções amostra, cada uma delas constituída de um certo número de observações. Tais funções amostra foram geradas por um modelo de mesma ordem que um dos dois modelos em consideração, por exemplo, de mesma ordem que o ARIMA (p_0, d_0, q_0) . Cada função amostra considerada foi então suposta como sendo gerada por um modelo ARIMA (p_0, d_0, q_0) , sendo então estimados os parâmetros relativos a esta hipótese. Foram em seguida estimados os parâmetros de um modelo de ordem (p_1, d_1, q_1) , associado a essa mesma função amostra. Uma vez estimados os parâmetros (ainda que em uma das hipóteses essa estimação tenha partido de uma suposição errada) o problema está reduzido a um teste de hipóteses com parâmetros conhecidos. A aplicação do TRV a essa situação, tratando os parâmetros estimados como verdadeiros, consiste no que se denomina o teste de razão de verossimilhança generalizado (TRVG). No que se refere à estimação dos parâmetros foram utilizadas estimativas de máxima verossimilhança^{1,3}.

Nesse ponto tornou-se necessário o estabelecimento de um critério, segundo o qual, se poderia determinar qual dos dois modelos associados à função amostra em questão melhor representaria o processo gerador. O critério escolhido foi o de se optar pelo modelo que tivesse menor erro médio quadrático na predição de um passo à frente dos n valores futuros da série, contados a partir do último valor observado. A fim de se poder avaliar o erro cometido em cada predição, foi necessário então que se gerasse funções amostra com n observações adicionais. Tais observações no entanto, não foram utilizadas nem na estimação dos parâmetros nem na implementação dos testes.

Verificou-se então o número de vezes em que os testes de razão de verossimilhança e das autocorrelações dos resíduos optavam por modelos em discordância com o critério descrito acima. Cada discordância foi então contada como um erro na escolha do modelo mais apropriado, erro este, condicionado à geração das funções amostra a partir de um modelo ARIMA (p_0, d_0, q_0) . Foi então possível chegar a estimativas das probabilidades de erro

$$P_{\text{TRVG}} | A_0$$

e

$$P_{\text{TARG}} | A_0$$

onde o evento A_0 é definido por

$$A_i = \left\{ \begin{array}{l} \text{As observações utilizadas foram geradas} \\ \text{pelo modelo ARIMA } (p_i, d_i, q_i) \end{array} \right\}, \quad i=0,1$$

Ressalte-se mais uma vez que o erro é aqui entendido como a escolha do pior modelo (de acordo com o critério descrito anteriormente) e não necessariamente como a escolha do modelo ARIMA de ordem diferente daquele utilizado na geração da função amostra em análise.

Gerando agora um determinado número de funções amostra a partir de um modelo ARIMA (p_1, d_1, q_1) , e procedendo-se de forma completamente análoga, é possível chegar às estimativas das probabilidades de erro

$$Pe_{TRVG|A_1}$$

e

$$Pe_{TARG|A_1}$$

com o evento A_1 previamente definido

Como as probabilidades a priori dos eventos A_0 e A_1 são iguais a 0,5, o que aliás foi assumido na implementação do TRV, tem-se que as probabilidades de erro associadas ao TRVG e ao TARG se escrevem respectivamente

$$Pe_{TRVG} = \frac{1}{2} \left[Pe_{TRVG}|A_1 + Pe_{TRVG}|A_0 \right]$$

e

$$Pe_{TARG} = \frac{1}{2} \left[Pe_{TARG}|A_1 + Pe_{TARG}|A_0 \right]$$

Exemplo 4.4 Este exemplo consiste na simulação dos desempenhos do TRVG e do TARG quando aplicados ao problema em que se quer decidir sobre dois modelos

H_0 : ARIMA (2,0,1)

H_1 : ARIMA (1,0,0)

passíveis de caracterizar um determinado processo cujas funções amostra, para efeito da simulação, foram obtidas através de dois modelos

A_0 : ARIMA (2,0,1)

A_1 : ARIMA (1,0,0)

completamente caracterizados pelos seguintes parâmetros

$$\phi_1^0 = 0,75$$

$$\phi_2^0 = - 0,125$$

$$\theta_1^0 = 0,5025$$

$$\theta_0^0 = 0$$

$$\sigma_{a_0}^2 = 20,0$$

$$\phi_1^1 = 0,25$$

$$\theta_0^1 = 0$$

e

$$\sigma_{a_1}^2 = 20,0$$

Esses modelos são precisamente aqueles utilizados no exemplo 4.1. Observe-se ainda que esses parâmetros foram empregados na geração das funções amostra correspondentes a cada uma das hipóteses. Entretanto na implementação dos testes eles foram supostos desconhecidos, sendo portanto necessário estimá-los, conforme descrito anteriormente.

Nas simulações foram utilizadas funções amostra com 80 observações, sendo os 40 primeiros valores utilizados na estimação dos parâmetros e implementação do teste. Os valores restantes foram empregados apenas na determinação do erro médio quadrático associado às previsões de um passo à frente, conforme já explicado.

Os resultados da simulação são apresentados nas figuras 4.7 e 4.8.

Figura 4.7 - Resultados da simulação dos desempenhos do teste de razão de verossimilhança gene ralizado e do teste das autocorrelações dos resíduos generalizado, condicionados à ocorrência dos eventos A_0 e A_1 , relativos ao exemplo 4.4

Figura 4.8 - Resultados da simulação dos desempenhos do teste de razão de verossimilhança $g_{\underline{g}}$ normalizado e do teste das autocorrelações dos resíduos generalizado, relativa ao exemplo 4.4

Na figura 4.9, apresenta-se a título de ilustração, uma função amostra particular (80 observações), gerada a partir do modelo ARIMA (1,0,0). São também apresentadas na figura as previsões de um passo à frente, obtidas para os 40 últimos valores da série, a partir dos modelos ARIMA (1,0,0) e ARIMA (2,0,1), que tiveram seus parâmetros estimados com base nas 40 primeiras observações. Resulta da figura 4.9 que as previsões obtidas são claramente superiores quando o modelo ARIMA (1,0,0) é empregado. Em verdade, esta foi a escolha do TRVG enquanto o TARG escolheu o modelo ARIMA (2,0,1). Para outras funções amostra as discrepâncias entre as previsões obtidas a partir de cada um dos modelos não foram necessariamente tão acentuadas, o que significa que a escolha do pior modelo nem sempre levaria a previsões tão inferiores como as mostradas na figura 4.9.

4.3 - Análise dos Resultados

Os resultados obtidos na seção 4.1 mostram que a utilização da soma dos quadrados das autocorrelações dos resíduos não é uma estatística adequada para decidir entre as duas hipóteses consideradas, pelo menos quando essa estatística não dá uma forte indicação de que uma das hipóteses deva ser rejeitada. Isto ocorrerá, sempre que os modelos correspondentes às duas hipóteses estejam suficientemente próximos para que ambos sejam aceitáveis como representativos da série de valores considerada. Além disso, o teste de razão de verossimilhança embora mais trabalhoso não é computacionalmente tão complexo a ponto de tornar-se desaconselhável. Assim sendo, no problema clássico de teste de hipóteses em que os parâmetros são desconhecidos o TAR é definitivamente desaconselhável. Em particular, para

Figura 4.9 - Comparação das previsões de um passo à frente para uma determinada função amostra, relativa ao exemplo 4.4

o exemplo 4.3, em que as duas hipóteses são facilmente distinguíveis a partir do TRV (probabilidade de erro essencialmente nula), o TARG apresenta resultados pouco satisfatórios. Mais do que isso, se cada modelo é examinado isoladamente não haveria, a partir da soma dos quadrados das autocorrelações dos resíduos, evidência de que qualquer dos dois deve-se ser rejeitado.

No que se refere à situação em que os parâmetros não são conhecidos observa-se que o TARG falha completamente. Tal situação corresponde precisamente ao problema de maior interesse prático em que se quer construir um modelo para um determinado processo estocástico. Assim sendo, a utilização do TRVG nesse caso é de toda a conveniência. Seria interessante ressaltar também que outros critérios (que não a qualidade das previsões) poderiam ser empregados para caracterizar o melhor modelo. De qualquer forma o critério empregado refere-se à uma utilização comum dos modelos construídos para séries temporais.

Em particular foi observado no exemplo 4.4 que o modelo ARIMA (1,0,0), contendo o menor número de parâmetros a serem estimados, conduzia (de acordo com o critério de mínimo erro médio quadrático nas previsões) ao melhor modelo, para a maioria das simulações. Da figura 4.7, observa-se que o TRVG se adapta perfeitamente bem a essa circunstância, no sentido de que a probabilidade de erro (escolha do pior modelo) é essencialmente independente do modelo gerador. Por outro lado, o TARG tem desempenho muito fraco, escolhendo na maioria das vezes o pior modelo, independentemente do modelo gerador.

Concluindo a presente seção, seria apropriado apresentar alguns comentários, interpretando mais explicitamente o procedimento adotado dentro do teste das autocorrelações dos resíduos. Observe-se inicialmente que a soma dos quadrados das autocorrelações dos resíduos é uma medida da proximidade entre a série de resíduos e uma série de ruído branco. Quanto maior essa proximidade menor deve ser essa soma. Box^{1,2} usa o fato de que essa soma tem uma distribuição χ^2 - quadrado para verificar se o valor obtido é suficientemente pequeno, de tal modo que não se possa rejeitar a hipótese de que a série de resíduos observada seja branca. Quando se compara dois modelos diferentes, o que se entende por "suficientemente pequeno" é traduzido em termos da probabilidade desse valor ser excedido na distribuição χ^2 - quadrado correspondente. Então o melhor modelo seria aquele correspondendo à soma com maior probabilidade de ser excedida. Os exemplos anteriores mostram que para modelos suficientemente bons esse critério não conduz a resultados satisfatórios.

Uma outra possibilidade seria encarar o problema de decidir entre dois modelos a partir da soma dos quadrados das autocorrelações dos resíduos, como um outro problema de teste de hipóteses. Nesse problema se quer decidir se é mais aceitável a soma associada ao primeiro modelo provir da χ^2 - quadrado correspondente ou se é mais aceitável a soma associada ao outro modelo provir da χ^2 - quadrado que a ele corresponde. Visto sob este aspecto é certamente discutível se o critério descrito acima tem alguma validade, principalmente se ambas as somas têm probabilidade grande de serem excedidas, nas respectivas distribuições. Por outro lado, torna-se difícil formular adequadamente esse problema de teste de hipóteses por que embora sejam conhecidas as distribuições das somas condicionadas às hipóteses verdadeiras, essas distribuições não são conhecidas quando condicionadas às hipóteses falsas.

No caso de parâmetros conhecidos apenas uma distribuição χ^2 - quadrado é envolvida². Então se poderia pensar em decidir pelo modelo cuja soma dos quadrados das autocorrelações dos resíduos correspondesse ao maior valor na densidade de probabilidade χ^2 -quadrado em questão. Observe-se que este critério coincide com o critério descrito anteriormente quando as somas são superiores ao valor máximo da distribuição χ^2 - quadrado, mas leva exatamente a resultado oposto se estas somas são inferiores a esse valor de máxima verossimilhança. Para os exemplos 4.1 até 4.3 foi examinada a possibilidade de utilização desse critério alternativo, resultando entretanto probabilidades de erro essencialmente com os mesmos valores que as probabilidades mostradas nas figuras 4.1 até 4.6. Para o exemplo 4.4 utilizou-se apenas o critério inicial, tendo em vista não só seu significado intrínseco (maior proximidade entre a série de resíduos e uma série de ruído branco) como também a dificuldade em adotar o outro ponto de vista agora sugerido.

CAPITULO 5

CONCLUSÕES

Uma primeira conclusão importante do presente trabalho se refere à possibilidade de implementar o teste de razão de verossimilhança, associado à escolha entre dois modelos ARIMA, a partir do filtro de Kalman-Bucy. Ainda que esse fato já tenha sido estabelecido anteriormente para diversas classes de problemas de teste de hipóteses, algumas características especiais do problema correspondente à escolha entre modelos ARIMA, não parece ter sido explicitamente considerada antes. Mais especificamente o problema de teste de hipóteses a que se chegou corresponde a se ter ruído branco em uma das hipóteses e ruído branco somado a um processo com ele correlatado na outra hipótese. Além disso os ruídos brancos aparecendo nas duas hipóteses são linearmente dependentes.

A segunda conclusão importante diz respeito à comparação entre o teste de razão de verossimilhança e um outro teste baseado nas estimativas das autocorrelações dos resíduos da série. Tanto considerando parâmetros conhecidos como empregando as versões generalizadas dos dois testes, em que os parâmetros são previamente estimados, o teste de razão de verossimilhança é definitivamente superior.

Uma possível generalização do trabalho a qui realizado consistiria na dedução e implementação do teste de razão de verossimilhança para decidir entre dois modelos passíveis de caracterizar o relacionamento entre duas séries temporais (modelos de função de transferência ¹).

Um outro aspecto que poderia ser examinado com maior profundidade diz respeito à possibilidade de definir um teste, ainda baseado nas autocorrelações dos resíduos, mas que não tomasse simplesmente a probabilidade da soma dos quadrados dessa autocorrelação ser excedida (de acordo com uma distribuição χ -quadrado conveniente) como a medida de maior ou menor adequação do modelo.

Finalmente, seria também apropriado investigar critérios, alternativos ao de previsões de um passo à frente, através dos quais se pudesse definir o melhor entre os diversos modelos em consideração.

APÊNDICE A

FILTRO LINEAR ÓTIMO

Neste apêndice são deduzidas algumas expressões relativas ao problema de filtragem linear ótima de um processo de parâmetro discreto, observado em presença de ruído. Estas expressões são naturalmente bem conhecidas^{7,9,10,11,12,13} sendo aqui revistas com a finalidade de serem apresentadas na forma exata em que foram usadas ao longo do texto.

Considere uma situação em que se observa

$$w(k) = s(k) + a(k) \quad , \quad k=1, \dots, N \quad (A.1)$$

onde $s(\cdot)$ e $a(\cdot)$ são processos de parâmetro discreto. O processo $a(\cdot)$ é um processo de ruído branco, de média nula e covariância

$$K_a(k, \ell) = E[a(k) a(\ell)] = \sigma_a^2 \delta(k, \ell) \quad , \quad (A.2) \\ \sigma_a^2 > 0$$

com $\delta(k, \ell)$ definido por

$$\delta(k, \ell) = \begin{cases} 1 & k=\ell \\ 0 & k \neq \ell \end{cases}$$

O processo $a(\cdot)$ não é necessariamente descorrelatado de $s(\cdot)$. O processo $s(\cdot)$ tem média nula e covariância finita, ou seja

$$\begin{cases} E[s(k)] = 0 \\ K_s(k, \ell) = E[s(k)s(\ell)] < \infty \end{cases} \quad (\text{A.3})$$

O problema de filtragem linear ótima consiste aqui em determinar a equação a ser satisfeita pelo filtro linear

$$h(m, j:i) \quad , \quad i=1, \dots, N ; j=1, \dots, N ; m \text{ inteiro} \quad (\text{A.4})$$

que fornece uma estimativa $\hat{s}(m|i)$ de $s(m)$ a partir das observações

$$w(k) \quad , \quad k=1, \dots, N \quad (\text{A.5})$$

de modo a minimizar o erro médio quadrático cometido. Nesta situação, a estimativa $\hat{s}(m|i)$ de mínimo erro médio quadrático de $s(m)$ seria obtida por

$$\hat{s}(m|i) = \sum_{j=1}^i h(m, j:i) w(j) \quad (\text{A.6})$$

O erro médio quadrático cometido seria dado por

$$\xi = E \left\{ \left[\hat{s}(m|i) - s(m) \right]^2 \right\} \quad (A.7)$$

que pode ser escrito como

$$\xi = E \left\{ \left[\sum_{j=1}^i h(m,j:i) w(j) - w(m) + a(m) \right]^2 \right\}$$

ou ainda

$$\begin{aligned} \xi = E & \left\{ \sum_{j=1}^i \sum_{k=1}^i h(m,j:i) h(m,k:i) w(j) w(k) \right\} + \\ & + 2E \left\{ \sum_{j=1}^i h(m,j:i) w(j) a(m) - \sum_{j=1}^i h(m,j:i) w(j) w(m) \right\} + \\ & + E[w^2(m)] + E[a^2(m)] - 2E[w(m)a(m)] \end{aligned} \quad (A.8)$$

Observe-se que se quer determinar os valores

$$h(m, \ell:i), \quad \ell = 1, \dots, i$$

que minimizam (A.8). Como se trata de uma função quadrática das i variáveis que se quer determinar tem-se que são condições necessárias e suficientes para a ocorrência de um mínimo

$$\frac{d\xi}{dh(m,\ell:i)} = 0 \quad , \quad \ell = 1, \dots, i \quad (\text{A.9})$$

e

$$\frac{d^2\xi}{dh^2(m,\ell:i)} > 0 \quad \ell = 1, \dots, i \quad (\text{A.10})$$

Derivando (A.8) em relação a $h(m,\ell:i)$ obtem-se

$$\frac{d\xi}{dh(m,\ell:i)} = 2K_{wa}(\ell,m) - 2K_w(\ell,m) + 2 \sum_{j=1}^i h(m,j:i) K_w(\ell,j) \quad (\text{A.11})$$

Desse modo, a condição (A.9) torna-se

$$K_{wa}(\ell,m) + \sum_{j=1}^i h(m,j:i) K_w(\ell,j) = K_w(\ell,m) \quad , \quad \ell = 1, \dots, i \quad (\text{A.12})$$

O cálculo da derivada segunda conduz facilmente a

$$\frac{d^2\xi}{dh^2(m,\ell:i)} = 2 K_w(\ell,\ell) \quad , \quad \ell = 1, \dots, i \quad (\text{A.13})$$

o que admitindo-se que $w(k)$ tem variância positiva para $k=1,2,\dots,N$, é uma quantidade estritamente positiva. Desse modo, o conjunto de equações a ser satisfeito pelos coeficientes do filtro ótimo seria

$$K_{wa}(\ell, m) + \sum_{j=1}^i h(m, j:i) K_w(\ell, j) = K_w(\ell, m) \quad , \quad \ell=1,\dots,i; i=1,\dots,N ;$$

m inteiro (A.14)

Em conexão com o problema de filtragem é, muitas vezes conveniente trabalhar-se com o processo de inovações $v(.)$ associado do processo $w(.)$ ^{12,14,15}. Para as observações do processo $w(.)$ dadas por (A.1) tem-se

$$v(k) = w(k) - \hat{s}(k|k-1) \quad , \quad k=1,\dots,N \quad (A.15)$$

Na equação (A.15), $\hat{s}(k|k-1)$ é obviamente dado por

$$\hat{s}(k|k-1) = \sum_{j=1}^{k-1} h(k, j:k-1) w(j) \quad , \quad k=1,\dots,N \quad (A.16)$$

Os conjuntos de equações (A.15) e (A.16) podem ainda ser escritos em notação matricial. Define-se

$$\underline{v} = \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(N) \end{bmatrix} \quad (\text{A.17})$$

cuja matriz covariância é dada por

$$\underline{K_v} = E \left[\underline{v} \underline{v}^T \right] = \begin{bmatrix} K_v(1,1) & K_v(1,2) & \dots & K_v(1,N) \\ K_v(2,1) & K_v(2,2) & \dots & K_v(2,N) \\ \vdots & & & \\ K_v(N,1) & K_v(N,2) & \dots & K_v(N,N) \end{bmatrix} \quad (\text{A.18})$$

onde

$$K_v(i,j) = E[v(i)v(j)] \quad , \quad i=1,\dots,N ; j=1,\dots,N$$

Define-se ainda

$$\underline{w} = \begin{bmatrix} w(1) \\ w(2) \\ \vdots \\ w(N) \end{bmatrix} \quad (\text{A.19})$$

e

$$\underline{H} = \begin{bmatrix} 0 & 0 & . & . & . & . & . & . & . & 0 \\ h(2,1:1) & 0 & . & . & . & . & . & . & . & 0 \\ h(3,1:2) & h(3,2:2) & . & . & . & . & . & . & . & 0 \\ \vdots & & & & & & & & & \\ h(N,1:N-1) & h(N,2:N-1) & . & . & . & . & . & . & . & 0 \end{bmatrix} \quad (\text{A.20})$$

Levando em consideração (A.16), o conjunto de equações (A.15) pode ser escrito como

$$\underline{v} = \left[\underline{I} - \underline{H} \right] \underline{w} \quad (\text{A.21})$$

onde $\underline{I}$ é a matriz identidade, ou ainda

$$\underline{v} = \underline{M} \underline{w} \quad (\text{A.22})$$

com

$$\underline{M} = \begin{bmatrix} 1 & 0 & . & . & . & . & . & . & . & 0 \\ -h(2,1:1) & 1 & . & . & . & . & . & . & . & 0 \\ -h(3,1:2) & -h(3,2:2) & 1 & . & . & . & . & . & . & 0 \\ \vdots & & & & & & & & & \\ -h(N,1:N-1) & -h(N,2:N-1) & . & . & . & . & . & . & . & 1 \end{bmatrix}$$

Desse modo tem-se

$$\underline{K_v} = \underline{M} \underline{K_w} \underline{M}^T \quad (\text{A.23})$$

onde

$$\underline{K_w} = E \left[\underline{w} \underline{w}^T \right]$$

A partir de (A.23), conclui-se facilmente que

$$\det \left[\underline{K_v} \right] = \det \left[\underline{K_w} \right] \quad (\text{A.24})$$

uma vez que

$$\det \left[\underline{M} \right] = \det \left[\underline{M}^T \right] = 1 \quad (\text{A.25})$$

Convém observar que o resultado encontrado em (A.24) é de extrema importância na implementação do teste de máxima verossimilhança apresentado no capítulo 2.

Um outro resultado interessante se obtém ao calcular a covariância cruzada entre o erro

$$\tilde{s}(m|i) = s(m) - \hat{s}(m|i) \quad (A.26)$$

e as observações. Tem-se

$$K_{\tilde{sw}}(m, \ell) = E\{\tilde{s}(m|i)w(\ell)\} = E\left\{s(m) - \sum_{j=1}^i h(m, j|i)w(j)\right\}w(\ell)$$

ou ainda

$$K_{\tilde{sw}}(m, \ell) = K_w(\ell, m) - \sum_{j=1}^i h(m, j|i)K_w(\ell, j) - K_{wa}(m, \ell) \quad (A.27)$$

Considerando (A.14), tem-se finalmente que

$$K_{\tilde{sw}}(m, \ell) = 0 \quad , \quad \ell=1, \dots, i \quad (A.28)$$

ou seja, o erro associado à estimativa $\hat{s}(m|i)$ é ortogonal às observações $w(\ell)$, $\ell=1, \dots, i$

APÊNDICE B

AUTOVALORES E AUTOFUNÇÕES

Frequentemente, ao longo do texto, são utilizadas expressões relativas à representação de um número finito de observações de um processo de parâmetro discreto na base de funções ortonormais correspondente às autofunções da matriz covariância associada ao vetor de observações. Esses resultados são bastante conhecidos^{7,8,9,16} e são aqui apresentados apenas para facilidade de referência. Eles aparecem ora em forma matricial ora na notação de funções definidas em um conjunto finito, dependendo da forma específica a ser utilizada no texto.

Considere as observações

$$w(k) \quad , \quad k=1, \dots, N \quad (B.1)$$

de um processo $w(\cdot)$ de parâmetro discreto. A função covariância das observações é dada por

$$K_w(k, \lambda) = E[w(k)w(\lambda)] \quad , \quad k=1, \dots, N ; \lambda=1, \dots, N \quad (B.2)$$

O objetivo deste apêndice é determinar uma base ortonormal de funções

$$\{\phi_i(k)\} \quad , \quad k=1,\dots,N \quad ; \quad i=1,\dots,N \quad (B.3)$$

tal que as componentes das observações do processo representadas nesta base, sejam estatisticamente independentes.

Define-se o produto escalar de duas destas funções por

$$\phi_i(k) \circ \phi_j(k) = \sum_{k=1}^N \phi_i(k) \phi_j(k) \quad (B.4)$$

A condição de ortonormalidade da base de funções requer

$$\phi_i(k) \circ \phi_j(k) = \delta(i,j) \quad (B.5)$$

com $\delta(i,j)$ definido por

$$\delta(i,j) = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} \quad (B.6)$$

As coordenadas das observações na base de funções $\{\phi_i(.)\}$ são dadas por

$$W(i) = \sum_{k=1}^N w(k) \phi_i(k) \quad , \quad i=1, \dots, N \quad (B.7)$$

Pretende-se então determinar esta base ortonormal de funções de modo a se ter

$$E[W(i)W(j)] = \lambda_i \delta(i,j) \quad (B.8)$$

Levando em conta (B.7), a condição (B.8) pode ser escrita como

$$\sum_{k=1}^N \phi_i(k) \sum_{\ell=1}^N \phi_j(\ell) K_w(k, \ell) = \lambda_i \delta(i,j) \quad , \quad i=1, \dots, N ; j=1, \dots, N \quad (B.9)$$

que é equivalente à equação

$$\sum_{\ell=1}^N \phi_j(\ell) K_w(k, \ell) = \lambda_j \phi_j(k) \quad j=1, \dots, N \quad (B.10)$$

Com efeito, admitindo-se que (B.10) é satisfeita tem - se que

$$\sum_{k=1}^N \phi_i(k) \sum_{\ell=1}^N \phi_j(\ell) K_w(k, \ell) = \lambda_j \sum_{k=1}^N \phi_j(k) \phi_i(k) \quad , \quad j=1, \dots, N ; i=1, \dots, N \quad (B.11)$$

que é obtida multiplicando-se ambos os membros de (B.10) por $\phi_i(k)$ e somando-se em relação a k . Considerando (B.5) resulta que (B.11) implica em (B.9)

Considere-se agora que (B.9) é satisfeita. Esta equação escreve também

$$\sum_{k=1}^N \phi_i(k) f_j(k) = \lambda_j(i, j) \quad , \quad i=1, \dots, N ; j=1, \dots, N \quad (B.12)$$

onde

$$f_j(k) = \sum_{\ell=1}^N \phi_j(\ell) K_w(k, \ell) \quad , \quad j=1, \dots, N \quad (B.13)$$

a equação (B.12) mostra que existe uma função $f_j(.)$ ortogonal a todas as funções $\phi_i(.)$ da base, com exceção de $\phi_j(.)$. Como as funções da base formam um conjunto completo, tal afirmativa só é verdadeira se a função $f_j(.)$ for proporcional a $\phi_j(.)$, o que é equivalente a (B.10), sendo λ_j a constante de proporcionalidade associada a $f_j(.)$. Conclui-se então que as funções da base que satisfazem a con

dição (B.8) devem satisfazer a equação (B.10). Convém observar que as constantes λ_j e as funções $\phi_j(\cdot)$ são denominadas, respectivamente, autovalores e autofunções da covariância $K_w(i,j)$ das observações do processo $w(\cdot)$ consideradas.

Considere-se agora o vetor $\underline{w}$ definido no apêndice A pela equação (A.19). Considere ainda as matrizes $\underline{K}_w$, $\underline{\phi}$ e $\underline{I}$ definidas por

$$\underline{K}_w = E \left[\underline{w} \underline{w}^T \right] = \begin{bmatrix} K_w(1,1) & K_w(1,2) & \dots & K_w(1,N) \\ K_w(2,1) & \dots & \dots & K_w(2,N) \\ \vdots & & & \\ K_w(N,1) & \dots & \dots & K_w(N,N) \end{bmatrix} \quad (\text{B.14})$$

$$\underline{\phi} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_N \end{bmatrix} \quad (\text{B.15})$$

onde

$$\phi_j = \begin{bmatrix} \phi_j(1) \\ \phi_j(2) \\ \vdots \\ \phi_j(N) \end{bmatrix}$$

e

$$\underline{L} = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ \vdots & & & & & \vdots \\ 0 & . & . & . & . & \lambda_N \end{bmatrix} \quad (\text{B.16})$$

Observe então que o conjunto de equações definido por (B.10) pode ser escrito matricialmente como

$$\underline{K}_w \underline{\phi}_j = \lambda_j \underline{\phi}_j \quad , \quad j=1, \dots, N \quad (\text{B.17})$$

ou ainda

$$\underline{K}_w \underline{\phi} = \underline{\phi} \underline{L} \quad (\text{B.18})$$

Considere agora a matriz $\underline{\phi}$ definida pela equação (B.15). A equação (B.5) permite escrever

$$\underline{\phi}^T \underline{\phi} = \underline{I} \quad (\text{B.19})$$

onde $\underline{I}$ é a matriz identidade. Como $\underline{I}^T = \underline{I}$, a partir de (B.19) tem-se

$$\underline{\phi} \underline{\phi}^T = \left[\underline{\phi}^T \underline{\phi} \right]^T = \underline{I} \quad (\text{B.20})$$

Em termos das autofunções $\phi_j(\cdot)$ da função covariância $K_w(\cdot, \cdot)$ das observações do processo $w(\cdot)$ consideradas, a equação (B.20) pode ser escrita como

$$\sum_{j=1}^N \phi_j(k) \phi_j(\ell) = \delta(k, \ell) \quad , \quad k=1, \dots, N \quad ; \quad \ell=1, \dots, N \quad (\text{B.21})$$

A partir de (B.18), e considerando (B.20), tem-se ainda

$$\underline{K}_w = \underline{\Phi} \underline{L} \underline{\Phi}^T \quad (\text{B.22})$$

que é equivalente ao conjunto de equações

$$K_w(k, \ell) = \sum_{j=1}^N \lambda_j \phi_j(k) \phi_j(\ell) \quad , \quad k=1, \dots, N \quad ; \quad \ell=1, \dots, N \quad (\text{B.23})$$

APÊNDICE C

EQUAÇÃO DE DEFINIÇÃO DE UM FILTRO ÚTIL

A implementação da razão de verossimilhança apresentada no capítulo 3 é feita de algum modo em função de um filtro linear $h_1(\cdot, \cdot, \cdot)$. O objetivo deste apêndice é então apresentar algumas equações a serem satisfeitas pelos coeficientes deste filtro.

Considere-se a mesma situação apresentada no apêndice A, ou seja

$$w(k) = s(k) + a(k) \quad , \quad k=1, \dots, i \quad (C.1)$$

onde $1 \leq i \leq N$ e

$$E[a(k)] = 0 \quad , \quad k=1, \dots, i$$

$$K_a(k, l) = E[a(k)a(l)] = \sigma_a^2 \delta(k, l) \quad , \quad k=1, \dots, i \quad ; \quad l=1, \dots, i$$

$$E[s(k)] = 0 \quad , \quad k=1, \dots, i$$

$$K_s(k, l) = E[s(k)s(l)] < \infty \quad , \quad k=1, \dots, i \quad ; \quad l=1, \dots, i$$

Os processos $s(\cdot)$ e $a(\cdot)$ não são necessariamente decorrelatados.

Define-se um filtro

$$h_1(m, j:i) \quad , \quad j=1, \dots, i \quad ; \quad i=1, \dots, N; \quad m \text{ inteiro} \quad (C.2)$$

dada por

$$h_1(m, j:i) = \sum_{k=1}^i \frac{\lambda_k(i) - \sigma_a^2}{\lambda_k(i)} \phi_k(m:i) \phi_k(j:i) \quad (C.3)$$

onde $\lambda_k(i)$ e $\phi_k(\cdot:i)$ são, respectivamente, os autovalores e as autofunções da função covariância das observações $w(k)$, $k=1, \dots, i$, do processo $w(\cdot)$

O objetivo deste apêndice é mostrar que os coeficientes $h_1(m, j:i)$ dados por (C.3) são soluções das equações

$$\sigma_a^2 \delta(\ell, m) + \sum_{j=1}^i h_1(m, j:i) K_w(\ell, j) = K_w(\ell, m), \quad \ell=1, \dots, i; \quad m=1, \dots, N; \quad (C.4)$$

$i=1, \dots, N$

Convém observar que os índices i aparecem de maneira ex
plícita nas notações dos autovalores $\lambda_k(i)$ e das autofun-
ções $\phi_k(\cdot:i)$ para que fique bem clara sua dependência com
o número de observações i .

Multiplicando-se ambos os membros da equa-
ção (C.3) por $K_w(\ell, j)$, e somando-se em relação a j , tem-se

$$\sum_{j=1}^i h_1(m, j:i) K_w(\ell, j) = \sum_{j=1}^i \sum_{k=1}^i \frac{\lambda_k(i) - \sigma_a^2}{\lambda_k(i)} \phi_k(m:i) \phi_k(j:i) K_w(\ell, j) \quad (C.5)$$

Considerando-se (B.23), tem-se

$$\sum_{j=1}^i h_1(m, j:i) K_w(\ell, j) = \sum_{j=1}^i \sum_{k=1}^i \frac{\lambda_k(i) - \sigma_a^2}{\lambda_k(i)} \phi_k(m:i) \phi_k(j:i)$$

$$= \sum_{n=1}^i \lambda_n(i) \phi_n(\ell:i) \phi_n(j:i) \quad (C.6)$$

ou ainda

$$\sum_{j=1}^i h_1(m, j:i) K_w(\ell, j) = \sum_{k=1}^i \sum_{n=1}^i \frac{\lambda_k(i) - \sigma_a^2}{\lambda_k(i)} \lambda_n(i) \phi_k(m:i) \phi_n(\ell:i)$$

$$= \sum_{j=1}^i \phi_k(j:i) \phi_n(j:i) \quad (C.7)$$

Levando em conta a equação (B.5), (C.7) pode ser escrita como

$$\sum_{j=1}^i h_1(m, j:i) K_w(\ell, j) = \sum_{k=1}^i \sum_{n=1}^i \frac{\lambda_k(i) - \sigma_a^2}{\lambda_k(i)} \lambda_n(i) \phi_k(m:i) \phi_n(\ell:i) \delta(k, n) \quad (C.8)$$

ou seja

$$\sum_{j=1}^i h_1(m, j:i) K_w(\ell, j) = \sum_{k=1}^i \left[\lambda_k(i) - \sigma_a^2 \right] \phi_k(m:i) \phi_k(\ell:i) \quad (C.9)$$

Finalmente, levando em consideração (B.21) e (B.23), chega-se à equação (C.4)

APÊNDICE D

FILTRO DE KALMAN-BUCY

Neste apêndice, são sumarizadas as equações do filtro de Kalman-Bucy para estimativas lineares de mínimo erro médio quadrático, envolvendo processos de parâmetro discreto, não estacionários e descritos por equações de estado.

Considere então o problema de estimação onde se observa um processo gerado por um modelo ARIMA (p,d,q), escrito sob a forma (2.36), ou seja

$$w(k) = s(k) + a(k) \quad (D.1)$$

onde

$$s(k) = \underline{H} \underline{x}(k) \quad (D.2)$$

sendo $\underline{x}(k)$ o vetor de estado de dimensão $n \times 1$; $\underline{H}$ uma matriz $1 \times n$; $a(k)$, $w(k)$ e $s(k)$ escalares. Sabe-se ainda que $a(\cdot)$ é um processo de ruído branco, Gaussiano, de média nula e covariância

$$K_a(k, \ell) = \sigma_a^2 \delta(k, \ell) \quad , \quad \sigma_a^2 > 0 \quad (D.3)$$

Por outro lado, de acordo com a equação (2.45)

$$K_{sa}(k, \ell) = 0 \quad \text{se} \quad k \leq \ell \quad (\text{D.4})$$

Seja $\underline{x}(\cdot)$ um processo descrito por uma equação de estado da forma (2.37), isto é

$$\underline{x}(k+1) = \underline{\phi} \underline{x}(k) + \underline{\Gamma} a(k) \quad (\text{D.5})$$

onde $\underline{\phi}$ é uma matriz $n \times n$, $\underline{\Gamma}$ é uma matriz $n \times 1$ e $a(k)$ é o mesmo ruído branco que aparece em (D.1)

A partir das equações (D.1) e (D.5), pode-se obter um modelo para o processo, conforme apresentado na figura D.1.

Figura D.1 - Modelo gerador do processo $w(\cdot)$

Define-se $\underline{P}(k|k)$ como

$$\underline{P}(k|k) = E \left\{ \left[\underline{x}(k) - \hat{\underline{x}}(k|k) \right] \left[\underline{x}(k) - \hat{\underline{x}}(k|k) \right]^T \right\} \quad (D.6)$$

ou seja, $\underline{P}(k|k)$ é a matriz covariância do erro cometido na estimação $\hat{\underline{x}}(k|k)$ do vetor de estado $\underline{x}(\cdot)$ no instante k a partir das observações $w(i)$, $i=1, \dots, k$.

As equações do filtro de Kalman-Bucy são então apresentadas a seguir:

- Equações do algoritmo

$$\hat{\underline{x}}(k+1|k+1) = \hat{\underline{x}}(k+1|k) + \underline{K}(k+1) \left[w(k+1) - \underline{H} \hat{\underline{x}}(k+1|k) \right] \quad (D.7)$$

e

$$\hat{\underline{x}}(k+1|k) = \underline{\Phi} \hat{\underline{x}}(k|k) + \underline{\Gamma} \left[w(k) - \underline{H} \hat{\underline{x}}(k|k) \right] \quad (D.8)$$

- Equação do ganho

$$\underline{K}(k+1) = \underline{P}(k+1|k) \underline{H}^T \left[\sigma_a^2 - \underline{H} \underline{P}(k+1|k) \underline{H}^T \right]^{-1} \quad (D.9)$$

- Equação da matriz covariância do erro, a priori

$$\begin{aligned} \underline{P}(k+1|k) &= E \left\{ \left[\underline{x}(k+1) - \hat{\underline{x}}(k+1|k) \right] \left[\underline{x}(k+1) - \hat{\underline{x}}(k+1|k) \right]^T \right\} = \\ &= \left[\underline{\Phi} - \underline{\Gamma} \underline{H} \right] \underline{P}(k|k) \left[\underline{\Phi} - \underline{\Gamma} \underline{H} \right]^T \end{aligned} \quad (D.10)$$

- Equação da matriz covariância do erro, a posteriori

$$\underline{P}(k+1|k+1) = \left[\underline{I} - \underline{K}(k+1)\underline{H} \right] \underline{P}(k+1|k) \quad (D.11)$$

Obviamente, a implementação do algoritmo de Kalman-Bucy fica condicionada à fixação de duas condições iniciais:

$$\hat{\underline{x}}(0|0)$$

e

$$(D.12)$$

$$\underline{P}(0|0)$$

Considerando a estimativa $\hat{s}(k+1|k)$ do processo $s(\cdot)$, feita através da equação

$$\hat{s}(k+1|k) = \underline{H} \hat{\underline{x}}(k+1|k) \quad (D.13)$$

pode-se definir um processo de inovações $v(\cdot)$ ^{12,14,15}, associado ao processo $w(\cdot)$. Desse modo tem-se

$$v(k) = w(k) - \hat{s}(k|k-1) \quad (D.14)$$

Observe-se que o processo de inovações definido desta maneira deve satisfazer às equações (A.23), (A.24) e (2.55). Isto significa que a função covariância do processo de inovações definida por (2.55) pode ser obtida através do algoritmo de Kalman-Bucy, uma vez que a partir de (D.10), (D.13), (2.49) e (2.56) tem-se

$$P_s(k|k-1) = \underline{H} \underline{P}(k|k-1) \underline{H}^T \quad (D.15)$$

ou seja

$$K_v(k, \ell) = \begin{cases} 0 & , \quad k > \ell \\ E[v^2(k)] = \sigma_a^2 + \underline{H} \underline{P}(k|k-1) \underline{H}^T, & k = \ell \\ 0 & , \quad k < \ell \end{cases} \quad (D.16)$$

REFERÊNCIAS

- 1 - G.E.P. Box and G.M. Jenkins, "Time Series Analysis , Forecasting and Control", Holden Day, San Francisco, 1970
- 2 - G.E.P. Box and D.A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", Jour. Amer. Stat. Assoc. 64, 1970
- 3 - R. Fletcher and C.M. Reeves, "Function Minimization by Conjugate Gradients", Computer Journal, vol. 7 , nº 2, 1964
- 4 - G.M. Jenkins and D.G. Watts, "Spectral Analysis and Its Applications", Holden Day, San Francisco, 1968
- 5 - J.M.P. Fortes e J.P.A. Albuquerque, "Identificação de Séries Temporais - Parte I", CETUC D-SC 201/75, 1975
- 6 - F. Schweppe, "Evaluation of Likelihood Functions for Gaussian Signals", IEEE Trans. IT-11, nº1, Jan. 1965
- 7 - H.L. Van Trees, "Detection, Estimation and Modulation Theory - Part I", John Wiley, New York, 1968
- 8 - H.L. Van Trees, "Detection, Estimation and Modulation Theory - Part III", John Wiley, New York, 1971
- 9 - F.C. Schweppe, "Uncertain Dynamic Systems", Prentice-Hall, New Jersey, 1975
- 10 - A.E. Bryson Jr. and Y.C. Ho, "Applied Optimal Control", Ginn and Company, Massachusetts, 1969
- 11 - A.P. Sage, "Optimum Systems Control", Prentice-Hall , New Jersey, 1968

- 12 - M.R. Gevers and T. Kailath, "An Innovations Approach to Least-Squares Estimation - Part IV: Discrete-Time Innovations Representations and Recursive Estimation", IEEE Trans. on Aut. Control, vol. AC-18, n°6, Dec. 1973
- 13 - T. Kailath, "A View of Three Decades of Linear Filtering Theory", IEEE Trans. on Inf. Theory, vol. IT-20, n°2, March 1974
- 14 - H.B. Aasnaes and T. Kailath, "An Innovations Approach to Least-Squares Estimation - Part VII: Some Applications of Vector Autoregressive-Moving Average Models", IEEE Trans. on Aut. Control, vol AC-18, n°6, Dec. 1973
- 15 - A.C.G. Vicira, "Método de Inovações Aplicado à Estimacão de Processos Gauss-Markov", Tese de Mestrado, PUC/RJ, Maio 1973
- 16 - C.T. Chen, "Introduction to Linear System Theory", Holt Rinehart and Winston Inc., 1970

Tese apresentada aos Senhores

José Paulo de Almeida e Albuquerque
José Paulo de Almeida e Albuquerque

João Célio Barros Brandão
João Célio Barros Brandão

Roberto Pereira d'Araújo
Roberto Pereira d'Araújo

Visto e permitida a impressão

Rio de Janeiro, 16/07/76.

Javier P. Soares
Coordenador dos Programas de Pós-Graduação
e Pesquisas do Centro Técnico Científico