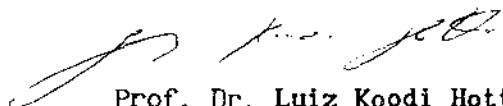


10 ABR 1990

AGREGAÇÃO TEMPORAL DE VARIÁVEL FLUXO EM MODELOS ARIMA

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. José Cardoso Neto e aprovada pela Comissão Julgadora.

Campinas, 9 de fevereiro de 1990.



Prof. Dr. Luiz Koodi Hotta

Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Estatística.

C179a

11805/BC

IMECC  
BIBLIOTECA CENTRAL

Para minha esposa

**Líbia**

e para meus filhos

**Cardosinho e**

**Flavinho.**

## AGRADECIMENTOS

À senhora Julia Martins Rego (MAINHA) que, cuidando dos meus filhos como se fossem seus, me deu tranqüillidade para realizar este trabalho.

Ao professor Dr. Luiz Koodi Hotta pela segura orientação e ensinamentos, mesmo nos encontros informais.

Aos professores Drs. Pedro Luiz Valls Pereira e Clélia Maria de Castro Tolói pelas sugestões e cujo contato poderia ter sido melhor aproveitado por mim.

Aos professores e funcionários do Departamento de Estatística do IMECC/UNICAMP e em particular aos professores Drs. José Antonio Cordeiro e Mauro Sérgio de Freitas Marques pelas discussões e sugestões.

Aos professores do Departamento de Estatística e Computação do ICE/FUA pelo apoio e incentivo, em particular aos professores Luiz Irapuan Pinheiro e Rosana Cristina Pereira Parente pelo apoio logístico.

Aos colegas da Pós-graduação do IMECC/UNICAMP pelo agradável convívio e momentos de descontração.

Aos meus pais e sogros que apesar da distância sempre estiveram do meu lado.

## ÍNDICE

I - INTRODUÇÃO	1
CAPÍTULO 1 - MODELOS ARIMA UNIVARIADOS	
1.1 - Introdução	5
1.2 - O modelo Linear Geral	11
1.3 - Modelos Autorregressivo	14
1.4 - Modelos Média Móvel	18
1.5 - Modelos Autorregressivo Média Móvel	23
1.6 - Modelos ARIMA	27
1.7 - Modelos Sazonais	29
1.8 - Previsão Baseada na Representação MA	32
1.9 - Previsão Baseada na Representação AR	36
1.10 - Previsão Baseada na Representação ARMA	41
1.11 - Critérios para Escolha do Modelo	42
CAPÍTULO 2 - O EFEITO DA AGREGAÇÃO SOBRE O MODELO	
2.1 - Introdução	45
2.2 - O Efeito sobre o Modelo MA	46
2.3 - O Efeito sobre o Modelo AR	50
2.4 - O Efeito sobre o Modelo ARMA	55
2.5 - O Efeito sobre Modelos Não Estacionários	56

## CAPÍTULO 3 - O EFEITO DA AGREGAÇÃO SOBRE A PREVISÃO

3.1 - Introdução	61
3.2 - Previsão com Modelo Conhecido	62
3.3 - Previsão com Processo ARMA com Coeficiente Estimado	68
3.3.1 - Processo com Média Zero	69
3.3.2 - Processo com Média Diferente de Zero	73
3.3.3 - Os EMQP Assintóticos dos Previsores	74
3.4 - Previsão com Processo AR de Ordem Desconhecida	77
3.5 - Previsão com Processo Não Estacionário	84

## CAPÍTULO 4 - ESTUDO DE MONTE CARLO

4.1 - Introdução	89
4.2 - Processo Autorregressivo	92
4.3 - Processo Média Móvel	95
4.4 - Processo Autorregressivo Média Móvel	97
4.5 - Conclusão	100

REFERÊNCIAS	101
-------------	-----

## INTRODUÇÃO

O problema de agregação em séries temporais é de considerável interesse, principalmente em economia. Por exemplo, suponha que o responsável por uma firma, com várias filiais, deseja estocar um certo produto para ser vendido na semana seguinte.

Considerando o espaço de armazenagem e o risco de faltar o produto no estoque, ele precisa da demanda do produto em todas as filiais para a próxima semana.

Na realidade, ele está interessado na previsão de uma variável que é agregada temporalmente, somando-se as demandas diárias durante a semana, e agregada contemporaneamente, somando-se as demandas semanais das filiais.

Nesse caso, a agregação temporal e contemporânea aparecem conjuntamente. Alguns autores como Lütkepohl (1986), tratam deste problema.

No caso de agregação temporal, tem-se dois tipos de variáveis: estoque e fluxo. Uma variável estoque, é a quantidade de alguma coisa em um particular ponto do tempo, enquanto uma variável fluxo, é a quantidade acumulada sobre um período de tempo.

Por exemplo, renda anual é uma variável fluxo, pois ela é a soma das rendas mensais durante um ano. A quantidade de certo produto em estoque, é uma variável estoque. Esta variável é algumas vezes

chamada de amostragem sistemática.

A agregação temporal de modelos ARIMA é bastante discutida na literatura estatística e muitos trabalhos apresentam estudos sobre estimação e previsão de modelos agregados temporalmente.

O modelo resultante da agregação temporal de variável fluxo, conhecendo-se o modelo da série original, é apresentado por Amemiya e Wu (1972), para o caso de modelos AR. Tiao (1972), obteve o modelo resultante da agregação temporal de variável fluxo no caso da série original ser um modelo IMA.

Brewer (1973), apresenta resultados para variáveis fluxo e estoque, no caso da série original ser um modelo ARMA. Para o caso de variável fluxo, Wei (1979) estendeu estes resultados para o caso da série original ser um modelo ARIMA sazonal multiplicativo.

Granger e Morris (1976), obtiveram importantes resultados para o caso de soma de processos ARMA independentes, por exemplo, se  $X_t$  e  $Z_t$  são duas séries estacionárias com média zero e independentes tais que  $X_t$  é  $ARMA(p_1, q_1)$  e  $Z_t$  é  $ARMA(p_2, q_2)$ , então  $Y_t = X_t + Z_t$  é  $ARMA(p, q)$ , com  $p \leq p_1 + p_2$  e  $q \leq \max(p_1 + q_2, p_2 + q_1)$ . Rose (1977), apresenta resultados para soma de processos ARIMA independentes.

Outras referências sobre o estudo de agregação temporal são: Tiao e Wei (1976), Abraham (1982), Abraham e Ledolter (1982), Kohn (1982), Engel (1984), Stram e Wei (1986), Pino, Morettin e Mentz (1987), entre outros.

A agregação temporal está relacionada com o problema de observações perdidas quando a série temporal é observada em dois períodos: um com dados na forma agregada e outro com dados na forma

desagregada.

Harvey e Pierse (1984), estudaram este caso usando o enfoque de espaço de estados.

O caso de agregação contemporânea de séries temporais foi considerado por Wei e Abraham (1981) usando o enfoque de espaço de Hilbert.

Neste trabalho, será estudada somente agregação temporal de variável fluxo, sendo analisada a relação dos modelos e comparados os previsores da variável agregada temporalmente.

A previsão feita com base nos dados desagregados, será comparada com a previsão baseada nos dados agregados, em termos de erro médio quadrático de previsão EMQP, para diferentes situações.

No capítulo 1, será apresentado um resumo da modelagem Box-Jenkins para séries temporais univariadas. No capítulo 2, será apresentado o efeito da agregação temporal sobre os vários tipos de modelos, considerando que os dados desagregados seguem um modelo totalmente conhecido.

No capítulo 3, serão comparados dois previsores nos seguintes casos: 1) quando os dados desagregados seguem um modelo totalmente conhecido; 2) quando os dados desagregados seguem um modelo  $ARMA(p,q)$  com as ordens  $p$  e  $q$  finitas e conhecidas e os coeficientes são estimados e; 3) quando ajustamos modelos autorregressivos de ordem desconhecida para os dados agregados e para os dados desagregados.

Nos dois últimos casos, os erros médios quadráticos de previsão são obtidos com base na teoria assintótica e no capítulo 4 será estudado o comportamento desses previsores em pequenas amostras



através de experimentos de Monte Carlo para modelos  $AR(p)$ ,  $MA(q)$  e  $ARMA(p,q)$  e apresentadas algumas conclusões.

## CAPÍTULO 1

### MODELOS ARIMA UNIVARIADOS

#### 1.1 - INTRODUÇÃO

Uma importante classe de modelos estocásticos usada para descrever séries temporais, são os chamados modelos autorregressivo - integrado - média móvel, ou abreviadamente modelos ARIMA.

Esta extensa classe de modelos, popularizada por Box & Jenkins (1976), produz uma variedade de modelos estacionários e não estacionários, que representam adequadamente muitas séries temporais encontradas na prática.

Neste capítulo serão apresentados os modelos desta classe com especial atenção para os modelos estacionários. Também serão apresentados alguns resultados, teoremas e definições necessários para o desenvolvimento desse e dos próximos capítulos.

**Série Temporal.** Uma série temporal é qualquer conjunto de observações geradas sequencialmente no tempo. Se o conjunto é discreto, a série temporal é dita discreta, se o conjunto é contínuo, a série temporal é dita contínua.

Serão consideradas somente as séries temporais discretas, medidas a intervalos iguais de tempo, denotando as  $T$  observações da série por  $y_1, y_2, \dots, y_T$ .

**Processo Estocástico.** Um fenômeno estatístico que evolui no tempo de acordo com leis probabilísticas é chamado um processo estocástico.

Precisamente, um processo estocástico é uma família  $Y = \{Y(t), t \in T\}$ , tal que para cada  $t \in T$ ,  $Y(t)$  é uma variável aleatória, onde  $T$  é um conjunto de índices.

Se  $T$  é finito ou enumerável, o processo diz-se com parâmetro discreto. Se  $T$  é um intervalo, o processo diz-se com parâmetro contínuo.

A função média do processo  $Y(t)$  é definida por

$$(1.1.1) \quad \mu(t) = E[Y(t)], \quad t \in T$$

e a função de autocovariância é definida por

$$(1.1.2) \quad \gamma(t,s) = \text{Cov}[Y(t), Y(s)], \quad t, s \in T.$$

**Processo Estocástico Estritamente Estacionário.** De maneira intuitiva, um processo estocástico estacionário evolui no tempo de tal modo que a escolha da origem dos tempos não é importante para a distribuição das observações.

Um processo estocástico é dito ser estritamente estacionário, se a distribuição de probabilidade conjunta das observações  $y_1, y_2, \dots, y_T$  é a mesma das observações  $y_{1+k}, y_{2+k}, \dots, y_{T+k}$ , para todo inteiro  $k$  e  $T \geq 1$ .

**Processo Estocástico Fracamente Estacionário.** Um processo estocástico é dito ser fracamente estacionário (ou estacionário de segunda ordem), se e somente se, as seguintes condições são satisfeitas:

- i)  $\mu(t) = \mu$  , constante para todo  $t \in T$ .
- ii)  $E[Y(t)^2] < \infty$  , para todo  $t \in T$ .
- iii)  $\gamma(t,s) = f(t-s)$  , é uma função somente de  $t-s$ .

Serão tratados aqui, somente processos fracamente estacionários que serão denominados de processos estacionários.

Um processo estocástico é dito ser Gaussiano, se as observações  $y_1, y_2, \dots, y_T$ , têm distribuição normal T-variada.

Se um processo for Gaussiano e fracamente estacionário, ele será estritamente estacionário.

Uma série temporal, é considerada como sendo uma possível realização de um processo estocástico.

**Funções de Autocovariância e Autocorrelação.** Suponha que  $y_t$  é um processo estocástico estacionário com média  $\mu = E(y_t)$  e variância finita  $\sigma^2 = V(y_t)$ .

A função de autocovariância no lag  $k$ , produz a covariância entre valores separados por um intervalo de tempo  $k$ , ou seja, entre  $y_t$  e  $y_{t+k}$  e é definida por

$$(1.1.3) \quad \gamma(k) = \text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)].$$

Como

$$\gamma(-k) = \text{Cov}(y_t, y_{t-k}) = \text{Cov}(y_{t-k}, y_t) = \text{Cov}(y_t, y_{t+k}) = \gamma(k),$$

isto mostra que esta função é simétrica em torno do zero.

Para  $k = 0$ ,

$$\gamma(0) = \text{Var}(y_t) = \sigma^2.$$

Também pode-se mostrar que  $|\gamma(k)| \leq \gamma(0)$  e que a função de autocovariância  $\gamma$ , é positiva definida.

A função de autocorrelação no lag  $k$ , produz a correlação entre  $y_t$  e  $y_{t+k}$  e é definida por

$$(1.1.4) \quad \rho(k) = \frac{\text{Cov}(y_t, y_{t+k})}{[\text{Var}(y_t)\text{Var}(y_{t+k})]^{1/2}} = \frac{\gamma(k)}{\gamma(0)}.$$

É claro que  $\rho(k)$  também é simétrica em torno do zero e  $\rho(0) = 1$ .

Se uma série temporal  $y_1, y_2, \dots, y_T$ , é observada, então a função de autocovariância pode ser estimada por

$$(1.1.5) \quad \hat{\gamma}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 0, 1, \dots, T-1, \text{ onde}$$

$$(1.1.6) \quad \bar{y} = \hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$$

é a média da serie temporal observada.

O fato de usar-se  $T$  em vez de  $T - k$  no denominador de (1.1.5), é que  $\hat{\gamma}(k)$  fornece estimativas positivas definidas, que é uma importante propriedade da função de autocovariância.

A função de autocorrelação é estimada por

$$(1.1.7) \quad \hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \quad k = 0, 1, \dots, T-1.$$

$\hat{\gamma}(k)$  é chamada função de autocovariância amostral e  $\hat{\rho}(k)$  função de autocorrelação amostral.

Define-se também

$$(1.1.8) \quad \hat{\gamma}(k) = \hat{\gamma}(-k) \quad \text{e} \quad \hat{\rho}(k) = \hat{\rho}(-k), \quad k = 0, 1, \dots, T-1.$$

Quando é observada uma única realização de um processo estocástico, a atenção volta-se de um ponto particular do tempo para as observações no tempo.

Isto é possível se a média, variância e função de autocovariância são independentes do tempo, ou seja, se o processo de geração dos dados é estacionário. Assim a média do processo pode ser estimada pela média amostral das observações como em (1.1.6).

O fato de obter-se estimadores consistentes de uma única realização de processos estacionários, é devido aos chamados teoremas ergódicos, os quais mostram que para muitos processos estacionários, que são comumente encontrados na prática, os momentos amostrais de uma realização observada de comprimento  $T$ , converge (em média quadrática)

para os correspondentes momentos populacionais (veja por exemplo, Chatfield 1980, página 65). Estes teoremas valem para modelos ARMA.

**Operadores.** Alguns operadores facilitam a notação e a maneira de representar séries temporais.

O operador atraso  $B$ , é definido por

$$By_t = y_{t-1}, \quad B^k y_t = y_{t-k}.$$

O operador avanço  $F$ , é o operador inverso de  $B$  e é definido

$$\text{por } Fy_t = y_{t+1}, \quad F^k y_t = y_{t+k}.$$

O operador diferença  $\nabla$ , é definido por

$$\nabla y_t = y_t - y_{t-1} = (1-B)y_t, \quad \text{o que implica que } \nabla = (1-B).$$

O operador inverso da diferença é o operador soma definido

$$\text{por } Sy_t = \sum_{i=0}^{\infty} y_{t-i}. \quad \text{Tem-se que}$$

$$Sy_t = (y_t + y_{t-1} + y_{t-2} + \dots) = (1+B+B^2 + \dots)y_t = (1-B)^{-1}y_t = \nabla^{-1}y_t,$$

$$\text{ou seja, } S = \nabla^{-1}.$$

**Ruído Branco.** Ruído branco é uma sequência de variáveis aleatórias não correlacionadas com média zero e variância constante.

Essa sequência será denotada por  $a_t$  de modo que  $E(a_t) = 0$ ,  $\text{Var}(a_t) = \sigma_a^2$  para todo  $t$  e  $E(a_t a_s) = 0$  para  $t \neq s$ .

Embora esta definição seja adotada, alguns autores definem os  $a_t$  como sendo uma sequência de variáveis aleatórias independentes e identicamente distribuídas, usualmente com distribuição normal.

A função de autocovariância do ruído branco é dada por

$$(1.1.9) \quad \gamma(k) = \begin{cases} \sigma_a^2 & \text{se } k = 0 \\ 0 & \text{se } k \neq 0 \end{cases}$$

e a função de autocorrelação é dada por

$$(1.1.10) \quad \rho(k) = \begin{cases} 1 & \text{se } k = 0 \\ 0 & \text{se } k \neq 0 \end{cases}.$$

## 1.2 - O MODELO LINEAR GERAL.

Suponha que  $y_t$  é um processo estocástico com média  $\mu = 0$ .

O modelo linear geral, é escrito como

$$y_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots, \text{ ou}$$

$$(1.2.1) \quad y_t = \theta(B)a_t, \text{ onde}$$

$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots$  e  $a_t$  é ruído branco.



A forma (1.2.1) expressa  $y_t$  como uma função linear dos ruídos presente e passados. Alternativamente  $y_t$  pode ser escrito em termos do ruído presente e todos os  $y_t$  passados, ou seja,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + a_t, \text{ ou}$$

$$(1.2.2) \quad \phi(B)y_t = a_t, \text{ onde}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots.$$

As duas formas de expressar  $y_t$ , (1.2.1) e (1.2.2), implicam que  $\theta(B) = \phi(B)^{-1}$ , isto é, os pesos  $\theta_1$  podem ser obtidos como função dos  $\phi_1$  e vice-versa.

A função de autocovariância do modelo linear geral é dada por

$$(1.2.3) \quad \gamma(k) = \sigma_a^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+k}, \text{ com } \theta_0 = 1,$$

em particular, para  $k = 0$  tem-se

$$(1.2.4) \quad \gamma(0) = \text{Var}(y_t) = \sigma_a^2 \sum_{i=0}^{\infty} \theta_i^2, \text{ com } \theta_0 = 1.$$

**Condições de Estacionaridade e Invertibilidade.** A convergência da série em (1.2.4) garante que o processo tem variância finita. Como a média é constante e (1.2.3) não depende do tempo, o modelo será estacionário se  $\sum_{i=0}^{\infty} \theta_i^2 < \infty$  e  $\sum_{i=0}^{\infty} \theta_i \theta_{i+k} < \infty$ .

Para o modelo linear geral, estas condições podem ser

substituídas por uma única condição, a de que a série  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots$  deve convergir para  $|B| \leq 1$  (Box e Jenkins 1976, seção 3.1).

Antes de enunciar as condições de invertibilidade, será apresentado um exemplo.

**EXEMPLO 1.2.1:** Considere o modelo linear,

$$(1.2.5) \quad y_t = a_t - \theta a_{t-1},$$

como  $\sum_{i=0}^{\infty} \theta^2 = 1 + \theta^2$ , é finita, tem-se que (1.2.5) é estacionário para qualquer valor de  $\theta$ .

Escrevendo  $y_t$  na forma (1.2.2), vem

$$(1.2.6) \quad y_t = \theta y_{t-1} + \theta^2 y_{t-2} + \dots + a_t,$$

portanto,

$$\phi(B) = 1 + \theta B + \theta^2 B^2 + \dots = \sum_{i=0}^{\infty} \theta^i B^i \text{ e } \phi_i = \theta^i, \text{ para } i \geq 1.$$

A sequência formada pelos pesos  $\phi_i$  será convergente se  $|\theta| < 1$  e neste caso diz-se que o modelo é invertível.

O modelo linear geral será invertível se a série  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots$  converge para  $|B| \leq 1$ .

### 1.3 - MODELOS AUTORREGRESSIVOS.

As propriedades do modelo geral  $AR(p)$ , serão indicadas a partir da análise dos casos particulares  $AR(1)$  e  $AR(2)$ .

**Modelo  $AR(1)$ .** Este modelo é da forma

$$(1.3.1) \quad y_t = \phi y_{t-1} + a_t .$$

Comparando com (1.2.2), tem-se que  $\phi(B) = 1 - \phi B$ , é finito, portanto este modelo é invertível para qualquer valor de  $\phi$ .

A condição para que o modelo seja estacionário, é que

$$\sum_{i=0}^{\infty} \theta_i = \sum_{i=0}^{\infty} \phi^i < \infty , \text{ o que é equivalente a } |\phi| < 1 .$$

Portanto, a condição de estacionaridade é equivalente a  $-1 < \phi < 1$ . Em termos do polinômio  $\phi(B)$ , isto é equivalente a  $1 - \phi B = 0$ , ter raiz  $B = \frac{1}{\phi}$  maior do que 1 em valor absoluto.

Escrevendo (1.3.1) na forma (1.2.1), e calculando a variância de  $y_t$ , obtem-se

$$\gamma(0) = E(y_t^2) = \sigma_a^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{1}{1 - \phi^2} \sigma_a^2 .$$

Para  $k \geq 1$ , tem-se que  $y_t$  e  $a_{t-k}$  são não correlacionados, portanto,

$$\gamma(k) = E(y_t y_{t-k}) = \phi E(y_{t-1} y_{t-k}) + E(a_t y_{t-k}), \text{ ou}$$

$$(1.3.2) \quad \gamma(k) = \phi \gamma(k-1), \text{ para } k \geq 1 .$$

A solução da equação de diferença de primeira ordem (1.3.2), é a função de autocovariância

$$(1.3.3) \quad \gamma(k) = \phi^k \gamma(0), \text{ para } k \geq 1 .$$

Portanto a função de autocorrelação é dada por

$$(1.3.4) \quad \rho(k) = \phi^k, \text{ para } k \geq 1 .$$

Note que  $\rho(0) = 1$  e  $\rho(1) = \phi$ . A função de autocorrelação decai exponencialmente para zero se  $\phi$  é positivo. Se  $\phi$  é negativo, a função decai exponencialmente para zero mas, oscilando.

**Modelo AR(2).** Este modelo é da forma

$$(1.3.5) \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t ,$$

como  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$  é finito, este modelo é sempre invertível.

Em termos do polinômio  $\phi(B)$ , a condição de estacionaridade é equivalente a exigir que as raízes de  $\phi(B) = 0$  estejam fora do círculo unitário, o que é equivalente a

$$\phi_1 + \phi_2 < 1 \quad , \quad \phi_2 - \phi_1 < 1 \quad \text{e} \quad -1 < \phi_2 < 1 .$$

Multiplicando (1.3.5) por  $y_{t-k}$  e tomando-se a esperança, obtem-se

$$(1.3.6) \quad E(y_t y_{t-k}) = \phi_1 E(y_{t-1} y_{t-k}) + \phi_2 E(y_{t-2} y_{t-k}) + E(a_t y_{t-k}) .$$

Como  $E(a_t y_{t-k}) = 0$  para  $k \geq 1$ , a função de autocovariância satisfaz a equação de diferença

$$(1.3.7) \quad \gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2), \text{ para } k \geq 1 .$$

Portanto, a função de autocorrelação satisfaz a equação de diferença

$$(1.3.8) \quad \rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \text{ para } k \geq 1 ,$$

que é conhecida como equações de Yule-Walker.

Pode-se escrever (1.3.8) como

$$0 = \phi(B)\rho(k) = (1 - \phi_1 B - \phi_2 B^2)\rho(k) ,$$

onde B opera sobre k, cuja solução depende das raízes de  $\phi(B) = 0$  que será portanto, soma de exponenciais e/ou senóides amortecidas.

Também de (1.3.8) tem-se a seguinte relação

$$\rho(1) = \frac{\phi_1}{1 - \phi_2} \quad \text{e} \quad \rho(2) = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 .$$

Fazendo  $k = 0$  em (1.3.6) obtem-se,

$$\text{Var}(y_t) = \gamma(0) = \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + \sigma_a^2,$$

e como  $\gamma(-k) = \gamma(k)$ , tem-se que

$$\gamma(0) = \frac{\sigma_a^2}{1 - \phi_1 \rho(1) - \phi_2 \rho(2)}, \text{ ou}$$

$$(1.3.9) \quad \gamma(0) = \frac{(1 - \phi_2) \sigma_a^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}.$$

**Modelo AR(p).** Este modelo é da forma

$$(1.3.10) \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t,$$

que também é sempre invertível, uma vez que  $\phi(B)$  é finito.

O modelo (1.3.10) pode ser escrito como

$$(1.3.11) \quad \phi(B)y_t = a_t, \text{ onde}$$

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

A condição de estacionaridade para este modelo é equivalente a exigir que as raízes de  $\phi(B) = 0$  estejam fora do círculo unitário.

A função de autocovariância satisfaz a seguinte equação de diferenças

$$(1.3.12) \quad \gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \dots + \phi_p \gamma(k-p), \text{ para } k \geq 1.$$

Portanto, a função de autocorrelação satisfaz as equações de Yule-Walker

$$(1.3.13) \quad \rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \dots + \phi_p \rho(k-p), \text{ para } k \geq 1.$$

A variância de  $y_t$  é dada por

$$(1.3.14) \quad \gamma(0) = \frac{\sigma_a^2}{1 - \phi_1 \rho(k-1) - \phi_2 \rho(k-2) - \dots - \phi_p \rho(k-p)}.$$

De maneira geral, a função de autocorrelação de um processo autorregressivo é uma mistura de polinômios, exponenciais e senóides amortecidas.

#### 1.4 - MODELOS MÉDIA MÓVEL.

Ao contrário dos modelos autorregressivos, estes modelos são sempre estacionários, porém algumas restrições aos parâmetros devem ser impostas para que o modelo seja invertível.

As propriedades do modelo geral  $MA(q)$ , serão indicadas a partir da análise dos casos particulares  $MA(1)$  e  $MA(2)$ .

**Modelo  $MA(1)$ .** Este modelo é da forma

$$(1.4.1) \quad y_t = a_t - \theta a_{t-1}.$$

Como  $\theta(B) = 1 - \theta$  é finito, este modelo é estacionário para qualquer valor de  $\theta$ .

A variância de  $y_t$  é dada por

$$(1.4.2) \quad \gamma(0) = (1 + \theta^2) \sigma_a^2.$$

Como

$$E[(a_t - \theta a_{t-1})(a_{t-k} - \theta a_{t-k-1})] = 0 \quad \text{para } k \geq 2,$$

tem-se que,  $\gamma(1) = -\theta \sigma_a^2$  e a função de autocorrelação é dada por

$$(1.4.3) \quad \rho(k) = \begin{cases} 1 & \text{se } k = 0 \\ -\theta & \text{se } k = 1 \\ \frac{-\theta}{1+\theta^2} & \text{se } k = 1 \\ 0 & \text{se } k \geq 2. \end{cases}$$

Para que o modelo seja invertível, é necessário que

$$\sum_{i=0}^{\infty} \phi_i = \sum_{i=0}^{\infty} \theta^i < \infty, \text{ o que é equivalente a } |\theta| < 1.$$

Esta condição é equivalente a  $-1 < \theta < 1$ . Em termos do polinômio  $\theta(B)$ , isto é equivalente a exigir que  $1 - \theta B = 0$ , tenha raiz  $B = \frac{1}{\theta}$  maior que 1 em valor absoluto.



**Modelo MA(2).** Este modelo é da forma

$$(1.4.4) \quad y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} .$$

A variância de  $y_t$  é dada por

$$(1.4.5) \quad \gamma(0) = E(y_t^2) = (1 + \theta_1^2 + \theta_2^2) \sigma_a^2 .$$

Como

$$E(y_t y_{t-k}) = E[(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})(a_{t-k} - \theta_1 a_{t-k-1} - \theta_2 a_{t-k-2})] = 0 ,$$

para  $k \geq 3$ , tem-se que

$$\gamma(1) = -\theta_1(1 - \theta_2) \sigma_a^2 , \quad \gamma(2) = -\theta_2 \sigma_a^2 \quad \text{e a função de autocorrelação é}$$

dada por

$$(1.4.6) \quad \rho(k) = \begin{cases} 1 & \text{se } k = 0 \\ \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} & \text{se } k = 1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & \text{se } k = 2 \\ 0 & \text{se } k \geq 3 \end{cases}$$

Em termos do polinômio  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$ , a condição de

invertibilidade é equivalente a exigir que as raízes de  $\theta(B) = 0$  estejam fora do círculo unitário, o que é equivalente a

$$\theta_2 + \theta_1 < 1, \quad \theta_2 - \theta_1 < 1 \quad \text{e} \quad -1 < \theta_2 < 1.$$

**Modelo MA(q).** Este modelo é da forma

$$(1.4.7) \quad y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q},$$

que também é sempre estacionário uma vez que  $\theta(B)$  é finito.

A variância de  $y_t$  é dada por

$$(1.4.8) \quad \gamma(0) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \sigma^2.$$

Como  $E(y_t y_{t-k}) = 0$  para  $k > q$ , a função de autocovariância satisfaz

$$(1.4.9) \quad \gamma(k) = \begin{cases} -\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_{q-k} \theta_q, & \text{se } k = 1, 2, \dots, q \\ 0 & \text{se } k > q \end{cases}.$$

Portanto a função de autocorrelação é dada por

$$(1.3.10) \quad \rho(k) = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & \text{se } k = 1, 2, \dots, q \\ 0 & \text{se } k > q \end{cases}.$$

A condição de invertibilidade em termos do polinômio  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ , é equivalente a exigir que as raízes de  $\theta(B) = 0$  estejam fora do círculo unitário.

**Invertibilidade e Identificabilidade.** Considere o modelo MA(1)

$$(1.4.11) \quad y_t = a_t - \theta a_{t-1} .$$

este modelo pode ser escrito como um modelo AR( $\infty$ )

$$(1.4.12) \quad y_t = \theta y_{t-1} + \theta^2 y_{t-2} + \dots + a_t .$$

Para que  $y_t$  não dependa do seu passado muito distante, é necessário que  $|\theta| < 1$  e neste caso o modelo (1.4.11) é invertível.

Agora seja  $\theta' = 1/\theta$  ( $|\theta'| > 1$ ) e considere o modelo MA(1)

$$(1.4.13) \quad y_t = a_t - \theta' a_{t-1} ,$$

que é estacionário e pode ser escrito na forma autorregressiva

$$(1.4.14) \quad y_t = \theta' y_{t-1} + \theta'^2 y_{t-2} + \dots + a_t .$$

A função de autocorrelação do modelo (1.4.13) é dada por

$$(1.4.15) \quad \rho(k) = \begin{cases} 1 & \text{se } k = 0 \\ \frac{-\theta'}{1+\theta'^2} = \frac{-\theta}{1+\theta^2} & \text{se } k = 1 \\ 0 & \text{se } k \geq 2 \end{cases},$$

que é exatamente a mesma para o modelo (1.4.11). No entanto, a série em (1.4.14) não converge para  $|\theta'| > 1$ , e o modelo em (1.4.13) não é invertível.

Portanto, à mesma função de autocorrelação correspondem dois modelos, um invertível e o outro não.

No caso geral de modelos MA(q), existe um máximo de  $2^q$  modelos com a mesma função de autocorrelação, mas somente um é invertível. Isto garante a unicidade de representação do modelo (veja por exemplo, Granger & Newbold (1986), seção 3.4).

Neste caso a condição de invertibilidade é equivalente a exigir que as raízes de  $\theta(B) = 0$  estejam fora do círculo unitário.

## 1.5 - MODELOS AUTORREGRESSIVO MÉDIA MÓVEL.

Estes modelos são da forma

$$(1.5.1) \quad y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

As propriedades do modelo geral ARMA(p,q), serão indicadas a partir da análise do caso particular ARMA(1,1).

Modelo ARMA(1,1). Este modelo é da forma

$$(1.5.2) \quad y_t = \phi_1 y_{t-1} + a_t - \theta_1 a_{t-1}.$$

Multiplicando (1.5.2) por  $y_{t-k}$  e tomando-se a esperança obtem-se

$$(1.5.3) \quad \gamma(k) = \phi_1 \gamma(k-1) + \gamma_{ya}(k) - \theta_1 \gamma_{ya}(k-1), \text{ onde}$$

$\gamma_{ya}(k) = E(y_{t-k} a_t)$ , é a covariância cruzada entre  $y$  e  $a$ .

Como  $y_{t-k}$  não depende de  $a_t$  se  $k > 0$ , tem-se que  $\gamma_{ya}(k) = 0$  para  $k > 0$  e  $\gamma_{ya}(k) \neq 0$  para  $k \leq 0$ .

Portanto, para  $k \geq 2$  (1.5.3) torna-se

$$(1.5.4) \quad \gamma(k) = \phi_1 \gamma(k-1).$$

Como  $\gamma_{ya}(0) = \sigma_a^2$  e  $\gamma_{ya}(-1) = (\phi_1 - \theta_1) \sigma_a^2$ , tem-se de (1.5.3) o seguinte sistema

$$\begin{cases} \gamma(0) = \phi_1 \gamma(1) + \sigma_a^2 - \theta_1 (\phi_1 - \theta_1) \sigma_a^2 \\ \gamma(1) = \phi_1 \gamma(0) - \theta_1 \sigma_a^2 \end{cases},$$

que tem solução

$$\gamma(0) = \frac{1 + \theta_1^2 - 2\phi_1 \theta_1}{1 - \phi_1^2} \sigma_a^2$$

$$\gamma(1) = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2} \sigma_a^2,$$

portanto a função de autocorrelação é dada por

$$(1.5.4) \quad \rho(k) = \begin{cases} 1 & \text{se } k = 0 \\ \frac{(1-\phi_1)(\phi_1-\theta_1)}{(1+\theta_1^2-2\phi_1\theta_1)} & \text{se } k = 1 \\ \phi_1\rho(k-1) & \text{se } k \geq 2. \end{cases}$$

Note que a partir do lag 2, a função de autocorrelação é afetada somente pela parte autorregressiva, enquanto a parte média móvel entra somente na determinação de  $\rho(1)$ .

O modelo em (1.5.2) pode ser escrito como

$$(1.5.5) \quad (1 - \phi_1 B)y_t = (1 - \theta_1 B)a_t,$$

ou seja, o modelo ARMA(1,1) é equivalente a um modelo AR( $\infty$ )

$$(1-\theta_1 B)^{-1}(1-\phi_1 B)y_t = a_t$$

com restrições nos parâmetros.

Também é equivalente a um modelo MA( $\infty$ )

$$y_t = (1 - \phi_1 B)^{-1}(1 - \theta_1 B)a_t$$

com restrições nos parâmetros.

A condição de estacionaridade é equivalente a  $|\phi_1| < 1$  e a condição de invertibilidade é equivalente a  $|\theta_1| < 1$ .

Em termos dos polinômios  $\phi(B)$  e  $\theta(B)$ , o modelo é estacionário se a raiz de  $1 - \phi_1 B = 0$  é  $B = 1/\phi_1$  maior que 1 em valor absoluto e é invertível se a raiz de  $1 - \theta_1 B = 0$  é  $B = 1/\theta_1$  maior que 1 em valor absoluto.

**Modelo ARMA(p,q).** Este modelo é da forma (1.5.1), que também é equivalente a um modelo AR( $\infty$ ) com restrições nos parâmetros e equivalente a um modelo MA( $\infty$ ) com restrições nos parâmetros.

Multiplicando (1.5.1) por  $y_{t-k}$  e tomando-se a esperança obtem-se

$$(1.5.6) \quad \gamma(k) = \phi_1 \gamma(k-1) + \dots + \phi_p \gamma(k-p) + \gamma_{ya}(k) - \theta_1 \gamma_{ya}(k-1) - \dots - \theta_q \gamma_{ya}(k-q).$$

Se  $k > q$ ,  $\gamma_{ya}(k) = 0$  e (1.5.6) torna-se

$$(1.5.7) \quad \gamma(k) = \phi_1 \gamma(k-1) + \dots + \phi_p \gamma(k-p), \text{ para } k > q,$$

portanto a função de autocorrelação é dada por

$$(1.5.8) \quad \rho(k) = \phi_1 \rho(k-1) + \dots + \phi_p \rho(k-p), \text{ para } k > q.$$

Note que as autocorrelações do lag 1 até o lag q, são afetadas pela parte média móvel, enquanto que nos lags maiores que q a função de autocorrelação se comporta como no modelo autorregressivo.

Pode-se verificar que se  $q < p$  esta função consiste de uma mistura de exponenciais e/ou senóides amortecidas, ao passo que se  $q \geq p$ , nos primeiros  $q - p - 1$  lags esta função não seguirá este padrão. Ver Morettin & Toloi (1985), seção 7.5.

Em termos dos polinômios  $\phi(B)$  e  $\theta(B)$ , a condição de estacionaridade é equivalente a exigir que as raízes de  $\phi(B) = 0$  estejam

fora do círculo unitário e a condição de invertibilidade é equivalente a exigir que as raízes de  $\theta(B) = 0$  estejam fora do círculo unitário.

## 1.6 - MODELOS ARIMA.

Os modelos apresentados anteriormente são apropriados para descrever séries estacionárias. Se a série é estacionária e  $\mu \neq 0$ , subtrai-se a média da série, isto é

$$(1.6.1) \quad \phi(B)(y_t - \mu) = \theta(B)a_t, \text{ ou}$$

$$(1.6.2) \quad \phi(B)y_t = \mu_0 + \theta(B)a_t, \text{ onde } \mu_0 = \phi(B)\mu = (1 - \phi_1 - \dots - \phi_p)\mu.$$

Quando a série não é estacionária, o que ocorre muito na prática, procura-se através de transformações torná-la estacionária.

Serão estudadas somente aqueles tipos de séries não estacionárias, que tornam-se estacionárias após a aplicação de um número finito de diferenças,  $d$ . Estas séries são chamadas não estacionárias homogêneas.



Se

$$(1.6.3) \quad x_t = \nabla^d y_t$$

é estacionária, então  $x_t$  pode ser representada por um modelo ARMA(p,q), ou seja

$$(1.6.4) \quad \phi(B)x_t = \theta(B)a_t.$$

Se  $x_t$  é uma diferença de  $y_t$ , então  $y_t$  é uma integral de  $x_t$ , por isso diz-se que  $y_t$  é um modelo autorregressivo integrado média móvel, ou modelo ARIMA(p,d,q) e escreve-se

$$(1.6.5) \quad \phi(B)\nabla^d y_t = \theta(B)a_t.$$

Como o modelo em (1.6.4) é estacionário, as raízes de  $\phi(B) = 0$  estão fora do círculo unitário.

Fazendo  $\varphi(B) = \phi(B)\nabla^d = \phi(B)(1-B)^d$ , o modelo em (1.6.5) é equivalente a

$$(1.6.6) \quad \varphi(B)y_t = \theta(B)a_t, \quad \text{onde}$$

$\varphi(B)$  é um polinômio em B de grau p+d, com d raízes sobre o círculo unitário, ou seja, iguais a 1, e as restantes fora do círculo unitário.

Portanto, o modelo em (1.6.6) supõe que a d-ésima diferença

de  $y_t$  pode ser representada por um modelo ARMA estacionário e invertível.

Usualmente  $d = 1$  ou  $d = 2$ , que correspondem a dois casos interessantes de não estacionaridade homogênea: séries não estacionárias quanto ao nível, que oscilam em torno de um nível médio e podem mudar para outros níveis temporários. Para torná-las estacionárias, é suficiente tomar uma diferença; séries não estacionárias quanto a inclinação, que oscilam numa direção e podem mudar para outras direções temporárias. Para torná-las estacionárias, é necessário tomar a segunda diferença.

Observe que

$$\nabla^d y_t = \nabla^d (y_t - \mu) \quad \text{se } d \geq 1.$$

### 1.7 - MODELOS SAZONAIS.

Estes modelos correspondem a generalização dos modelos ARIMA para séries com forte componente sazonal.

Em muitos casos, é mais realístico considerar esta componente como um processo estocástico, como é feito aqui, e não determinística.

Se a série tem período de sazonalidade  $s$ , os modelos a seguir são generalizações imediatas dos modelos ARIMA e levam em consideração a dependência de observações afastadas  $s$  períodos.

**Modelos AR(P)s.** Neste caso tem-se

$$(1.7.1) \quad y_t = \phi_1 y_{t-s} + \phi_2 y_{t-2s} + \dots + \phi_p y_{t-ps} + b_t, \text{ ou}$$

$$\phi(B^s)y_t = b_t, \text{ onde}$$

$\phi(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}$  e satisfaz a condição de estacionaridade.

A função de autocorrelação decresce exponencialmente para zero nos lags  $s, 2s, 3s, \dots$ , se  $b_t$  é ruído branco e é zero nos lags intermediários.

**Modelos MA(Q)s.** Neste caso tem-se

$$(1.7.2) \quad y_t = b_t - \theta_1 b_{t-s} + \theta_2 b_{t-2s} + \dots + \theta_q b_{t-qs}, \text{ ou}$$

$$y_t = \theta(B^s)b_t, \text{ onde}$$

$\theta(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_q B^{qs}$  e satisfaz a condição de

invertibilidade.

A função de autocorrelação será diferente de zero somente nos lags  $s, 2s, \dots, qs$  se  $b_t$  é ruído branco, o que significa que não há dependência nos períodos adjacentes  $t - 1, t$  e  $t + 1$ , o que em muitos casos não é razoável.

**Modelos ARMA(P,Q)s.** Neste caso tem-se

$$(1.7.3) \quad \Phi(B^s)y_t = \Theta(B^s)b_t, \quad \text{onde}$$

$\Phi(B^s)$  e  $\Theta(B^s)$  são polinômios em  $B$  como nos modelos AR e MA respectivamente, satisfazendo as condições de estacionaridade e invertibilidade.

Se  $b_t$  é ruído branco, a função de autocorrelação é não nula somente nos lags múltiplos de  $s$ .

**Modelos Sazonais Multiplicativos.** Nos modelos anteriores, tem-se que, se  $b_t$  é ruído branco, o modelo não é razoável pois supõe que períodos adjacentes são não correlacionados.

O mais razoável é supor os  $b_t$  correlacionados e modelar em geral por um modelo ARIMA.

Portanto o modelo será

$$(1.7.4) \quad \Phi(B^s)(1-B^s)^D y_t = \Theta(B^s)b_t,$$

onde  $b_t$  é dado por

$$(1.7.5) \quad \phi(B)(1-B)^d b_t = \theta(B)a_t$$

e  $a_t$  é ruído branco. Ou seja,

$$(1.7.6) \quad \Phi(B^s)\phi(B)(1-B^s)^D(1-B)^d y_t = \Theta(B^s)\theta(B)a_t, \quad \text{onde}$$

$(1-B^s)^D$  é o operador diferença sazonal de grau D,  $(1-B)^d$  é o operador diferença não sazonal de grau d,  $\Phi(B^s)$ ,  $\Theta(B^s)$ ,  $\phi(B)$  e  $\theta(B)$  são polinômios definidos anteriormente e satisfazendo as condições de estacionaridade e invertibilidade. Estes modelos são chamados de ARIMA sazonal multiplicativos, modelos  $ARIMA(p,d,q) \times (P,D,Q)s$ .

A função de autocorrelação do modelo tem o mesmo comportamento do modelo ARIMA para os lags até s, e nos lags múltiplos de s. Nos lags adjacentes aos lags múltiplos de s, cada modelo tem um comportamento diferente.

### 1.8 - PREVISÃO BASEADA NA REPRESENTAÇÃO MA.

Seja  $y_t$  um processo estocástico estacionário com média zero e representação MA

$$(1.8.1) \quad y_t = \sum_{i=0}^{\infty} \theta_i a_{t-i} = \theta(B)a_t, \text{ onde } \theta_0 = 1,$$

as raízes de  $\theta(B) = 0$  estão fora do círculo unitário e  $a_t$  é ruído branco.

Considere o previsor linear de horizonte h na origem t, baseado em  $S_t = \{a_s | s \leq t\}$  dado por

$$(1.8.2) \quad y_t(h) = \sum_{i=0}^{\infty} \theta_{h+i} a_{t-i} = \sum_{i=h}^{\infty} \theta_i a_{t+h-i}.$$

O erro de previsão associado a  $y_t(h)$  é dado por

$$e_y(h) = y_{t+h} - y_t(h) = \sum_{i=0}^{\infty} \theta_i a_{t+h-i} - \sum_{i=h}^{\infty} \theta_i a_{t+h-i}, \text{ ou}$$

$$(1.8.3) \quad e_y(h) = \sum_{i=0}^{h-1} \theta_i a_{t+h-i},$$

considerando-se que os parâmetros são conhecidos e que a série é suficientemente longa.

Portanto,  $y_t(h)$  é não viciado com erro médio quadrático de previsão (EMQP) dado por

$$(1.8.4) \quad \sigma_y^2(h) = \sigma_a^2 \sum_{i=0}^{h-1} \theta_i^2.$$

**TEOREMA 1.6.1:** O previsor  $y_t(h)$  definido em (1.8.2) é ótimo no sentido de EMQP, isto é, (1.8.4) é mínimo entre os previsores lineares.

**PROVA:** Seja

$$y_t^*(h) = \sum_{i=0}^{\infty} \lambda_{h+i} a_{t-i} = \sum_{i=h}^{\infty} \lambda_i a_{t+h-i},$$

qualquer previsor linear de  $y_{t+h}$ .

Então o erro de previsão associado a  $y_t^*(h)$  é

$$\begin{aligned} e_{y^*}(h) &= y_{t+h} - y_t^*(h) = \sum_{i=0}^{\infty} \theta_i a_{t+h-i} - \sum_{i=0}^{\infty} \lambda_i a_{t+h-i} \\ &= \sum_{i=0}^{h-1} \theta_i a_{t+h-i} + \sum_{i=h}^{\infty} (\theta_i - \lambda_i) a_{t+h-i}. \end{aligned}$$

Assim, o EMQP de  $y_t^*(h)$  está dado por

$$\sigma_{y^*}^2(h) = \sigma_a^2 \sum_{i=0}^{h-1} \theta_i^2 + \sigma_a^2 \sum_{i=h}^{\infty} (\theta_i - \lambda_i)^2,$$

portanto  $\sigma_{y^*}^2(h) \geq \sigma_y^2(h)$ .

■

O EMQP de  $y_t(h)$  pode ser escrito como

$$\sigma_y^2(h) = \sigma_a^2 \sum_{i=0}^{h-1} \theta_i^2 = \sigma_a^2 \sum_{i=0}^{h-2} \theta_i^2 + \sigma_a^2 \theta_{h-1}^2 = \sigma_y^2(h-1) + \sigma_a^2 \theta_{h-1}^2,$$

ou seja, o EMQP é não decrescente com o crescimento do horizonte de previsão  $h$ , isto é,

$$(1.8.5) \quad \sigma_y^2(s) \geq \sigma_y^2(r), \text{ para } s \geq r.$$

O erro de previsão de 1 passo à frente é dado por

$$e_y(1) = y_{t+1} - y_t(1) = a_{t+1},$$

o que significa que os erros de previsão de 1 passo à frente são não correlacionados.

Se os  $a_t$  são independentes e identicamente distribuídos segundo uma distribuição normal de média zero e variância  $\sigma_a^2$ , então  $y_t(h)$  tem distribuição normal com média zero e variância  $\sigma_y^2(h)$ . Isto pode ser usado para construir intervalos de confiança para as previsões.

Por exemplo, um intervalo de  $(1-\alpha)100\%$  de confiança para a previsão de  $h$  passos à frente na origem  $t$ , está dado por

$$(1.8.6) \quad y_t(h) - z_{\alpha/2} \sigma_y(h) \leq y_{t+h} \leq y_t(h) + z_{\alpha/2} \sigma_y(h),$$

onde  $z_{\alpha/2}$  é o valor crítico da distribuição normal padrão.

Se  $y_t$  é um processo estocástico estacionário de média  $\mu = E(y_t)$ , então  $z_t = y_t - \mu$  tem média zero e neste caso o previsor linear ótimo para  $y_t$  será  $y_t(h) = \mu + z_t(h)$ , onde  $z_t(h)$  é o previsor ótimo para  $z_t$ . O EMQP de  $y_t$  não é afetado, isto é,  $\sigma_y^2(h) = \sigma_z^2(h)$ , onde  $\sigma_z^2(h)$  é o EMQP de  $z_t(h)$ .

Muitas vezes é de interesse a matriz de EMQP conjunta para diferentes horizontes de previsão. Denotando a matriz de EMQP de 1 até  $h$  passos à frente por  $\Sigma_y(h)$ , tem-se

$$\Sigma_y(h) = E \begin{bmatrix} y_{t+1} - y_t(1) \\ y_{t+2} - y_t(2) \\ \vdots \\ y_{t+h} - y_t(h) \end{bmatrix} \begin{bmatrix} y_{t+1} - y_t(1) \\ y_{t+2} - y_t(2) \\ \vdots \\ y_{t+h} - y_t(h) \end{bmatrix},$$

$$= E \begin{bmatrix} 0 \\ \sum_{i=0}^1 \theta_i a_{t+1-i} \\ 1 \\ \sum_{i=0}^1 \theta_i a_{t+2-i} \\ \vdots \\ h-1 \\ \sum_{i=0}^{h-1} \theta_i a_{t+h-i} \end{bmatrix} \begin{bmatrix} 0 \\ \sum_{i=0}^1 \theta_i a_{t+1-i} \\ 1 \\ \sum_{i=0}^1 \theta_i a_{t+2-i} \\ \vdots \\ h-1 \\ \sum_{i=0}^{h-1} \theta_i a_{t+h-i} \end{bmatrix}, \text{ ou}$$



$$(1.8.7) \quad \Sigma_y(h) = \begin{bmatrix} \sigma_a^2 & \theta_1 \sigma_a^2 & \dots & \theta_{h-1} \sigma_a^2 \\ \theta_1 \sigma_a^2 & \sum_{l=0}^1 \theta_l^2 \sigma_a^2 & \dots & \sum_{l=0}^1 \theta_l \theta_{h-2+l} \sigma_a^2 \\ \vdots & \vdots & & \vdots \\ \theta_{h-1} \sigma_a^2 & \sum_{l=0}^1 \theta_{h-2+l} \theta_l \sigma_a^2 & \dots & \sum_{l=0}^{h-1} \theta_l^2 \sigma_a^2 \end{bmatrix}$$

Note que os elementos da diagonal de  $\Sigma_y(h)$  são os EMQP de 1 até h passos à frente.

### 1.9 - PREVISÃO BASEADA NA REPRESENTAÇÃO AR.

Sendo a representação MA (1.8.1) de  $y_t$  invertível, isto é, se as raízes de  $\theta(B) = 0$  estão fora do círculo unitário, então o previsor ótimo pode ser calculado usando a representação autorregressiva

$$(1.9.1) \quad y_t = \sum_{l=1}^{\infty} \phi_l y_{t-l} + a_t, \text{ ou } \phi(B)y_t = a_t, \text{ onde}$$

as raízes de  $\phi(B) = 0$  estão fora do círculo unitário.

Note que  $\phi(B) = \theta(B)^{-1}$  e portanto os coeficientes de  $\phi(B)$  podem ser determinados em função dos coeficientes de  $\theta(B)$  e vice-versa.

De (1.8.2) o previsor linear ótimo de horizonte 1 na origem t, baseado em  $S_t = \{a_s | s \leq t\}$  ou equivalentemente  $S_t = \{y_s | s \leq t\}$  é dado por

$$y_t(1) = \sum_{i=1}^{\infty} \theta_i a_{t+i-1} = y_{t+1} - a_{t+1} = \sum_{i=0}^{\infty} \phi_i y_{t+i-1}.$$

Analogamente, o previsor linear ótimo de horizonte 2 na origem  $t$ , baseado em  $S_t$  é dado por

$$\begin{aligned} y_t(2) &= \sum_{i=2}^{\infty} \theta_i a_{t+i-1} = y_{t+2} - a_{t+2} - \theta_1 a_{t+1} \\ &= \sum_{i=1}^{\infty} \phi_i y_{t+i-1} + a_{t+2} - a_{t+2} - \theta_1 a_{t+1} \\ &= \sum_{i=1}^{\infty} \phi_i y_{t+i-1} - \phi_1 a_{t+1} \\ &= \phi_1 y_t(1) - \phi_1 y_{t+1} + \sum_{i=1}^{\infty} \phi_i y_{t+i-1}, \end{aligned}$$

onde foi usado o fato de que  $\phi_1 = \theta_1$  e  $y_t(1) = y_{t+1} - a_{t+1}$ .

Procedendo desta maneira, tem-se para  $h > 2$

$$(1.9.2) \quad y_t(h) = \sum_{i=1}^{h-1} \phi_i y_t(h-i) + \sum_{i=h}^{\infty} \phi_i y_{t+h-i},$$

o qual pode ser usado recursivamente para calcular previsões.

É assumido que a soma é zero se o limite inferior no índice do somatório for menor que o limite superior. Assim (1.9.2) pode também ser usado para  $h = 1$ .

Se os  $a_t$  são independentes, (1.9.2) pode ser obtido tomando-se a esperança condicional de

$$y_{t+h} = \sum_{i=1}^{\infty} \phi_i y_{t+h-i} + a_{t+h}, \text{ dado } S_t = \{y_s | s \leq t\},$$

$$(1.9.3) \quad E(y_{t+h} | S_t) = \sum_{i=1}^{h-1} \phi_i E(y_{t+h-i} | S_t) + \sum_{i=h}^{\infty} \phi_i y_{t+h-i},$$

e por indução em  $h$ , tem-se que  $E(y_{t+h} | S_t) = y_t(h)$ .

**TEOREMA 1.9.1:** O previsor  $y_t(h)$  definido em (1.9.2) é ótimo no sentido de erro médio quadrático de previsão.

**PROVA.** Considere  $y_t^*(h)$  qualquer previsor de  $y_{t+h}$  baseado em  $S_t$ , então

$$E[y_{t+h} - y_t^*(h)]^2 = E\{[y_{t+h} - E(y_{t+h} | S_t)] + [E(y_{t+h} | S_t) - y_t^*(h)]\}^2, \text{ ou}$$

$$(1.9.4) \quad E[y_{t+h} - y_t^*(h)]^2 = E[y_{t+h} - E(y_{t+h} | S_t)]^2 + E[E(y_{t+h} | S_t) - y_t^*(h)]^2,$$

pois  $[y_{t+h} - E(y_{t+h} | S_t)]$  é uma ponderação dos  $a_s$  tal que  $s > t$ , enquanto

$[E(y_{t+h} | S_t) - y_t^*(h)]$  é uma ponderação dos  $a_s$  tal que  $s \leq t$ , portanto

$$E\{[y_{t+h} - E(y_{t+h} | S_t)][E(y_{t+h} | S_t) - y_t^*(h)]\} = 0.$$

Assim, (1.9.4) é minimizada para  $y_t^*(h) = E(y_{t+h} | S_t)$ . Logo se os  $a_t$  são independentes,  $y_t(h)$  é ótimo entre todos os previsores de  $y_{t+h}$  baseado em  $S_t$  e não apenas entre os lineares.

Mostra-se no exemplo a seguir que a condição de independência dos  $a_t$  é crítica (veja Lütkepohl 1987, página 56).

EXEMPLO 1.9.1: Considere o AR(1)  $y_t = \phi y_{t-1} + a_t$ , onde  $|\phi| < 1$  e

$$a_t = \begin{cases} \varepsilon_t & \text{se } t = 0, \pm 2, \pm 4, \dots \\ \frac{\varepsilon_t^2 - 1}{\sqrt{2}} & \text{se } t = \pm 1, \pm 3, \dots \end{cases}$$

e os  $\varepsilon_t$  são variáveis aleatórias independentes e  $N(0,1)$ .

Note que os  $a_t$  são não correlacionados, mas não são independentes.

O previsor de  $y_{t+1}$  baseado em  $S_t$  é  $y_t(1) = \phi y_t$ , de acordo com (1.9.2). Entretanto, este não é sempre o valor esperado condicional de  $y_{t+1}$  dado  $S_t$ .

Para  $t$  ímpar

$$E(y_{t+1} | S_t) = E(\phi y_t + a_{t+1} | S_t) = \phi y_t + \frac{\varepsilon_t^2 - 1}{\sqrt{2}}.$$

□

Se  $y_t$  é um processo estocástico estacionário com média  $\mu = E(y_t)$ , então

$$(y_t - \mu) = \sum_{i=1}^{\infty} \phi_i (y_{t-i} - \mu) + a_t, \text{ ou}$$

$$(1.9.5) \quad y_t = \mu_0 + \sum_{i=1}^{\infty} \phi_i y_{t-i} + a_t, \text{ onde}$$

$$\mu_0 = (1 - \phi_1 - \phi_2 - \dots) \mu.$$

Usando (1.9.2), o previsor ótimo de  $y_{t+h}$  é dado por

$$\begin{aligned}
 y_t(h) &= \mu + [y_t(h) - \mu] \\
 &= \mu + \sum_{i=1}^{h-1} \phi_i [y_t(h-i) - \mu] + \sum_{i=h}^{\infty} \phi_i [y_{t+h-i} - \mu] \\
 &= \mu - \sum_{i=1}^{\infty} \phi_i \mu + \sum_{i=1}^{h-1} \phi_i y_t(h-i) + \sum_{i=h}^{\infty} \phi_i y_{t+h-i}, \text{ ou} \\
 (1.9.6) \quad y_t(h) &= \mu_0 + \sum_{i=1}^{h-1} \phi_i y_t(h-i) + \sum_{i=h}^{\infty} \phi_i y_{t+h-i}.
 \end{aligned}$$

É assumido que o conjunto  $S_t$  contém infinitas informações. Na prática, tem-se somente uma série finita,  $y_1, y_2, \dots, y_T$ .

Se  $T$  não é muito pequeno e as raízes do polinômio MA não estão próximas do círculo unitário, uma boa aproximação para (1.9.2) é

$$(1.9.7) \quad y_T^*(h) = \sum_{i=1}^{h-1} \phi_i y_T^*(h-i) + \sum_{i=h}^{T+h-1} \phi_i y_{T+h-i},$$

desde que  $\phi_i$  se aproxime de zero quando  $i$  vai para infinito.

Do ponto de vista prático, a suposição de processo gerador de dados ser conhecido é menos realística do que  $S_t$  conter infinitas informações.

### 1.10 - PREVISÃO BASEADA NA REPRESENTAÇÃO ARMA.

Considere  $y_t$  um processo estocástico estacionário com média zero e representação ARMA

$$(1.10.1) \quad y_t = \sum_{i=1}^p \phi_i y_{t-i} + a_t - \sum_{i=1}^q \theta_i a_{t-i}, \text{ ou}$$

$$\phi(B)y_t = \theta(B)a_t, \text{ onde}$$

as raízes de  $\phi(B) = 0$  e  $\theta(B) = 0$  estão fora do círculo unitário.

Como na seção anterior, o preditor linear ótimo na origem  $t$  e horizonte  $h$  baseado em  $S_t = \{a_s | s \leq t\}$ , pode ser calculado recursivamente usando-se

$$(1.10.2) \quad y_t(h) = \sum_{i=1}^{h-1} \phi_i y_{t-h+i} + \sum_{i=h}^p \phi_i y_{t-h+i} + \sum_{i=h}^q \theta_i a_{t-h+i},$$

onde o primeiro somatório é definido como sendo zero se o limite inferior no índice do somatório for maior que o limite superior,  $\phi_i = 0$  se  $i > p$  e  $\theta_i = 0$  se  $i > q$ .

Se  $y_t$  tem média  $\mu = E(y_t) \neq 0$ , então  $z_t = y_t - \mu$  tem média zero e o preditor ótimo para  $y_t$  é

$$(1.10.3) \quad y_t(h) = \mu + z_t(h),$$

onde  $z_t(h)$  é o preditor ótimo de  $z_t$ .

Se uma série de comprimento finito  $y_1, y_2, \dots, y_T$  é observada, os  $a_t$  em geral, não podem ser obtidos exatamente.

Uma aproximação para os  $a_t$  serem usados em (1.10.2) pode ser obtida da recursão

$$(1.10.4) \quad a_t^* = y_t - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{i=1}^q \theta_i a_{t-i}^*, \text{ para } t = p+1, p+2, \dots, T,$$

onde  $a_t^* = 0$  para  $t \leq p$ . Na prática, os  $a_t$  em (1.10.2) são sempre trocados pelos resíduos estimados.

Até aqui, os EMQP foram obtidos supondo-se que o modelo é totalmente conhecido, o que na prática não é razoável. Em geral, o processo gerador dos dados é sempre desconhecido.

Ainda que as ordens do processo ARMA sejam conhecidas, os coeficientes devem ser estimados e o EMQP é obtido aproximadamente.

Em muitas situações o processo gerador dos dados pode ser aproximado por um processo autorregressivo de ordem desconhecida. Alguns critérios para escolha da ordem serão apresentados na próxima seção.

### 1.11 - CRITÉRIOS PARA A ESCOLHA DO MODELO.

A ordem do processo gerador de um particular conjunto de dados é sempre desconhecida. Na prática, deseja-se ter um critério razoável para escolha do modelo no caso de amostras finitas.

Nesta seção serão descritos os seguintes critérios:

$$(1.11.1) \quad AIC(p) = \text{Ln} \hat{\sigma}_p^2 + \frac{2p}{T-p},$$

(Akaike 1974),

$$(1.11.2) \quad HQ(p) = \text{Ln} \hat{\sigma}_p^2 + \frac{2p \text{LnLn}(T-P)}{T-P},$$

(Hanna & Quinn 1979, Quinn 1980),

$$(1.11.3) \quad SC(p) = \text{Ln} \hat{\sigma}_p^2 + \frac{p \text{Ln}(T-P)}{T-P},$$

(Schwarz 1978, Rissanen 1978),

onde  $\hat{\sigma}_p^2$  é a estimativa da variância do ruído branco quando é ajustado um modelo AR(p) aos dados, mas usando-se T-P resíduos. T é o tamanho da amostra e P é a ordem máxima especificada para os AR ajustados.

Para cada um dos três critérios, modelos AR(p) são ajustados para  $p = 0, 1, \dots, P$  e a ordem do AR,  $\tilde{p}$ , é escolhida de modo que o critério é minimizado, por exemplo,

$$\tilde{p}(\text{AIC}) = \min_p \{ \text{AIC}(p) | p = 0, 1, \dots, P \}.$$

Esses critérios podem ser interpretados como consistindo de um termo que decresce quando a ordem do AR cresce e outro que penaliza pelo crescimento da ordem do AR.

Um critério atinge seu mínimo quando esses termos são balanceados. O termo que penaliza é diferente para cada critério.

HQ e SC são critérios consistentes, no sentido de que  $\tilde{p}(\text{HQ})$  e  $\tilde{p}(\text{SC})$  convergem em probabilidade, quando  $T \rightarrow \infty$ , para a verdadeira ordem p do AR, desde que  $y_t$  seja realmente de ordem finita, estacionário e a ordem máxima escolhida seja tal que  $p \leq P$ .

O critério AIC superestima a verdadeira ordem



assintoticamente com probabilidade positiva se  $p < P$ .

Outros critérios para escolha de modelo tem sido proposto na literatura. Muitos deles são, em princípio, similares ao AIC, HQ ou SC, ou tem propriedades em pequenas amostras inferiores, no sentido de que eles produzem modelos com capacidade inferior de previsão (veja Lütkepohl 1986).

## CAPÍTULO 2

### O EFEITO DA AGREGAÇÃO SOBRE O MODELO

#### 2.1 - INTRODUÇÃO

Suponha que  $y_t$  é um processo estocástico estacionário que satisfaz um modelo ARMA(p,q) e seja  $Y_\tau$  o processo agregado temporalmente com  $t = m\tau$ , onde  $m$  é o período de agregação e  $\tau$  é o tempo para o processo agregado, por exemplo, se  $t$  refere-se ao mês e  $m = 3$ , então  $\tau$  refere-se ao trimestre.

Neste capítulo serão apresentados alguns teoremas que irão mostrar o comportamento do modelo após a agregação.

Em todos os casos é assumido que o modelo básico ou desagregado é totalmente conhecido, isto é, são conhecidas as ordens e os parâmetros do modelo.

Será considerado,

(2.1.1)  $y_t, t = 0, \pm 1, \pm 2, \dots$ , um processo estocástico e

$$(2.1.2) \quad Y_\tau = y_t + y_{t-1} + \dots + y_{t-m+1} = \left[ \sum_{i=0}^{m-1} B^i \right] y_t, \text{ com } t = m\tau,$$

e  $\tau = 0, \pm 1, \pm 2$ , o processo agregado temporalmente com período de agregação  $m$ .

Para  $t = 0, \pm 1, \pm 2, \dots$ ,  $a_t$  são consideradas variáveis aleatórias não correlacionadas com média zero e variância constante

$$\sigma_a^2, \text{ isto é, } E(a_t) = 0 \text{ para todo } t \text{ e } E(a_t a_1) = \begin{cases} \sigma_a^2 & \text{se } t = 1 \\ 0 & \text{se } t \neq 1 \end{cases}. B \text{ é o}$$

operador atraso tal que  $B^1 y_t = y_{t-1}$ .

Para a demonstração dos teoremas apresentados neste capítulo, será usada a identidade abaixo conforme Pereira, P. L. V. (1981) e (1982)

$$(1 - \alpha_j^m B^m) \equiv (1 - \alpha_j B)(1 + \alpha_j B + \dots + \alpha_j^{m-1} B^{m-1}), \text{ para } j = 1, 2, \dots, p,$$

através das expressões:

$$(2.1.3) \quad \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) (1 - \alpha_j B)^{-1} \right] \left[ \sum_{i=0}^{m-1} B^i \right] \text{ e}$$

$$(2.1.4) \quad \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) (1 - \alpha_j B)^{-1} \right] \left[ \sum_{i=0}^{m-1} B^i \right]^{d+1}.$$

## 2.2 - O EFEITO SOBRE O MODELO MA

O resultado obtido nesta seção, é um caso particular de uma situação mais geral apresentada por Brewer (1973), para modelos ARMA, que será visto na seção 2.4. Aqui será apresentado formalmente.

**TEOREMA 2.2.1 :** Seja  $y_t$  um modelo  $MA(q)$ , isto é,

$$(2.2.1) \quad y_t = \theta_q(B) a_t$$

onde  $\theta_q(B)$  é um polinômio em  $B$  de grau  $q$ , cujas raízes estão todas

fora do círculo unitário. Então,  $Y_\tau$  é um modelo média móvel de ordem  $q^* = \left[ \frac{q+m-1}{m} \right]$ , onde  $[x]$  significa a parte inteira de  $x$ .

PROVA : Multiplicando (2.2.1) por  $\left[ \sum_{i=0}^{m-1} B^i \right]$  obtém-se

$$(2.2.2) \quad Y_\tau = \theta_{q_1}(B) a_t, \text{ onde}$$

$$\theta_{q_1}(B) a_t = \left[ \sum_{i=0}^{m-1} B^i \right] \theta_q(B) a_t.$$

$$\text{Como, } E \left[ \theta_{q_1}(B) a_t \cdot \theta_{q_1}(B) a_{t-mk} \right] = \begin{cases} \neq 0 & \text{para } mk \leq q + m - 1 \\ = 0 & \text{caso contrário} \end{cases},$$

o lado direito de (2.2.2) pode ser representado por um processo média móvel de ordem  $q^*$ , ou seja,

$$Y_\tau = \theta_{q^*}^*(B) a_\tau^*, \text{ onde } B = B^m, \text{ tal que } BY_\tau = Y_{\tau-1},$$

e os coeficientes de  $\theta_{q^*}^*(B)$  e a variância de  $a_\tau^*$  estão relacionados com os coeficientes do processo desagregado através da relação

$$(2.2.3) \quad \text{Cov} \left[ \theta_{q^*}^*(B) a_\tau^*, \theta_{q^*}^*(B) a_{\tau+k}^* \right] = \text{Cov} \left[ \theta_{q_1}(B) a_t, \theta_{q_1}(B) a_{t-mk} \right],$$

com  $t = m\tau$ .

■

Os coeficientes de  $\theta_{q^*}^*(B)$ , podem ser determinados através de um procedimento em Anderson (1971) seção 5.7.

No exemplo a seguir, ilustra-se como determinar estes

coeficientes e a variância de  $a_{\tau}^*$ .

EXEMPLO 2.2.1: Suponha que  $y_t$  segue um modelo MA(1),  $y_t = a_t - \theta a_{t-1}$ .

Então, pelo teorema 2.2.1, com  $m = 2$ ,  $Y_{\tau}$  segue um modelo MA(1), isto é  $Y_{\tau} = a_{\tau}^* - \theta^* a_{\tau-1}^*$ , e serão determinados  $\theta^*$  e  $\text{Var}(a_{\tau}^*) = \sigma_{\tau}^2$ .

Para  $t = 2\tau$ ,

$$\begin{aligned} Y_{\tau} &= y_t + y_{t-1} \\ &= a_t - \theta a_{t-1} + a_{t-1} - \theta a_{t-2} \\ &= a_t + (1-\theta)a_{t-1} - \theta a_{t-2}. \end{aligned}$$

Assim,

$$\text{Cov}(a_t + (1-\theta)a_{t-1} - \theta a_{t-2}, a_{t+2k} + (1-\theta)a_{t+2k-1} - \theta a_{t+2k-2}) =$$

$$= \begin{cases} [1+(1-\theta)^2+\theta^2]\sigma_a^2, & \text{para } k = 0 \\ -\theta\sigma_a^2, & \text{para } k = 1 \\ 0, & \text{para } k \geq 2, \text{ ou} \end{cases}$$

$$(2.2.4) \quad \sigma_Y(v) = \text{Cov}(Y_{\tau}, Y_{\tau+v}) = \begin{cases} [1+(1-\theta)^2+\theta^2]\sigma_a^2, & \text{para } v = 0 \\ -\theta\sigma_a^2, & \text{para } v = 1 \\ 0, & \text{para } v \geq 2. \end{cases}$$

Como  $Y_{\tau}$  é um MA(1) tem-se que,

$$\text{Cov}(Y_\tau, Y_{\tau+v}) = \begin{cases} (1 + \theta^{*2})\sigma_\tau^2, & \text{para } v = 0 \\ -\theta^* \sigma_\tau^2 & , \text{ para } v = 1 \\ 0 & , \text{ para } v \geq 2. \end{cases}$$

Portanto, de (2.2.3) obtém-se,

$$\begin{cases} (1 + \theta^{*2})\sigma_\tau^2 = [1 + (1-\theta)^2 + \theta^2]\sigma_a^2 \\ \theta^* \sigma_\tau^2 = \theta \sigma_a^2 \end{cases} \Rightarrow \begin{cases} \theta \sigma_a^2 \theta^{*2} - 2\sigma_a^2(1-\theta+\theta^2)\theta^* + \theta \sigma_a^2 = 0 \\ \sigma_\tau^2 = \theta \sigma_a^2 / \theta^* \end{cases}$$

daí deve-se escolher  $\theta^*$  tal que  $|\theta^*| < 1$ .

Para  $\theta = 0.2$  e  $\sigma_a^2 = 1$  encontra-se,  $\theta^* = 0.1208$  e  $\sigma_\tau^2 = 1.6556$ .

Isto é equivalente ao procedimento dado em Anderson (1971), para determinar os coeficientes de  $\theta_{q^*}^*(B)$ , como segue:

A função geratriz de autocovariância de um processo MA(1) é dada por  $\sigma_Y(-1) + \sigma_Y(0)z + \sigma_Y(1)z^2 = 0$  (Anderson, 1971 p. 224-225), a qual tem duas raízes:  $z_1$  com  $|z_1| < 1$  e a outra é  $1/z_1$ , então  $z - z_1 = \alpha_0^* z + \alpha_1^* = z + \theta^*$ , e tem-se que

$$\theta^* = -z_1 = \frac{\sigma_Y(0) \mp \{\sigma_Y^2(0) - 4\sigma_Y^2(1)\}^{1/2}}{2\sigma_Y(1)},$$

a qual é menor que um em valor absoluto e  $\sigma_Y(v)$  é dado em (2.2.4).

Para  $\theta = 0.2$  e  $\sigma_a^2 = 1$ , encontra-se  $\theta^* = 0.1208$ .

□

Note que a agregação neste caso não alterou o modelo, pois ainda continua-se com um modelo média móvel de ordem 1.

### 2.3 - O EFEITO SOBRE O MODELO AR

Telser(1967) e Amemiya & Wu (1972), trabalharam com processos puramente autorregressivos, ou modelos AR(p), concluindo que após a agregação obtem-se um modelo ARMA(p,q). Amemiya & Wu ainda especificaram a ordem q como sendo  $q = \left[ \frac{(p+1)(m-1)}{m} \right]$  se  $m < p + 1$  e  $q = p$  se  $m \geq p + 1$ , onde  $[x]$  significa a parte inteira de x. Eles também mostraram que a parte média móvel do modelo agregado é invertível.

**TEOREMA 2.3.1:** Suponha que  $y_t$  é um modelo autorregressivo de ordem p, ou AR(p), isto é,

$$(2.3.1) \quad \phi_p(B)y_t = a_t,$$

onde  $\phi_p(B)$  é um polinômio em B de grau p, cujas raízes estão todas fora do círculo unitário. Então,  $Y_T$  é um modelo autorregressivo média móvel de ordens máximas p e q respectivamente, com  $q = \left[ \frac{(p+1)(m-1)}{m} \right]$ , onde  $[x]$  significa a parte inteira de x.

**PROVA :** Sejam  $\alpha_1^{-1}, \dots, \alpha_p^{-1}$  as raízes de  $\phi_p(B)$ , então (2.3.1) pode ser escrito como

$$(2.3.2) \quad \left[ \prod_{j=1}^p (1 - \alpha_j B) \right] y_t = a_t.$$

Multiplicando (2.1.3) por (2.3.2) obtem-se,

$$(2.3.3) \quad \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) \right] Y_\tau = \theta_{q_2}(B) a_t, \text{ onde}$$

$$\theta_{q_2}(B) = \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) (1 - \alpha_j B)^{-1} \right] \left[ \sum_{i=0}^{m-1} B^i \right] = \prod_{j=0}^p \left[ \sum_{i=0}^{m-1} \alpha_j^i B^i \right], \text{ com } \alpha_0 = 1.$$

Note que, quando os polinômios  $\left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) \right]$  e  $\theta_{q_2}(B)$  têm

raízes comum, o modelo resultante terá ordens menores que  $p$  e  $q$  especificadas no teorema.

Em particular, esta situação se verifica quando o polinômio  $\phi_p(B)$  é da forma  $\sum_{i=1}^{p/m} \phi_{i_m} B^{im}$ , onde  $m$  é o período de agregação. Será visto mais adiante um exemplo desta situação.

Fazendo  $B^m = B$  de modo que  $BY_\tau = Y_{\tau-1}$ ,

$$\phi_p^*(B) = \left[ \prod_{j=1}^p (1 - \alpha_j^m B) \right] \text{ e notando que}$$

$$E[\theta_{q_2}(B) a_t \cdot \theta_{q_2}(B) a_{t-mk}] = \begin{cases} \neq 0 & \text{para } mk \leq (p+1)(m-1) \\ = 0 & \text{caso contrário} \end{cases},$$

tem-se que o lado direito de (2.3.3) pode ser representado por um processo média móvel da ordem especificada no teorema, ou seja,

$$\phi_p^*(B) Y_\tau = \theta_q^*(B) a_\tau^*.$$

Observe que as raízes de  $\phi_p^*(B)$  são iguais a  $m$ -ésima potência das correspondentes raízes de  $\phi_p(B)$ . Os coeficientes de  $\theta_q^*(B)$  e a variância de  $a_\tau^*$  estão relacionados com os coeficientes do processo



desagregado através da relação

$$(2.3.4) \quad \text{Cov} \left[ \theta_q^*(B) a_\tau^*, \theta_q^*(B) a_{\tau+k}^* \right] = \text{Cov} \left[ \theta_{q_2}(B) a_t, \theta_{q_2}(B) a_{t+mk} \right],$$

com  $t = m\tau$ .

■

No exemplo a seguir, ilustra-se a determinação dos coeficientes dos polinômios  $\phi_p^*(B)$  e  $\theta_q^*(B)$  e a variância de  $a_\tau^*$ .

**EXEMPLO 2.3.1:** Suponha que  $y_t$  segue um modelo AR(1),  $y_t = \phi y_{t-1} + a_t$ . Então, pelo teorema 2.3.1, com  $m = 2$ ,  $Y_\tau$  segue um modelo ARMA(1,1) e serão determinados os coeficientes de  $\phi_p^*(B)$  e  $\theta_q^*(B)$  e a  $\text{Var}(a_\tau^*) = \sigma_\tau^2$ .

Para  $t = 2\tau$ ,

$$\begin{aligned} Y_\tau &= y_t + y_{t-1} \\ &= \phi y_{t-1} + a_t + \phi y_{t-2} + a_{t-1} \\ &= \phi(\phi y_{t-2} + a_{t-1}) + a_t + \phi(\phi y_{t-3} + a_{t-2}) + a_{t-1} \\ &= \phi^2(y_{t-2} + y_{t-3}) + \phi(a_{t-1} + a_{t-2}) + a_t + a_{t-1} \\ &= \phi^2 y_{\tau-1} + a_t + (1+\phi)a_{t-1} + \phi a_{t-2}. \end{aligned}$$

Assim,  $\phi^* = \phi^2$  e

$$\text{Cov}(a_t + (1+\phi)a_{t-1} + \phi a_{t-2}, a_{t+2k} + (1+\phi)a_{t+2k-1} + \phi a_{t+2k-2}) =$$

$$= \begin{cases} [1+(1+\phi)^2+\phi^2]\sigma_a^2, & \text{para } k = 0 \\ \phi\sigma_a^2 & , \text{ para } k = 1 \\ 0 & , \text{ para } k \geq 2 \end{cases} , \text{ ou}$$

$$\sigma_Y(v) = \text{Cov}(Y_\tau, Y_{\tau+v}) = \begin{cases} 2\sigma_a^2/(1-\phi) , & \text{para } v = 0 \\ \phi\sigma_a^2(1+\phi)/(1-\phi), & \text{para } v = 1 \\ \phi^2\sigma_a^2(v-1) , & \text{para } v = 2 \end{cases}$$

Sendo  $Y_\tau$  um modelo ARMA(1,1), tem-se que

$$\sigma_Y(v) = \text{Cov}(Y_\tau, Y_{\tau+v}) = \begin{cases} (1+\theta^{*2} - 2\phi^*\theta^*)\sigma_\tau^2 / (1 - \phi^{*2}) , & v = 0 \\ (1 - \phi^*\theta^*)(\phi^* - \theta^*)\sigma_\tau^2 / (1 - \phi^{*2}) , & v = 1 \\ \phi^*\sigma_\tau^2(\kappa - 1) & , v \geq 2. \end{cases}$$

Portanto, seguindo o exemplo anterior e usando (2.3.4) com  $\phi = 0.8$ , encontra-se  $\phi^* = 0.64$ ,  $\theta^* = -0.1686$  e  $\sigma_\tau^2 = 4.7451$ .

□

É importante notar neste caso, que houve alteração no modelo causada pela agregação. No exemplo a seguir apresenta-se um caso de raízes comum aos polinômios AR e MA.

**EXEMPLO 2.3.2:** Considere o modelo AR(2),  $y_t = \phi^2 y_{t-2} + a_t$ . Então, pelo teorema 2.3.1, com  $m = 2$ ,  $Y_\tau$  segue um modelo ARMA(p,q) com  $p \leq 2$  e  $q \leq 1$ .

Para  $t = 2\tau$ ,

$$\begin{aligned} Y_\tau &= y_{2\tau} + y_{2\tau-1} \\ &= \phi^2 y_{2\tau-2} + a_{2\tau} + \phi^2 y_{2\tau-3} + a_{2\tau-1} \end{aligned}$$

$$= \phi^2(y_{2\tau-2} + y_{2\tau-3}) + a_{2\tau} + a_{2\tau-1}$$

$$= \phi^2 y_{\tau-1} + u_{\tau}, \quad \text{com } u_{\tau} = a_{2\tau} + a_{2\tau-1}.$$

Logo  $Y_{\tau}$  segue um modelo AR(1). Isto se verifica porque os polinômios AR e MA têm raízes comum, como segue-se.

$$(2.3.5) \quad y_t = \phi^2 y_{t-2} + a_t \Rightarrow (1-\phi B)(1+\phi B)y_t = a_t,$$

assim tem-se,  $\alpha_1^{-1} = -\phi$  e  $\alpha_2^{-1} = \phi$ .

Multiplicando  $\left[ \prod_{j=1}^2 (1-\alpha_j^2 B^2)(1-\alpha_j B)^{-1} \right]$  por (2.3.5) e após alguns cálculos obtém-se,

$$(1-\phi^2 B)(1-\phi^2 B)Y_{\tau} = (1-\phi^2 B)u_{\tau}, \quad \text{com } B = B^2 \text{ e } u_{\tau} = a_{2\tau} + a_{2\tau-1}.$$

Assim,  $1/\phi^2$  é raiz comum aos polinômios AR e MA.

□

Os teoremas seguintes, que envolvem polinômios autorregressivo no modelo desagregado, também podem apresentar problemas de raízes comum no modelo agregado, valendo portanto as mesmas considerações feitas no exemplo 2.3.2.

## 2.4 - O EFEITO SOBRE O MODELO ARMA

O resultado desta seção é mais geral que os das seções anteriores e foi obtido por Brewer(1973).

**TEOREMA 2.4.1:** Suponha que  $y_t$  é um modelo autorregressivo média móvel, ARMA(p,q), isto é,

$$(2.4.1) \quad \phi_p(B)y_t = \theta_q(B)a_t,$$

onde  $\theta_q(B)$  e  $\phi_p(B)$  são como nos teoremas 2.2.1 e 2.3.1 respectivamente. Então  $y_\tau$  é um modelo ARMA de ordens máximas p e  $q^*$ , com  $q^* = \left[ \frac{(p+1)(m-1)+q}{m} \right]$  e  $\{x\}$  é a parte inteira de x.

**PROVA :** Sejam  $\alpha_1^{-1}, \dots, \alpha_p^{-1}$  as raízes de  $\phi_p(B)$ , então (2.4.1) pode ser escrito como

$$(2.4.2) \quad \left[ \prod_{j=1}^p (1 - \alpha_j B) \right] y_t = \theta_q(B) a_t.$$

Multiplicando (2.1.3) por (2.4.2), obtem-se

$$(2.4.3) \quad \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) \right] y_\tau = \theta_{q_3}(B) a_t, \quad \text{onde}$$

$$\theta_{q_3}(B) = \prod_{j=0}^p \left[ \sum_{i=0}^{m-1} B^i \right] \theta_q(B).$$

Fazendo  $B^m = B$ , de maneira que  $B y_\tau = y_{\tau-1}$ ,

$$\phi_p^*(B) = \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) \right] \text{ e observando-se que}$$

$$E[\theta_{q_3}(B)a_t \cdot \theta_{q_3}(B)a_{t-mk}] = \begin{cases} \neq 0 & \text{para } mk \leq (p+1)(m-1)+q \\ = 0 & \text{caso contrário} \end{cases},$$

pode-se notar que o lado direito de (2.4.3) pode ser representado por um processo média móvel de ordem  $q^*$ , ou seja,

$$\phi_p^*(B)Y_\tau = \theta_{q^*}^*(B)a_\tau^*.$$

Os coeficientes de  $\theta_{q^*}^*(B)$  e a variância de  $a_\tau^*$  estão relacionados com os coeficientes do processo desagregado através da relação

$$\text{Cov}[\theta_{q^*}^*(B)a_\tau^*, \theta_{q^*}^*(B)a_{\tau+k}^*] = \text{Cov}[\theta_{q_3}(B)a_t, \theta_{q_3}(B)a_{t+mk}], \text{ com } t = m\tau.$$

■

As raízes de  $\phi_p^*(B)$  ficam determinadas pela  $m$ -ésima potência das correspondentes raízes de  $\phi_p(B)$ , enquanto que os coeficientes de  $\theta_{q^*}^*(B)$  são determinados como no exemplo anterior.

A seguir, será visto o comportamento do modelo após a agregação para processos não estacionários.

## 2.5 - O EFEITO SOBRE MODELOS NÃO ESTACIONÁRIOS

Nas seções anteriores, os processos estudados são todos estacionários. Tiao (1972), investigou o efeito da agregação temporal sobre o processo integrado média móvel, ou modelos IMA(d,q).

Wei (1979), estendeu os resultados Brewer e Tiao para o caso geral de processo autorregressivo integrado média móvel, ou modelos ARIMA(p,d,q), e modelos sazonais multiplicativos ARIMA(p,d,q)x(P,D,Q)<sub>s</sub>. Estes resultados serão apresentados nesta seção.

**TEOREMA 2.5.1:** Seja  $y_t$  um modelo autorregressivo integrado média móvel, ou seja um ARIMA(p,d,q), isto é

$$(2.5.1) \quad \phi_p(B)(1-B)^d y_t = \theta_q(B) a_t$$

onde  $\phi_p(B)$  e  $\theta_q(B)$  são como no teorema anterior e d é o grau da diferença necessária para tornar o processo estacionário. Então  $Y_T$  é um modelo ARIMA com ordens máximas p, d e  $q^*$ , com  $q^* = \left\lceil \frac{(m-1)(p+d+1)+q}{m} \right\rceil$  e  $[x]$  é a parte inteira de x.

**PROVA :** Sejam  $\alpha_1^{-1}, \dots, \alpha_p^{-1}$  as raízes de  $\phi_p(B)$ , então (2.5.1) pode ser escrito como

$$(2.5.2) \quad \left[ \prod_{j=1}^p (1-\alpha_j B) \right] (1-B)^d y_t = \theta_q(B) a_t, \text{ ou}$$

$$(2.5.3) \quad \left[ \prod_{j=1}^{p+d} (1-\alpha_j B) \right] y_t = \theta_q(B) a_t, \text{ com } \alpha_{p+1} = \dots = \alpha_{p+d} = 1.$$

Multiplicando (2.5.3) por

$$\left[ \prod_{j=1}^{p+d} (1-\alpha_j^{-m} B^m) (1-\alpha_j B)^{-1} \right] \left[ \sum_{i=0}^{m-1} B^i \right] \text{ com } \alpha_{p+1} = \dots = \alpha_{p+d} = 1 \text{ vem,}$$

$$(2.5.4) \quad \left[ \prod_{j=1}^{p+d} (1 - \alpha_j^m B^m) \right] Y_\tau = \theta_{q_4}(B) a_t, \quad \text{onde}$$

$$\theta_{q_4}(B) = \prod_{j=0}^{p+d} \left[ \sum_{i=0}^{m-1} \alpha_j^i B^i \right] \theta_q(B), \quad \text{com } \alpha_0 = 1.$$

$$\text{Agora, } \left[ \prod_{j=1}^{p+d} (1 - \alpha_j^m B^m) \right] = \left[ \prod_{j=1}^p (1 - \alpha_j^m B^m) \right] (1 - B^m)^d = \phi_p^*(B) (1 - B)^d \quad \text{com}$$

$B^m = B$  de maneira que  $BY_\tau = Y_{\tau-1}$ , e como

$$E[\theta_{q_4}(B) a_t \cdot \theta_{q_4}(B) a_{t-mk}] = \begin{cases} \neq 0 & \text{para } mk \leq (m-1)(p+d+1)+q \\ = 0 & \text{caso contrário} \end{cases},$$

então, o lado direito de (2.5.4) pode ser representado por um processo média móvel de ordem  $q^*$ , ou seja,

$$\phi_p^*(B) (1-B)^d Y_\tau = \theta_{q^*}^*(B) a_\tau^*.$$

Os coeficientes de  $\theta_{q^*}^*(B)$  e a variância de  $a_\tau^*$ , estão relacionados com o processo desagregado através da relação

$$\text{Cov}[\theta_{q^*}^*(B^m) a_\tau^*, \theta_{q^*}^*(B^m) a_{\tau+k}^*] = \text{Cov}[\theta_{q_4}(B) a_t, \theta_{q_4}(B) a_{t+mk}], \quad \text{com } t = m\tau.$$

■

Os coeficientes dos polinômios autorregressivo e média móvel, são obtidos como no teorema anterior. Note que o grau de diferença  $d$  não se altera.

No caso de modelos com sazonalidade, haverá mudança na ordem do polinômio média móvel para a parte não sazonal e no período

de sazonalidade, como será visto no seguinte teorema.

**TEOREMA 2.5.2:** Seja  $y_t$  um modelo autorregressivo integrado média móvel sazonal multiplicativo ou  $ARIMA(p,d,q) \times (P,D,Q)_S$ , isto é,

$$(2.5.5) \quad \phi_p(B)\phi_p(B^S)(1-B)^d(1-B^S)^D y_t = \theta_q(B)\theta_q(B^S)a_t,$$

onde  $\phi_p(B)$ ,  $\theta_q(B)$  e  $d$ , são como no teorema anterior,  $\phi_p(B^S)$  e  $\theta_q(B^S)$  são polinômios em  $B^S$  de graus  $P$  e  $Q$  respectivamente, cujas raízes estão todas fora do círculo unitário e,  $D$  é o grau da diferença sazonal. Então,  $Y_\tau$  é um modelo  $ARIMA$  sazonal multiplicativo com ordens máximas  $p$ ,  $d$ ,  $q^*$ ,  $P$ ,  $D$ ,  $Q$  e  $\Delta$ , com  $q^* = \left[ \frac{(m-1)(p+d+1)+q}{m} \right]$  e  $\Delta = s/m$ , onde  $[x]$  é a parte inteira de  $x$ .

**PROVA :** Sejam  $\alpha_1^{-1}, \dots, \alpha_p^{-1}$  e  $\delta_1^{-1}, \dots, \delta_P^{-1}$  as raízes de  $\phi_p(B)$  e  $\phi_p(B^S)$  respectivamente, então (2.5.5) pode ser escrito como

$$(2.5.6) \quad \left[ \prod_{j=1}^p (1-\alpha_j B) \right] \left[ \prod_{j=1}^P (1-\delta_j B^S) \right] (1-B)^d (1-B^S)^D y_t = \theta_q(B)\theta_q(B^S)a_t,$$

Multiplicando (2.1.4) por (2.5.6), tem-se

$$(2.5.7) \quad \left[ \prod_{j=1}^p (1-\alpha_j^m B^m) \right] \left[ \prod_{j=1}^P (1-\delta_j B^S) \right] (1-B^m)^d (1-B^S)^D Y_\tau = \theta_{q_S}(B)a_t, \text{ onde}$$

$$\theta_{q_S}(B) = \left[ \sum_{i=0}^{m-1} B^i \right]^{d+1} \left[ \prod_{j=1}^p \left( \sum_{i=0}^{m-1} \alpha_j^i B^i \right) \right] \theta_q(B)\theta_q(B^S).$$

Notando que,



$$E[\theta_{q_s}(B)a_t \cdot \theta_{q_s}(B)a_{t-mk}] = \begin{cases} \neq 0 & \text{para } mk \leq (m-1)(p+d+1)+q+sQ \\ = 0 & \text{caso contrário} \end{cases},$$

tem-se que o lado direito de (2.5.7) pode ser representado por um processo média móvel de ordem  $q_1^* = \left[ \frac{(m-1)(p+d+1)+q}{m} + \frac{s}{m} Q \right]$ , o qual é equivalente a um processo média móvel de ordem  $q^* = \left[ \frac{(m-1)(p+d+1)}{m} \right]$  para a parte não sazonal e ordem  $sQ$  para a parte sazonal com período de sazonalidade  $s = s/m$ , ou seja, com  $B = B^m$  (2.5.7) torna-se

$$\phi_p^*(B)\Phi_p^*(B^\Delta)(1-B)^d(1-B^\Delta)^D Y_\tau = \theta_q^*(B)\Theta_q^*(B^\Delta)a_\tau^*.$$

Os coeficientes de  $\theta_{q^*}^*(B)$  e a variância de  $a_\tau^*$  estão relacionados com os coeficientes do processo desagregado através da relação

$$\text{Cov}[\theta_{q^*}^*(B^m)a_\tau^*, \theta_{q^*}^*(B^m)a_{\tau+k}^*] = \text{Cov}[\theta_{q_s}(B)a_t, \theta_{q_s}(B)a_{t+mk}], \text{ com } t = m\tau.$$

As raízes de  $\phi_p^*(B)$ , são a  $m$ -ésima potência das respectivas raízes de  $\phi_p(B)$ . As raízes de  $\Phi_p^*(B^\Delta)$  e  $\Theta_q^*(B^\Delta)$ , são as correspondentes raízes de  $\Phi_p(B^s)$  e  $\Theta_q(B^s)$  respectivamente. Os coeficientes de  $\theta_q^*(B)$  e a variância de  $a_\tau^*$  são obtidos como no teorema anterior.

Note que o período físico não muda, isto é, se por exemplo, os dados são mensais com sazonalidade de um ano,  $s = 12$ , e o interesse é o trimestre, então  $m = 3$  e neste caso o novo período passa a ser  $s = 4$ , porém o período físico continua sendo de um ano.

## CAPÍTULO 3

### O EFEITO DA AGREGAÇÃO SOBRE A PREVISÃO

#### 3.1 - INTRODUÇÃO.

Existem várias situações onde a previsão de uma variável fluxo agregada temporalmente é de interesse.

Supondo que se dispõe dos dados desagregados, serão comparadas as duas maneiras de prever o processo agregado, isto é, prever o processo desagregado e agregar as previsões e prever o processo agregado diretamente.

O processo desagregado será denotado por  $y_t$  e o processo agregado temporalmente, no qual está-se interessado, por  $Y_t$ , onde  $t = m\tau$  e  $m$  é o período de agregação.

O previsor baseado no processo desagregado será denotado por  $y_t^0(h)$  e o previsor baseado no processo agregado por  $Y_t(h)$ , onde  $h$  é o horizonte de previsão.

Na seção 3.2 estes previsores serão comparados supondo que o modelo para o processo desagregado é totalmente conhecido.

Na seção 3.3 estes previsores serão comparados no caso de coeficientes estimados e na seção 3.4, serão comparados quando é ajustado um modelo AR de ordem finita para cada um dos processos, agregado e desagregado.

### 3.2 - PREVISÃO COM MODELO CONHECIDO.

Considere o processo desagregado  $y_t$ , com representação MA como na seção 1.6, e seja  $Y_\tau$  o processo agregado temporalmente com período de agregação  $m$  e  $t = m\tau$ , tal que

$$(3.2.1) \quad Y_\tau = y_{m\tau} + y_{m\tau-1} + \dots + y_{m\tau-m+1} = \left[ \sum_{i=0}^{m-1} B^i \right] y_{m\tau}.$$

Se a representação MA de  $y_t$  é invertível, isto é, se as raízes de  $\theta(B) = 0$  estão fora do círculo unitário, então  $Y_\tau$  tem representação MA (capítulo 2), ou seja,

$$(3.2.2) \quad Y_\tau = \sum_{i=0}^{\infty} \beta_i u_{\tau-i} = \beta(B) u_\tau,$$

onde as raízes de  $\beta(B) = 0$  caem fora do círculo unitário e  $u_\tau$  é ruído branco, isto é, são variáveis aleatórias não correlacionadas com média zero e variância constante  $\sigma_\tau^2$ .

Da seção 1.6, o previsor linear ótimo  $h$  passos à frente de  $y_t$  baseado em  $S_t = \{a_s | s \leq t\}$ , é dado por

$$(3.2.3) \quad y_t(h) = \sum_{i=0}^{\infty} \theta_{h+1-i} a_{t-i} = \sum_{i=h}^{\infty} \theta_i a_{t+h-i},$$

com erro médio quadrático de previsão dado por

$$(3.2.4) \quad \sigma_y^2(h) = \sigma_a^2 \sum_{i=0}^{h-1} \theta_i^2.$$

Da mesma maneira, o preditor linear ótimo  $h$  passos à frente de  $Y_\tau$  baseado em  $S_\tau = \{u_s | s \leq m\tau\}$  é dado por

$$(3.2.5) \quad Y_\tau(h) = \sum_{i=0}^{\infty} \beta_{h+1} u_{\tau-1} = \sum_{i=h}^{\infty} \beta_i u_{\tau+h-1},$$

com erro médio quadrático de previsão dado por

$$(3.2.6) \quad \sigma_Y^2(h) = \sigma_\tau^2 \sum_{i=0}^{h-1} \beta_i^2.$$

Se os dados desagregados estão disponíveis, o preditor  $Y_\tau(h)$  não é em geral o preditor ótimo do processo agregado  $Y_\tau$ , como será visto no teorema seguinte (veja Lütkepohl 1987, seção 4.2.1).

**TEOREMA 3.2.1:** O preditor linear ótimo no sentido de erro médio quadrático de previsão, do processo agregado  $Y_\tau$ ,  $h$  passos à frente na origem  $\tau$  baseado em  $S_\tau = \{a_s | s \leq m\tau\}$ , é dado por

$$(3.2.7) \quad Y_\tau^0(h) = y_{m\tau}(mh) + y_{m\tau}(mh-1) + \dots + y_{m\tau}(mh-m+1),$$

onde  $y_{m\tau}(j)$  é o preditor linear ótimo  $j$  passos à frente na origem  $m\tau$  do processo desagregado  $y_t$ .

**PROVA:** Considere  $Y_\tau^0(h) = \sum_{j=0}^{\infty} \pi_j a_{m\tau-j}$  o preditor linear ótimo do processo agregado e serão determinados os pesos  $\pi_j$  que minimizam o EMQP de  $Y_\tau^0(h)$ .

O erro de previsão associado a  $Y_\tau^0(h)$  é dado por

$$\begin{aligned}
e_0(h) &= Y_{\tau+h} - Y_{\tau}^0(h) = \left[ \sum_{i=0}^{m-1} B^i \right] y_{m(\tau+h)} - \sum_{j=0}^{\infty} \pi_j a_{m\tau-j} \\
&= \left[ \sum_{i=0}^{m-1} B^i \right] \sum_{j=0}^{\infty} \theta_j a_{m(\tau+h)-j} - \sum_{j=0}^{\infty} \pi_j a_{m\tau-j} \\
&= \left[ \sum_{i=0}^{m-1} B^i \right] \left[ \sum_{j=0}^{mh-1} \theta_j a_{m(\tau+h)-j} + \sum_{j=mh}^{\infty} \theta_j a_{m(\tau+h)-j} \right] - \sum_{j=0}^{\infty} \pi_j a_{m\tau-j} \\
&= \left[ \sum_{i=0}^{m-1} B^i \right] \left[ \sum_{j=0}^{mh-1} \theta_j a_{m(\tau+h)-j} + \sum_{j=0}^{\infty} \theta_{mh+j} a_{m\tau-j} \right] - \sum_{j=0}^{\infty} \pi_j a_{m\tau-j} \\
&= \left[ \sum_{i=0}^{m-1} B^i \right] \sum_{j=0}^{mh-1} \theta_j a_{m(\tau+h)-j} + \sum_{j=0}^{\infty} \left[ \left[ \sum_{i=0}^{m-1} B^i \right] \theta_{mh+j} - \pi_j \right] a_{m\tau-j} .
\end{aligned}$$

Assim  $\sigma_0^2(h) = E[e_0(h)]^2$  será minimizada para

$$\pi_j = \left[ \sum_{i=0}^{m-1} B^i \right] \theta_{mh+j} , \quad j = 0, 1, 2, \dots ,$$

portanto, substituindo os  $\pi_j$  tem-se

$$\begin{aligned}
Y_{\tau}^0(h) &= \sum_{j=0}^{\infty} \pi_j a_{m\tau-j} = \sum_{j=0}^{\infty} \left[ \sum_{i=0}^{m-1} B^i \right] \theta_{mh+j} a_{m\tau-j} \\
&= \sum_{j=0}^{\infty} \theta_{mh+j} a_{m\tau-j} + \sum_{j=0}^{\infty} \theta_{mh-1+j} a_{m\tau-j} + \dots + \sum_{j=0}^{\infty} \theta_{mh-m+1+j} a_{m\tau-j} \\
&= y_{m\tau}(mh) + y_{m\tau}(mh-1) + \dots + y_{m\tau}(mh-m+1).
\end{aligned}$$

■

Sendo  $\Sigma_y(h)$  a matriz de EMQP conjunta de 1 até h passos à frente definida em (1.6.7), o erro médio quadrático de previsão de

$Y_{\tau}^0(h)$  é dado por

$$(3.2.8) \quad \sigma_0^2(h) = v' \Sigma_y(mh) v, \text{ onde}$$

$v' = (0, \dots, 0, \dots, 1, \dots, 1)$  é um vetor de dimensão  $mh$  cujas últimas  $m$  componentes são todas iguais a 1.

O teorema a seguir da condições para a igualdade dos previsores do processo agregado e é uma versão do caso multivariado dado por Lütkepohl (1987), proposição 8.1.

**TEOREMA 3.2.2:** Considere os previsores do processo agregado  $Y_{\tau}(h)$  e  $Y_{\tau}^0(h)$  como em (3.2.5) e (3.2.7) respectivamente. Então,

$$Y_{\tau}(h) = Y_{\tau}^0(h), \text{ para } h = 1, 2, \dots \text{ se, e somente se,}$$

$$(3.2.9) \quad (1 + B + \dots + B^{m-1}) \left[ \sum_{i=0}^{\infty} \theta_i B^i \right] = \\ = \left[ \sum_{j=0}^{\infty} (\theta_{jm} + \dots + \theta_{j-m+1}) B^{jm} \right] \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_i) B^i \right],$$

onde  $\theta_0 = 1$  e  $\theta_k = 0$  para  $k < 0$ .

**PROVA:** Como a igualdade dos previsores para  $h = 1$  implica na igualdade para  $h > 1$ , é suficiente mostrar que ,

$$u_{\tau+1} = Y_{\tau+1} - Y_{\tau}(1) = Y_{\tau+1} - Y_{\tau}^0(1), \text{ ou}$$

$$(3.2.10) \quad u_{\tau+1} = \left[ \sum_{i=0}^{m-1} (1+\theta_1+\dots+\theta_1) B^i \right] a_{m(\tau+1)}.$$

Suponha que vale a igualdade em (3.2.9), então

$$\begin{aligned} \sum_{j=0}^{\infty} \beta_j u_{\tau-j} &= Y_{\tau} = (1+B+\dots+B^{m-1}) y_{m\tau} \\ &= (1+B+\dots+B^{m-1}) \left[ \sum_{i=0}^{\infty} \theta_1 B^i \right] a_{m\tau}, \text{ logo por suposição,} \end{aligned}$$

$$\begin{aligned} \sum_{j=0}^{\infty} \beta_j u_{\tau-j} &= \left[ \sum_{j=0}^{\infty} (\theta_{jm} + \dots + \theta_{j(m-m+1)}) B^{jm} \right] \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_1) B^i \right] a_{m\tau} \\ &= \sum_{j=0}^{\infty} (\theta_{jm} + \dots + \theta_{j(m-m+1)}) \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_1) B^i \right] a_{m(\tau-j)} \\ &= \sum_{j=0}^{\infty} (\theta_{jm} + \dots + \theta_{j(m-m+1)}) u_{\tau-j}, \end{aligned}$$

pois a representação MA é única.

Assim,

$$u_{\tau} = \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_1) B^i \right] a_{m\tau}$$

e portanto a igualdade dos previsores segue de (3.2.10).

Agora se

$$u_{\tau+1} = \left[ \sum_{i=0}^{m-1} (1+\theta_1+\dots+\theta_1) B^i \right] a_{m(\tau+1)}, \text{ então}$$

$$Y_{\tau} = \sum_{j=0}^{\infty} \beta_j u_{\tau-j} = \sum_{j=0}^{\infty} \beta_j \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_i) B^i \right] a_{m(\tau-j)} , \text{ ou}$$

$$(3.2.11) \quad Y_{\tau} = \sum_{j=0}^{\infty} \beta_j \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_i) \right] a_{m(\tau-j)-1} ,$$

portanto o coeficiente de  $a_{m(\tau-j)}$  é  $\beta_j$ .

Por outro lado,

$$(3.2.12) \quad Y_{\tau} = \sum_{k=0}^{m-1} y_{m\tau-k} = \sum_{k=0}^{m-1} \left[ \sum_{i=0}^{\infty} \theta_i a_{m\tau-k-i} \right] ,$$

portanto  $a_{m\tau-k-i} = a_{m(\tau-j)}$  se  $i = jm - k$ , então o coeficiente de

$$a_{m(\tau-j)} \text{ em (3.2.12) é } \sum_{k=0}^{m-1} \theta_{jm-k} .$$

Das igualdades em (3.2.11) e (3.2.12), tem-se que

$$\beta_j = \sum_{k=0}^{m-1} \theta_{jm-k} , \text{ e substituindo } \beta_j \text{ em (3.2.11), obtem-se}$$

$$Y_{\tau} = \sum_{j=0}^{\infty} (\theta_{jm} + \dots + \theta_{jm-m+1}) \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_i) \right] a_{m(\tau-j)-1} , \text{ ou}$$

$$Y_{\tau} = \left[ \sum_{j=0}^{\infty} (\theta_{jm} + \dots + \theta_{jm-m+1}) B^{jm} \right] \left[ \sum_{i=0}^{m-1} (\theta_0 + \theta_1 + \dots + \theta_i) \right] a_{m\tau} , \text{ e como}$$

$$Y_{\tau} = (1+B+\dots+B^{m-1}) \left[ \sum_{i=0}^{\infty} \theta_i B^i \right] a_{m\tau} , \text{ segue-se a igualdade desejada.}$$

■

Uma interessante consequência deste teorema, é que se  $y_t$  é



um processo MA puramente sazonal, com período de sazonalidade  $m$ , isto é, se  $\theta_i = 0$  a menos que  $i$  seja um inteiro múltiplo de  $m$ , então (3.2.9) é satisfeita.

Ou seja, se

$$y_t = \sum_{i=0}^{\infty} \theta_{im} a_{t-im},$$

então  $y_{\tau}(h) = Y_{\tau}^0(h)$ , para  $h = 1, 2, \dots$ .

Outro resultado interessante é que  $\sigma^2(h) - \sigma_0^2(h) \rightarrow 0$  quando  $h \rightarrow \infty$ , isto é, os dois previsores tornam-se idênticos quando o horizonte de previsão vai para infinito.  $\sigma^2(h)$  e  $\sigma_0^2(h)$  convergem para a variância do processo  $Y_{\tau}$ .

### 3.3 - PREVISÃO COM PROCESSO ARMA COM COEFICIENTES ESTIMADOS.

Nesta seção será assumido que  $a_t$  é uma sequência de variáveis aleatórias independentes identicamente distribuídas com média zero e variância constante  $\sigma_a^2$ .

$y_t$  é um processo estocástico Gaussiano, estacionário e invertível, seguindo um modelo ARMA com ordens finitas  $p$  e  $q$  conhecidas e com coeficientes estimados por máxima verossimilhança.

O vetor de todos os parâmetros (exceto  $\sigma_a^2$ ) será denotado por  $\Theta = (\theta_1, \theta_2, \dots, \theta_k)'$  e seu estimador de máxima verossimilhança por  $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)'$ .

Sabe-se da teoria estatística, que sob certas condições de regularidade esse estimador é consistente e tem distribuição normal

assintótica, ou seja,

$$\sqrt{T} (\hat{\theta} - \theta) \text{ é assintoticamente } N(0, \Sigma_{\theta}^{\wedge}),$$

onde  $T$  é o tamanho da amostra usada para estimação e  $\Sigma_{\theta}^{\wedge}$  é a inversa da matriz de informação, isto é,

$$\Sigma_{\theta}^{\wedge} = T \left[ -E \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right]^{-1},$$

onde  $L$  é o logaritmo da função de verossimilhança e as derivadas são calculadas no verdadeiro vetor de parâmetros (veja Serfling 1980 seção 4.2.2).

Para obter os resultados assintóticos, será assumido que estimação e previsão são baseadas em processos independentes mas com mesma estrutura estocástica.

Esta suposição facilita os cálculos mas é sempre violada na prática, pois uma única amostra é usada para estimação e previsão.

Se esta suposição é trocada pela suposição de que a origem de previsão é o último período usado para estimação, os resultados a seguir ainda são válidos (veja Reinsel (1980), Fuller & Hasza (1981)).

Previsores e EMQP obtidos sob estas condições serão simbolizados por um (^).

### 3.3.1 - PROCESSO COM MÉDIA ZERO.

Seja  $y_t$  um processo estocástico estacionário com média

zero. O teorema a seguir permite obter-se uma aproximação para os EMQP dos previsores de interesse (veja Lütkepohl 1987, seção 3.2.1).

**TEOREMA 3.3.1.1:** Seja  $\hat{y}_t(h)$  o predictor obtido trocando-se os verdadeiros parâmetros em (1.10.2) por seus estimadores de máxima verossimilhança.

Então,

$\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  é assintoticamente  $N(0, \omega_y(h))$ , onde

$$\omega_y(h) = E \left[ \frac{\partial y_t(h)}{\partial \theta} \Sigma_{\theta} \frac{\partial y_t(h)'}{\partial \theta} \right], \text{ e}$$

$$\frac{\partial y_t(h)}{\partial \theta} = \left[ \frac{\partial y_t(h)}{\partial \theta_j} \right] \quad j = 1, 2, \dots, K$$

é um vetor  $(1 \times K)$  e  $\frac{\partial y_t(h)'}{\partial \theta} = \left[ \frac{\partial y_t(h)'}{\partial \theta} \right]'$ .

**PROVA:** Considere o vetor aleatório

$$Z_t = (y_t, y_{t-1}, \dots, y_{t-p+1}, a_t, \dots, a_{t-q+1})'.$$

Sabe-se que  $\sqrt{T} \hat{\theta}$  é assintoticamente normal com média  $\theta$  e matriz de covariância  $\Sigma_{\theta}$ .

Dada uma realização de  $Z_t$ ,  $\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  é uma função diferenciável de  $\sqrt{T} \hat{\theta}$ .

Portanto, da proposição 6.8 de James Barry, R. (1981) página 249, tem-se que  $\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  dado  $Z_t$  é assintoticamente

$$(3.3.1.1) \quad N \left[ 0, \frac{\partial y_t(h)}{\partial \theta} \Sigma_{\hat{\theta}} \frac{\partial y_t(h)'}{\partial \theta} \right]$$

desde que nenhuma linha de  $\partial y_t(h)/\partial \theta$  seja identicamente nula.

Note que

$$\frac{\partial \hat{y}(h)}{\partial \hat{\theta}} \bigg|_{\hat{\theta} = \theta} = \frac{\partial y(h)}{\partial \theta},$$

pois  $\hat{y}_t(h)$  tem a mesma forma funcional de  $y_t(h)$ .

Assim, a distribuição normal em limite vale para quase toda realização  $Z_t$ .

Como  $Z_t$  e  $\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  dado  $Z_t$  são normais

multivariadas, a distribuição conjunta assintótica de  $Z_t$  e

$\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  é normal multivariada e consequentemente, a

distribuição marginal assintótica de  $\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  é normal com

média média zero e variância dada por

$$\text{Var} \left[ \sqrt{T} [\hat{y}_t(h) - y_t(h)] \right] = \text{Var} \left[ E \left[ \sqrt{T} [\hat{y}_t(h) - y_t(h)] \middle| Z_t \right] \right] +$$

$$+ E \left[ \text{Var} \left[ \sqrt{T} \left[ \hat{y}_t(h) - y_t(h) \right] \middle| Z_t \right] \right] =$$

$$= E \left[ \frac{\partial y_t(h)}{\partial \Theta} \Sigma_{\Theta}^{\wedge} \frac{\partial y_t(h)'}{\partial \Theta} \right] ,$$

já que  $E \left[ \sqrt{T} \left[ \hat{y}_t(h) - y_t(h) \right] \middle| Z_t \right] = 0$ .

O erro de previsão h passos à frente pode ser escrito como

$$y_{t+h} - \hat{y}_t(h) = y_{t+h} - y_t(h) + y_t(h) - \hat{y}_t(h)$$

$$= \sum_{i=0}^{m-1} \lambda_i a_{t+h-i} + y_t(h) - \hat{y}_t(h)$$

então, a suposição de processos independentes para estimação e previsão, permite escrever o EMQP assintótico de  $\hat{y}_t(h)$  como

$$\hat{\sigma}_y^2(h) = E[y_{t+h} - \hat{y}_t(h)]^2$$

$$= E[y_{t+h} - y_t(h)]^2 + E[y_t(h) - \hat{y}_t(h)]^2 , \text{ ou}$$

$$(3.3.1.4) \quad \hat{\sigma}_y^2(h) = \sigma_y^2(h) + \frac{1}{T} \omega_y(h) ,$$

onde a última esperança é tomada com respeito a distribuição

assintótica de  $\sqrt{T} \left[ \hat{y}_t(h) - y_t(h) \right]$ . Desta forma, para um dado tamanho

de amostra finito, (3.3.1.4) é apenas uma aproximação para o EMQP assintótico de  $\hat{y}_t(h)$ .

O resultado em (3.3.1.4) mostra que sob as suposições desta seção, o EMQP assintótico de  $\hat{y}_t(h)$  é composto de duas partes: uma parte devida a variabilidade do processo e outra devida a variabilidade da estimativa, que vai para zero quando T vai para infinito. Isto não surpreende, uma vez que os parâmetros são estimados consistentemente.

### 3.3.2 - PROCESSO COM MÉDIA DIFERENTE DE ZERO.

Se  $y_t$  tem média  $\mu = E(y_t) \neq 0$ , esta média pode ser estimada pela média amostral

$$(3.3.2.1) \quad \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t,$$

onde T é o tamanho da amostra.

O vetor de parâmetros  $\Theta$  contendo todos os coeficientes AR e MA, podem ser estimados por máxima verossimilhança usando-se os dados ajustados  $(y_1 - \bar{y}), (y_2 - \bar{y}), \dots, (y_T - \bar{y})$ .

Sabe-se também da teoria estatística que  $\bar{y}$  é um estimador consistente de  $\mu$  e que

$$(3.3.2.2) \quad \sqrt{T}(\bar{y} - \mu) \text{ é assintoticamente } N(0, \sigma_y^2).$$

Alem disso,  $\bar{y}$  é assintoticamente independente de  $\hat{\Theta}$ , então o

EMQP assintótico de  $\hat{y}_t(h)$  é

$$(3.3.2.3) \quad \hat{\sigma}_y^2(h) = \sigma_y^2(h) + \frac{1}{T} \omega_y(h) + \frac{1}{T} \pi_y(h),$$

onde a suposição de processos independentes para estimação e previsão é mantida e

$$(3.3.2.4) \quad \pi_y(h) = E \left[ \frac{\partial y_t(h)}{\partial \mu} \sigma_y^2 \frac{\partial y_t(h)'}{\partial \mu} \right].$$

Neste caso, a matriz de covariância de  $\sqrt{T} [\hat{y}_t(h) - y_t(h)]$  é dada por

$$(3.3.2.5) \quad E \left[ \begin{bmatrix} \frac{\partial y_t(h)}{\partial \mu} & \frac{\partial y_t(h)}{\partial \theta} \end{bmatrix} \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \Sigma_\theta \end{bmatrix} \begin{bmatrix} \frac{\partial y_t(h)}{\partial \mu} & \frac{\partial y_t(h)}{\partial \theta} \end{bmatrix}' \right] =$$

$$= E \left[ \frac{\partial y_t(h)}{\partial \mu} \sigma_y^2 \frac{\partial y_t(h)'}{\partial \mu} \right] + E \left[ \frac{\partial y_t(h)}{\partial \theta} \Sigma_\theta \frac{\partial y_t(h)'}{\partial \theta} \right].$$

Portanto, se a média do processo é diferente de zero, o EMQP assintótico de  $\hat{y}_t(h)$ , é acrescido de um termo que vai para zero quando o tamanho da amostra  $T$  vai para infinito.

### 3.3.3 - OS EMQP ASSINTÓTICOS DOS PREVISORES.

A matriz de EMQP assintótica conjunta das previsões de 1 até  $h$  passos à frente é dada por

$$(3.3.3.1) \quad \hat{\Sigma}_y(h) = \Sigma_y(h) + \frac{1}{T} \Omega_y(h) + \frac{1}{T} \Pi_y(h).$$

Denotando o previsor com coeficientes estimados baseado no processo desagregado por  $\hat{Y}_\tau^0(h)$ , o seu EMQP assintótico é dado por

$$\begin{aligned} (3.3.3.2) \quad \hat{\sigma}_0^2(h) &= v' \hat{\Sigma}_y(mh) v \\ &= v' \Sigma_y(mh) v + \frac{1}{T} v' \Omega_y(mh) v + \frac{1}{T} v' \Pi_y(mh) v \\ &= \sigma_0^2(h) + \frac{1}{T} v' [\Omega_y(mh) + \Pi_y(mh)] v, \end{aligned}$$

onde  $v$  é definido em (3.2.8).

Se  $y_t$  é Gaussiano, então  $Y_\tau$  também é Gaussiano e portanto o EMQP assintótico de  $\hat{Y}_\tau(h)$ , o previsor com coeficientes estimados baseado no processo agregado, é dado por

$$(3.3.3.3) \quad \hat{\sigma}^2(h) = \sigma^2(h) + \frac{m}{T} [\omega_y(h) + \pi_y(h)],$$

onde é assumido que o tamanho da amostra para a estimação dos coeficientes em  $Y_\tau(h)$  é  $J = T/m$ .  $\omega_y(h)$  é devido a estimação dos parâmetros ARMA e  $\pi_y(h)$  representa a variação na estimação da média.

Observe que os termos  $\sigma_0^2(h)$  e  $\sigma^2(h)$  em (3.3.3.2) e (3.3.3.3) dominam os EMQP assintóticos para grandes valores de  $T$ .

Portanto, se a diferença  $\sigma^2(h) - \sigma_0^2(h)$  é positiva, de maneira que  $Y_\tau^0(h)$  domina completamente  $Y_\tau(h)$  para o processo



conhecido, o mesmo vale para processos com coeficientes estimados quando o tamanho da amostra usada para estimação é grande.

Além disso, se o horizonte de previsão vai para infinito, os dois previsores se aproximam da média do processo  $Y_\tau$  e assim,  $v' \Omega_y(mh)v$  e  $\omega_y(h)$ , tendem para zero, enquanto  $v' \Pi_y(mh)v$  e  $\pi_y(h)$  convergem para  $m^2 \sigma_y^2$  e  $\sigma_y^2$  respectivamente, onde  $\sigma_y^2$  é a variância da distribuição assintótica de  $\sqrt{J} (\bar{Y} - \mu_y)$  e

$$(3.3.3.4) \quad \bar{Y} = \frac{1}{J} \sum_{\tau=1}^J Y_\tau$$

é a média do processo agregado. Também,  $\sigma_y^2$  é a variância assintótica da distribuição de  $\sqrt{T} (\bar{y} - \mu_y)$  com  $\bar{y}$  definido em (3.3.2.1).

Note que para  $h \rightarrow \infty$ ,  $\hat{Y}_\tau^0(h) \rightarrow m\bar{y}$  e portanto a parte do EMQP devido a estimação se aproxima de  $m^2 \sigma_y^2 / T$ .

Como

$$(3.3.3.5) \quad \bar{Y} = \frac{m}{T} \sum_{t=1}^T y_t = m\bar{y},$$

segue-se que

$$\sqrt{J} (\bar{Y} - \mu_y) = \sqrt{T/m} (m\bar{y} - m\mu_y) = \sqrt{m} \sqrt{T} (\bar{y} - \mu_y), \text{ tal que}$$

$$(3.3.3.6) \quad m\sigma_y^2 = m^2 \sigma_y^2,$$

e assim,  $\hat{\sigma}^2(h) - \hat{\sigma}_0^2(h) \rightarrow 0$  quando  $h \rightarrow \infty$ . Ou seja, os dois previsores são idênticos para grandes horizontes de previsão.

### 3.4 - PREVISÃO COM PROCESSO AR DE ORDEM DESCONHECIDA.

Suponha que o processo gerador de uma série pode ser aproximado por um autorregressivo de ordem finita mas, desconhecida.

Se um limite superior  $P$  para esta ordem, for especificado, um modelo  $AR(P)$  pode ser ajustado aos dados por mínimos quadrados.

Porém, se a verdadeira ordem do processo gerador da série,  $p$ , for menor que  $P$ , estes estimadores não serão assintoticamente eficientes, pois não levam em conta a restrição  $\phi_i = 0$  para  $p < i \leq P$ , onde  $\phi_i$  são os coeficientes do modelo  $AR(P)$  ajustado.

Considere o processo estacionário  $y_t$  com representação MA como em (3.2.1) e suponha que este processo tem representação AR de ordem possivelmente infinita,

$$(3.4.1) \quad y_t = \sum_{i=1}^{\infty} \phi_i y_{t-i} + a_t, \text{ ou } \phi(B)y_t = a_t,$$

onde as raízes de  $\phi(B) = 0$  estão fora do círculo unitário. Com esta suposição o processo agregado  $Y_\tau$  é estacionário e tem representação AR e MA (Lütkepohl 1986).

Será assumido que somente processos AR de ordem finita são ajustados aos dados, embora isto seja um erro de especificação se o processo é realmente de ordem infinita.

Nesta situação, uma teoria assintótica devida a Lewis & Reinsel (1982) e Lütkepohl (1985), pode ser usada para derivar o erro médio quadrático de previsão para os previsores  $Y_\tau(h)$  e  $Y_\tau^0(h)$ .

Para isso suponha que as seguintes suposições são

satisfeitas pelo processo desagregado  $y_t$ :

(1)  $\sum_{i=1}^{\infty} |\theta_i| < \infty$ , onde os  $\theta_i$  são os coeficientes da representação MA.

(2) Os  $a_t$  são independentes e identicamente distribuídos com  $E(a_t^4) < \infty$ .

(3) A ordem  $\tilde{p}(T)$  do modelo AR ajustado para a série  $y_t$ , é uma função do tamanho da amostra  $T$  tal que, quando  $T \rightarrow \infty$ , tem-se que.

$$\tilde{p}(T) \rightarrow \infty, \quad \frac{\tilde{p}(T)^3}{T} \rightarrow 0 \quad \text{e} \quad \sqrt{T} \left[ \sum_{j=\tilde{p}(T)+1}^{\infty} |\phi_j| \right] \rightarrow 0,$$

onde  $\phi_i$  são os coeficientes da representação AR.

É assumido que estas condições são também satisfeitas pelo processo agregado  $Y_T$ , mas a ordem do processo ajustado para  $Y_T$  depende do tamanho da amostra reduzido,  $J$ , o qual é usualmente aproximado por  $T/m$ .

Os parâmetros dos processos AR ajustados, são estimados por mínimos quadrados usando uma realização de um processo que é independente do processo usado para previsão mais que tem a mesma estrutura estocástica.

A condição (1) não é muito restritiva e vale, por exemplo, se  $y_t$  tem uma representação ARMA finita. A condição (2) é satisfeita se por exemplo, a distribuição dos  $a_t$  é normal, e leva em conta muitos processos de ruído branco uma vez que só exige quarto momento finito além de independentes e identicamente distribuídos.

A condição (3) exige que a ordem dos processos ajustados cresça com o tamanho da amostra. Entretanto, esta ordem deve crescer muito mais lentamente que o tamanho da amostra.

Esta condição é sempre satisfeita se  $y_t$  é realmente um processo AR de ordem finita e  $\tilde{p}(T) \rightarrow \infty$ .

Para um processo AR de ordem infinita, esta condição pode impor um limite inferior sobre a velocidade pela qual  $\tilde{p}(T) \rightarrow \infty$ . Para um exemplo veja Lütkepohl (1987), p. 33.

$$\text{Em geral a condição } \sqrt{T} \left[ \sum_{j=\tilde{p}(T)+1}^{\infty} \phi_j \right] \xrightarrow{T \rightarrow \infty} 0 \text{ implica num}$$

limite inferior para a velocidade pela qual  $\tilde{p}(T) \rightarrow \infty$  quando  $T \rightarrow \infty$ ,

enquanto que  $\frac{\tilde{p}(T)^3}{T} \rightarrow 0$  quando  $T \rightarrow \infty$  implica num limite superior.

Previsor e erro médio quadrático de previsão obtidos sob as condições desta seção serão simbolizados com um ( $\sim$ ).

Os resultados a seguir são derivados do seguinte teorema dado em Lütkepohl (1985).

**Teorema:** Seja  $y_t = (y_{1t}, \dots, y_{kt})'$  um processo estocástico discreto  $k$ -dimensional com representação MA

$$y_t = \sum_{i=0}^{\infty} \theta_i a_{t-i}, \quad \theta_0 = I_k, \text{ onde}$$

$\Theta_i$  são matrizes de coeficientes  $(k \times k)$ ,  $I_k$  é a matriz identidade de ordem  $k$  e  $a_t = (a_{1t}, \dots, a_{kt})$  é ruído branco independente, isto é,  $Ea_t = 0$ ,  $Ea_t a_t' = \Xi$ ,  $a_t$  e  $a_s$  são independentes e identicamente distribuídos para  $t \neq s$  e é assumido que

$$E|a_{it} a_{jt} a_{lt} a_{mt}| < \infty, \text{ para } i, j, l, m \leq k,$$

$$\sum_{i=0}^{\infty} \sqrt{\text{tr}(\Theta_i \Theta_i')} < \infty.$$

Suponha que a ordem do  $AR(p)$  ajustado aos dados é uma função do tamanho da amostra,  $T$ , tal que, quando  $T \rightarrow \infty$ ,

$$p \rightarrow \infty, \quad \frac{p^3}{T} \rightarrow 0 \quad \text{e} \quad \sum_{i=p+1}^{\infty} \sqrt{\text{tr}(\Phi_i \Phi_i')} < \infty, \text{ onde}$$

$\Phi_i$  são matrizes de coeficientes  $(k \times k)$  da representação  $AR$  de  $y_t$ .

Então, para  $h \geq 1$ ,

$$\sqrt{\frac{T}{p}} \begin{bmatrix} \tilde{y}_t(1) - y_t(1) \\ \vdots \\ \tilde{y}_t(h) - y_t(h) \end{bmatrix} \text{ é assintoticamente } N\left[0, \theta(h)\right], \text{ onde}$$

$\tilde{y}_t(j)$  é o preditor  $j$  passos à frente baseado no  $AR(p)$  ajustado,  $T$  é o tamanho da amostra e  $y_t(j)$  é o preditor linear ótimo de  $y_t$   $j$  passos à frente e

$$\theta(h) = E \begin{bmatrix} y_{t+h} - y_t(1) \\ \vdots \\ y_{t+h} - y_t(h) \end{bmatrix} \begin{bmatrix} y_{t+h} - y_t(1) \\ \vdots \\ y_{t+h} - y_t(h) \end{bmatrix}' =$$

$$= \begin{bmatrix} 0 & & 0 \\ \sum_{i=0}^h \theta_i \Xi \theta'_i & \cdots & \sum_{i=0}^h \theta_i \Xi \theta'_{i+h-1} \\ \vdots & & \vdots \\ 0 & & h-1 \\ \sum_{i=0}^h \theta_{i+h-1} \Xi \theta'_i & \cdots & \sum_{i=0}^h \theta_i \Xi \theta'_i \end{bmatrix}$$

**Prova:** Lütkepohl (1985).

Do teorema acima, tem-se para  $h \geq 1$ ,

$$(3.4.2) \quad \sqrt{\frac{T}{p}} \begin{bmatrix} \tilde{y}_t(1) - y_t(1) \\ \vdots \\ \tilde{y}_t(h) - y_t(h) \end{bmatrix} \text{ é assintoticamente } N\left(0, \Sigma_y(h)\right),$$

onde  $\tilde{y}_t(j)$  é o predictor  $j$  passos à frente baseado no AR(p) ajustado,  $T$  é o tamanho da amostra e  $y_t(j)$  é o predictor linear ótimo de  $y_t$   $j$  passos à frente e  $\Sigma_y(h)$  é dada em (1.8.7).

Deste resultado, pode ser obtida uma aproximação para o EMQP assintótico dos previsores.

Escrevendo

$$(3.4.3) \quad \begin{bmatrix} y_{t+1} - \tilde{y}_t(1) \\ \vdots \\ y_{t+h} - \tilde{y}_t(h) \end{bmatrix} = \begin{bmatrix} a_{t+1} \\ a_{t+h} + \theta_1 a_{t+h-1} + \cdots + \theta_{h-1} a_{t+1} \end{bmatrix} + \begin{bmatrix} y_t(1) - \tilde{y}_t(1) \\ \vdots \\ y_t(h) - \tilde{y}_t(h) \end{bmatrix}$$

e sob a suposição de que as componentes do lado direito de (3.4.3) são independentes, pois, os  $a_{t+1}$  são independentes de  $y_t$  e dos coeficientes estimados, uma aproximação para a matriz de EMQP

assintótico conjunta para as previsões de 1 até h passos à frente está dada por

$$(3.4.4) \quad \tilde{\Sigma}_y(h) = E \begin{bmatrix} y_{t+1} - \tilde{y}_t(1) \\ \vdots \\ y_{t+h} - \tilde{y}_t(h) \end{bmatrix} \begin{bmatrix} y_{t+1} - \tilde{y}_t(1) \\ \vdots \\ y_{t+h} - \tilde{y}_t(h) \end{bmatrix}' \approx \left(1 + \frac{p}{T}\right) \Sigma_y(h) .$$

Deste resultado, o EMQP assintótico de  $Y_T^0(h)$  é dado por

$$(3.4.5) \quad \tilde{\sigma}_0^2(h) = v' \tilde{\Sigma}_y(mh) v = \left(1 + \frac{\tilde{p}(T)}{T}\right) v' \Sigma_y(mh) v ,$$

ou de (3.2.8)

$$(3.4.6) \quad \tilde{\sigma}_0^2(h) = \left(1 + \frac{\tilde{p}(T)}{T}\right) \sigma_0^2(h) .$$

Da mesma maneira, o erro médio quadrático de previsão assintótico de  $\tilde{Y}_T(h)$  é dado por

$$(3.4.4) \quad \tilde{\sigma}^2(h) = \left(1 + \frac{\tilde{p}(J)}{J}\right) \sigma^2(h) ,$$

onde J é o tamanho da amostra reduzido para o processo agregado.

Algumas consequências surgem de imediato, por exemplo, se  $\sigma_0^2(h) < \sigma^2(h)$ , então  $\tilde{\sigma}_0^2(h) < \tilde{\sigma}^2(h)$ , uma vez que T é suficientemente maior que J e pela condição (3).

Se  $y_t$  satisfaz a condição (3.2.9) do teorema 3.2.2, então

$\tilde{\sigma}_0^2(h) = \tilde{\sigma}^2(h)$ , para  $h = 1, 2, \dots$  e neste caso a diferença entre os EMQP assintóticos com  $J = T/m$  é dado por,

$$(3.4.7) \quad \tilde{\sigma}^2(h) - \tilde{\sigma}_0^2(h) = \left[ \frac{\tilde{p}(J)}{J} - \frac{\tilde{p}(T)}{T} \right] \sigma^2(h) = \frac{\sigma^2(h)}{T} \left[ m\tilde{p}(T/m) - \tilde{p}(T) \right].$$

Assumindo que  $\tilde{p}(T) = T^{1/(3+n)}$ , onde  $n > 0$ , (3.4.5) torna-se

$$\frac{\sigma^2(h)T^{1/(3+n)}}{T} \left[ m^{(2+n)/(3+n)} - 1 \right] \text{ que é sempre positiva uma vez que } m \geq 2.$$

Isto significa que mesmo sob condições mais favoráveis para o processo agregado temporalmente, se os EMQP dos previsores são iguais para o processo conhecido, o processo desagregado ainda deve ser preferido se a ordem e os coeficientes devem ser estimados.

Estes resultados, entretanto, apoiam-se na teoria assintótica e na escolha de  $\tilde{p}(\cdot)$ . Se  $\tilde{p}(\cdot)$  fosse uma função linear do seu argumento e  $J = T/m$ , então  $m\tilde{p}(J) - \tilde{p}(T) = 0$  e assim  $\tilde{\sigma}_0^2(h) = \tilde{\sigma}^2(h)$ .

Entretanto, a condição (3) exclui esta escolha de  $\tilde{p}(\cdot)$ , pois ela não satisfaz a condição  $\frac{\tilde{p}(T)^3}{T} \rightarrow 0$  quando  $T \rightarrow \infty$ .

Por outro lado, a condição (3) é uma exigência assintótica e portanto, em pequenas amostras a diferença  $m\tilde{p}(J) - \tilde{p}(T)$  pode ser negativa.

Consequentemente, a condição acima depende da escolha de  $p(\cdot)$ .

Critérios para escolha de modelos que estimam a ordem consistentemente, tais como o SC, satisfazem as condições desta seção e podem ser utilizados para escolher a ordem dos modelos ajustados.



### 3.5 - PREVISÃO COM PROCESSO NÃO ESTACIONÁRIO.

Nesta seção, serão considerados somente os processos que podem ser tornados estacionários após a aplicação de um operador diferença, isto é,

$$(3.5.1) \quad \nabla(B)y_t = x_t$$

é estacionário, onde

$$\nabla(B) = \begin{cases} (1-B)^d \text{ ou} \\ (1-B^s)^D \text{ ou} \\ (1-B)^d(1-B^s)^D \end{cases}.$$

Aqui,  $d$  e  $D$  são inteiros não negativos e representam o grau de diferença não sazonal e sazonal respectivamente, necessárias para tornar o processo estacionário e  $s$  é um inteiro positivo representando o período de sazonalidade.

Se  $\nabla(B) = (1-B)^d$ , então com  $B = B^m$  tal que,  $BY_\tau = Y_{\tau-1}$ , tem-se que,

$$(1-B)^d Y_\tau = (1-B^m)^d \left[ \sum_{i=0}^{m-1} B^i \right] y_{m\tau} = \left[ \sum_{i=0}^{m-1} B^i \right]^{d+1} (1-B)^d y_{m\tau} = \left[ \sum_{i=0}^{m-1} B^i \right]^{d+1} x_{m\tau},$$

que é estacionário, pois é uma transformação linear de processo estacionário.

Se  $\nabla(B) = (1-B^s)^D$ , o período de sazonalidade  $s$ , é considerado como um múltiplo do período de agregação  $m$  (o que na prática é razoável), então com  $d = s/m$  e  $B = B^m$ ,  $(1-B^s)^D Y_\tau$  é

estacionário.

Da mesma forma se  $(1-B)^d(1-B^s)^D y_t$  é estacionário, então  $(1-B)^d(1-B^s)^D Y_\tau$  é estacionário.

Isto significa que para tornar o processo agregado estacionário, o operador diferença simples é aplicado tantas vezes quanto no processo desagregado, enquanto que o operador diferença sazonal deve ser ajustado para o período causado pela agregação.

Por exemplo, se  $y_t$  é um processo mensal e  $Y_\tau$  é um processo trimestral, o operador diferença sazonal  $(1-B^{12})$  para  $y_t$  corresponde ao operador  $(1-B^4)$  para  $Y_\tau$ .

Suponha que o processo desagregado  $y_t$  é conhecido e estacionário após uma diferença, isto é,

$$(3.5.2) \quad x_t = y_t - y_{t-1} \quad \text{ou} \quad y_t = x_t + y_{t-1}, \quad \text{onde}$$

$x_t$  é um processo estacionário com representação MA

$$(3.5.3) \quad x_t = \sum_{i=0}^{\infty} \theta_i a_{t-i} = \theta(B)a_t, \quad \text{onde}$$

as raízes de  $\theta(B) = 0$  estão fora do círculo unitário.

De (3.5.2) pode-se escrever,

$$(3.5.4) \quad y_t = y_{t-m} + \sum_{j=1}^m x_{t-m+j}, \quad \text{onde}$$

$x_t$  é a parte estacionária de  $y_t$ .

O processo agregado por um período  $m$  é dado por,

$$\begin{aligned}
Y_{\tau} &= y_{m\tau} + y_{m\tau-1} + \dots + y_{m\tau-m+1} \\
&= \left[ y_{m\tau-m} + \sum_{j=1}^m x_{m\tau-m+j} \right] + \dots + \left[ y_{m\tau-m+1-m} + \sum_{j=1}^m x_{m\tau-m+1-m+j} \right] \\
&= Y_{\tau-1} + \sum_{i=0}^{m-1} \sum_{j=1}^m x_{m\tau-m-1+j} \\
&= Y_{\tau-1} + X_{\tau},
\end{aligned}$$

onde  $X_{\tau}$  é a parte estacionária do processo agregado  $Y_{\tau}$ .

O processo obtido agregando-se temporalmente o processo estacionário  $x_t$  é  $x_{m\tau} + x_{m\tau-1} + \dots + x_{m\tau-m+1}$ , que é diferente de  $X_{\tau}$ .

Portanto, para comparar os previsores  $Y_{\tau}(h)$  e  $Y_{\tau}^0(h)$ , não é suficiente comparar suas partes estacionárias.

De (3.5.2)

$$(3.5.5) \quad y_{t+h} = y_t + \sum_{j=1}^h x_{t+j},$$

portanto, da seção 1.6 tem-se que o preditor ótimo  $h$  passos do processo  $y_t$  baseado em  $S_t = \{ a_s | s \leq t \}$  é dado por

$$(3.5.6) \quad y_t(h) = y_t + \sum_{j=1}^h x_t(j), \text{ onde}$$

$x_t(j)$  é o preditor ótimo  $j$  passos de  $x_t$ , isto é,

$$(3.5.7) \quad x_t(j) = \sum_{i=0}^{\infty} \theta_{j+1-i} a_{t-i}.$$

O erro de previsão associado a  $y_t(h)$  é dado por

$$e(h) = y_{t+h} - y_t(h) = \sum_{j=1}^h \left[ x_{t+j} - x_t(j) \right] ,$$

que é a soma de erros de previsão do processo estacionário.

Assim o EMQP de  $y_t(h)$  é

$$(3.5.8) \quad \sigma_y^2(h) = V' \Sigma_x(h) V ,$$

onde  $\Sigma_x(h)$  é a matriz de EMQP de 1 até h passos para o processo estacionário  $x_t$  é dada por (1.8.7) e  $V'$  é um vetor de 1's de dimensão h.

É importante notar que

$$y_t(1) = y_t + x_t(1) \quad \text{e} \quad \sigma_y^2(1) = \Sigma_x(1) = \sigma_a^2 .$$

Se a parte estacionária de  $y_t$  tem média diferente de zero, isto é, se

$$(3.5.9) \quad x_t = \mu + \sum_{i=0}^{\infty} \theta_i a_{t-i} ,$$

não haverá mudança em (3.5.6), basta adicionar  $\mu$  em (3.5.7).

O previsor de  $Y_\tau$  baseado no processo totalmente agregado h passos à frente é dado por

$$(3.5.10) \quad Y_\tau(h) = Y_\tau + \sum_{j=1}^h X_\tau(j) ,$$

com EMQP dado por

$$(3.5.11) \quad \sigma_Y^2(h) = V' \Sigma_X(h) V ,$$

onde  $\Sigma_X(h)$  é a matriz de EMQP de 1 até h passos do processo agregado estacionário  $X_\tau$ .

Por outro lado, o previsor baseado no processo desagregado é dado por

$$(3.5.12) \quad Y_\tau^0(h) = Y_\tau + \sum_{j=1}^h X_\tau^0(j) ,$$

onde  $X_\tau^0(j) = \left[ \sum_{l=0}^{m-1} B^l \right] x_{m\tau}(j)$ , com EMQP dado por

$$(3.5.13) \quad \sigma_0^2(h) = v' \Sigma_X(h) v ,$$

onde  $v$  é como em (3.2.8).

O previsor baseado no processo desagregado é pelo menos tão bom quanto o previsor baseado no processo agregado, se o processo pode ser tornado estacionário por diferença.

Além disso, eles não se tornam idênticos com o crescimento do horizonte de previsão.

## CAPÍTULO 4

### ESTUDO DE MONTE CARLO

#### 4.1 - INTRODUÇÃO.

Os resultados obtidos no capítulo anterior, são baseados na teoria assintótica e portanto, um estudo de Monte Carlo foi conduzido para verificar o comportamento dos dois previsores,  $Y_{\tau}(h)$  e  $Y_{\tau}^0(h)$  em pequenas amostras.

O experimento de Monte Carlo foi realizado somente para o caso da seção 3.4, onde modelos autorregressivos são ajustados aos dados agregados e desagregados.

Este estudo é similar ao estudo de Monte Carlo feito por Lütkepohl (1987), seção 8.5, onde foram usados modelos AR e MA satisfazendo a condição 3.2.9 do teorema 3.2.2 e previsões de 1 e 5 passos à frente.

Aqui foi feito também um estudo para um modelo ARMA que não satisfaz a condição acima e as previsões são de 1 e 3 passos à frente, uma vez que, para grandes horizontes de previsão os EMQP se aproximam da variância do processo.

Realizações de processos AR, MA e ARMA foram geradas e os dois previsores com seus respectivos erro médio quadrático de previsão foram calculados.

A componente de ruído branco para o processo desagregado  $y_t$ , foi gerada como uma normal padrão independente usando o gerador de Wichmann & Hill (1982).

Para cada processo, 100 realizações foram geradas para três tamanhos de amostra,  $T = 60$ ,  $T = 100$  e  $T = 200$  e os previsores calculados para dois horizontes de previsão,  $h = 1$  e  $h = 3$ .

Para cada realização, mais de 50 valores pré-amostrais foram gerados para reduzir o impacto dos valores iniciais.

Alguns valores pré-amostrais foram usados para estimação. Por exemplo, para ajustar um modelo AR(2) aos dados, numa amostra de tamanho 60, são usados na realidade 62 valores, ou seja, são incluídos dois valores pré-amostrais na estimação.

Estimação e previsão foram baseadas na mesma amostra e a origem de previsão é sempre o último período amostral usado para estimação.

Note que neste aspecto, o estudo de Monte Carlo difere da suposição básica da teoria assintótica de processos independentes para estimação e previsão.

Um único processo para estimação e previsão é mais realístico do ponto de vista prático. Também na prática, é usualmente dada uma única amostra de tamanho  $T$ , e uma ordem para o AR deve ser escolhida com base nessa amostra.

Nesta situação, em vez de especificar uma função arbitrária  $\tilde{p}(T)$  para a ordem do AR, é mais razoável determinar a ordem através de um critério de especificação de modelos. Os critérios usados neste estudo foram descritos na seção 1.11.

A estatística usada para comparar os EMQP, é  $\hat{\rho}$  dada em Granger & Newbold (1986), seção 9.3.

Suponha que  $e^{(1)}$  e  $e^{(2)}$  são erros de previsão obtidos

usando-se dois previsores distintos e não viciados e assuma que  $(e^{(1)}, e^{(2)})$  tem distribuição normal bivariada com variâncias  $EMQP_1$  e  $EMQP_2$ .

Além disso, suponha que  $(e_1^{(1)}, e_1^{(2)})$ ,  $i = 1, 2, \dots, N$ , é uma amostra aleatória de tamanho  $N$  da distribuição de  $(e^{(1)}, e^{(2)})$ .

Definindo

$$(4.1.1) \quad \hat{\rho} = \frac{\sum_{i=1}^N (e_i^{(1)} + e_i^{(2)})(e_i^{(1)} - e_i^{(2)})}{\left[ \sum_{i=1}^N (e_i^{(1)} + e_i^{(2)})^2 \sum_{i=1}^N (e_i^{(1)} - e_i^{(2)})^2 \right]^{1/2}},$$

pode-se mostrar que, assintoticamente  $\hat{\rho}\sqrt{N}$  tem uma distribuição normal padrão, isto é,

$$(4.1.2) \quad \hat{\rho}\sqrt{N} \xrightarrow{D} N(0, 1),$$

se  $EMQP_1 = EMQP_2$ .

Então, neste estudo é usada a estatística  $\hat{\rho}\sqrt{N}$  para testar a hipótese

$$(4.1.3) \quad H_0: \sigma^2(h) = \sigma_0^2(h) \text{ contra } H_1: \sigma^2(h) \neq \sigma_0^2(h).$$

Para avaliar a significância da diferença entre os EMQP usando a estatística  $\hat{\rho}\sqrt{N}$ , os erros  $(e^{(1)}, e^{(2)})$  usualmente não tem uma distribuição normal exata.

Por exemplo, se os previsores são baseados em processos AR



estimados, os erros de previsão somente terão distribuição normal assintótica quando o tamanho da amostra vai para infinito.

Portanto, para  $\hat{\rho}\sqrt{N}$  ter distribuição normal assintótica, tanto o tamanho da amostra  $T$  como o número de replicações de experimentos de Monte Carlo  $N$ , devem ir para infinito.

#### 4.2 - PROCESSO AUTORREGRESSIVO.

Inicialmente foi investigado o processo autorregressivo de ordem 2, AR(2).

$$(4.2.1) \quad y_t = 1 + 0.8y_{t-2} + a_t, \text{ com } \sigma_a^2 = 1.$$

Para um período de agregação  $m = 2$ , o processo agregado temporalmente  $Y_\tau = y_{2\tau} + y_{2\tau-1}$ , é o AR(1)

$$(4.2.2) \quad Y_\tau = 2 + 0.8Y_{\tau-1} + u_\tau, \text{ com } \sigma_\tau^2 = 2,$$

onde  $u_\tau = a_{2\tau} + a_{2\tau-1}$ .

O processo desagregado  $y_t$ , satisfaz a condição (3.2.9) do teorema 3.2.2, tal que, os EMQP são idênticos para o processo conhecido.

As ordens máximas consideradas para os modelos AR ajustados foram,  $P = 8$  para o processo desagregado, e  $P = 4$  para o processo agregado.

A tabela 4.2.1, apresenta a distribuição de frequência das ordens escolhidas pelos três critérios. As ordens máximas não são

escolhidas com muita frequência para nenhum dos critérios, portanto esta ordens não são restritivas.

Como foi mencionado na seção 1.11, os critérios HQ e SC estimam a ordem corretamente com alta probabilidade, enquanto que o AIC superestima a ordem mesmo para  $T = 200$  (ou  $J = 100$  para o processo agregado).

Tabela 4.2.1: Distribuição de frequência das ordens estimadas.

Processo Desagregado					Processo Agregado				
T	p	p(AIC)	p(HQ)	p(SC)	J	p	p(AIC)	p(HQ)	p(SC)
60	0	0	0	0	30	0	3	5	5
	1	0	0	0		1	77	82	87
	2	73	85	94		2	11	7	7
	3	12	9	4		3	4	4	1
	4	6	4	2		4	5	2	0
	5	3	2	0					
	6	3	0	0					
	7	2	0	0					
	8	1	0	0					
100	0	0	0	0	50	0	0	0	0
	1	0	0	0		1	69	83	91
	2	69	85	93		2	22	12	9
	3	11	6	4		3	4	1	0
	4	6	6	3		4	5	4	0
	5	5	3	0					
	6	1	0	0					
	7	3	0	0					
	8	5	0	0					
200	0	0	0	0	100	0	0	0	0
	1	0	0	0		1	70	87	95
	2	71	90	95		2	14	8	3
	3	15	7	5		3	8	1	1
	4	6	3	0		4	8	4	1
	5	3	0	0					
	6	3	1	0					
	7	1	0	0					
	8	1	0	0					

As diferenças dos EMQP entre as séries agregada e desagregada ( $\sigma^2(h) - \sigma_0^2(h)$ ), são apresentadas na tabela 4.2.2.

**Tabela 4.2.2:** Diferenças de EMQP (  $\sigma^2(h) - \sigma_0^2(h)$  ).

Tamanho da amostra para o processo desagregado T	Horizonte de previsão h	Critério de escolha		
		AIC	HQ	SC
60	1	0.05	0.23**	0.32**
	3	0.12	0.40**	0.35**
100	1	0.08	0.05	0.05
	3	0.19	0.15	-0.01
200	1	-0.04	-0.04	-0.07
	3	-0.00	0.02	0.05

\*\* significativo ao nível de 1% .

Para quase todos os tamanhos de amostra e horizontes de previsão, os EMQP usando os dados desagregados são menores que os EMQP usando os dados agregados, isto é,  $\sigma^2(h) - \sigma_0^2(h)$  é positivo de acordo com os resultados assintóticos da seção 3.4.

Além disso, usando a estatística de teste da seção 4.1, apenas diferenças positivas são significantes aos níveis usuais.

Para  $T = 100$  e  $T = 200$ , não foi obtida diferença significativa entre os EMQP.

Para amostra de tamanho 60, com  $m = 2$  é reduzida para 30, que é um tamanho de amostra muito pequeno e a diferença entre os EMQP são significantes, como esperado, devido a forte influência das estimativas.

Para este caso, somente o critério AIC não detectou

diferença significativa.

As conclusões em pequenas amostras estão de acordo com os resultados assintóticos da seção 3.4.

Na seção 3.4 é admitido que os processos tenham uma representação AR de ordem infinita, portanto, foram também geradas realizações de um processo MA de ordem finita.

#### 4.3 - PROCESSO MÉDIA MÓVEL.

Outro processo usado neste estudo de Monte Carlo foi o MA(2)

$$(4.4.1) \quad y_t = 1 + a_t - 0.8a_{t-2}, \text{ com } \sigma_a^2 = 1.$$

Com período de agregação  $m = 2$ , o processo agregado temporalmente  $Y_\tau = y_{2\tau} + y_{2\tau-1}$  é o MA(1)

$$(4.4.2) \quad Y_\tau = 2 + u_\tau - 0.8u_{\tau-1}, \text{ com } \sigma_\tau^2 = 2, \text{ onde}$$

$$u_\tau = a_{2\tau} + a_{2\tau-1}.$$

Novamente  $y_t$  satisfaz a condição (3.2.9) do teorema 3.2.2 e portanto os EMQP são idênticos para os processos conhecidos.

As realizações do processo (4.4.1) foram geradas da mesma maneira como na seção anterior e as ordens máximas consideradas foram:  $P = 8$  para o processo desagregado e  $P = 4$  para o processo agregado.

A distribuição de frequência das ordens escolhidas pelos

três critérios, é apresentada na tabela 4.3.1.

Neste caso, a ordem do AR escolhida pelos três critérios crescem em média de acordo com a teoria da seção 3.4.

Tabela 4.3.1: Distribuição de frequência das ordens estimadas.

Processo Desagregado					Processo Agregado				
T	p	p(AIC)	p(HQ)	p(SC)	J	p	p(AIC)	p(HQ)	p(SC)
60	0	0	1	4	30	0	1	2	6
	1	0	0	0		1	10	12	28
	2	14	24	40		2	41	45	45
	3	0	0	0		3	19	18	11
	4	33	47	46		4	29	23	10
	5	3	1	1					
	6	21	15	6					
	7	4	1	1					
	8	25	11	2					
100	0	0	0	1	50	0	0	0	0
	1	0	0	0		1	5	8	17
	2	0	5	22		2	33	41	52
	3	0	0	0		3	23	23	18
	4	16	38	52		4	39	28	13
	5	0	1	1					
	6	37	38	18					
	7	5	0	0					
	8	42	18	6					
200	0	0	0	0	100	0	0	0	0
	1	0	0	0		1	0	0	1
	2	0	0	2		2	11	22	39
	3	0	0	0		3	22	31	30
	4	1	8	42		4	67	47	30
	5	0	0	0					
	6	15	29	34					
	7	2	3	2					
	8	82	60	20					

A tabela 4.3.2 apresenta as diferenças entre os EMQP dos dados agregado e desagregado. Algumas dessas diferenças são positivas e estão de acordo com a teoria assintótica da seção 3.3.

Tabela 4.3.2: Diferenças de EMQP (  $\sigma^2(h) - \sigma_0^2(h)$  )

Tamanho da amostra para o processo desagregado T	Horizonte de previsão h	Critério de escolha		
		AIC	HQ	SC
60	1	-0.17	-0.10	0.05
	3	-0.15	-0.12	0.03
100	1	0.15	0.23*	0.16
	3	-0.02	-0.06	-0.03
200	1	-0.05	-0.06	-0.02
	3	-0.02	-0.01	-0.02

\* significativo ao nível de 5%.

Para  $T = 100$  e  $h = 1$ , o modelo desagregado escolhido pelo HQ, produz previsão significativamente melhor que o correspondente modelo agregado.

Em muitos casos, a diferença é negativa, porém não é significativa ao nível de 5% e o ganho no EMQP é muito pequeno.

Portanto, os resultados para o processo (4.3.1) estão de acordo com a teoria assintótica.

#### 4.4 - PROCESSO AUTORREGRESSIVO MÉDIA MÓVEL.

Foi ainda considerado neste estudo de Monte Carlo, o processo ARMA(1,1)

$$(4.4.1) \quad y_t = 1 + 0.8y_{t-1} + a_t - 0.2a_{t-1}, \text{ com } \sigma_a^2 = 1.$$

O processo agregado temporalmente com período de agregação  $m = 2$  é, pelo teorema 2.4.1, o ARMA(1,1)

$$(4.4.2) \quad Y_{\tau} = 3.6 + 0.64Y_{\tau-1} + u_{\tau} + 0.5362u_{\tau-1}, \text{ com } \sigma_{\tau}^2 = 3.4316.$$

Ao contrário dos processos estudados anteriormente, o processo (4.4.1) não satisfaz a condição (3.2.9) do teorema 3.2.2, e mesmo para o caso de processos desconhecidos, é esperado que o processo desagregado produza melhores previsões.

As realizações do processo desagregado (4.4.1), foram geradas como nas seções 4.2 e 4.3.

As ordens máximas usadas para ajustar os AR ainda foram  $P = 8$  para o processo desagregado e  $P = 4$  para o processo agregado.

A distribuição de frequência das ordens escolhidas pelos três critérios são apresentadas na tabela 4.4.1.

Novamente as ordens máximas não são restritivas, pois, não são atingidas com muita frequência.

**Tabela 4.4.1:** Distribuição de frequência das ordens estimadas.

Processo Desagregado					Processo Agregado				
T	p	p(AIC)	p(HQ)	p(SC)	J	p	p(AIC)	p(HQ)	p(SC)
60	0	0	0	1	30	0	3	6	9
	1	60	68	88		1	75	79	86
	2	25	23	9		2	14	10	4
	3	5	4	2		3	3	2	0
	4	3	3	0		4	5	3	1
	5	5	2	0					
	6	1	0	0					
	7	0	0	0					
	8	1	0	0					
100	0	0	0	0	50	0	0	0	0
	1	45	63	81		1	69	77	85
	2	31	30	15		2	22	19	14
	3	8	4	4		3	4	2	1
	4	11	2	0		4	5	2	0
	5	1	1	0					
	6	1	0	0					
	7	2	0	0					
	8	1	0	0					
200	0	0	0	0	100	0	0	0	0
	1	28	47	64		1	69	84	94
	2	48	46	34		2	18	14	5
	3	9	4	1		3	6	2	1
	4	9	3	1		4	7	0	0
	5	3	0	0					
	6	2	0	0					
	7	0	0	0					
	8	1	0	0					

A tabela 4.4.2 apresenta as diferenças entre os EMQP dos dados agregados e desagregados.

Em muitos casos, essa diferença é significativa e positiva como esperado.

Para  $T = 100$  e  $h = 3$ , o modelo agregado escolhido pelo SC, tem menor EMQP que o correspondente modelo desagregado, embora a diferença seja não significativa.



Tabela 4.4.2: Diferenças de EMQP (  $\sigma^2(h) - \sigma_0^2(h)$  )

Tamanho da amostra para o processo desagregado T	Horizonte de previsão h	Critério de escolha		
		AIC	HQ	SC
60	1	0.46	0.63**	0.75**
	3	0.38	0.29	0.52*
100	1	0.44*	0.41*	0.45*
	3	0.08	0.05	-0.07
200	1	0.76**	0.64**	0.67**
	3	0.52**	0.23	0.18

\* significativo ao nível de 5%. \*\* significativo ao nível de 1%.

#### 4.5 - CONCLUSÃO.

Foi mostrado no capítulo 3 que, se o processo gerador dos dados é conhecido, o EMQP do previsor que usa o processo agregado, é maior ou igual ao EMQP do previsor baseado no processo desagregado.

Sob certas suposições, eles são iguais para qualquer horizonte de previsão.

Quando as ordens e/ou os coeficientes do processo gerador dos dados devem ser estimados de uma amostra avaliada, como é usualmente o caso prático, os resultados obtidos sugerem que ainda sejam usados os dados desagregados.

Os resultados das seções anteriores, mostram que a teoria assintótica da seção 3.3 pode ser um bom indicador da performance relativa dos previsores em pequenas amostras, quando a ordem do AR é escolhida por um dos três critérios apresentados na seção 1.11.

## REFERÊNCIAS

- Abraham, B. (1982) *Temporal aggregation and time series*. Int. Statist. Rev. 50, 285-291.
- Abraham, B. e Ledolter, J. (1982) *Forecast efficiency of systematically sampled time series*. Comm. Statist. - Theor. Meth. 51, 2857-2868.
- Akaike, H. (1974) *A new look at the statistical model identification*. IEEE Transactions on Automatica, AC-19, 716-723.
- Amemiya, T. e Wu, R. Y. (1972) *The effect of aggregation on prediction in the autoregressive model*. J. of the Amer. Statist. Assoc. N<sup>o</sup> 339, Vol 67, 628-632.
- Anderson, T. W. (1971) *The Statistical Analysis of Time Series*. New York ; Wiley.
- Box, G. E. P. e Jenkins, G. M (1976) *Time Series Analysis, Forecasting and Control*. Holden-Day ; San Francisco.
- Brewer, K. R. W. (1973) *Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models*. J. of Econometrics. Vol 1, 133-154.

- Chatfield, C. (1980) *The Analysis of Time Series: An Introduction*. 2<sup>a</sup> edição. Chapman and Hall. London.
- Engel, E. M. R. A. (1984) *A unified approach to the study of sums, products, time-aggregation and other functions of ARMA processes*. J. Time Series Anal. 5, 159-171.
- Fuller, W. A. e Hasza, D. P. (1981) *Properties of predictors for autoregressive time series*. Journal of the American Statistical Association, Vol 76, N<sup>o</sup> 373, 155-161.
- Granger, C. W. J. e Morris, M. J. (1976) *Time series modelling and interpretation*. J. R. Statist. Soc. A, 139, 246-257.
- Granger, C. W. J. e Newbold, P. (1986) *Forecasting Economic Time Series*. 2<sup>a</sup> Ed. Academic Press.
- Hannan, E. J. e Quinn, B. G. (1979) *The determination of the order of an autoregression*. Journal of the Royal Statistical Society B. Vol41, 190-195.
- Harvey, A. C. e Pierse, R. G. (1984) *Estimating missing observations in economic time series*. J. of the Amer. Statist. Assoc. N<sup>o</sup> 385, Vol 79, 125-131.
- James, Barry R. (1981) *Probabilidade : un curso em nível intermediário*. Projeto Euclides. Rio de Janeiro.

- Kohn, R. (1982) *When is an aggregate of a time series efficiently forecast by its past?* J. of Econometrics. 18, 337-349.
- Lewis, R. e Reinsel, G. (1982) *Prediction of multivariate time series by autoregressive model fitting.* Amer. Statist. Assoc., Proceedings of the Business and Economic Statistical Section, 149-154.
- Lütkepohl, H (1985) *The joint asymptotic distribution of multistep prediction erros of estimated vector autoregressions.* Economics Letters. 17, 103-106.
- Lütkepohl, H. (1986) *Comparison of predictors for temporally and contemporaneously aggregated time series.* Int. Jour. of Forecasting. 2, 461-475.
- Lütkepohl, H. (1987) *Forecasting Aggregated Vector ARMA Processes.* Springer-Verlag.
- Morettin, P. A. e Tolói, C. M. C. (1985) *Previsão de Séries Temporais.* Ed. Atual.
- Pereira, P. L. V. (1981) *Loss of efficiency in temporal aggregation of economic time series I : the noseasonal case.* II Latin American Meeting of the Econometric Society. Julho de 1981. Rio de Janeiro - RJ.
- Pereira, P. L. V. (1982) *Loss of efficiency in temporal aggregation of*

*economic time series II : the seasonal case.* III Latin American Meeting of the Econometric Society. Julho de 1982. Cidade do Mexico - Mexico.

Pino, F. A., Morettin, P. A. e Mentz, R. P. (1987) *Modelling and forecasting linear combinations of time series.* Int. Statist. Rev. N<sup>o</sup> 55, Vol 3, 295-313.

Quinn, B. G. (1980) *Order determination for a multivariate autoregression.* Journal of the Royal Statistical Society B. Vol 42, 182-185.

Reinsel, G. (1980) *Assymptotic properties of prediction errors for the multivariate autoregressive model using estimated parameters.* Journal of the Royal Statistical Society. Vol 42, N<sup>o</sup> 3, 328-333.

Rissanen, J. (1978) *Modeling by shortest data description.* Automatica, 14, 465-471.

Rose, D. E. (1977) *Forecasting aggregates of independent ARIMA processes.* J. of Econometrics. 5, 323-345.

Schwarz, G. (1978) *Estimating the dimension of a model.* Annals of Statistics, 6, 461-464.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics.* John Wiley e Sons.

- Stram, D. O. e Wei, W. W. S. (1986) *Temporal aggregation in the ARIMA process*. J. of Time Series Anal. N<sup>o</sup> 4, Vol 7, 279-292.
- Telser, L. G. (1967) *Discrete samples and moving sums in stationary stochastic processes*. J. of the Amer. Statist. Assoc. N<sup>o</sup> 318, Vol 62, 482-499.
- Tiao, G. C. (1972) *Asymptotic behaviour of temporal aggregates of time series*. Biometrika. N<sup>o</sup> 59, Vol 3, 525-531.
- Tiao, G. C. e Wei, W. S. (1976) *Effect of temporal aggregation on the dynamic relationship of two time series variables*. Biometrika. 63, 513-523.
- Wei, W. W. S. (1979) *Some consequences of temporal aggregation in seasonal time series models*. In Seasonal Analysis of Economic Time Series. 443-448.
- Wei, W. W. S. e Abraham, B. (1981) *Forecasting contemporal time series aggregates*. Comm. Statist. A, Vol 10, 1335-1344.
- Wichmann, B. A. e Hill, I. D. (1982) *An efficient and portable pseudo-random number generator (Algorithm AS 183)*. Applied Statistics. Vol 31, 188-190.