

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
DEPARTAMENTO DE ESTATÍSTICA

Bade  
OK

**Modelos de Previsão de Séries de Tempo:  
ARIMA e Modelos Estruturais**

**Patrícia Sorgatto Kuyven**

**Orientador: Marcelo Savino Portugal**

UFRGS  
SISTEMA DE BIBLIOTECAS  
BIBLIOTECA SETORIAL DE MATEMÁTICA

**Porto Alegre, Fevereiro de 1999 .**

UFRGS  
SISTEMA DE BIBLIOTECAS  
BIBLIOTECA SETORIAL DE MATEMÁTICA

# ÍNDICE

<b>1. INTRODUÇÃO.....</b>	<b>1</b>
<b>1.1 PREVISÕES DE SÉRIES TEMPORAIS.....</b>	<b>1</b>
<b>1.2 OBJETIVOS.....</b>	<b>2</b>
<b>1.3 ESTRUTURA DO TRABALHO.....</b>	<b>2</b>
<b>2. MODELOS ARIMA: METODOLOGIA DE BOX &amp; JENKINS.....</b>	<b>4</b>
<b>2.1 DEFINIÇÕES INICIAIS.....</b>	<b>4</b>
<b>2.2 MODELOS LINEARES ESTACIONÁRIOS.....</b>	<b>8</b>
<b>2.2.1 Modelo Linear Geral.....</b>	<b>8</b>
<b>2.2.2 Condições de Estacionariedade e Invertibilidade de Modelos Lineares.....</b>	<b>9</b>
<b>2.2.3 Modelos Autoregressivos.....</b>	<b>10</b>
<b>2.2.4 Modelos Médias Móveis.....</b>	<b>14</b>
<b>2.2.5 Modelos Mistos Autoregressivos – Médias Móveis.....</b>	<b>15</b>
<b>2.3 MODELOS LINEARES NÃO ESTACIONÁRIOS.....</b>	<b>16</b>
<b>2.4 SAZONALIDADE.....</b>	<b>16</b>
<b>2.5 IDENTIFICAÇÃO DO MODELO.....</b>	<b>17</b>
<b>2.6 ESTIMAÇÃO DOS PARÂMETROS DO MODELO.....</b>	<b>19</b>
<b>2.7 VERIFICAÇÃO DO MODELO ESTIMADO.....</b>	<b>20</b>
<b>2.8 PREVISÃO DE VALORES FUTUROS.....</b>	<b>21</b>
<b>2.8.1 Cálculo da Previsão de Erro Quadrático Mínimo.....</b>	<b>21</b>
<b>3. MODELOS ESTRUTURAIS.....</b>	<b>24</b>
<b>3.1 Introdução.....</b>	<b>24</b>
<b>3.1.1 Média Móvel Ponderada Exponencialmente.....</b>	<b>24</b>

3.1.2 Procedimento de Previsão Holt-Winters para Tendência Linear Local.....	26
3.1.3 Mínimos Quadrados Ponderados (DLS).....	27
3.1.4 Sazonalidade.....	29
3.2 A ESTRUTURA DOS MODELOS DE SÉRIES TEMPORAIS.....	32
3.2.1 Tendência.....	34
3.2.2 Ciclos.....	36
3.2.3 Sazonalidade.....	37
3.2.4 Modelos Estruturais.....	40
3.3 TESTES E SELEÇÃO DE MODELOS.....	44
3.3.1 Seleção de Modelos.....	44
3.3.2 Testes de especificação baseados nos resíduos.....	46
3.3.3 Qualidade do ajuste.....	48
3.3.4 Comparação entre modelos.....	50
3.3.5 Testes preditivos e avaliação do modelo.....	51
3.4 A FORMA DE ESPAÇO DE ESTADO E FILTRO DE KALMAN.....	53
3.4.1 A forma de espaço de estado.....	53
3.4.2 O filtro de Kalman.....	54
4. APLICAÇÃO DAS METODOLOGIAS ARIMA E MODELOS ESTRUTURAIS.....	57
4.1 METODOLOGIA ARIMA.....	57
4.2 MODELOS ESTRUTURAIS.....	59
4.3 COMPARAÇÃO DOS RESULTADOS EMPÍRICOS.....	64
5. CONCLUSÃO.....	68
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	69

# 1. INTRODUÇÃO

## 1.1 Previsões de Séries de Tempo

A Econometria é o ramo da Ciência Econômica que, através da análise estatística da realidade, busca estabelecer proposições econômicas, de caráter quantitativo, que possibilitam não somente compreender e analisar o comportamento das variáveis sob uma determinada base econômica teórica, como também permitem a previsão de seus comportamentos futuros.

Tendo em vista que a maioria das decisões no âmbito gerencial são tomadas levando em consideração fatores que não podem ser obtidos no momento da decisão, os modelos de previsão tornam-se ferramentas indispensáveis, já que, a partir de séries históricas pode-se *projetar* o comportamento destes fatores. O conjunto das previsões passa diretamente por todos os campos da administração, finanças, marketing, produção, pesquisas operacionais, economia, e administração pública.

Historicamente, as primeiras previsões de séries econômicas eram feitas a partir de procedimentos *ad hoc* (Holt-Winters). Box & Jenkins, em 1976, deram origem aos modelos ARIMA, nos quais as séries de tempo eram ajustadas através de um modelo estatístico, porém sem a estrutura descritiva que os antigos procedimentos continham. Em 1989, Harvey introduziu os Modelos Estruturais, os quais conseguiram descrever os componentes das séries com todo o embasamento estatístico necessário para que fossem cientificamente aceitos.

A obtenção dos modelos de previsão de séries de tempo, neste trabalho, estão baseadas nos dois métodos estatísticos citados acima: metodologia ARIMA (Box & Jenkins, 1976) e metodologia de Modelos Estruturais (Harvey, 1989).

## **1.2 Objetivos**

O presente trabalho tem por objetivos:

a) Reunir os principais aspectos de duas metodologias para a obtenção de modelos de previsão de séries temporais: ARIMA e Modelos Estruturais.

b) Aplicar as duas metodologias a fim de obter modelos de previsões para duas séries históricas de indicadores econômicos.

c) A partir dos resultados das previsões fazer uma comparação da eficiência das duas metodologias analisadas para cada uma das séries trabalhadas.

## **1.3 Estrutura do Trabalho**

Este trabalho é composto dos seguintes capítulos:

1. Introdução: uma apresentação dos objetivos e organização do trabalho.

2. Modelos ARIMA: Esta capítulo descreve detalhadamente a metodologia ARIMA.

3. Modelos Estruturais: Neste capítulo são discorridos mais profundamente a técnica de Modelos Estruturais para previsão de séries de tempo univariadas e sua aplicabilidade.

4. Aplicação das metodologias: O capítulo 4 é uma aplicação prática das duas metodologias mencionadas acima. Foram analisadas as séries do PIB brasileiro e do IDI/RS visando obter previsões, as quais foram comparadas quanto ao nível de aproximação destes mesmos valores ocorridos na realidade. Para a Metodologia de Box & Jenkins foi utilizado o software ITSM, enquanto para a análise através de Modelos Estruturais, o software utilizado foi o STAMP 5.0 .

## 2. MODELOS ARIMA: METODOLOGIA DE BOX & JENKINS

### 2.1 Definições Iniciais

Séries Temporais: Uma série temporal é um conjunto de observações no tempo. Neste trabalho serão consideradas apenas as séries temporais discretas onde as observações são feitas em intervalos de tempo iguais:  $y_0, y_1, y_2, \dots, y_n$ . Se os valores futuros da série ( $y_{n+1}, y_{n+2}, \dots$ ) puderem ser descritos somente em termos de uma distribuição probabilística, a série de tempo é dita não determinística.

Processos Estocásticos: são fenômenos estatísticos que se desenvolvem no tempo de acordo com um processo probabilístico.

Processos Estocásticos Estacionários: Um processo estocástico é dito estritamente ( ou fortemente ) estacionário se as suas propriedades não forem afetadas por diferentes tempo de origem; isto é, se todos os momentos são constantes no tempo

$$F(y_{t+h}, y_{t+h+1}, \dots, y_{t+h+m}) = F(y_t, y_{t+1}, \dots, y_{t+m}), \text{ para } \forall h \text{ e } \forall m .$$

Quando  $m=0$ , a estacionariedade implica que a distribuição de probabilidade  $p(y_t)$  é a mesma para todos os tempos  $t$  e pode ser escrita como  $p(y)$ . Assim, o processo tem uma média constante

$$\mu = E[y_t] = \int_{-\infty}^{+\infty} yp(y)dy \quad (2.1)$$

a qual define o nível da série, e tem uma variância constante

$$\sigma^2 = E[(y_t - \mu)^2] = \int_{-\infty}^{+\infty} (y - \mu)^2 p(y) dy \quad (2.2)$$

Como a distribuição de probabilidade  $p(y)$  é a mesma em todos tempos  $t$ , este modelo pode ser representado por um histograma das observações  $y_1, y_2, \dots, y_N$ , produzindo as observações da série temporal. Então, a média  $\mu$  do processo estocástico pode ser estimada por

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t \quad (2.3)$$

e a variância  $\sigma_y^2$  do processo, pode ser estimada por

$$\sigma_y^2 = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^2. \quad (2.4)$$

Coeficientes de autocovariância e autocorrelação: A hipótese de estacionariedade implica que a distribuição conjunta  $p(y_{t_1}, y_{t_2})$  é a mesma para todos  $t_1$  e  $t_2$ .

A covariância entre  $y_t$  e seu valor  $y_{t+k}$ , separados por  $k$  intervalos de tempo, é chamada de autocovariância de lag  $k$  e definida por

$$\gamma_k = cov[y_t, y_{t+k}] = E[(y_t - \mu)(y_{t+k} - \mu)] \quad (2.5)$$

Similarmente, a autocorrelação de lag  $k$  é

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2] E[(y_{t+k} - \mu)^2]}} = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sigma_y^2} \quad (2.6)$$

Já que, para um processo estacionário, a variância  $\sigma_y^2 = \gamma_0$  é a mesma tanto para o tempo  $t+k$  como para o tempo  $t$ , a autocorrelação no lag  $k$  é

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (2.7)$$



O que implica que  $\rho_0 = 1$ .

Propriedades da autocorrelação dos processos estacionários:

$$i) \gamma_k \leq \gamma_0, \forall k$$

$$ii) \gamma_k = \gamma_{-k}, \forall k$$

$$iii) \gamma_0 \geq 0$$

$$iv) -1 \leq \rho_k \leq 1$$

Ruídos Brancos: Um processo  $y_t$  será um Ruído Branco (RB) se satisfizer as seguintes condições:

$$i) E[y_t] = \mu, \text{ para } \forall t$$

$$ii) \text{var}[y_t] = \sigma^2, \text{ para } \forall t$$

$$iii) \text{cov}[y_i, y_j] = 0, \text{ para } \forall i \neq j$$

Estacionariedade Fraca: Um processo estacionário será fracamente estacionário ou estacionário de segunda ordem se

$$i) E[y_t] = \mu, \text{ para } \forall t$$

$$ii) E[y_t^2] < \infty, \text{ para } \forall t$$

$$iii) \text{cov}[y_i, y_j] = \gamma_{|j-i|} = \gamma_k, \text{ para } \forall i \text{ e } \forall j$$

Funções de autocovariância e autocorrelação: O gráfico de  $\gamma_k$  versus o lag  $k$ , é chamado de função de autocovariância  $\{\gamma_k\}$  do processo estocástico. Similarmente o gráfico de  $\rho_k$  versus o lag  $k$ , é chamado função de autocorrelação do processo. Já que  $\gamma_k = \sigma_y^2 \rho_k$ , o conhecimento da função de autocorrelação

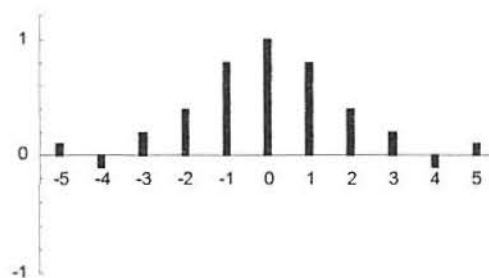
$\{\rho_k\}$  e da variância  $\sigma_y^2$  é equivalente ao conhecimento da função de autocovariância  $\{\gamma_k\}$ .

Exemplo: Seja o processo estacionário  $y_t$ , com os seguintes coeficientes de autovariância e de autocorrelação:

Lag (k)	Coefficiente autocorrelação
-5	0.1
-4	-0.1
-3	0.2
-2	0.4
-1	0.8
0	1
1	0.8
2	0.4
3	0.2
4	-0.1
5	0.1

Sua função de autocorrelação é representada por:

#### Função de autocorrelação



Estimação das Funções de Autocovariância e Autocorrelação: Sob a suposição de que a série  $y_t$  seja estacionária e observada nos pontos  $t = 1, 2, \dots, N$ , a estimativa da função de autocorrelação no lag  $k$  é dada por

$$r_k = \frac{c_k}{c_0} \text{ onde } c_k = \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 0, 1, 2, \dots, K$$

é a estimativa da função de autocovariância no lag  $k$ , e  $\bar{y}$  é a média da série temporal.

Obs.: Na prática, para obter uma boa estimativa da função de autocorrelação, é necessário utilizar no mínimo 50 observações da série temporal e calcular  $r_k$  para  $k = 0, 1, \dots, K$ , onde  $K$  não deve ultrapassar  $N/4$ .

Erro Padrão das Estimativas de Autocorrelação: Para identificar um modelo para uma série temporal, utilizando os métodos que serão descritos mais adiante, é necessário checar se  $\rho_k$  é efetivamente zero num determinado lag. Para isso, usa-se para a variância do coeficiente de autocorrelação uma aproximação de um processo estacionário Normal (Bartlett-1946).

$$\text{var}[r_k] \cong \frac{1}{N} \sum_{v=-\infty}^{+\infty} \{ \rho_v^2 + \rho_{v+k} \rho_{v-k} - 4\rho_k \rho_v \rho_{v-k} + 2\rho_v^2 \rho_k^2 \} \quad (2.8)$$

## 2.2 Modelos Lineares Estacionários

### 2.2.1 Modelo Linear Geral

Um modelo estocástico linear geral é descrito quando supõe-se que uma série temporal pode ser generalizada por um agregado linear de choques aleatórios.

Uma das representações desse processo estocástico, é a saída de um filtro linear (ou sistema linear) e tem como entrada um ruído branco ( $a_t$ ):

$$\tilde{y}_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \quad (2.9)$$

onde  $\tilde{y}_t = y_t - \mu$  é o afastamento do processo de sua origem, ou de sua média, se o processo é estacionário. O modelo linear geral permite-nos representar  $\tilde{y}_t$

como uma soma do valor presente e dos pesos dos valores passados de um processo ruído branco ( $a_t$ ).

O modelo descrito acima implica que, sob certas condições,  $\tilde{y}_t$  é uma soma ponderada de valores passados de  $\tilde{y}_t$  mais um choque aleatório adicional, ou seja

$$\tilde{y}_t = \pi_1 \tilde{y}_{t-1} + \pi_2 \tilde{y}_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j \tilde{y}_{t-j} + a_t \quad (2.10)$$

A relação entre os pesos  $\psi$  e os pesos  $\pi$  pode ser obtida usando o operador de defasagem para o passado, denotado por  $B$  e definido por

$$By_t = y_{t-1} \quad B^j y_t = y_{t-j} \quad (2.11)$$

E também, usando o operador de defasagem para o futuro  $F = B^{-1}$ , definido por

$$Fy_t = y_{t+1} \quad F^j y_t = y_{t+j} \quad (2.12)$$

Veja que (9) pode ser expressa como  $\tilde{y}_t = \psi(B)a_t$  e (10) pode ser expressa como  $\pi(B)\tilde{y}_t = a_t$ , então:

$$\psi(B)\pi(B)\tilde{y}_t = \psi(B)a_t = \tilde{y}_t$$

Consequentemente,

$$\psi(B)\pi(B) = 1 \quad \text{ou} \quad \pi(B) = \psi^{-1}(B) \quad (2.13)$$

A relação (2.13) pode ser usada para obter os pesos  $\pi$ , em função dos pesos  $\psi$  e vice-versa.

## 2.2.2 Condições de Estacionariedade e Invertibilidade de Modelos Lineares

Estacionariedade: Um modelo linear é estacionário quando o operador  $\psi(B)$  converge para  $|B| \leq 1$ . Isto é, dentro e sobre o círculo unitário.

Invertibilidade: Um modelo linear é invertível quando o operador  $\pi(B)$  converge para  $|B| \leq 1$ . Isto é, dentro e sobre o círculo unitário.

### 2.2.3 Modelos Autoregressivos

O processo  $\tilde{y}_t$  será um processo Autoregressivo de ordem  $p$ , denotado por AR(p), quando satisfizer a relação

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \dots + \phi_p \tilde{y}_{t-p} + a_t \text{ ou } (1 - \phi_1 B - \dots - \phi_p B^p) \tilde{y}_t = \phi(B) \tilde{y}_t = a_t \quad (2.14)$$

Este processo será estacionário se as raízes do polinômio  $\phi(B) = 0$  estiverem fora do círculo unitário.

E visto que,

$$\pi(B) = \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.15)$$

é finito, não há restrições para que um processo autoregressivo seja invertível.

Para obter a função de autocorrelação de um processo autoregressivo, primeiramente multiplicamos  $\tilde{y}_t$  por  $\tilde{y}_{t-k}$  e verificamos a sua esperança:

$$\gamma_k = E[\tilde{y}_{t-k} \tilde{y}_t] = E[\phi_1 \tilde{y}_{t-k} \tilde{y}_{t-1} + \dots + \phi_p \tilde{y}_{t-k} \tilde{y}_{t-p} + \tilde{y}_{t-k} a_t] \quad (2.16)$$

Como  $E[\tilde{y}_{t-k} a_t] = 0$  para  $k > 0$ , podemos escrever (16) por

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

Agora, dividindo (2.17) por  $\gamma_0$ , obtemos a equação da função de autocorrelação:

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

Podemos escrever os parâmetros autoregressivos em termos das autocorrelações (Equações de Yule-Walker) substituindo  $k = 1, 2, 3, \dots, p$  em (2.17):

$$\begin{aligned} \rho_1 &= \phi_1 & + \phi_2 \rho_1 & + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 & + \phi_2 & + \dots + \phi_p \rho_{p-2} \\ \vdots & \vdots & \vdots & \dots \vdots \\ \rho_p &= \phi_1 \rho_{p-1} & + \phi_2 \rho_{p-2} & + \dots + \phi_p \end{aligned} \quad (2.19)$$

Para  $k=0$ , o termo  $E[\tilde{y}_{t-k} a_t]$  é igual a  $E[a_t^2] = \sigma_a^2$ , então para  $k=0$ ,

$$\sigma_y^2 = \frac{\sigma_a^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2 - \dots - \rho_k \phi_k} \quad (2.20)$$

Exemplo 1: O modelo autoregressivo de 1ª ordem, AR(1) é dado por

$$\begin{aligned} \tilde{y}_t &= \phi_1 \tilde{y}_{t-1} + a_t \\ &= a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \dots \end{aligned} \quad (2.21)$$

onde  $-1 < \phi_1 < +1$  para que o processo seja estacionário.

Pelas equações de Yule-Walker temos:

$$\begin{aligned} \rho_1 &= \phi_1 \\ \rho_2 &= \phi_1 \rho_1 = \phi_1^2 \\ \rho_3 &= \phi_1 \rho_2 = \phi_1^3 \\ &\dots \\ \rho_k &= \phi_1 \rho_{k-1} = \phi_1^k \quad \text{para } k > 0 \end{aligned}$$

E como  $\rho_0 = 1$ , então para um modelo autoregressivo de primeira ordem, AR(1), temos que

$$\rho_k = \phi_1^k, \quad \text{para } k \geq 0 \quad (2.22)$$

com variância

$$\sigma_y^2 = \frac{\sigma_a^2}{1 - \phi_1^2} \quad (2.23)$$

Exemplo 2: O modelo autoregressivo de 2ª ordem, AR(2) é dado por

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + a_t \quad (2.24)$$

onde  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0$  para que o processo seja estacionário, ou seja,

$$\phi_2 + \phi_1 < 1$$

$$\phi_2 - \phi_1 < 1$$

$$-1 < \phi_2 < 1$$

Pelas equações de Yule-Walker temos:

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

Então, para um modelo autoregressivo de 2ª ordem, AR(2), temos

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad \text{e} \quad \rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \quad (2.25)$$

com variância

$$\sigma_y^2 = \left( \frac{1 - \phi_2}{1 + \phi_2} \right) \frac{\sigma_a^2}{[(1 - \phi_2)^2 - \phi_1^2]} \quad (2.26)$$

Função de Autocorrelação Parcial: Box & Jenkins (1970) propõem a utilização da função de autocorrelação parcial para facilitar o processo de identificação do modelo de uma série temporal. A f.a.c.p. é denotada por  $\phi_{kj}$ , o  $j$ ésimo coeficiente de um modelo AR(p), de tal modo que  $\phi_{kk}$  seja o último coeficiente. Para um modelo AR(p),  $\phi_{kk}$  é obtida por

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (2.27)$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

Em geral, para  $\phi_{kk}$ , o determinante do numerador tem os mesmos elementos do denominador, porém a última coluna é substituída pelo vetor de autocorrelações.

Para um processo autoregressivo de ordem  $p$ , a função de autocorrelação parcial  $\phi_{kk}$  será diferente de zero para  $k$  menores ou igual a  $p$  e zero para  $k$  maior que  $p$ . Ou seja, a facp num processo autoregressivo de ordem  $p$  tem um corte a partir do lag  $p$ .

A estimativa da facp pode ser obtida de duas formas: a primeira consiste em estimar, sucessivamente, modelos autoregressivos de ordens  $p = 1, 2, 3, \dots$  por mínimos quadrados e tomar as estimativas do último coeficiente de cada ordem; a outra maneira consiste resolver as equações abaixo para  $k = 1, 2, 3, \dots$

$$r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{k(k-1)}r_{j-k+1} + \hat{\phi}_{kk}r_{j-k} \quad j = 1, 2, \dots, k \quad (2.28)$$

onde  $r$  é a estimativa da fac.

Sob a hipótese de que um processo é AR( $p$ ), Quenouille(1949) mostra que as facp estimadas de ordem  $p+1, p+2, \dots$ , são aproximadamente independentemente distribuídas, com

$$\text{var}(\hat{\phi}_{kk}) \cong \frac{1}{N} \quad k \geq p+1 \quad (2.29)$$



Deste modo, se o número de observações  $N$  é suficientemente grande,  $\phi_{kk}$  tem distribuição aproximadamente normal, o que permite a construção de intervalos de confiança para  $\phi_{kk}$ .

#### 2.2.4 Modelos de Médias Móveis

O processo  $\tilde{y}_t$  será um processo de médias móveis de ordem  $q$ , denotado por MA( $q$ ), quando satisfizer a relação

$$\begin{aligned}\tilde{y}_t &= a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \\ \tilde{y}_t &= (1 - \theta_1 B - \dots - \theta_q B^q) a_t \\ \tilde{y}_t &= \theta(B) a_t\end{aligned}\tag{2.30}$$

Este processo será invertível se as raízes do polinômio  $\theta(B) = 0$  estiverem fora do círculo unitário.

E visto que,

$$\psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p\tag{2.31}$$

é finito, não há restrições para que um processo de médias móveis seja estacionário.

Para obter a função de autocorrelação de um processo médias móveis, usamos a função de autocovariância do processo

$$\gamma_k = E[(a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \dots - \theta_q a_{t-k-q})]\tag{2.32}$$

Então, a variância do processo é

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2\tag{2.33}$$

e

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q)\sigma_a^2 & k=1,2,\dots,q \\ 0 & k > q \end{cases} \quad (2.34)$$

Deste modo, a  $\hat{\rho}_k$  é

$$\hat{\rho}_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k=1,2,\dots,q \\ 0 & k > q \end{cases} \quad (2.35)$$

Nota-se então que  $\hat{\rho}_k$  num processo de médias móveis de ordem  $q$  tem um corte a partir do lag  $q$ .

Exemplo 1: O modelo médias móveis de 1ª ordem, MA(1) é dado por

$$\begin{aligned} \tilde{y}_t &= a_t - \theta_1 a_{t-1} \\ &= (1 - \theta_1 B)a_t \end{aligned} \quad (2.36)$$

onde  $-1 < \theta_1 < +1$  para que o processo seja invertível.

Usando (33), a variância do processo é  $\gamma_0 = (1 + \theta_1^2)\sigma_a^2$  e usando (35), a função de autocorrelação é

$$\hat{\rho}_k = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2} & k=1 \\ 0 & k \geq 2 \end{cases} \quad (2.37)$$

De onde obtém-se, com  $k=1$ ,

$$\theta_1^2 + \frac{\theta_1}{\hat{\rho}_1} + 1 = 0 \quad (2.38)$$

### 2.2.5 Modelos Mistos Autoregressivos - Médias Móveis

O processo  $\tilde{y}_t$  será um processo autoregressivo e de médias móveis de ordem  $q$ , denotado por ARMA(p,q), quando satisfizer a relação

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \dots + \phi_p \tilde{y}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

isto é,  $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{y}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$

ou 
$$\phi(B) \tilde{y}_t = \theta(B) a_t \quad (2.39)$$

Este processo será invertível se as raízes do polinômio  $\theta(B) = 0$  estiverem fora do círculo unitário, e será estacionário se as raízes do polinômio  $\phi(B) = 0$  estiverem fora do círculo unitário.

### 2.3 Modelos Lineares Não Estacionários

Box & Jenkins introduziram não estacionariedade em seus modelos assumindo que sucessivas diferenças na série original produzem uma série estacionária, isto é, séries que se desenvolvem no tempo ao redor de uma média constante.

Por exemplo, a série  $y_t = y_{t-1} + a_t - \theta a_{t-1}$ , é uma série não estacionária, mas a 1ª diferença  $w_t = \nabla y_t = y_t - y_{t-1} = a_t - \theta a_{t-1}$  é um processo de médias móveis estacionário.

Em geral, se assumimos que a d-ésima diferença  $w_t = \nabla^d \tilde{y}_t = \nabla^d y_t = (1 - B)^d y_t$  é um processo ARMA (p,q) estacionário, então  $\phi(B) \nabla^d y_t = \theta(B) a_t$  define um processo chamado Modelo ARIMA (p,d,q) (*Autoregressivo Integrado Médias Móveis de ordem p,d,q*), onde p e q são as ordens de  $\phi(B)$  e de  $\theta(B)$ , respectivamente.

### 2.4 Sazonalidade

Podem ocorrer em uma série temporal, componentes periódicos que

tendem a repetir-se com um período de S unidades de tempo. Isto significa que o comportamento estrutural subjacente do processo pode ser descrito olhando a série indexada em incrementos de algum período sazonal subjacente S. Nestes casos, pode-se ajustar as equações de diferenças em múltiplos deste componente sazonal. Estas equações podem incorporar ao modelo, o operador autoregressivo sazonal de ordem P:

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS}, \text{ estacionário}; \quad (2.40)$$

o operador de médias móveis sazonal de ordem Q:

$$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_q B^{qS}, \text{ invertível}; \quad (2.41)$$

e o operador diferença sazonal:

$$\nabla_S^D = (1 - B^S)^D. \quad (2.42)$$

Este modelo multiplicativo é chamado *Autoregressivo Integrado Médias Móveis com Sazonalidade* e representado por SARIMA (p,d,q)<sub>x</sub>(P,D,Q)<sub>S</sub>.

Assim, o modelo completo de Box & Jenkins é dado por

$$\Phi_p(B^S)\phi_p(B)\nabla_S^D\nabla^d y_t = c + \Theta_q(B^S)\theta_q(B)a_t \quad (2.43)$$

## 2.5 Identificação do Modelo

Quando trabalhamos com uma série temporal, utilizando a metodologia de Box & Jenkins, precisamos identificar o modelo que melhor se ajusta à série. Pode haver um grande número de modelos na classe ARIMA(p,d,q)(P,D,Q)<sub>S</sub> que sejam plausíveis para um determinado conjunto de dados experimentais. O

problema é desenvolver métodos para escolher entre modelos “competitivos” e escolher o “melhor” modelo para determinada série de interesse.

As técnicas de Box & Jenkins são baseadas nas propriedades das funções de autocorrelação e de autocorrelação parcial de modelos ARIMA.

As seguintes propriedades qualitativas fornecem sugestões sobre a estrutura do modelo:

Propriedade 1: Não Estacionariedade

Quando a  $\text{fac}$  decai muito vagarosamente e a  $\text{facp}$  tem um grande pico no lag 1 temos uma série não estacionária. Possivelmente pode-se chegar à estacionariedade através de diferenças, de modo que o processo  $\nabla^d y_t$  seja reduzido a um ARMA(p,q). O número de diferenças,  $d$ , necessárias para que o processo se torne estacionário é alcançado quando a  $\text{fac}$  amostral de  $w_t = \nabla^d y_t$  decresce rapidamente para zero.

Obs.: na prática,  $d=0, 1$  ou  $2$  é suficiente para se obter a estacionariedade, quando isto não ocorre, a causa pode ser uma tendência determinística e a técnica a ser utilizada é a Análise de Regressão no tempo.

Propriedade 2: Não Estacionariedade Sazonal

Ocorre quando a  $\text{fac}$  é muito próximo de zero, exceto nos lags  $S, 2S, 3S, \dots$ , decaindo muito vagarosamente, e a  $\text{facp}$  tem um pico no lag  $S$ .

Obtém-se a estacionariedade tomando diferenças da forma  $\nabla_S^D y_t$ . Normalmente o número de diferenças,  $D$ , necessário para se alcançar a estacionariedade é 1.

Propriedade 3: Comportamento Autoregressivo

Quando a  $\text{facp}$  é diferente de zero apenas para os lags 1, 2, ...e p, e a  $\text{fac}$  decai exponencialmente, ajusta-se um modelo  $\text{AR}(p)$ .

Propriedade 4: Comportamento Autoregressivo Sazonal

Quando a  $\text{facp}$  é diferente de zero apenas para os lags S, 2S, ...e PS, e a  $\text{fac}$  decai exponencialmente, ajusta-se um modelo  $\text{AR}_S(p)$ .

Propriedade 5: Comportamento de Médias Móveis

Quando a  $\text{fac}$  é diferente de zero apenas para os lags 1, 2, ...e q, e a  $\text{facp}$  tem um decaimento exponencial dominante, ajusta-se um modelo  $\text{MA}(q)$ .

Propriedade 6: Comportamento Autoregressivo Sazonal

Quando a  $\text{fac}$  é diferente de zero apenas para os lags S, 2S, ...e QS, e a  $\text{facp}$  tem um decaimento exponencial dominante, ajusta-se um modelo  $\text{MA}_S(Q)$ .

## 2.6 Estimação dos Parâmetros do Modelo

Como os procedimentos de estimação são na maioria das vezes numéricos e iterativos, quanto maior a proximidade dos valores iniciais em relação ao máximo da função a ser maximizada, mais rápida e garantida será a convergência. Estas estimativas iniciais para os modelos ARIMA são obtidas através da autocorrelações amostrais da série já estacionária.

O princípio básico sobre o qual a estimação conduzir-se-á é o da máxima verossimilhança.

Denote  $\underline{\theta} = (\phi_0, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2)$  o vetor de parâmetros populacionais.

Esta abordagem requer a especificação de uma particular distribuição para o Ruído Branco  $a_t$ . Tipicamente, assumiremos que  $a_t$  é R.B. gaussiano, ou seja,  $a_t \sim \text{i.i.d. } N(0; \sigma_a^2)$ . Embora esta suposição seja forte, as estimativas de  $\theta$  frequentemente serão consistentes para processos não gaussianos.

### Distribuição Assintótica dos Estimadores de Máxima Verossimilhança

Se o tamanho da série temporal  $n$  é suficientemente grande, frequentemente a distribuição dos MLE pode ser aproximada por

$$\hat{\theta} \sim N(\theta; n^{-1} \tau^{-1})$$

onde  $\theta$  denota o “verdadeiro” vetor paramétrico e  $\tau$  é conhecida como matriz de informação, cuja estimação pode ser feita através da quantidade

$$\hat{\tau}_{2D} = -n^{-1} \left. \frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'} \right|_{\theta = \hat{\theta}}$$

a qual é o estimador pela Segunda derivada (2D).

## 2.7 Verificação do Modelo Estimado

Após estimar o modelo, é necessário verificar se ele representa ou não adequadamente os dados. Esta verificação é feita a partir da análise dos resíduos do modelo estimado: se este modelo for coerente, os “erros verdadeiros”  $a_t = \theta^{-1}(B)\phi(B)\tilde{w}_t$ , constituirão um Ruído Branco. Assim, estimados todos  $\phi$ 's e todos  $\theta$ 's, as quantidades  $\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)\tilde{w}_t$ , serão chamados resíduos estimados.

Observando as funções  $f_{ac}$  e  $f_{acp}$  amostrais dos resíduos estimados, estas devem mostrar um Ruído Branco.

Sob a suposição de que o modelo ajustado é apropriado, os  $\hat{a}_t$  estimados deverão estar próximos dos  $a_t$  verdadeiros e portanto deverão ser aproximadamente não correlacionados. Em particular, teremos, aproximadamente,

$$r_{k_n} \approx N\left(0, \frac{1}{n}\right) \quad (2.44)$$

Quando observa-se que algum  $r_{k_n}$  está fora do limite  $\pm 2 \frac{1}{\sqrt{n}}$ , acrescentamos a ordem do modelo  $k$  unidades, obtendo um novo modelo. Então o ciclo de identificação, estimação e verificação deve ser continuado, até que um modelo satisfatório seja encontrado.

## 2.8 Previsão de Valores Futuros

Supondo que existem observações  $\dots, y_{t-2}, y_{t-1}, y_t$ , até o instante  $t$  de uma série, é interessante prever um valor  $y_{t+h}$ ,  $h \geq 1$ . A previsão de origem  $t$  e horizonte  $h$  será denotada por  $\hat{y}_t(h)$ .

### 2.8.1 Cálculo da Previsão de Erro Quadrático Médio Mínimo

Num modelo ARIMA é razoável supor que  $\hat{y}_t(h)$  seja uma função das observações até o instante  $t$ , conseqüentemente será uma função de  $a_t, a_{t-1}, \dots$

Se indicarmos a melhor previsão por



$$\hat{y}_t(h) = \psi_h^* a_t + \psi_{h+1}^* a_{t-1} + \psi_{h+2}^* a_{t-2} + \dots \quad (2.45)$$

devemos então determinar os pesos  $\psi_i^*$  que minimizem o EQM de previsão.

$$EQM_{min} = E[y_{t+h} - \hat{y}_t(h)]^2 = E\left[\sum_{i=0}^{h-1} \psi_i a_{t+h-i} + \sum_{i=0}^{+\infty} (\psi_{h+i} - \psi_{h+i}^*) a_{t-i}\right]^2 \quad (2.46)$$

Pela equação acima vemos que se tomarmos  $\psi_i^* = \psi_i$ ,  $i = h, h+1, \dots$  minimizaremos a equação.

Calculando o  $EQM_{min}$ , obtemos

$$EQM_{min} = \sum_{i=0}^{h-1} \psi_i^2 \sigma_a^2 = (1 + \psi_1^2 + \dots + \psi_{h-1}^2) \sigma_a^2 \quad (2.47)$$

Lembrando que  $y_{t+h} = \hat{y}_t(h) + e_t(h)$ , com  $h \geq 1$ , e denotado por  $[y_{t+h}] = E[y_{t+h} / y_t, y_{t-1}, y_{t-2}, \dots]$  temos as seguintes conclusões:

- i)  $\hat{y}_t(h) = [y_{t+h}]$ , ou seja, a previsão de  $EQM_{min}$  é a esperança condicional de  $y_{t+h}$ , dadas as observações passadas da série.
- ii)  $E[e_t(h)] = 0$  e  $\text{var}[e_t(h)] = (1 + \psi_1^2 + \dots + \psi_{h-1}^2) \sigma_a^2$  (2.48)
- iii) O erro de previsão a um passo é  $e_t(1) = y_{t+1} - \hat{y}_t(1) = a_{t+1}$ , isto é, os erros de previsão a um passo são não correlacionados, o que não acontece para  $h > 1$ .

A equação de previsão, considerando o modelo ARIMA (p,d,q) é dada substituindo-se  $t$  por  $t+h$ :  $\phi(B)(1-B)^d y_{t+h} = \theta(B) a_{t+h}$ .

Podemos escrevê-la como

$$\phi(B) y_{t+h} = \theta(B) a_{t+h} \quad (2.49)$$

ou ainda

$$y_{t+h} = \varphi_1 y_{t+h-1} + \dots + \varphi_{p+d} y_{t+h-p-d} - \theta_1 a_{t+h-1} - \dots - \theta_q a_{t+h-q} + a_{t+h} \quad (2.50)$$

Tomando a esperança condicional em (2.50), obtemos

$$\hat{y}(h) = \varphi_1 [y_{t+h-1}] + \dots + \varphi_{p+d} [y_{t+h-p-d}] - \theta_1 [a_{t+h-1}] - \dots - \theta_q [a_{t+h-q}] + [a_{t+h}] \quad (2.51)$$

Observa-se que cada  $\hat{y}_t(h)$  é calculado recursivamente, pois depende de

$$\hat{y}_t(h-1), \hat{y}_t(h-2), \dots$$

Sob a suposição de que  $a_t$  é Ruído Branco e  $a_t \sim N(0; V(h))$  para todo  $t$ , então, dados  $y_t, y_{t-1}, y_{t-2}, \dots$  a distribuição de  $y_{t+h} / y_t, y_{t-1}, y_{t-2}, \dots$  é  $N(\hat{y}_t(h); V(h))$ , ou seja,

$$\frac{y_{t+h} - \hat{y}_t(h)}{\sqrt{V(h)}} \sim N(0;1) \quad (2.52)$$

onde  $V(h) = \text{var}[e_t(h)]$ .

Logo, um I.C. com  $(1-\varepsilon)(100\%)$  de confiança para  $y_{t+h}$  condicional à informação até o instante  $t$  é dado por

$$\left( \hat{y}_t(h) - u_{\frac{\varepsilon}{2}} \sqrt{V(h)}; \hat{y}_t(h) + u_{\frac{\varepsilon}{2}} \sqrt{V(h)} \right) \quad (2.53)$$

onde  $u_{\varepsilon/2}$  é o limite da distribuição Normal(0;1).

### 3. MODELOS ESTRUTURAIS

#### 3.1 Introdução

O presente capítulo tem como meta descrever os principais aspectos dos Modelos Estruturais. Esta forma de modelagem de séries de tempo tem o grande atrativo de oferecer uma descrição das principais características da série, através da decomposição da série em componentes não observáveis, tais como: tendência, sazonalidade, ciclo e componente irregular. Além do que, permite ainda a análise de mudança no comportamento destas componentes. Por permitir uma interpretação mais direta dos componentes da série este tipo de modelo pode ser considerado mais interessante que os tradicionais modelos ARIMA.

##### 3.1.1 Média móvel ponderada exponencialmente

O modo mais simples de calcular o nível de uma série de observações é usar a média amostral,  $\bar{y}$ . Porém, se o propósito de calcular este nível é usá-lo como base para prever observações futuras, deve-se por mais peso nas observações mais recentes. Assim a estimativa do nível atual de uma série será

$$m_t = \sum_{j=0}^{T-t} w_j y_{T-j} \quad (3.1)$$

onde os  $w_j$ 's são um conjunto de pesos que somam uma unidade. Esta estimativa deverá então ser a previsão de observações futuras. Assim

$$\hat{y}_{T+l/T} = m_T, \quad l = 1, 2, \dots \quad (3.2)$$

e a função de previsão é uma linha horizontal.

Um modo de pôr mais peso nas observações mais recentes é deixar o pesos decaírem exponencialmente. Assim

$$m_t = (1 - \lambda)m_{t-1} + \lambda y_t, \quad t=2, \dots, T \quad (3.3)$$

onde  $\lambda$  é uma constante suavizadora entre  $0 < \lambda \leq 1$ , com  $m_1 = y_1$ .

Visto que  $m_t$  é a previsão apropriada de  $y_{t+1}$ , a recursão em (3.3) é escrita freqüentemente como

$$\hat{y}_{t+1/t} = (1 - \lambda)\hat{y}_{t/t-1} + \lambda y_t \quad (3.4)$$

Assim a previsão a um passo é uma média ponderada da observação atual e a previsão da observação atual feita no período anterior. Alternativamente

$$\hat{y}_{t+1/t} = \hat{y}_{t/t-1} + \lambda(y_t - \hat{y}_{t/t-1}) \quad (3.5)$$

de forma que as previsões são modificadas de acordo com o tamanho do erro de previsão no período atual. Se o valor de  $\lambda$  igual a zero for admissível, significaria que nenhuma atualização aconteceria. Um valor de unidade para  $\lambda$  significaria que toda a informação necessária para previsão está contida na observação atual.

A recursão em (3.3) ou (3.4) é conhecida como *Média móvel ponderada exponencialmente* (EWMA). O procedimento de construção de uma função de previsão como a descrita acima é conhecido como alisamento exponencial simples. Até que ponto são descontadas observações passadas formando (3.2) depende do valor de  $\lambda$ . Isto pode ser fixado em uma priori a qual está normalmente entre 0.05 e 0.30. Porém, se um número razoável de observações está disponível,  $\lambda$  pode ser escolhido como aquele

valor o qual minimiza a soma de quadrados dos erros de previsão um passo à frente, i.e.

$$S(\lambda) = \sum \hat{v}_t^2 \text{ onde } \hat{v}_t = y_t - \hat{y}_{t|t-1} \quad (3.6)$$

Se a recursão em (3.4) começa com  $\hat{y}_{2|1}$  igual a  $y_1$  que é equivalente a fixar  $m_1$  igual a  $y_1$ , os erros de previsão são definidos para  $t = 2, 3, \dots, T$ .

### 3.1.2 Procedimento de previsão de Holt-Winters para a tendência linear local

A função de previsão para o procedimento de EWMA é uma linha horizontal. Produzindo uma inclinação,  $b_t$ , na função de previsão temos

$$\hat{y}_{T+l|T} = m_T + b_T l, \quad l = 1, 2, \dots \quad (3.7)$$

Holt (1957) e Winters (1960) introduziram um esquema de atualização para calcular  $m_t$  e  $b_t$  no qual são descontadas observações passadas por meio de duas constantes suavizadoras,  $\lambda_0$  e  $\lambda_1$  onde  $0 < \lambda_0, \lambda_1 < 1$ . Deixando que  $m_{t-1}$  e  $b_{t-1}$  denotam as estimativas do nível e da inclinação no momento  $t-1$ . A previsão um passo a frente é então

$$\hat{y}_{t|t-1} = m_{t-1} + b_{t-1} \quad (3.8)$$

Como no EWMA, a estimativa atualizada do nível,  $m_t$ , é uma combinação linear de  $\hat{y}_{t|t-1}$  e  $y_t$ . Então

$$m_t = \lambda_0 y_t + (1 - \lambda_0)(m_{t-1} + b_{t-1}) \quad (3.9)$$

Desta nova estimativa de  $m_t$ , uma estimativa da inclinação pode ser construída

como  $m_t - m_{t-1}$ . Isto sugere que uma estimativa atualizada,  $b_t$ , seja formada por uma combinação linear de  $m_t - m_{t-1}$  e a estimativa anterior, assim

$$b_t = \lambda_1(m_t - m_{t-1}) + (1 - \lambda_1)b_{t-1} \quad (3.10)$$

Juntando (3.9) e (3.10) forma-se a recursão de Holt. Com o argumento dado pelo EWMA, começamos a recursão utilizando as observações iniciais  $m_2 = y_2$  e  $b_2 = y_2 - y_1$ . Portanto a recursão vale para  $t=3$  até  $t = T$  ao que a função de previsão (3.8) é fixa para além de T.

Reorganizando (3.9 e 3.10) em termos do erro de previsão um passo à frente,  $\hat{v}_t$ , definido em (3.6). Temos  $m_t = m_{t-1} + b_{t-1} + \lambda_0 \hat{v}_t$  (3.11)

$$b_t = b_{t-1} + \lambda_0 \lambda_1 \hat{v}_t$$

Como com o procedimento EWMA, estas constantes suavizadoras são calculadas minimizando a função soma de quadrados, analogamente a (3.6).

### 3.1.3 Mínimos quadrados ponderados (DLS)

Uma tendência global pode ser ajustada através de mínimos quadrados. Assim uma tendência linear global é obtida regressando as observações em uma constante e tempo. Um nível global é simplesmente a média amostral,  $\bar{y}$ , entretanto, é claro que, este também é um estimator OLS. Por outro lado, alisamentos exponenciais simples podem ser obtidos introduzindo um fator de ponderação,  $\omega$ , na função de soma de quadrados. Este é o exemplo mais simples da técnica conhecida como *Mínimos quadrados ponderados* (DLS).

A estimação de um nível local está baseada em achar o valor de  $m$  que minimiza

$$S(m; \omega) = \sum_{j=0}^{T-1} \omega^j (y_{T-j} - m)^2 \quad (3.12)$$

onde  $0 \leq \omega \leq 1$ . Diferenciando (3.12) em relação a  $m$  dá-se a estimativa de Mínimo quadrado ponderado

$$m_T = \left[ \sum_{j=0}^{T-1} \omega^j \right]^{-1} \sum_{j=0}^{T-1} \omega^j y_{T-j} \quad (3.13)$$

Isto reduz à média amostral se  $\omega$  é igual a um. Para  $\omega$  estritamente menor que um,

$$m_T = [(1-\omega)(1-\omega^T)]^{-1} \sum_{j=0}^{T-1} \omega^j y_{T-j} \quad (3.14)$$

Se um termo de inclinação é introduzido na função de previsão, mínimos quadrados ponderados equivale a encontrar os valores de nível,  $m$ , e de inclinação,  $b$ , que minimizam,

$$S(m, b; \omega) = \sum_{j=0}^{T-1} \omega^j (y_{T-j} - m + bj)^2 \quad (3.15)$$

Estes valores são

$$\begin{bmatrix} m_T \\ b_T \end{bmatrix} = \begin{bmatrix} \sum \omega^j & -\sum j\omega^j \\ -\sum j\omega^j & \sum j^2\omega^j \end{bmatrix}^{-1} \begin{bmatrix} \sum \omega^j y_{T-j} \\ -\sum j\omega^j y_{T-j} \end{bmatrix} \quad (3.16)$$

Se  $T$  for grande, e forem adicionadas algumas manipulações (veja Abraham e Ledolter/1983, p.105 ou Montgomery e Johnson/1976, pp.85-6), temos:

$$m_t = m_{t-1} + b_{t-1} + (1 - \omega^2) \hat{v}_t \quad (3.17)$$

$$b_t = b_{t-1} + (1 - \omega)^2 \hat{v}_t$$

onde  $\hat{v}_t$  é o erro de previsão um passo a frente.

O uso da recursão (3.17) junto com a função de previsão linear (3.7) é conhecido como *alisamento exponencial duplo*. Normalmente é sugerido que o fator de ponderação seja fixado razoavelmente perto de um. Por exemplo, Brown (1963) sugere um valor entre  $\sqrt{0.70} = 0.84$  e  $\sqrt{0.95} = 0.97$ . Uma comparação de (3.17) com (3.11) mostra um alisamento exponencial duplo para um caso especial do procedimento de Holt no qual  $\lambda_0 = 1 - \omega^2$  e  $\lambda_1 = (1 - \omega)/(1 + \omega)$

### 3.1.4 Sazonalidade

Os efeitos sazonais podem ser dirigidos de dois modos. O primeiro é o ajuste sazonal dos dados, no qual é construída uma função de previsão satisfatória para a tendência e então é somado o efeito sazonal. O segundo é construir o efeitos sazonais diretamente no procedimento de previsão.

Ajuste sazonal. O modo mais simples para ajustar sazonalmente uma série é primeiramente alisá-la por meio de uma média móvel. Se há um componente periódico que tende a repetir-se com um período de  $s$  unidades de tempo, a série alisada é determinada por

$$y_t^* = \sum_{j=-s/2}^{s/2} w_j y_{t-j}, \quad t = (s/2) + 1, \dots, T - (s/2) - 1 \quad (3.18)$$



onde  $w_j = 1/s$  para  $j = 0, \pm 1, \dots, \pm((1/2)s - 1)$  e  $w_j = w_{-j} = 1/(2s)$  para  $j = \pm s/2$ . São calculadas então, as médias dos resíduos  $y_t = y_t^*$  de cada unidade dos períodos sazonais  $c_h^*$ ,  $h = 1, \dots, s$ . Subtraindo estes fatores da série original, é obtida a série ajustada sazonalmente,  $y_t^*$ ,  $t = 1, \dots, T$ .

A função de previsão da série ajustada sazonalmente pode ser construída conforme os procedimentos descritos nas seções anteriores.

Mínimos quadrados ponderados. Efeitos sazonais podem ser incorporados em uma função de previsão por um conjunto de seno e cossenos definidos pelas frequências  $\lambda_j = 2\pi j/s, j=1, \dots, [s/2]$ . O efeito sazonal em  $l$  passos a frente é

$$c_{T+l/T} = \sum_{j=1}^{[s/2]} (d_{j,T} \cos \lambda_j l + d_{j,T}^* \text{sen } \lambda_j l), \quad l = 1, 2, \dots \quad (3.19)$$

onde  $d_{j,T}$  e  $d_{j,T}^*$  são coeficientes estimados.

A função de previsão sazonal tem duas propriedades importantes:

i) Ela se repete todos os anos, i.e.

$$c_{T+s+h/T} = c_{T+h/T}, \quad h = 1, 2, \dots \quad (3.20)$$

ii) O efeito sazonal soma zero em qualquer período de tempo com  $s$  unidades de tempo consecutivas, i.e.

$$\sum_{h=k+1}^{k+s} c_{T+h/T} = 0, \quad k = 0, 1, 2, \dots \quad (3.21)$$

Se a função de previsão consiste em uma tendência linear e um efeito sazonal, pode ser escrita como

$$\hat{y}_{T+l|T} = m_T + b_T l + c_{T+l|T}, \quad l = 1, 2, \dots \quad (3.22)$$

Os coeficientes  $m_T$ ,  $b_T$ , e  $d_{j,T}, j = 1, \dots, [s/2]$  pode ser determinado DLS.

A desvantagem de usar DLS neste contexto é que o mesmo fator de ponderação é aplicado à tendência e aos componentes sazonais. A experiência prática mostra que este arranjo é tipicamente insatisfatório visto que o padrão sazonal muda muito mais lentamente que a tendência. Harrison e Akram (1983) sugeriram um modo no qual podem ser aplicados fatores de ponderação diferentes à tendência e aos componentes sazonais.

Holt-Winters. O procedimento de previsão para tendência linear local de Holt-Winters pode ser estendido para a sazonalidade.

O procedimento Holt-Winters usa três constantes suavizadoras,  $\lambda_0, \lambda_1, \text{ e } \lambda_s$ , todas entre zero e um. A recursão para o nível, inclinação e componentes sazonais, estão descritos a seguir:

$$m_t = \lambda_0 (y_t - c_{t-s}) + (1 - \lambda_0)(m_{t-1} + b_{t-1}) \quad (3.23.a)$$

$$b_t = \lambda_1 (m_t - m_{t-1}) + (1 - \lambda_1)b_{t-1} \quad (3.23.b)$$

$$c_t = \lambda_s (y_t - m_t) + (1 - \lambda_s)c_{t-s} \quad (3.23.c)$$

A função de previsão é da forma (3.22), porém com  $c_{T+l|T}$  sendo o fator sazonal obtido de (3.23.c).

### 3.2 A Estrutura dos Modelos de Séries Temporais

O ponto de partida para o desenvolvimento de modelos de série de tempo é um modelo de regressão no qual as variáveis explicativas são funções do tempo e os parâmetros variam no tempo. Por exemplo, assim um modelo com tendência e componentes sazonais pode ser formulado como

$$y_t = \alpha + \beta t + \sum_j \gamma_j z_{jt} + \varepsilon_t \quad (3.24)$$

onde  $\alpha$  e  $\beta$  são os coeficientes associados com a tendência e os  $\gamma_j$ 's são coeficientes sazonais que somam zero. Esta condição de somar zero é obtida montando os  $z_{jt}$ 's como variáveis dummy definidas tal que para  $j = 1, \dots, s-1$ ,

$$z_{jt} = \begin{cases} 1, & t = j, j+s, j+2s, \dots \\ 0 & t \neq j, j+s, j+2s, \dots \\ -1, & t = s, 2s, 3s, \dots \end{cases} \quad (3.25)$$

A tendência poderia ser modificada somando um termo como, por exemplo,  $t^2$ , enquanto outros componentes, representando características como ciclos também podem ser adicionados. A característica em (3.24), entretanto, é que a única parte estocástica do modelo é o componente irregular,  $\varepsilon_t$ . Este é assumido como um termo aleatório normalmente distribuído com média zero e variância  $\sigma^2$ , isto é  $\varepsilon_t \sim NID(0, \sigma^2)$ .

As conseqüências da formulação em (3.24) é que os coeficientes podem ser calculados pelo usual método dos mínimos quadrados (OLS) e previsões são feitas por uma função de previsão semelhante a forma de (3.22). Também podem ser construídos intervalos de confiança para estas previsões, algo o qual não é possível com a técnica *ad*

*hoc* descrita na seção anterior. Por exemplo, se o modelo contém uma tendência linear, simplesmente, isto é,  $y_t = \alpha + \beta t + \varepsilon_t$ , a predição um passo a frente é determinada por

$$\tilde{y}_{T+1/T} = a + b(T+1)$$

onde  $a$  e  $b$  são estimadores de OLS de  $\alpha$  e  $\beta$ . O erro quadrado médio de previsão é

$$MSE(y_{T+l/T}) = \sigma^2 \left\{ 1 + \frac{2[(2T-1)(T-1) + 6l(T+l-1)]}{T(T^2-1)} \right\} \quad (3.26)$$

e isto pode ser calculado substituindo  $\sigma^2$  pelo seu estimador imparcial, isto é, a soma de quadrados dos resíduos dividido por  $T-2$ . O intervalo de  $100(1 - \alpha)\%$  de confiança para  $y_{T+l}$  é então

$$\tilde{y}_{T+l/T} \pm t_{T-2}^{(\alpha/2)} \cdot RMSE(\tilde{y}_{T+l/T}) \quad (3.27)$$

onde  $t_{T-2}^{(\alpha/2)}$  é o ponto na distribuição  $t$  com  $T-2$  graus de liberdade tal que a probabilidade está à direita dele é  $\alpha/2$ .

A desvantagem em usar modelos como (3.24) é que eles raramente são apropriados. A tendência e componentes sazonais são funções determinísticas do tempo e todas as observações recebem a mesma ponderação quando são feitas predições. É então necessário formular modelos que conduzam ao desconto de observações passadas. No caso de (3.24) isto pode ser feito deixando  $\alpha$ ,  $\beta$  e os  $\gamma_j$ 's evoluírem com o passar do tempo. Um modelo estrutural de série de tempo pode ser então considerado como um modelo de regressão onde as variáveis explicativas são funções do tempo e os parâmetros destas funções também variam no tempo.

A formulação essencialmente estática de (3.24) é convertida num modelo de série de tempo dinâmico incorporando certos processos estocásticos . Estes processos estocásticos não só provêm de modelos que permitem que os parâmetros evoluam com o passar do tempo mas também conduz à consideração de outros efeitos dinâmicos que necessariamente não seriam sugeridos por um modelo de regressão estático.

### 3.2.1 Tendência

A tendência representa o movimento a longo prazo de uma série que pode ser extrapolada no futuro. O modelo de série temporal estrutural mais simples consiste em um componente de tendência mais um ruído branco. O termo de choques fortuitos pode ser interpretado como um componente irregular ou um erro de medida. De qualquer modo o modelo pode ser escrito como

$$y_t = \mu_t + \varepsilon_t, \quad t = 1, \dots, T \quad (3.28)$$

onde  $\mu_t$  é a tendência e  $\varepsilon_t$  é o ruído branco que é assumido como não correlacionado com qualquer elemento estocástico em  $\mu_t$ .

A tendência pode ter uma variedade de formas. Esta seção se concentra no caso quando é linear. A tendência linear determinística é

$$\mu_t = \alpha + \beta t, \quad t = 1, \dots, T \quad (3.29)$$

e ao substituir em (3.28), o modelo obtido é

$$y_t = \alpha + \beta t + \varepsilon_t \quad t = 1, \dots, T \quad (3.30)$$

A tendência determinística em (3.29) pode se tornar estocástica considerando  $\alpha$  e  $\beta$  como passeios ao acaso. Porém, isto conduziria a um padrão um pouco descontínuo para  $\mu_t$ ; Um modelo mais satisfatório é obtido trabalhando diretamente com o nível atual,  $\mu_t$ , em lugar do seu intercepto,  $\alpha$ . Visto que  $\mu_t$  pode ser obtida recursivamente por

$$\mu_t = \mu_{t-1} + \beta \quad t = 1, \dots, T \quad (3.31)$$

com  $\mu_0 = \alpha$ , podem ser introduzidas condições estocásticas como segue:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t, \quad t = \dots, -1, 0, 1, \dots, \end{aligned} \quad (3.32)$$

onde  $\eta_t$  e  $\zeta_t$  são ruídos brancos não correlacionados com média zero e variâncias  $\sigma_\eta^2$  e  $\sigma_\zeta^2$ . O efeito de  $\eta_t$  é permitir que o nível da tendência possa variar para cima e para baixo, enquanto  $\zeta_t$  permite uma mudança de inclinação. Quanto maior as variâncias, maior os movimentos estocásticos na tendência. Se  $\sigma_\eta^2 = \sigma_\zeta^2 = 0$ , a equação em (3.32) se desfaz o que mostra uma tendência determinística limitada.

Embora o efeito de (3.32) seja permitir alterações dos parâmetros da tendência no tempo, a suposição adicional que os três ruídos têm médias independentes, mostram que a função de previsão ainda tem a forma linear de (3.28). O processo que gera o ajuste é um vetor autoregressivo de primeira ordem

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} \quad (3.33)$$

Fazendo as esperanças condicionais para os lags  $l = 1, 2, \dots$  e resolvendo a equação de diferença temos

$$\tilde{\mu}_{T+l|T} = \tilde{\mu}_T + \tilde{\beta}_T l \quad (3.34)$$

onde  $\tilde{\mu}_T$  e  $\tilde{\beta}_T$  são as esperanças condicionais de  $\mu_T$  e  $\beta_T$  no momento T. Visto que o modelo (3.28) especifica que as observações são iguais a uma tendência mais um termo de um ruído branco, (3.34) também é a função de previsão para a própria série, isto é,

$$\tilde{y}_{T+l|T} = \tilde{\mu}_T + \tilde{\beta}_T l, \quad l = 1, 2, \dots \quad (3.35)$$

Quando os ruídos brancos  $\eta_t$  e  $\zeta_t$  são normalmente distribuídos, o cálculo de  $\tilde{\mu}_T$  e  $\tilde{\beta}_T$  podem ser obtidos pondo o modelo em forma de estado de espaços e aplicando o filtro de Kalman. Sem a suposição de normalidade, a não correlação entre  $\eta_t$  e  $\zeta_t$  é suficiente para que o filtro de Kalman possa obter ótimas previsões lineares.

### 3.2.2 Ciclos

Considere que  $\psi_t$  seja uma função cíclica do tempo com frequência  $\lambda_c$ , a qual é medida em radianos. O período do ciclo que é o tempo levado para passar por uma sucessão completa de valores, é  $2\pi/\lambda_c$ . Um ciclo pode ser expresso como uma onda de senos ou uma onda de cossenos com parâmetros adicionais que representam a amplitude e o período. Assim, se uma onda de cossenos é usada,

$$\psi_t = A \cos(\lambda_c t - \theta) \quad t = 1, \dots, T \quad (3.36)$$

onde  $A$  é amplitude e  $\theta$  é o período. Uma fórmula mais conveniente é obtida escrevendo o ciclo como uma mistura de ondas de senos e de cossenos. A amplitude e o período são substituídos através de dois novos parâmetros,  $\alpha$  e  $\beta$ , e

$$\psi_t = \alpha \cos \lambda_c t + \beta \sin \lambda_c t \quad (3.37)$$

onde  $(\alpha^2 + \beta^2)^{\frac{1}{2}}$  é a amplitude e  $\tan^{-1}(\alpha/\beta)$  é o período.

Se as observações fossem determinadas por

$$y_t = \psi_t + \varepsilon_t \quad t = 1, \dots, T \quad (3.38)$$

onde  $\varepsilon_t$  é um ruído branco e  $\psi_t$  está definido em (3.37) com  $\lambda_c$  conhecida, OLS seria uma técnica de estimação apropriada. Porém, para introduzi-las no sistema, o ciclo precisa de ser estocástico permitindo que os parâmetros  $\alpha$  e  $\beta$  evoluam com o passar do tempo. Como com a tendência linear, a continuidade é preservada fazendo uma recursão para construir  $\psi_t$  antes de introduzir os componentes estocásticos no modelo.

Esta recursão é

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix}, \quad t = 1, \dots, T \quad (3.39)$$

com  $\psi_0 = \alpha$  e  $\psi_0^* = \beta$ . Os parâmetros novos são  $\psi_t$ , que é o valor atual do ciclo, e  $\psi_t^*$  que aparece através da construção de  $\psi_t$ .

Introduzindo dois ruídos brancos,  $k_t$  e  $k_t^*$ , os quais assumimos que tenham variâncias iguais ou sejam não correlacionados, para que o modelo seja identificável; e ainda introduzindo um fator de amortecimento,  $\rho$ , para se obter flexibilidade, temos



$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \text{sen } \lambda_c \\ -\text{sen } \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} k_t \\ k_t^* \end{bmatrix} \quad (3.40)$$

onde  $0 \leq \rho \leq 1$ . Este modelo é um processo AR(1). E  $0 \leq \rho \leq 1$  é condição de estacionariedade.

Se  $\tilde{\psi}_t$  e  $\tilde{\psi}_t^*$  denotam os valores esperados de  $\psi_t$  e  $\psi_t^*$ , no momento T; a previsão  $l$  passos a frente de  $\psi_t$  sob a condição de ruídos brancos independentes é

$$\tilde{\psi}_{T+l/T} = \rho^l (\tilde{\psi}_T \cos \lambda_c l + \tilde{\psi}_T^* \text{sen } \lambda_c l), \quad l = 1, 2, \dots \quad (3.41)$$

### 3.2.3 Sazonalidade

Ao formular um componente sazonal determinístico, a exigência é que a soma dos efeitos sazonais sejam zero. Neste caso, para  $t = s, 2s, 3s, \dots$ , o efeito sazonal médio no tempo  $t$  é

$$\sum_{j=0}^{s-1} \gamma_{t-j} = 0 \quad (3.42)$$

Introduzindo um ruído branco nesta equação, permite-se que os efeitos sazonais possam mudar com o passar do tempo. Então teremos

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad (3.43)$$

onde  $\omega_t$  é ruído branco com média zero e variância  $\sigma_\omega^2$ . Quanto maior o valor da variância  $\sigma_\omega^2$  em relação a  $\sigma_c^2$ , mais observações passadas são descontadas construindo

um padrão sazonal para a função de previsão. Para ruídos com variâncias independentes as previsões ótimas satisfazem a recursão

$$\tilde{\gamma}_{T+l/T} = -\sum_{j=1}^{s-1} \tilde{\gamma}_{T+l-j/T}, \quad l = 1, 2, \dots \quad (3.44)$$

onde os valores são determinados pelas esperanças condicionais dos efeitos sazonais,  $\gamma_T, \dots, \gamma_{T-s+2}$ , no momento T. Assim o padrão sazonal projetado no futuro é fixo, e seus componentes somam zero para qualquer período de s meses sucessivos.

Uma das maneiras de permitir que as variáveis dummy sazonais evoluam com o tempo é supor que o efeito de cada sazonalidade evolui como um passeio ao acaso. Considerando  $\gamma_{jt}$  o efeito sazonal j no tempo t, temos

$$\gamma_{jt} = \gamma_{j,t-1} + \omega_{jt}, \quad j = 1, \dots, s \quad (3.45)$$

onde  $\omega_{jt}$  é um ruído branco com média zero e variância  $\sigma_\omega^2$ . Embora todos s componentes sazonais estejam evoluindo continuamente, somente um afeta a observação num tempo particular, isto é

$$\gamma_t = \gamma_{jt}, \quad t = 1, \dots, T \quad (3.46)$$

quando a sazonalidade j é prevaiente no tempo t. A exigência que os componentes sazonais somem zero na função de previsão em qualquer período de s tempos consecutivos é reforçada pela restrição que num tempo particular, o componente sazonal e consequentemente o ruído, somem zero, ou seja

$$\sum_{j=1}^s \gamma_{jt} = 0 = \sum_{j=1}^s \omega_{jt} \quad (47)$$

Uma maneira alternativa de escrever a expressão (3.43) é feita utilizando um conjunto de termos trigonométricos nas frequências sazonais,  $\lambda_j = 2\pi j / s$ ,  $j=1, \dots, [s/2]$ .

O efeito sazonal no momento  $t$  é então

$$\gamma_t = \sum_{j=1}^{[s/2]} (\gamma_j \cos \lambda_j t + \gamma_j^* \operatorname{sen} \lambda_j t) \quad (3.48)$$

Desta maneira a sazonalidade padrão descrita em (3.43) é uma soma de  $[s/2]$  componentes cíclicos e permite-se que evolua com o passar do tempo do mesmo modo como é feito com o *ciclo* na seção anterior. Assim

$$\left. \begin{aligned} \gamma_{jt} &= \gamma_{j,t-1} \cos \lambda_j + \gamma_{j,t-1}^* \operatorname{sen} \lambda_j + \omega_{jt} \\ \gamma_{jt}^* &= -\gamma_{j,t-1} \operatorname{sen} \lambda_j + \gamma_{j,t-1}^* \cos \lambda_j + \omega_{jt}^* \end{aligned} \right\} \quad j = 1, \dots, [s/2] \quad (3.49)$$

onde  $\omega_{jt}$  e  $\omega_{jt}^*$  são ruídos brancos com média zero os quais são não-correlacionados entre si, com variância comum  $\sigma_j^2$  para  $j = 1, \dots, [s/2]$ .

### 3.2.5 Modelos Estruturais

Até o momento foi descrito como poderia ser permitido que componentes como tendência e sazonalidade pudessem evoluir com o passar do tempo de tal modo que elas satisfizessem certas propriedades desejáveis. A sub-seção presente discute como estes componentes são reunidos em *modelos estruturais*. Para isto é mostrado também como cada um dos modelos pode ser expresso em termos de uma única equação.

Todos os modelos abaixo descritos são tais que os vários componentes entram no modelo como uma adição. E ainda, a especificação de um modelo tem como suposição que as distribuições dos ruídos brancos são todas normais.

Nível local / passeio aleatório mais ruído. O nível local do modelo foi introduzido anteriormente onde tínhamos

$$y_t = \mu_t + \varepsilon_t, \quad t = 1, \dots, T \quad (3.50.a)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad t = \dots, -1, 0, 1, \quad (3.50.b)$$

onde o componente de tendência,  $\mu_t$ , é simplesmente um nível que oscila de acordo com um passeio aleatório. Nenhum valor inicial precisa de ser especificado para  $\mu_t$  desde que seja assumido que tenha iniciado em algum ponto no passado distante.

A forma reduzida do modelo é obtida adotando a seguinte convenção: que um processo não estacionário tal qual em (3.50.a) pode ser escrito como  $\mu_t = \eta_t / \Delta$ , onde  $\Delta$  é o operador de diferença de primeira ordem. Substituindo em (3.50.a) e acrescentando (quando houver) um termo de inclinação teremos

$$y_t = \frac{\eta_t}{\Delta} + \beta_t + \varepsilon_t, \quad t = 1, \dots, T \quad (3.51)$$

Tendência linear local. O modelo de tendência linear local consiste no sistema (3.50.a ; 3.50.b) mais a tendência gerada por  $\beta_t = \beta_{t-1} + \zeta_t$ . Substituindo por  $\beta_{t-1} = \zeta_{t-1} / \Delta$  em (3.50.a) temos

$$y_t = \frac{\eta_t}{\Delta} + \frac{\zeta_{t-1}}{\Delta^2} + \varepsilon_t, \quad t = 1, \dots, T \quad (3.52)$$

Modelos cíclicos. O ciclo mais um ruído é representado por

$$y_t = \mu + \psi_t + \varepsilon_t, \quad t = 1, \dots, T \quad (3.53)$$

com  $\psi_t$  gerado pelo processo estocástico:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \text{sen } \lambda_c \\ -\text{sen } \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} k_t \\ k_t^* \end{bmatrix} \quad (3.54)$$

e o termo irregular  $\varepsilon_t$  é assumido como não correlacionado com os ruídos  $k_t$  e  $k_t^*$ , e  $\mu$  é um termo constante.

Ao acrescentarmos o componente cíclico na equação de tendência obtemos

$$y_t = \frac{\eta_t}{\Delta} + \frac{(1 - \rho \cos \lambda_c L)k_{t-1} + (\rho \text{sen } \lambda_c L)k_{t-1}^*}{\Delta(1 - 2\rho \cos \lambda_c L + \rho^2 L^2)} + \frac{\zeta_{t-1}}{\Delta^2} + \varepsilon_t, \quad t = 1, \dots, T \quad (3.55)$$

onde L é o lag do operador.

Tendência amortecida. O modelo de tendência linear local pode ser modificado introduzindo-se um fator de amortecimento no componente de inclinação, desta maneira,  $\beta_t = \rho\beta_{t-1} + \zeta_t$ , com  $0 \leq \rho \leq 1$ . A equação do modelo de tendência amortecida fica

$$y_t = \frac{\eta_t}{\Delta} + \frac{\zeta_{t-1}}{\Delta(1 - \rho L)} + \varepsilon_t, \quad t = 1, \dots, T \quad (3.56)$$

Modelo estrutural básico. O modelo estrutural básico (BSM) é uma generalização estocástica de (3.24), isto é

$$y_t = \mu_t + \gamma_t + \varepsilon_t \quad (3.57)$$

onde  $\mu_t$  é a tendência linear local de (3.32) e  $\gamma_t$  é a variável dummy do componente sazonal de (3.43) ou o componente sazonal trigonométrico de (3.48). O componente irregular,  $\varepsilon_t$ , é assumido como aleatório, e os ruídos dos três componentes são mutuamente não correlacionados.

A forma da equação para o componente sazonal variável é vista como

$$y_t = \frac{\eta_t}{\Delta} + \frac{\zeta_{t-1}}{\Delta^2} + \frac{\omega_t}{S(L)} + \varepsilon_t \quad (3.58)$$

onde  $S(L)$  é o operador de adição sazonal, pois esta expressão pode ser escrita como  $S(L)\gamma_t = \omega_t$ .

Como respeito a sazonalidade trigonométrica, segue de (3.54) e (3.55) que

$$\gamma_{jt} = \frac{(1 - \cos \lambda_j L)\omega_{jt} + (\text{sen } \lambda_j L)\omega_{jt}^*}{1 - 2 \cos \lambda_j L + L^2}, \quad j = 1, \dots, [s/2] \quad (3.59)$$

onde, para  $s$  par, o último componente cai para

$$\gamma_{\frac{s}{2}-1,t} = \omega_{\frac{s}{2},t} / (1+L) \quad (3.60)$$

O componente sazonal consiste na soma dos  $\gamma_{jt}$ 's, então, para  $s$  par

$$y_t = \sum_{j=1}^{\frac{s}{2}-1} \frac{(1 - \cos \lambda_j L)\omega_{jt} + (\text{sen } \lambda_j L)\omega_{jt}^*}{\gamma_j(L)} + \frac{\omega_{\frac{s}{2},t}}{\gamma_{\frac{s}{2}}(L)} \quad (3.61)$$

onde  $\gamma_j(L)$  e  $\gamma_{s/2}(L)$  são os operadores trigonométricos. Se  $s$  é ímpar, a adição em (3.61) corre de  $j = 1$  para  $(s - 1)/2$  e o último termo do lado direito da equação não aparece.

### 3.3 Testes e Seleção de Modelos

#### 3.3.1 Seleção de modelos

A fase inicial da seleção de um modelo é uma análise preliminar dos dados e a formulação de um possível modelo ou modelos a serem estimados. Uma vez estimado, um modelo está sujeito a testes de especificação e verificação. Um modelo que satisfaz estes procedimentos podem ser validados internamente, ou externamente por testes de predição e comparação da qualidade de seu ajuste com o de outros modelos.

*Análise preliminar e especificação de modelo* O passo essencial na análise preliminar é a observação gráfica da série. Pode ser útil observar gráficos com transformações da série através de diferenças, logaritmos, etc. Estes gráficos devem dar uma indicação sobre se um modelo é invariante no tempo, ou seja, apropriado, ou se a série está sujeita a mudanças estruturais, ou ao longo da série ou em pontos particulares do tempo.

Se um modelo é invariante no tempo geralmente será possível reduzi-lo a uma série estacionária fazendo diferenças apropriadas. Podem ser calculadas as medidas de assimetria e curtose, e então ter uma indicação de não-normalidade.

Se é visto que a série pode ser reduzida a um modelo estacionário através de transformações adequadas, o correlograma pode prover uma útil informação adicional. Assim, não faz sentido selecionar um modelo estrutural no qual a função de autocorrelação teórica é incompatível com o correlograma. Por exemplo, seria impróprio tentar ajustar um passeio ao acaso mais um termo irregular se a autocorrelação amostral de primeira-ordem das observações diferenciadas for fortemente positiva.

Por via de regra, um modelo geral deveria ser escolhido primeiramente. O que pode ser reduzido subsequente a uma forma mais parcimoniosa. A única exceção para formular um modelo geral é quando tal uma formulação pode conduzir a problemas de identificabilidade. Assim um componente cíclico só deve ser incluído em um modelo se há fortes razões para se acreditar na presença deste componente.

Testes de especificação e Diagnósticos São úteis principalmente para ver se o modelo pode ser simplificado. Por exemplo, pode ser possível fixar um hiperparâmetro com variância igual a zero. Além de testes estatísticos formais, podem ser examinados valores de certos parâmetros, para verificar se eles são consistentes com propriedades conhecidas da série. Por exemplo, isto se aplica ao componente cíclico, onde a análise preliminar ou conhecimento anterior podem dar alguma indicação sobre o período, neste caso, pode-se verificar se o parâmetro corresponde a informação prévia.

São feitos testes de diagnóstico para determinar se os resíduos são serialmente não correlacionados e normalmente distribuídos com média zero e variância constante. Deficiência dos diagnósticos podem indicar uma variedade de má especificações. Isto inclui a possibilidade que um modelo univariado invariante no tempo poderá não ser adequado.

Uma vez que um modelo obtido satisfaz os procedimentos de diagnóstico para os dados amostrais, é feito um teste de erro de predição para determinar se a qualidade é a mesma fora da amostra. Os testes formais estão baseado em erros de predição um passo à frente. Porém a função de previsão multi-passos no período após a amostra pode ser tão informativo quanto um passo a frente. Pois, dão uma demonstração mais convincente da efetividade de previsão de um modelo.

Comparação de modelos concorrentes Pode acontecer que dois ou mais modelos satisfaçam os vários critérios de especificação interna de uma mesma série. As



estatísticas de qualidade  $R_D^2$  e  $R_S^2$  comparam modelos ajustados com especificações mais simples. Se eles são menores que zero, pode-se reverter para um modelo mais simples. Também pode ser levado em conta o desempenho em um determinado período após os dados da amostra, especialmente a respeito dos erros de predições.

### 3.3.2 Testes de especificação baseados nos resíduos

Correlação Serial As autocorrelações dos resíduos amostrais são definidos por

$$r_v(\tau) = \frac{\sum_{t=d+1+\tau}^T (\tilde{v}_t - \bar{\tilde{v}})(\tilde{v}_{t-\tau} - \bar{\tilde{v}})}{\sum_{t=d+1}^T (\tilde{v}_t - \bar{\tilde{v}})^2} \quad \tau = 1, 2, \dots \quad (3.62)$$

e o correlograma resultante nos dá uma indicação de alguma correlação serial.

Um teste de significância para as primeiras P autocorrelações residuais é determinado por uma estatística baseada nas somas de quadrados destas autocorrelações. O forma da estatística de teste Box-Ljung é

$$Q^* = T^*(T^* + 2) \sum_{\tau=1}^P (T^* - \tau)^{-1} r_v^2(\tau), \quad \text{onde } T^* = T - d \quad (3.63)$$

Esta estatística de teste é uma versão modificada da estatística Box-Pierce; Ljung e Box (1978) afirmam que ela tem propriedades melhores em amostra pequenas. Quando um ARIMA(p, d, q) é ajustado,  $Q^*$  é assintoticamente  $\chi_{P-(p+q)}^2$ . Este resultado assintótico requer que P aumente com T, de modo que, por exemplo, P pode ser determinado como  $P = \sqrt{T}$  ou  $\log T$ . Em um modelo estrutural pode ser conjecturado que  $Q^*$  também é assintoticamente  $\chi^2$ , mas com graus de liberdade dados por  $P - n^*$ .

Não-linearidade Pode ser útil calcular a autocorrelação amostral para determinadas funções dos resíduos. Em particular, a autocorrelação amostral dos

resíduos quadrados, denotada por  $rsq_v(\tau)$ , a qual pode prover uma indicação de efeitos não-lineares que podem estar presentes. Esta autocorrelação amostral,

$$rsq_v(\tau) = \frac{\sum (\tilde{v}_t^2 - \tilde{\sigma}_*^2)(\tilde{v}_{t-\tau}^2 - \tilde{\sigma}_*^2)}{\sum (\tilde{v}_t^2 - \tilde{\sigma}_*^2)^2}, \quad \tau = 1, 2, \dots \quad (3.64)$$

pode ser usada para construir um tipo de teste Box-Ljung para não-linearidade substituindo  $r_v(\tau)$  por  $rsq_v(\tau)$  em (3.63). É mostrado em McLeod e Li (1983), que a distribuição assintótica dessa estatística é  $\chi_{\rho}^2$ .

Heteroscedasticidade Um teste para diagnosticar heteroscedasticidade pode ser construído a partir dos resíduos. Supondo que  $h$  é o inteiro mais próximo de  $T^*/3$ . A estatística do teste é

$$H(h) = \frac{\sum_{t=T-h+1}^T \tilde{v}_t^2}{\sum_{t=d+1}^{d+1+h} \tilde{v}_t^2} \quad (3.65)$$

Se o único parâmetro desconhecido no modelo fosse  $\sigma_*^2$ , esta estatística teria uma distribuição  $F(h, h)$  sob a hipótese nula. É sugerido então que  $H(h)$  seja testada com uma distribuição  $F(h, h)$  em situações mais gerais. A alternativa é basear o teste na distribuição assintótica de  $hH(h)$  a qual é  $\chi_h^2$  sob a hipótese nula. Se esta suspeita-se que as variâncias podem ser maiores (menores) no o fim da amostra, então o canto direito (esquerdo) da distribuição  $F$  deve ser usada como a região crítica em um teste unilateral. Um teste bilateral permite ambas as possibilidades mas às custas de uma perda em poder.

Normalidade Os terceiros e quartos momentos padronizados dos resíduos sobre a média são

$$\sqrt{b_1} = \hat{\sigma}_*^3 \sum (\tilde{v}_t - \bar{\tilde{v}})^3 / T^* \quad e \quad (3.66)$$

$$b_2 = \hat{\sigma}_*^4 \sum (\tilde{v}_t - \bar{\tilde{v}})^4 / T^* \quad (3.67)$$

onde

$$\hat{\sigma}_*^2 = (T - d - 1)^{-1} \sum_{t=d+1}^T (\tilde{v}_t - \bar{\tilde{v}})^2. \quad (3.68)$$

Estas estatísticas são as medidas básicas de assimetria e curtose, e para uma distribuição normal elas devem estar centradas respectivamente ao redor de zero e três. Elas são assintoticamente normais, com

$$\sqrt{b_1} \sim AN(0,6 / T^*) \text{ e } b_2 \sim AN(3,24 / T^*)$$

quando o modelo é especificado corretamente; detalhes podem ser vistos em Bowman e Shenton (1975). Um teste para não-normalidade pode estar então baseado na estatística

$$N = (T^* / 6)b_1 + (T^* / 24)(b_2 - 3)^2 \quad (3.69)$$

Sob a hipótese nula, esta estatística tem uma distribuição  $\chi_2^2$  para amostras grandes. Como é mostrado em Jarque e Bera (1980), o teste pode ser obtido usando o princípio da LM.

### 3.3.3 Qualidade do Ajuste

Variância do erro de previsão A variância de erro de previsão de um modelo de tempo invariante é a variância dos erros de previsão um passo à frente no estado fixo. Assim se as variâncias no modelo são expressas como  $\sigma_*^2$ , a variância de erro de previsão é definida por

$$\sigma^2 = \sigma_*^2 \bar{f} \quad (3.70)$$

onde  $\bar{f}$  é o valor de  $f_t$  no estado fixo.

A variância de erro de previsão é igual a variância do termo irregular na forma reduzida e é uma medida natural de qualidade do ajuste. E o estimador correspondente a  $\sigma^2$  pode ser escrito como

$$\tilde{\sigma}^2 = s^2 = \tilde{\sigma}_*^2 \bar{f} \quad (3.71)$$

A distribuição assintótica do estimador da variância de erro de previsão pode ser obtida notando que no modelo ARMA de forma reduzida, o estimador de máxima verossimilhança de  $\sigma^2$  é assintoticamente independente dos estimadores de ML dos parâmetros AR e MA.

Na forma reduzida de um modelo estrutural, os parâmetros AR e MA estarão normalmente sujeitos a restrições. Porém, como será visto mais adiante, a distribuição de  $\sigma^2$  não é afetada por estas restrições.

Coeficiente de determinação Em análise de regressão o coeficiente de determinação,  $R^2$ , é apresentado quase sempre como uma medida de qualidade de ajuste. Normalmente observações da série no tempo mostram fortes movimentos para cima ou para baixo e qualquer modelo que é capaz de selecionar razoavelmente bem uma tendência terá  $R^2 = 1$ . A estatística é então de pouco valor, a não ser quando a série é estacionária. Mas ainda assim, o uso de  $R^2$  é bastante freqüente, pois não só é citado invariavelmente mas é geralmente usado como evidência de que um modelo é bom.

As medidas apresentadas são aplicáveis a modelos de regressão como também a modelos univariados de série de tempo.

A soma de quadrados residual para um modelo fixo no tempo pode ser definida como

$$SSE = \bar{f} \sum_t \tilde{v}_t^2 = \bar{f} S(\psi_*) = (T - d) \tilde{\sigma}^2 \quad (3.72)$$

onde  $\tilde{\sigma}^2$  é o estimador da variância de erro de previsão baseado em (3.70). O  $R^2$  tradicional é obtido dividindo SSE pela soma dos quadrados das observações em torno da média e subtraindo-se de um. Uma medida melhor para dados de série de tempo,  $R_D^2$ , é obtida substituindo as observações pelas primeiras diferenças

$$R_D^2 = 1 - \frac{SSE}{\sum_{t=2}^T (\Delta y_t \overline{\Delta y})^2} \quad (3.73)$$

A medida adotada em (3.73) é para o passeio ao acaso mais inclinação,

$$y_t = y_{t-1} + \beta + \eta_t, \quad t = 2, \dots, T \quad (3.74)$$

Este é um modelo muito simples no qual a previsão do próximo será a observação atual mais o incremento médio no período de amostra. Apesar de sua simplicidade o modelo pode apresentar previsões bastante boas.

Quando as observações são mensais ou trimestrais, efeitos sazonais provavelmente estão presentes nos dados não ajustados, e até mesmo uma simples previsão os levará em conta. É feita então uma simples extensão em (3.74), na qual Adiciona-se s dummies sazonais ao lado direito da equação. Isto sugere o uso de um critério de qualidade de ajuste

$$R_s^2 = 1 - SSE / SSDSM \quad (3.75)$$

onde *SSDSM* é a soma dos quadrados da primeira diferença ao redor das médias sazonais.

### 3.3.4 Comparação entre modelos diferentes

Dois modelos diferentes podem ser comparados com base em suas respectivas variâncias de erro de previsão. As funções de máxima verossimilhança também podem ser usadas como uma base para comparação. Porém, isto é o mesmo que comparar o p.e.v.'s, quando a amostra é grande.

Se o modelo concorrente tem diferente número de parâmetros, a comparação pode ser feita pelo critério Akaike (AIC) ou o critério Bayes (BIC). Expressando o AIC em função do p.e.v.

$$AIC^+ = \tilde{\sigma}^2 \exp[2(n+d) / T] \quad (3.76)$$

onde  $d$  é o número de elementos não-estacionários no vetor estatal e  $n$  é o número de hiperparâmetros. O critério  $BIC^+$  é definido de modo semelhante, onde o “2” é substituído por “logT”. A razão para incluir o número de elementos não-estacionários no vetor de estados na função do critério é permitir comparações que envolvem modelos com componentes determinísticos.

### 3.3.5 Testes preditivos e avaliação do modelo

Até este momento só foram vistos testes referentes a seleção de um modelo apropriado aos dados. Uma vez que um modelo satisfatório foi especificado, seus parâmetros são calculados para o respectivo conjunto de dados. Considerando que mais de um modelo pode ser especificado para a mesma amostra é interessante que alguns valores finais da amostra sejam reservados. Isto é, ao selecionar e estimar um modelo para uma amostra com  $T+l$  observações, só utilizar as primeiras  $T$  observações para a seleção e estimação. Assim, as observações  $y_{T+1}, y_{T+2}, \dots, y_{T+l}$  serão usadas para comparar as previsões obtidas pelo modelo baseado na amostra  $y_1, y_2, \dots, y_T$ .

Avaliação das previsões um passo à frente Supondo que um modelo é especificado corretamente e que todos os hiperparâmetros, à parte de  $\sigma_*^2$ , são conhecidos. Os erros de previsão um passo à frente nos lags posteriores a amostra serão então normalmente distribuídos com média zero e variância  $\sigma_*^2 f_{T+j}$ . Os erros de previsão padronizados são definidos como

$$\tilde{v}_{T+j} = v_{T+j} / \sqrt{f_{T+j}}, j = 1, \dots, l \quad (3.77)$$

deste modo

$$\tilde{v}_t \sim NID(0, \sigma_*^2), \quad t = d+1, \dots, T+l$$

Portanto, sob estas suposições, a distribuição do teste estatístico para previsão,

$$\xi(l) = \sum_{j=1}^l \tilde{v}_{T+j}^2 / (l\tilde{\sigma}_*^2) \quad (3.78)$$

é  $F(l, T-d)$ . Nos casos mais comuns, quando parâmetros diferentes de  $\sigma_*^2$  são estimados,  $l\xi(l)$  é assintoticamente  $\chi_1^2$ .

Quando um modelo é especificado, e é feito um ajuste ruim obtido no período de amostra mas que aparenta ser satisfatório, o valor de  $\xi(l)$  será grande. O teste então sugere uma rejeição do modelo. Porém, um valor insignificante de  $\xi(l)$  não é necessariamente uma prova de que o modelo está bem ajustado.

Predições l passos à frente As predições usadas para avaliar o desempenho do modelo até agora se referiram a predições um passo a frente. Porém, também é possível extrapolar o momento T fazendo predições de até l passos a frente da série. O erro de predição será

$$v_{T+j/T} = y_{T+j} - \tilde{y}_{T+j/T}, \quad j = 1, \dots, l \quad (3.79)$$

Os erros de predição de mais que um passo à frente não provêm nenhuma informação adicional para avaliar a validade interna de um modelo. Porém, eles são úteis provendo uma medida de desempenho de predição que pode ser usada como uma base para comparação de modelos concorrentes. A estatística natural para esta medida então pode ser

$$SQEP(l) = \sum_{j=1}^l v_{T+j/T}^2 \quad (3.80)$$

ou alternativamente

$$SAEP(l) = \sum_{j=1}^l |v_{T+j/T}|. \quad (3.81)$$

### 3.4 A forma de Espaço de Estado e o Filtro de Kalman

#### 3.4.1 A forma de espaço de estado

A forma de espaços de estado geral (SSF) é aplicada a uma série de tempo multivariada,  $y_t$ , com  $N$  elementos. Estas variáveis observáveis são relacionadas a um vetor  $m \times 1$ , denotado por  $\alpha_t$ , conhecido como vetor de estados, por uma *equação de medida*,

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad t = 1, \dots, T \quad (3.82)$$

onde  $Z_t$  é uma matriz  $N \times m$ ,  $d_t$  é vetor  $N \times 1$  e  $\varepsilon_t$  é um vetor  $N \times 1$  de ruídos brancos com média zero e matriz de covariância  $H_t$ .

Num modelo univariado,  $N = 1$  e a equação é escrita como

$$y_t = Z_t' \alpha_t + d_t + \varepsilon_t, \quad \text{Var}(\varepsilon_t) = h_t, \quad t = 1, \dots, T \quad (3.83)$$

Em geral os elementos de  $\alpha_t$  não são observáveis. Porém, eles são conhecidos por serem gerados por processo Markoviano de primeira ordem,

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \quad t = 1, \dots, T \quad (3.84)$$

onde  $T_t$  é uma matriz  $m \times m$ ,  $c_t$  é um vetor  $m \times 1$ ,  $R_t$  é uma matriz  $m \times g$  e  $\eta_t$  é um vetor  $g \times 1$  de ruídos serialmente não correlacionados com média zero e matriz de covariância,  $Q_t$ .

A equação (3.84) é a *equação de transição*. A inclusão da matriz  $R_t$  na frente do termo de ruídos é, até certo ponto, arbitrário. A especificação do sistema de espaço de estados é completada através de duas suposições adicionais:

$$\text{a) } E(\alpha_0) = a_0 \quad \text{e} \quad \text{Var}(\alpha_0) = P_0 \quad (3.85)$$

$$\text{b) } E(\varepsilon_t, \eta_t) = 0, \quad E(\varepsilon_t, \alpha'_0) = 0 \quad \text{e} \quad E(\eta_t, \alpha'_0) = 0 \quad \text{para } t=1, \dots, T \quad (3.86)$$

As matrizes  $Z_t$ ,  $d_t$  e  $H_t$  da equação de medidas e as matrizes  $T_t$ ,  $c_t$ ,  $R_t$  e  $Q_t$  da equação de transição formam um sistema de matrizes. A menos que se declare caso



contrário, será assumido que elas não são estocásticas. Assim, embora elas possam mudar com tempo, isto deve ocorrer de modo predeterminado. Como resultado, o sistema é linear e para qualquer valor de  $t$ ,  $y_t$ , pode ser expresso como uma combinação linear de  $\varepsilon_t$ 's e  $\eta_t$ 's atuais e passados e do vetor de estado inicial,  $\alpha_0$ .

Se o sistema de matrizes  $Z_t$ ,  $d_t$ ,  $H_t$ ,  $T_t$ ,  $c_t$ ,  $R_t$  e  $Q_t$  não muda com o passar do tempo, o modelo é dito invariante no tempo. Modelos estacionários são um caso especial. Porém, embora a classe de modelos invariantes no tempo seja muito maior que a classe de modelos estacionários, muitos modelos invariantes no tempo têm uma forma estacionária que pode ser obtida aplicando diferenças adequadas.

O sistema de matrizes  $Z_t$ ,  $d_t$ ,  $H_t$ ,  $T_t$ ,  $R_t$  e  $Q_t$  pode depender de um conjunto de parâmetros desconhecidos, e será necessário a estimação destes parâmetros. Assim no passeio ao acaso mais ruído, os parâmetros  $\sigma_\eta^2$  e  $\sigma_\varepsilon^2$  vão, em geral, ser desconhecidos, enquanto nos AR(2) e MA(1) parâmetros de MA e AR serão desconhecidos. Estes parâmetros serão denotados por um vetor  $n \times 1$   $\psi$  e referidos como hiperparâmetros para se distinguir de parâmetros que podem entrar no modelo por  $c_t$  ou  $d_t$ . O hiperparâmetros determinam as propriedades estocásticas do modelo, e os parâmetros que aparecem em  $c_t$  e  $d_t$  só afetam o valor esperado do estado e as observações de um modo de determinístico. Porém, deve ser notado que esta distinção pode ser obscurecida, por exemplo se  $d_t$  é uma função de um valor passado de  $y_t$ .

Se  $c_t$  ou  $d_t$  é uma função linear de parâmetros desconhecidos, estes parâmetros podem ser tratados como variáveis de estado.

### 3.4.2 O filtro de Kalman

Uma vez que um modelo é posto em forma de espaço de estado, está preparado para a aplicação de vários algoritmos importantes. No centro destes está o filtro de

Kalman. O filtro de Kalman é um procedimento recursivo para computar o estimador ótimo do vetor de estado no momento  $t$ , baseado na informação disponível no momento  $t$ . Esta informação consiste nas observações até e inclusive  $y_t$ . O sistema de matrizes mais  $a_0$  e  $P_0$  são assumidos como conhecidos em todos os períodos e assim não precisam de ser incluídos explicitamente no conjunto de informações.

Em certas aplicações o filtro de Kalman é importante por causa da estimação on-line. O valor atual do vetor de estado é de grande interesse e o filtro de Kalman habilita a estimativa do vetor de estado ser atualizada continuamente. O vetor de estado nem sempre tem uma interpretação econômica e, em alguns casos, é mais apropriado estimá-lo em um ponto particular no tempo usando toda a informação da amostra, não só uma parte dela. Estes dois problemas são conhecidos como predição e *smoothing* respectivamente. Harvey(1989) mostra que o filtro de Kalman provê a base para a solução de ambos.

Outra razão para a tamanha importância do filtro de Kalman é que quando os ruídos e o vetor de estado inicial são distribuídos normalmente, habilita o que é conhecido como a decomposição do erro de predição, o que permite a estimação de qualquer parâmetro desconhecido no modelo. Também provê a base para testes estatísticos e especificação de modelos.

A utilização do filtro de Kalman é feita sob a suposição que os ruídos e o vetor de estado inicial é normalmente distribuído. Um resultado padrão da distribuição normal multivariada é usado para mostrar como é possível calcular recursivamente a distribuição condicional de  $\alpha_t$  fixada na informação do momento  $t$ , para  $t = 1, \dots, T$ . Estas distribuições condicionais são normais e conseqüentemente são completamente especificadas por matrizes de covariâncias. São estas quantidades que o filtro de Kalman calcula, e ainda pode ser mostrado que a distribuição condicional média de  $\alpha_t$

é um estimador ótimo para  $\alpha_t$ , já que minimiza o erro quadrático médio (MSE). Quando não existe a suposição de normalidade não há mais nenhuma garantia que o filtro de Kalman fornecerá a média condicional do vetor de estado. Porém, ainda é um ótimo estimador, visto que minimiza o MSE dentro da classe de todos estimadores lineares.

No modelo Gaussiano,(1) a (3), o filtro de Kalman produz  $a_T$ , o estimador de mínimos quadrados médios(MMSE) de  $\alpha_T$  baseado em todas as observações

$$a_{T+1/T} = T_{T+1/T}a_T + c_{T+1} \quad (3.87)$$

simultaneamente com a previsão um passo a frente

$$\tilde{y}_{T+1/T} = Z_{T+1}a_{T+1/T} + d_{T+1} \quad (8.88)$$

A distribuição condicional de  $\alpha_{T+i}$  é normal com matriz de covariância  $P_{T+i/T}$ . Assim as equações de previsão multi-passos são

$$a_{T+i/T} = T_{T+i/T}a_{T+i-1/T} + c_{T+i} \quad (3.89)$$

$$P_{T+i/T} = T_{T+i}P_{T+i-1/T}T'_{T+i} + R_{T+i}Q_{T+i}R'_{T+i} \quad t = 1, 2, 3, \dots \quad (3.90)$$

## 4. APLICAÇÃO DAS METODOLOGIAS:

### ARIMA E MODELOS ESTRUTURAIS

A fim de comparar a eficiência da metodologia de Box & Jenkins e de Modelos Estruturais para previsão de séries temporais, estes dois métodos foram aplicados a duas séries históricas de indicadores econômicos: o Produto Interno Bruto (PIB) trimestral brasileiro, coletado junto ao Instituto Brasileiro de Geografia e Estatística (IBGE) e o Índice de Desempenho Industrial (IDI) do Rio Grande do Sul, obtido pela FIERGS. Para a Metodologia de Box & Jenkins foi utilizado o software ITSM, enquanto para a análise por Modelos Estruturais, o software utilizado foi o STAMP 5.0.

#### 4.1 Metodologia ARIMA

Foram analisados os dados trimestrais de 1980/1 até 1997/4 do índice do Produto Interno Bruto Brasileiro com base fixa em 1990, porém só foram utilizados para estimação do modelo os dados até 1994/4, para que ao se fazer previsões a partir do modelo estimado, fosse possível comparar estas previsões com os dados que ocorreram na realidade.

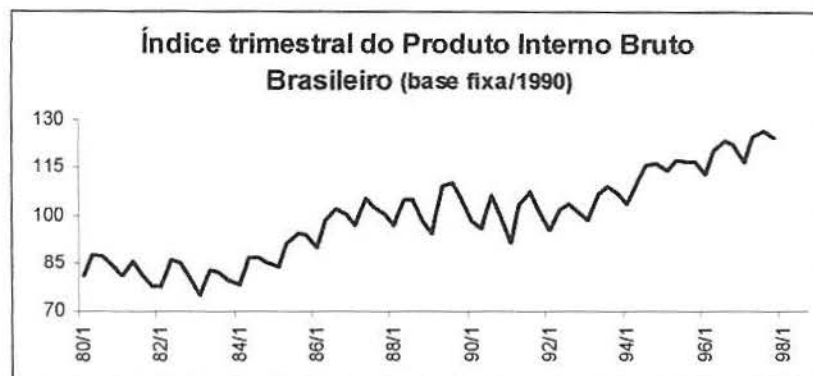


Fig. 1 – Série: Índice trimestral do PIB

Através do Software ITSM foram analisadas as fac e facp da série original: por ser uma série não estacionária foi necessário fazer a 1ª diferença e

ainda assim, foi necessário fazer também uma diferença sazonal com  $s = 4$  para que a série se tornasse estacionária. Ao experimentar alguns modelos e verificar os seus respectivos resíduos <sup>(1)</sup>, foi escolhido o modelo:

$$(1 + 0,1779 L^{12})(1 - L)(1 - L^4)y_t = (1 - 0,2196 L^2 - 0,7809 L^4)\varepsilon_t$$

Tabela 1  
Previsões do modelo ARIMA para o índice do PIB trimestral

Trimestre	Valor real	Um passo a frente	erro(%)	Até doze passos a frente	erro(%)
95/1	114,17	111,51	2,33	111,51	2,33
95/2	117,44	121,56	-3,51	118,85	-1,20
95/3	116,98	120,53	-3,04	122,56	-4,77
95/4	116,63	114,93	1,46	119,76	-2,68
96/1	112,96	112,56	0,35	114,24	-1,14
96/2	120,48	118,25	1,85	121,21	-0,61
96/3	123,25	122,45	0,65	124,90	-1,34
96/4	122,17	121,31	0,70	122,05	0,10
97/1	116,80	118,30	-1,29	116,73	0,06
97/2	124,81	122,96	1,49	123,93	0,70
97/3	126,43	126,77	-0,27	127,03	-0,48
97/4	124,63	124,11	0,42	123,66	0,78

A tabela 1 mostra os resultados das previsões um passo a frente durante doze meses e mostra também as previsões até doze passos a frente de 94/4. Estes resultados serão posteriormente comparados com os resultados obtidos por um modelo estrutural para esta mesma série.

A metodologia ARIMA também foi aplicada a série do Índice do Desempenho Industrial do RS (fig.2), a qual foi analisada com dados mensais de janeiro de 1991 a outubro de 1998, sendo que os últimos 12 meses não foram utilizados para estimação para que se pudesse comparar os valores previstos pelo modelo com os valores reais da série.

(1) Neste trabalho a decisão de aplicar a diferença de primeira ordem e a diferença sazonal de primeira ordem no lag 4 foi baseada na visualização das funções  $f_{ac}$  e  $f_{acp}$ , porém existem testes estatísticos para verificar a necessidade destas diferenças: para a não-estacionariedade pode ser usado o Dickey-Fuller ou Dickey-Fuller aumentado; para a não-estacionariedade sazonal pode ser usada uma extensão do D.F.



Fig.2 – Série: IDI do Rio Grande do Sul

Após experimentar alguns modelos para esta série o que apresentou melhores resultados foi

$$(1 + 0.260L + 0.255L^2 + 0.253L^3 + 0.230L^{12})z_t = 1,226 + \varepsilon_t$$

onde  $z_t = \ln(y_t)$

Ao ajustar este modelo aos dados foram obtidas as previsões descritas na tabela 2.

Tabela 2  
Previsões do modelo ARIMA para o IDI/RS

Período	Valor real	Um passo a frente	erro(%)	Até doze passos a frente	erro(%)
Nov/97	131,95	131,48	-0,36	131,48	-0,36
Dez/97	126,00	126,55	0,44	128,00	1,59
Jan/98	119,36	125,33	5,01	124,49	4,30
Fev/98	118,11	113,95	-3,52	119,17	0,89
Mar/98	135,60	119,08	-12,18	120,74	-10,96
Abr/98	131,67	125,38	-4,78	121,22	-7,94
Mai/98	132,90	127,02	-4,42	122,25	-8,01
Jun/98	135,52	131,35	-3,08	122,99	-9,25
Jul/98	132,72	132,18	-0,40	124,28	-6,36
Ago/98	131,47	129,04	-1,85	123,38	-6,15
Set/98	135,28	131,00	-3,17	124,17	-8,22
Out/98	133,40	132,59	-0,61	125,81	-5,69

## 4.2 Modelos estruturais

A mesma análise realizada na subseção anterior foi feita aqui porém, utilizando outro tipo de modelagem: a série de tempo é vista como o resultado do

somatório de vários componentes não diretamente observados, ou seja, na forma de um modelo estrutural.

Para a série do índice do Produto Interno Bruto Brasileiro, foi utilizado o modelo estrutural básico, no qual se verificou que  $\sigma_{\zeta}^2 = 0$  (tab.3), indicando que a inclinação é fixa. O padrão sazonal varia com o tempo, o que é mostrado na tabela 3, e os fatores sazonais finais são mostrados na tabela 5.

Tabela 3

Desvio padrão estimado dos componentes	
Componente	$\sigma$
Irregular	0,8510
Nível local	2,5084
Inclinação	0,0000
Sazonal	0,1506

Tabela 4

Vetor de estado estimado no último ponto no tempo			
	Coefficiente	RMSE	t - value
Nível local	116,0100	1,0763	107,79**
Inclinação	0,5209	0,3276	1,59
Sazonal_1	-0,8191	0,6474	-1,27
Sazonal_2	-4,4774	0,6528	-6,86
Sazonal_3	0,7641	0,4778	1,60**

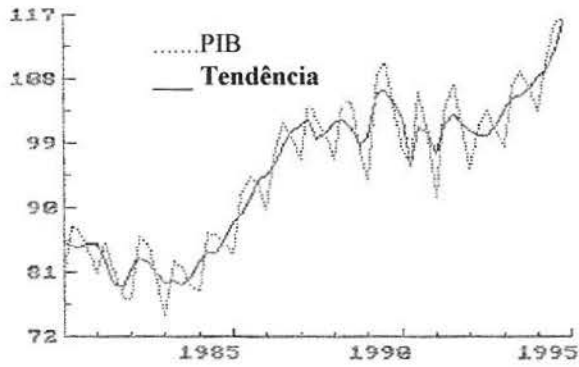
Variância do erro de previsão: 8,66774

Tabela 5

Fatores Sazonais			
1º trimestre	2º trimestre	3º trimestre	4º trimestre
-5,241	1,583	3,713	-0,055

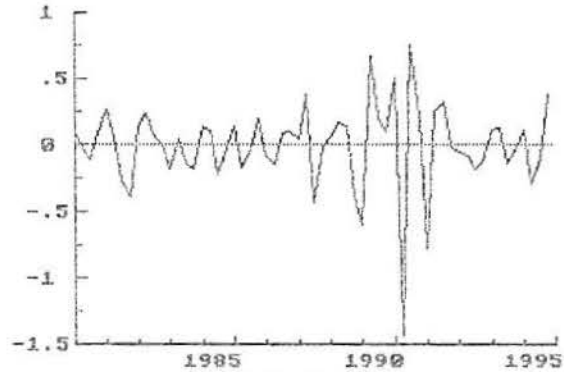
Os gráficos a seguir mostram cada componente da série separadamente:

Tendência



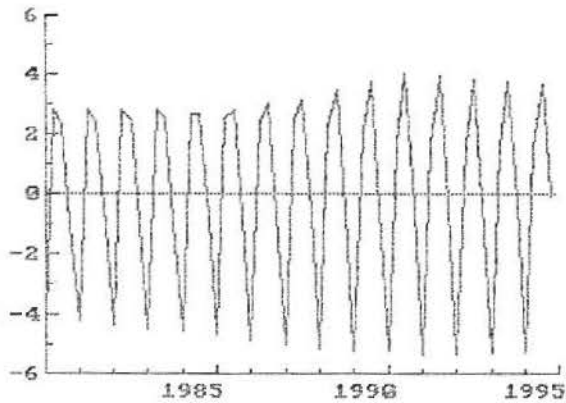
(Fig.3)

Componente Irregular



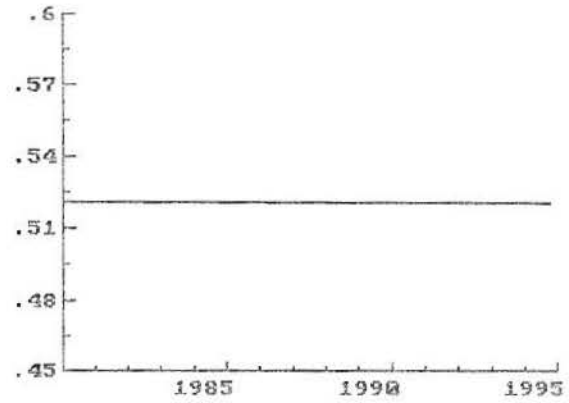
(Fig.4)

Componente Sazonal



(Fig.5)

Inclinação



(Fig.6)

Os resultados das previsões do modelo estrutural estão expostos na tabela 6:

Tabela 6

Previsões do Modelo Estrutural para o índice do PIB trimestral

Trimestre	Valor real	Um passo a frente	erro(%)	Até doze passos a frente	erro(%)
95/1	114,17	111,32	2,49	111,32	2,49
95/2	117,44	121,07	-3,09	118,70	-1,07
95/3	116,98	120,74	-3,22	121,34	-3,73
95/4	116,63	114,10	2,17	118,05	-1,22
96/1	112,96	112,31	0,58	113,41	-0,40
96/2	120,48	119,50	0,82	120,79	-0,26
96/3	123,25	122,43	0,66	123,43	-0,15
96/4	122,17	120,44	1,42	120,14	1,66
97/1	116,80	117,81	-0,87	115,49	1,12
97/2	124,81	123,73	0,87	122,87	1,55
97/3	126,43	126,68	-0,20	125,51	0,73
97/4	124,63	124,15	0,39	122,22	1,93



Para o IDI/RS primeiramente foi ajustado um modelo estrutural básico ao logaritmo natural da série, porém após examinar os hiperparâmetros estimados notou-se que o componente irregular tinha  $\sigma_{\varepsilon}^2 = 0$ , então foi adicionado ao modelo um componente cíclico que gerou os resultados mostrados nas tabelas 7, 8, e 9.

Tabela 7

Desvio Padrão Estimado	
Componente	$\sigma$
Irregular	0,00441
Nível local	0,00000
Inclinação	0,00000
Cíclico	0,01379
Sazonal	0,00302

Tabela 8

Vetor de Estado estimado no último ponto no tempo			
variável	coeficiente	RMSE	t-value
lvl	4,8776	0,0104	467,25**
slp	0,0040	0,0002	17,714**
cy_1	0,0151	0,0208	
cy_2	0,0010	0,0276	
sea1	0,0050	0,0148	0,3400
sea2	-0,0397	0,0152	-2,6093*
sea3	0,0289	0,0093	3,0986**
sea4	0,0274	0,0098	2,7684**
sea5	-0,0061	0,0083	-0,7335
sea6	0,0133	0,0087	1,5238
sea7	0,0164	0,0079	2,0697*
sea8	-0,0059	0,0084	-0,6993
sea9	0,0044	0,0079	0,5593
sea10	0,0115	0,0085	1,3538
sea11	-0,0108	0,0069	-1,5712
Variância do erro de previsão :	0,001221		
Variância do ciclo :	0,001451		
coeficiente $\rho$ :	0,93245		
Frequência :	0,33212		
Amplitude do ciclo :	0,01512		
Período do ciclo : 18,919	(1,577 anos)		

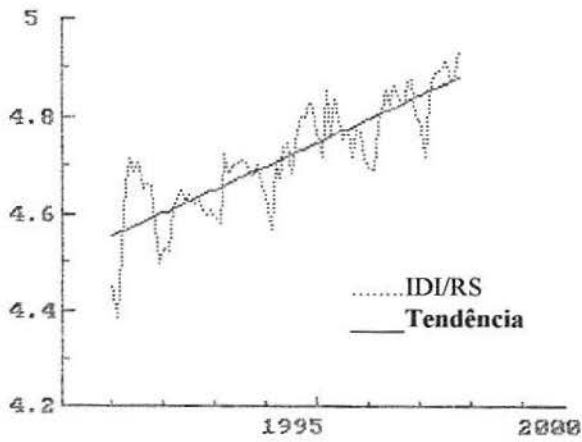
Tabela 9

Fatores Sazonais

jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
-0,043	-0,117	-0,003	0,031	0,036	0,029	0,040	-0,007	0,000	0,038	0,035	-0,038

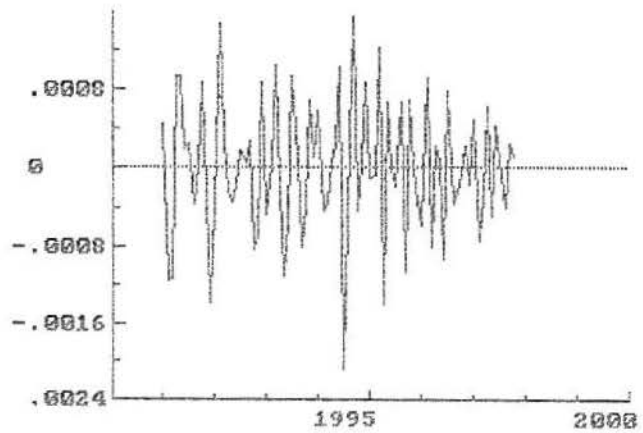
Os gráficos a seguir mostram cada componente da série separadamente:

Tendência



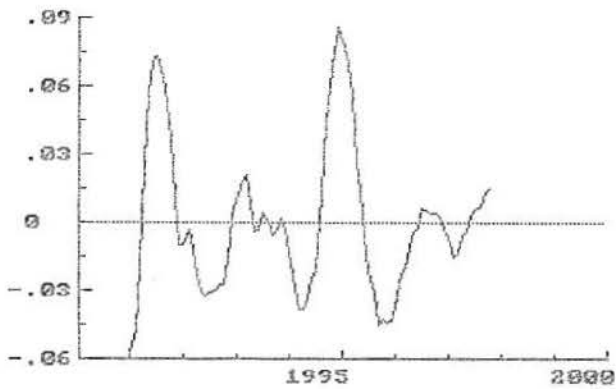
(Fig.7)

Componente Irregular



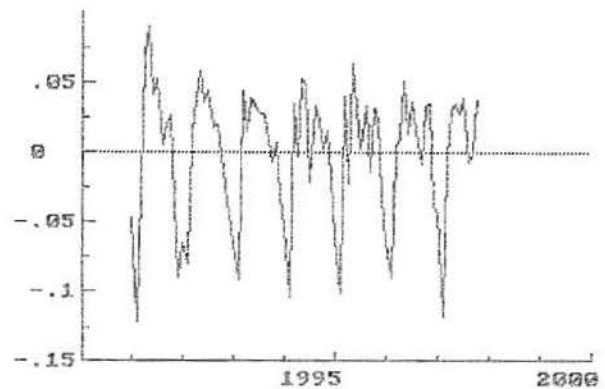
(Fig.8)

Componente Cíclica



(Fig.9)

Componente Sazonal



(Fig.10)

Pelos gráficos das componentes pode-se verificar uma tendência determinística na série, enquanto as componentes cíclica e sazonal são estocásticas.

As previsões geradas pelo modelo descrito acima estão na tabela 10 e posteriormente serão comparadas com as previsões geradas pelo modelo ARIMA para a mesma série.

Tabela 10  
Previsões do Modelo Estrutural para o IDI/RS

Período	Valor real	Um passo a frente	erro(%)	Até doze passos a frente	erro(%)
Nov/97	131,95	138,45	4,93	138,45	4,93
Dez/97	126,00	126,08	0,06	128,78	2,21
Jan/98	119,36	125,47	5,12	128,25	7,45
Fev/98	118,11	114,84	-2,77	119,11	0,84
Mar/98	135,60	130,65	-3,65	133,60	-1,48
Abr/98	131,67	139,66	6,06	138,30	5,04
Mai/98	132,90	137,07	3,14	139,16	4,72
Jun/98	135,52	135,54	0,01	138,51	2,20
Jul/98	132,72	138,59	4,42	140,58	5,93
Ago/98	131,47	131,82	0,27	134,75	2,50
Set/98	135,28	134,48	-0,59	136,39	0,82
Out/98	133,40	143,45	7,53	142,47	6,79

### 4.3 Comparação dos resultados empíricos

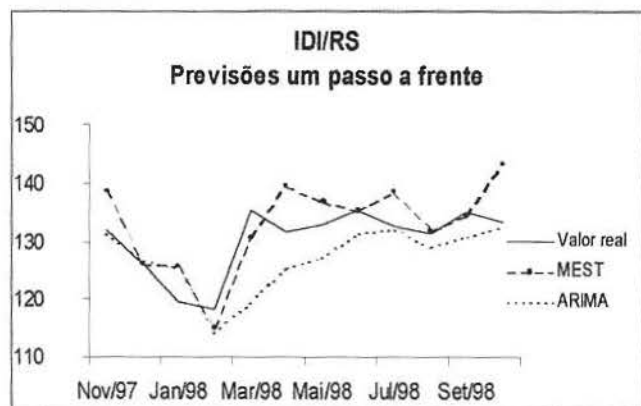
Visto que tanto a série do índice do PIB brasileiro quanto a série do IDI do RS foram modeladas por um modelo ARIMA e um modelo estrutural (MEST), puderam ser feitas comparações quanto a capacidade de cada metodologia para prever estas duas séries. Esta comparação foi feita através da análise do Erro Percentual Absoluto Médio (EPAM) para previsões um passo a frente durante doze períodos e também para previsões doze passos a frente do último período estimado.

Previsões um passo a frente. Pelos resultados das tabelas 1 e 6 e pelos gráficos das figuras 11 e 12 verifica-se que para a série do índice do PIB brasileiro as previsões um passo a frente mostraram-se bastante parecidas para os dois modelos estimados (ARIMA e MEST), sendo que para seis das doze

previsões (incluindo todo ano de 97) o MEST teve erros percentuais menores que o modelo ARIMA, e nas outras seis, o modelo ARIMA se aproximou mais dos valores reais. Já para as previsões da série do IDI/RS, em sete das doze previsões um passo a frente, o MEST foi mais eficaz que o modelo ARIMA onde destaca-se o período mar/98 onde o EPA do modelo ARIMA foi de  $-12,18\%$  enquanto para o mesmo período o EPA do MEST foi de  $-3,65\%$ . Em contrapartida, entre as quatro previsões em que o modelo ARIMA teve menor EPA destaca-se o período de nov/97, no qual o modelo ARIMA teve EPA de  $-0,36\%$ , para o MEST este erro foi de  $4,93\%$ .



(Fig.11)



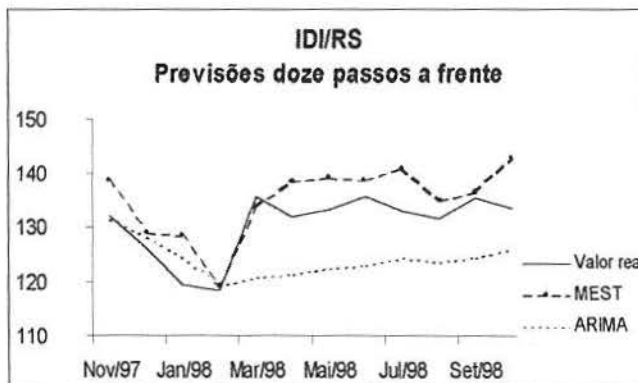
(Fig.12)

Previsões doze passos a frente. Como pode ser observado nas tabelas 2 e 10 e nas figuras 13 e 14, nas previsões até doze passos a frente (o último período conhecido era out/97) o MEST estimado para a série do PIB foi mais eficaz nos períodos entre dez/97 a maio/98. O modelo ARIMA previu melhor o período de nov/97 e os períodos entre abril/98 até out/98. Nas previsões da série do IDI/RS, o MEST teve erros percentuais menores que o modelo ARIMA em oito das doze previsões sendo que estas oito compreendem todo o período de fev/98 a set/98.

Novamente destaca-se o período de mar/98 onde o EPA do modelo ARIMA foi de -10,96% e do MEST foi de -1.48%.



(Fig.13)



(Fig.14)

Tabela 11  
Comparação dos EPAM percentuais

Série	Previsão	Média dos EPA		
		ARIMA	MEST	Prev. Ingênua
PIB	um passo	1,447	1,398	2,645
IDI/RS	um passo	3,317	3,213	3,527
PIB	até doze passos	1,350	1,360	-
IDI/RS	até doze passos	5,809	3,741	-

Um resumo da eficiência de cada modelo para cada série pode ser visto na tabela 11, onde foi calculada a média do EPA dos doze períodos estimados de cada modelo. Foi acrescentado ainda, para fins de comparação, o cálculo da média do EPA dos doze períodos estimados de um modelo de previsão ingênua, ou seja,  $1 \xi (1)$ .

Olhando para as médias dos EPA nota-se que somente na previsão doze passos a frente do PIB o MEST teve média menor que o modelo ARIMA e ainda assim, por uma diferença pequena, nas outras o MEST teve sempre, em média,

menor diferença entre os valores previstos e os valores reais, o que mostra um indício de que o Modelo Estrutural pode ser mais eficaz do que os modelos ARIMA.

## 5. CONCLUSÃO

O presente trabalho teve como meta reunir os principais aspectos de duas metodologias para previsão de séries de tempo: ARIMA e Modelos Estruturais, e ainda fazer uma comparação empírica onde procurou-se investigar a eficiência relativa destes modelos ao fazer a previsão de duas séries de tempo reais.

O que se nota a respeito das duas metodologias é que, os Modelos Estruturais têm a grande vantagem de isolar cada um dos componentes de uma determinada série, permitindo uma descrição das principais características pertinentes a esta série. Neste sentido, os modelos ARIMA deixam a desejar na maneira como são escritos, pois ao contrário dos MEST, eles não têm uma interpretação direta..

Pelos resultados empíricos do trabalho, de uma maneira geral, os Modelos Estruturais obtiveram melhores resultados quanto as previsões feitas para as séries do PIB brasileiro e do IDI do Rio Grande do Sul, mostrando que para estas séries é realmente mais vantajoso trabalhar com Modelos Estruturais.

Os modelos ARIMA tiveram seu grande momento nas décadas de setenta e oitenta e os Modelos Estruturais praticamente os substituíram a partir do início da década de noventa. Hoje em dia, os Modelos Estruturais estão começando a ser questionados quanto a sua eficiência pelas redes neurais desenvolvidas originalmente no campo da inteligência artificial, o que pode servir de tema para trabalhos futuros.

## 6. BIBLIOGRAFIA

- BOX, G. E. P., JENKINS, G. M e REISEL, G.C (1994), Time Series Analysis, Forecasting and Control, 3ª ed., Prentice Hall, New Jersey.
- FERNANDES, L. G. F., PORTUGAL, M. S. e NAVAUX, P. O. A. (1996), “Previsão de Séries de Tempo: Redes Neurais e Modelos Estruturais”, Pesquisa e Planejamento Econômico, 26(2), p. 253-276.
- HARVEY, A. C. (1989), Forecasting Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge.
- NELSON, C. (1973), Applied Time Series Analysis, Holden-Day, San Francisco
- PORTUGAL, M. S. (1995), Neural Networks versus Time Series Methods: a Forecasting Exercise, Revista Brasileira de Economia, 49(4), p. 611-629.
- CORRÊA, W.R. e PORTUGAL, M.S. (1998), Previsão de Séries de Tempo na Presença de Mudança Estrutural: Redes Neurais e Modelos Estruturais, Revista Economia Aplicada, FIPE/FEA-USP