

## A Previsão com a Metodologia de Box-Jenkins

### 1. Introdução

Suponha que exista um sistema que atue como um filtro (Figura 1), que é estimulado por uma série de ruídos brancos, resultantes de um processo de geração de números aleatórios, e que com esse estímulo seja gerada pelo sistema uma seqüência de valores observados seguindo um padrão, que corresponde à série temporal  $Y_t$ .



Figura 1 – Geração de série temporal  $Y_t$

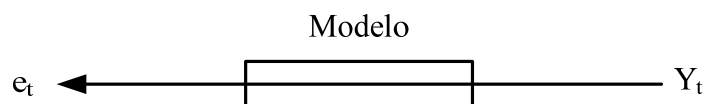


Figura 2 – Associação de modelo à série de observações  $Y_t$

Em situações da realidade, tem-se o caminho inverso ou seja, conhece-se o conjunto de observações seqüenciais  $Y_t$  geradas pelo sistema em questão, ao qual busca-se associar um modelo que corresponda aos processos internos ao sistema que as gerou (Figura 2). Uma vez que se estabeleça um modelo operacional para essa representação, a série aleatória  $e_t$  de valores em torno de zero corresponde à seqüência de valores (resíduos) que resulta ao extrair de  $Y_t$  os valores obtidos com o modelo ajustado à essa mesma série de valores observados  $Y_t$ .

A metodologia de Box-Jenkins para a previsão se baseia no ajuste de modelos tentativos denominados ARIMA a séries temporais de valores observados de forma que a diferença entre os valores gerados pelos modelos e os valores observados resulte em séries de resíduos de comportamento aleatório em torno de zero.

Os modelos ARIMA (autoregressivos integrados e de médias móveis) são capazes de descrever os processos de geração de uma variedade de séries temporais para os previsores (que correspondem aos filtros) sem precisar levar em conta as relações econômicas, por exemplo, que geraram as séries.

Segundo a sistemática da metodologia de Box-Jenkins os modelos ARIMA descrevem tanto o comportamento estacionário como o não-estacionário. Dessa forma, pode-se afirmar que essa é uma metodologia de modelagem flexível em que as previsões com base nesses modelos são feitas a partir dos valores correntes e passados dessas séries.

Pode-se associar o conceito inicial de um filtro estimulado por uma série aleatória do tipo ruído branco à metodologia de Box-Jenkins (Figura 3).

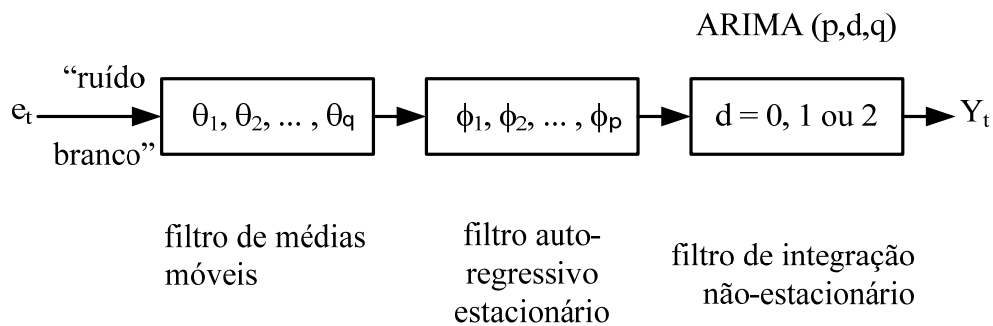


Figura 3 – Os filtros de médias móveis, autoregressivo e de integração não-estacionária

Na Figura 3 é representado um conjunto de sucessivos filtros aos quais associam-se os parâmetros dos modelos ARIMA (p,d,q) que representam os sistemas estimulados pela série  $e_t$  que geraram a série temporal  $Y_t$ : o filtro de médias móveis (parâmetro q), o filtro autoregressivo estacionário (parâmetro p) e o filtro de integração não-estacionário (parâmetro d).

## 2. A Estacionariedade da Série Temporal

Os modelos ARIMA ou Box-Jenkins são excelentes modelos de previsão de curto prazo (Granger & Newbold, 1977). Resultados de análises com esses modelos mostram que os melhores resultados (previsões) são obtidos com de 5 a 10 anos de informação (mensal), particularmente na presença de sazonalidade. Como já visto, a importância do processo observado ser estacionário é a possibilidade de fixar parâmetros do modelo válidos para previsão do futuro a partir do passado. Assim, como primeiro passo para essa modelagem são realizados procedimentos para a remoção da não-estacionariedade.

A análise das estatísticas básicas das séries estacionárias permite separar a estacionariedade em dois grupos:

- (1) estacionariedade no amplo senso: médias, variâncias e covariâncias constantes no tempo.
- (2) estacionariedade no estrito senso: probabilidade de uma dada flutuação no processo em torno da média é a mesma em qualquer momento do processo.

Na prática, aceita-se que as séries observadas sejam séries fracamente estacionárias, situações nas quais garante-se apenas médias e variâncias invariantes no tempo. Existe uma equivalência entre os modelos ARIMA e os modelos ARMA (autoregressivos e de médias móveis). Esses últimos são ajustados a séries já estacionárias transformadas pelo método das diferenças de ordem  $d$  ou seja, cujas séries originais são séries não-estacionárias homogêneas (assim denominadas por ter sido possível obter a estacionariedade com um número finito de diferenciações), como a série do exemplo a seguir (Figuras 4, 5, 6 e 7).

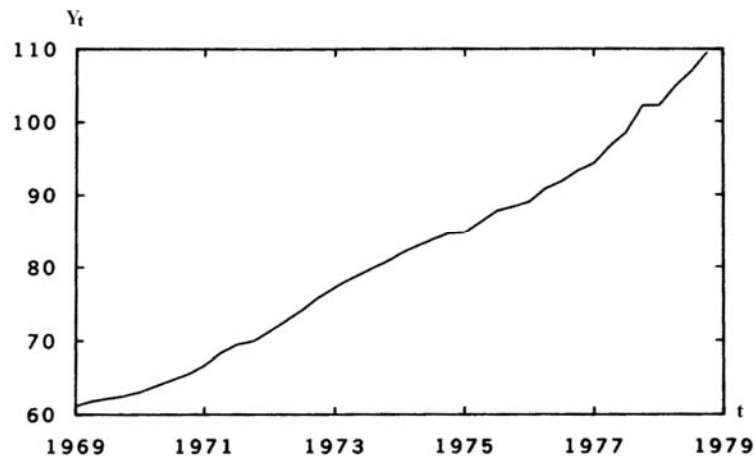
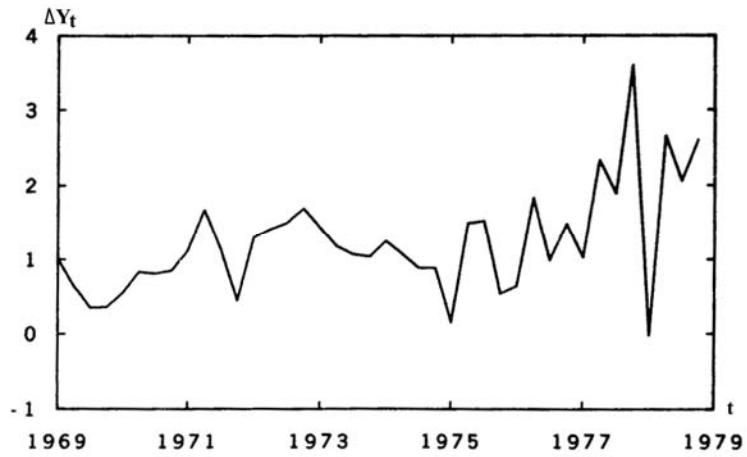


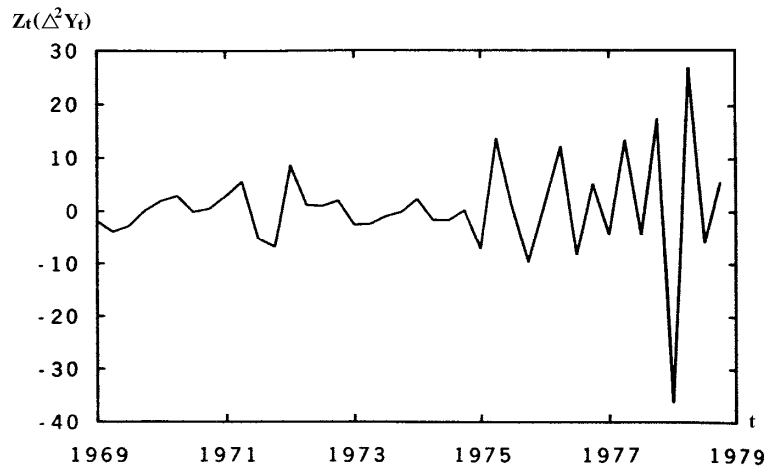
Figura 4 – Série de oferta monetária ajustada sazonalmente



$t = 2, \dots, n.$

Variância não-constante no tempo.

Figura 5 – Série das 1<sup>as</sup> diferenças

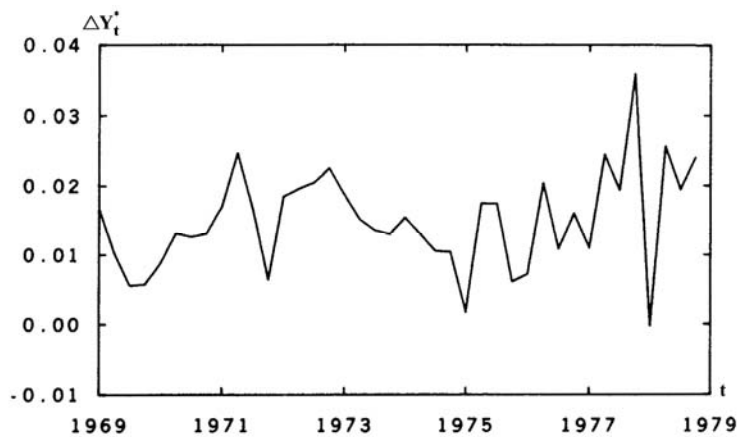


$t = 3, \dots, n.$

$$Z_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}.$$

Variância não-constante no tempo.

Figura 6 – Série das 2<sup>as</sup> diferenças.



$$Y_t^* = \log Y_t$$

Variância  $\pm$  constante.

Série pode ser dita estacionária.

A série original é dita homogênea de ordem  $d = 1.$

Figura 7 – Série das primeiras diferenças dos dados transformados.

Nas figuras 4, 5 e 6 são mostrados os gráficos de uma série original de oferta monetária ajustada sazonalmente, onde existe clara tendência, e das respectivas séries transformadas de primeiras e segundas diferenças. Observa-se que a variância é não-constante no tempo em todas elas. Nesse caso é feita a transformação logarítmica nos dados originais e em seguida obtida a série das primeiras diferenças da série  $Y_t^* = \log Y_t$  ou seja:  $\Delta Y_t^*$ , para a qual pode-se aceitar que a variância é constante e, assim, a série original é dita homogênea de primeira ordem ( $d = 1$ ) (Figura 7).

À luz da análise de autocorrelação, seguem exemplos do uso do correlograma para identificar séries temporais estacionárias para dados não sazonais:

1. Série pode se considerada estacionária quando há truncamento abrupto ou padrão de decaimento rápido (Figuras 8 e 9).

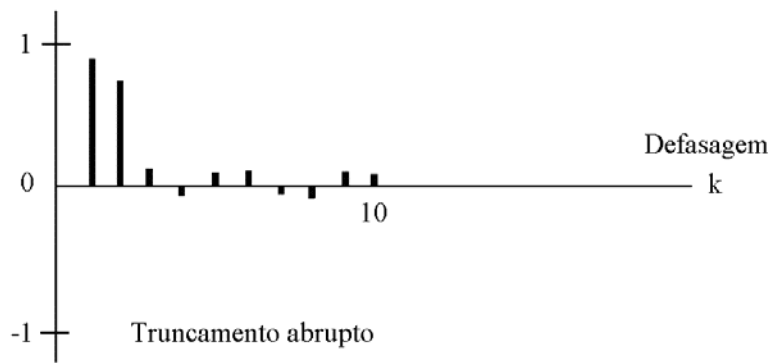


Figura 8 – Valores das autocorrelações significativos nos períodos iniciais de defasagem e truncamento após

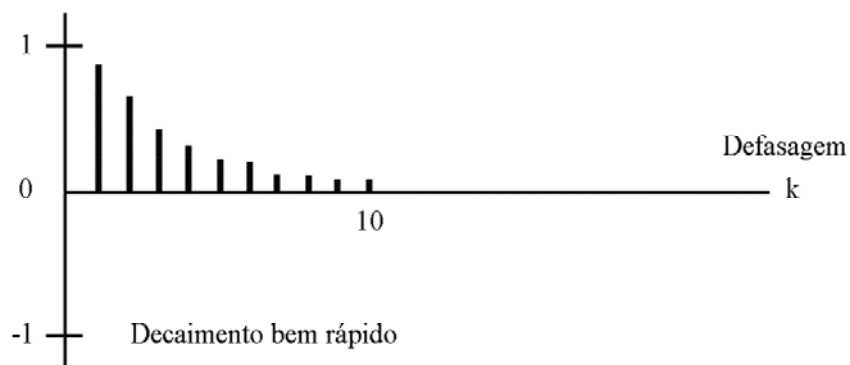


Figura 9 – Valores das autocorrelações significativos decaindo bem rápido para valores próximos de zero

2. Série pode ser considerada não-estacionária quando há padrão de decaimento lento para zero (Figura 10).

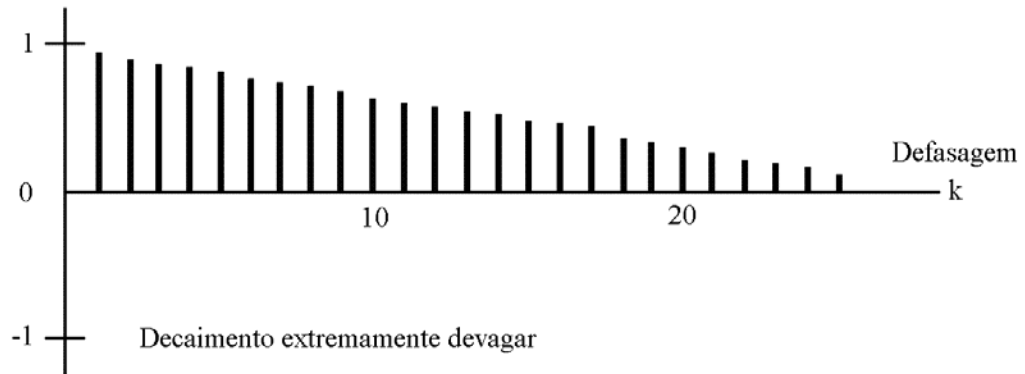


Figura 10 – Valores das autocorrelações significativos decaindo extremamente devagar segundo padrões de decaimento

Como exemplos de padrão de decaimento podem ser observados os padrões representados na Figura 11:

- a) decaimento exponencial sem oscilações
- b) decaimento exponencial com oscilações
- c) decaimento em onda senoidal

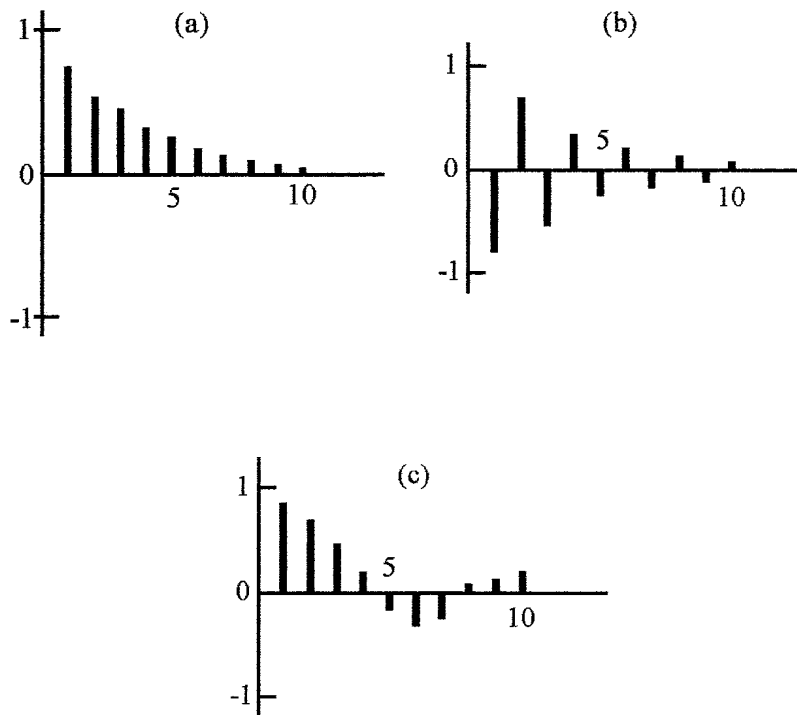


Figura 11 – Exemplos de padrão de decaimento

## 2.1 Um Teste para Verificar a Estacionariedade: O Teste da Raiz Unitária de Dickey-Fuller

Seja o processo estocástico de raiz unitária:

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \quad -1 \leq \rho \leq 1,$$

onde  $\varepsilon_t$  é um termo de erro de ruído branco.

Quando  $\rho = 1$  (raiz unitária), o processo gerador da série  $Y_t$  é o passeio aleatório e a série é não-estacionária.

Subtraindo-se  $Y_{t-1}$  dos dois lados da igualdade, escreve-se:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \varepsilon_t$$

$$\therefore \Delta Y_{t-1} = \delta Y_{t-1} + \varepsilon_t, \quad (1)$$

onde  $\delta = \rho - 1$ .

Um procedimento de teste da raiz unitária pode ser aplicado à equação (1). Dessa forma, estima-se os parâmetros da equação (1) e testa-se se  $\delta = 0$ .

Se  $\delta = 0, \rho = 1 \Rightarrow$  raiz unitária e a série original é não-estacionária.

Entretanto, a estatística  $t$  de Students não se aplica nesse caso (Gujarati, 2006). Usa-se o teste de Dickey-Fuller, que avalia se o valor da estatística  $t$  estimado para o parâmetro de  $Y_{t-1}$  segue a distribuição da estatística  $\tau$  (tau). Dessa forma, o teste de Dickey-Fuller verifica a hipótese nula ( $\delta = 0$ ) é rejeitada ou não em determinados níveis de significância estatística, conforme valores tabulados (Gujarati, 2006, pág. 791). Se o valor em módulo de  $t < |\text{estatística } \tau|$ , não é possível rejeitar a hipótese nula ou seja, a hipótese de não-estacionariedade.

O procedimento do teste deve ser mais sofisticado, entretanto. À estimação da equação (1) (e conseqüente teste de hipótese sobre  $\delta = 0$ ), devem se juntar a estimação das variantes possíveis do processo de passeio aleatório e os respectivos testes de hipóteses sobre  $\delta = 0$ :

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \varepsilon_t \quad (2)$$

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \varepsilon_t \quad (3)$$

onde  $t$  é o tempo ou a variável de tendência.

Nessas duas variantes [(2) e (3)],  $\delta = 0$  significa que há raiz unitária ou seja, a série  $Y_t$  é não-estacionária. A hipótese alternativa é que  $\delta < 0$ , significando  $\rho - 1 < 0$  ou seja  $\rho < 1$  (observe que  $\rho > 1$  não é aceita por implicar em explosão de valores da série).

Se a hipótese nula for rejeitada, isso significa que (Gujarati, 2006):

- $Y_t$  é uma série temporal estacionária com média zero: equação (1)
- $Y_t$  é uma série temporal estacionária com média diferente de zero: equação (2)
- $Y_t$  é uma série estacionária em torno de tendência determinística: equação (3)

## 2.2 O Teste F de Dickey-Fuller

Suponha-se que a variável  $Y_t$  possa ser descrita por:

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t \quad (4)$$

Dickey e Fuller derivaram a distribuição para o estimador  $\hat{\rho}$  e geraram um teste para a hipótese de  $Y_t$  se comportar como o passeio aleatório isto é, sob a hipótese que  $\beta = 0$  e  $\rho = 1$ .

O teste pode ser aplicado para uma versão geral da equação (4) ou seja, sob a suposição de que  $Y_t$  possa ser descrita por:

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \lambda_1 \Delta Y_{t-1} + \varepsilon_t,$$

onde  $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$ .

Outras defasagens  $\Delta Y_{t-i}$  podem ser incluídas na equação, mas o teste seria o mesmo.

Com base nos Mínimos Quadrados Ordinários, são estimados os parâmetros das regressões irrestritas (5) e restrita (6) seguintes:

$$Y_t - Y_{t-1} = \alpha + \beta t + (\rho - 1) Y_{t-1} + \lambda_1 \Delta Y_{t-1} \quad (5)$$

$$Y_t - Y_{t-1} = \alpha + \lambda_1 \Delta Y_{t-1} \quad (6)$$



O procedimento do teste calcula a estatística F como a razão:

$$F = \frac{(N - K) (ESS_R - ESS_{IR})}{q (ESS_{IR})}$$

onde:

$ESS_R$  e  $ESS_{IR}$  são a soma do quadrado dos resíduos nas regressões restrita e irrestrita, respectivamente,

$N$  é o número de observações,

$K$  é o número de parâmetros estimados na regressão irrestrita, e

$q$  é o número de restrições de parâmetros (no caso:  $\beta = 0$ ;  $\rho = 1$ )

O valor calculado de F é comparado com os valores críticos da estatística tabulada em uma tabela apropriada desenvolvida por Dickey e Fuller (Pindyck e Rubinfeld, 1991, pág. 461). Assim, se  $F_{\text{calculado}} > F_{\text{crítico}}$  de Dickey-Fuller, a hipótese do passeio aleatório pode ser rejeitada.

### 3. A Formulação dos Modelos ARMA (ou ARIMA)

A diferença básica entre a regressão clássica e os modelos de séries temporais é que nos modelos de séries temporais ARMA (ou ARIMA) não se pode assumir independência entre observações. Ao contrário, os modelos autoregressivos e de médias móveis vão modelar o grau de autocorrelação entre desvios e observações defasadas.

Em suma, de forma geral, quando faz-se referência a modelos ARIMA esses modelos estão sendo ajustados à série original. Já ao fazer-se referência a modelos ARMA, considera-se que a série é uma série diferenciada. Supondo  $Y_t$  a série já diferenciada, os modelos ARMA em sua forma geral se escrevem:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

O melhor modelo deve ser parcimonioso ou seja, deve-se  $\min \{p, q\}$  (o que quer dizer que se deve tentar utilizar o menor conjunto de parâmetros possível para seu

ajustamento à série de dados observados). Os parâmetros  $p$  e  $q$  representam o número de parâmetros relativos aos comprimentos de defasagem em que observa-se valores significativos das autocorrelações e que correspondem a particularidades do sistema de geração das séries que devem ser explicadas pelo modelo (pois correspondem a um padrão de geração). O processo gerador dos dados da série é dito aleatório linear se o modelo ajustado  $Y_t$  pode ser descrito como uma combinação linear de valores defasados de  $Y_t$  e  $\varepsilon_t$ .

Esses modelos podem, para facilitar a compreensão, ser separados em dois modelos complementares (conforme Figura 3): os modelos de médias móveis e os modelos autoregressivos. Os primeiros correspondem aos processos de médias móveis de ordem  $q$  em que cada observação  $Y_t$  é gerada por uma média ponderada dos erros aleatórios  $q$  períodos no passado. Denota-se esse processo por MA( $q$ ), cuja equação é:  $Y_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ , onde os parâmetros  $\theta_1, \dots, \theta_q$  podem ser positivos ou negativos. O sinal negativo no terceiro termo em diante corresponde a uma convenção. Assume-se que os erros aleatórios são gerados por um processo de geração de ruídos brancos ou seja:  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  e covariância  $\gamma_k = 0$  para defasagens  $k \neq 0$ . Observa-se que  $Y_t$  são dependentes, ou melhor, uma função linear da ponderação dos erros.

Seja exemplo, inicialmente, o modelo de processos aleatórios simples:

$Y_t = \varepsilon_t$ ,  $t = 0, \pm 1, \pm 2, \dots$ , válido para série puramente aleatória, nos quais, se  $\varepsilon_t$  são normalmente distribuídos, as observações são ditas corresponderem ao ruído branco.

Por outro lado, é exemplo de modelo de um processo aleatório linear simples o modelo de médias móveis. Por exemplo, o modelo de médias móveis de primeira ordem escreve-se:  $Y_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ , cujos parâmetros são  $\alpha, \theta_1, \sigma_\varepsilon^2$ . Como exemplo de processo aleatório linear simples (MA(1)), tome-se exemplo de MA(1) com  $\theta_1 = 0,25$  (Figura 12).

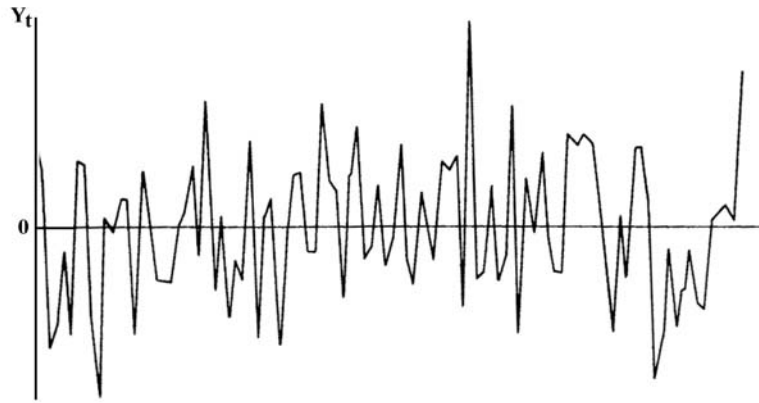


Figura 12 – Exemplo de processo de médias móveis MA(1) com  $\theta_1 = 0,25$  (Levenbach e Cleary, 1984).

Outro exemplo de processo aleatório linear simples é o processo AR(p) (auto-regressivo) em que  $p = 1$ . O modelo AR(1) correspondente ao processo autoregressivo de 1ª ordem é também conhecido por processo de Markov, onde:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \varepsilon_t, \text{ cujos parâmetros são } \alpha, \phi_1, \sigma_\varepsilon^2.$$

Na Figura 13 é representado exemplo de AR(1) com  $\phi_1 = +0,9$  ( $Y_t = 0,9Y_{t-1} + \varepsilon_t$ ).

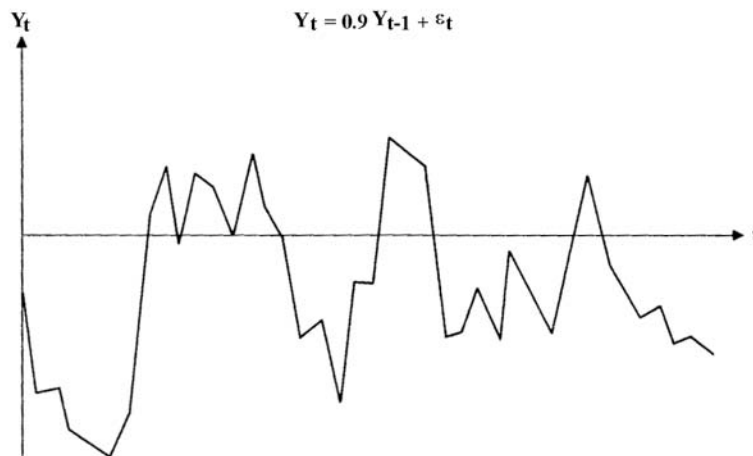


Figura 13 – Exemplo de processo autoregressivo AR(1) com  $\phi_1 = 0,9$

Outro exemplo, é o modelo AR(1) com  $\phi_1 = -0,9$  ( $Y_t = -0,9Y_{t-1} + \varepsilon_t$ ) (Figura 14).

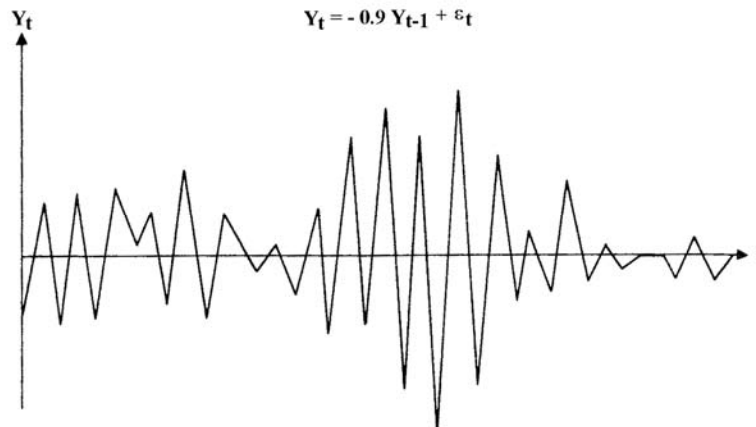


Figura 14 – Exemplo de processo autoregressivo AR(1) com  $\phi_1 = -0,9$

O modelo autoregressivo genérico AR(p) escreve-se:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t.$$

Na verdade, como o nome indica, AR(p) modela uma autoregressão da variável  $Y_t$  com essa mesma variável, defasada ( $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ ), para os p períodos de defasagem em que a autocorrelação parcial entre as variáveis é significativa.

Na Figura 15 é representado o gráfico de exemplo de AR(2) com  $\phi_1 = +1,0$  e  $\phi_2 = -0,5$  ( $Y_t = 1,0Y_{t-1} - 0,5Y_{t-2} + \epsilon_t$ ).

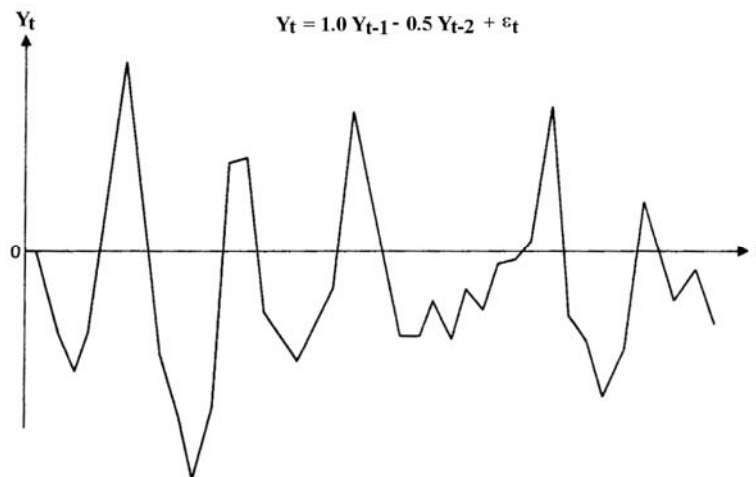


Figura 15 – Exemplo de processo autoregressivo AR(2) com  $\phi_1 = +1,0$  e  $\phi_2 = -0,5$

Existe um princípio de dualidade entre os modelos do tipo MA e AR de forma que existe a seguinte correspondência entre eles:

MA de ordem finita  $\Leftrightarrow$  AR de ordem infinita

AR de ordem finita  $\Leftrightarrow$  MA de ordem infinita

Quando o processo gerador de série temporal corresponde a modelos compostos por parte autoregressiva e parte médias móveis, os modelos utilizados são o ARMA(p,q). Como exemplo de ARMA(1,1), tem-se:  $Y_t = \alpha + \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ .

#### 4. Estágios da Metodologia de Box-Jenkins

A metodologia de Box-Jenkins corresponde a três estágios principais: (1) identificação de modelos tentativos (e de seus parâmetros), (2) estimação, e (3) teste de adequação, aos quais se segue a aplicação do modelo para a previsão ou controle do sistema de geração dos valores observados  $Y_t$  (Figura 16).

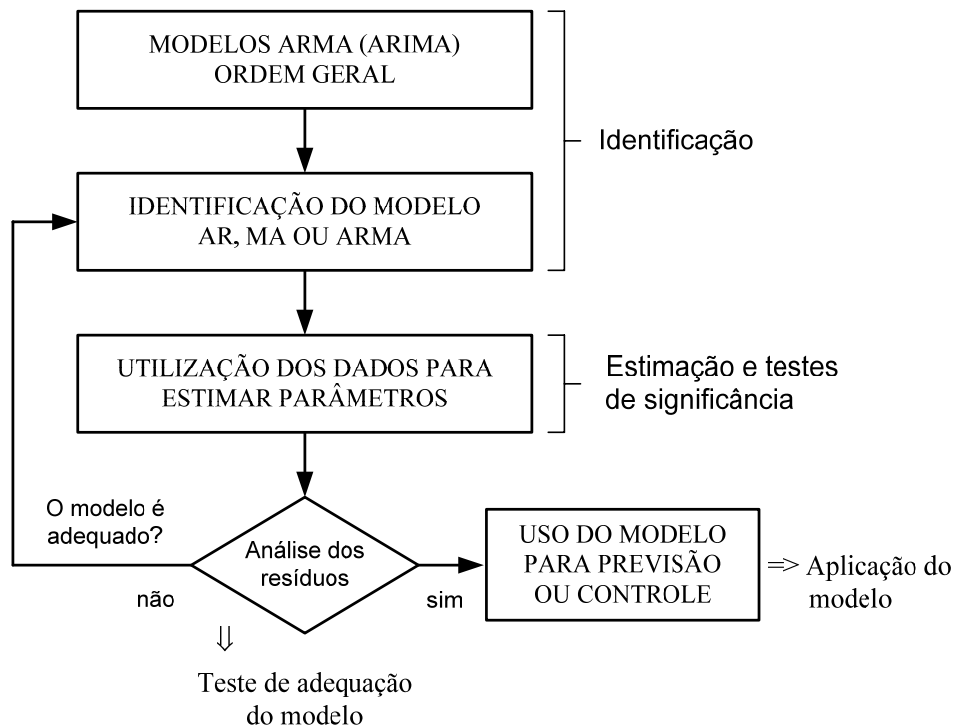


Figura 16 – Estágios da metodologia de Box-Jenkins

Assim, como tarefa inicial é preciso determinar p e q para a identificação de modelos tentativos. Para isso, procede-se ao exame dos coeficientes de autocorrelação e dos coeficientes de autocorrelação parcial, que permitem medir a força relativa de interação entre as variáveis  $Y_t$  defasadas.

A combinação dos termos ponderados por esses dois coeficientes, na ausência de aleatoriedade, poderiam revelar o modelo exato ARMA(p,q). Mas é fato que a aleatoriedade está presente na amostra, o que leva a desvios dos verdadeiros valores observados. Logo, podem haver enganos na identificação dos coeficientes de autocorrelação com base na análise de dados amostrais. Esses enganos são revelados no teste de adequação do modelo ou análise dos resíduos.

## 5. A Autocorrelação Parcial

A autocorrelação parcial pode ser definida com o último termo autoregressivo de um modelo AR(m). Assim,  $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \dots, \hat{\phi}_{m-1}, \hat{\phi}_m$  são as m autocorrelações parciais de qualquer processo AR(m), como pode ser visto nas equações seguintes:

$$\begin{array}{l} \text{Equações} \left\{ \begin{array}{l} X_t = \hat{\phi}_1 X_{t-1} + \varepsilon_t \quad (7) \\ X_t = \phi_1 X_{t-1} + \hat{\phi}_2 X_{t-2} + \varepsilon_t \quad (8) \\ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \hat{\phi}_3 X_{t-3} + \varepsilon_t \quad (9) \\ \dots \\ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \hat{\phi}_{m-1} X_{t-m+1} + \varepsilon_t \quad (10) \\ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \hat{\phi}_m X_{t-m} + \varepsilon_t \quad (11) \end{array} \right. \end{array}$$

$\hat{\phi}_i$  são usadas para medir o grau de associação entre  $X_t$  e  $X_{t-i}$ , os demais efeitos constantes, em sistemas de equações do tipo das equações (7)-(11) acima, que correspondem à um sistema de solução trabalhosa.

As estimativas de  $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_m$  podem ser obtidas com base na transformação do sistema de equações. Seja exemplo:

$$X_t = \hat{\phi}_1 X_{t-1} + \varepsilon_t \quad \text{modelo autoregressivo de 1ª ordem}$$

Multiplicando o primeiro e segundo termos da equação por  $X_{t-1}$ , obtém-se:

$$X_{t-1} \cdot X_t = \hat{\phi}_1 X_{t-1} \cdot X_{t-1} + X_{t-1} \varepsilon_t,$$

onde:

$$E(X_{t-1} \cdot X_t) = \hat{\phi}_1 E(X_{t-1}^2) + E(X_{t-1} \varepsilon_t)$$

$\gamma_1$

$\gamma_0$

$0$

$$\therefore \gamma_1 = \hat{\phi}_1 \gamma_0, \frac{\gamma_1}{\gamma_0} = \hat{\phi}_1 \Rightarrow \rho_1 = \hat{\phi}_1, \text{ ou melhor, que: } \hat{\phi}_1 = \rho_1 = r_1, \text{ sendo } \rho_1 \text{ a}$$

autocorrelação teórica (população) e  $r_1$  o valor amostral estimado. Sucessivamente,

$$\text{pode-se derivar } \rho_k = \frac{\gamma_k}{\gamma_0}.$$

Pode-se trabalhar as equações do sistema de equações (7)-(11) e obter as equações de Yule-Walker para obtenção das autocorrelações defasadas e das autocorrelações parciais. Mas no caso de modelos complexos como o modelo ARMA (p,q), opta-se por simplificação nas estimativas dos valores dessas autocorrelações. Assim, obtém-se o correlograma da função da autocorrelação parcial (ou correlograma parcial, que equivale ao gráfico das autocorrelações parciais para diferentes defasagens), com base no cálculo de  $r_{kk} \equiv$  coeficiente de autocorrelação parcial amostral e na realização de testes de hipótese para avaliação dos intervalos de defasagem entre as variáveis observadas em que essa autocorrelação é significativa. Para isso, deve-se observar:

$$1) \quad r_{kk} = \begin{cases} r_1, & \text{se } k=1 \\ \frac{(r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j})}{(1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j)} & \text{se } k = 2, 3, \dots \end{cases}$$

$$\text{onde } r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad \text{para } j = 1, 2, \dots, k-1$$

2) O desvio padrão de  $r_{kk}$

$$S_{r_{kk}} = \frac{1}{(n-b+1)^{1/2}}$$

$$3) \quad t_{r_{kk}} = \frac{r_{kk}}{S_{r_{kk}}}$$

- 4) Na prática, se  $|t_{r_{kk}}| > 2$ , significa que autocorrelações parciais são significativas.

## 6. Expressão dos Modelos ARIMA na Forma Compactada

Define-se o operador B como  $BY_t = Y_{t-1}$ .

Assim:  $B^2 Y_t = B(BY_t) = B(Y_{t-1}) = Y_{t-2}$

$$B^k Y_t = Y_{t-k}$$

$$Y_t - Y_{t-1} = Y_t - BY_t = (1-B)Y_t$$

Dessa forma, o modelo AR(1) se escreve, na forma compactada:

$$AR(1) \Rightarrow (1 - \phi_1 B)Y_t = \alpha + \varepsilon_t$$

Por simplificação,  $Y_t$  incluirá  $\alpha$ , sendo  $(Y_t - \alpha)$ , o que corresponde a modelar a série ajustada à média amostral ou seja:  $(Y_t - \bar{Y})$ ,  $\bar{Y}$  sendo a média amostral da série.

Assim, reescreve-se AR(1) como  $(1 - \phi_1 B) Y_t = \varepsilon_t$ .

Para defasagens de ordem sazonal pode-se utilizar o operador  $B^L$ , sendo L o comprimento de sazonalidade, de forma que:  $Y_t - Y_{t-12} = Y_t - B^{12}Y_t = (1-B^{12})Y_t$ .

Escrevendo os modelos na forma compactada é possível explorar a dualidade entre os modelos do tipo MA e AR. Para isso, seja o modelo AR(1):  $(1 - \phi_1 B) Y_t = \varepsilon_t$ ,

que pode ser reescrito  $Y_t = \frac{\varepsilon_t}{1 - \phi_1 B}$ .

Expandindo  $Y_t$  como série de potência da forma:

$$(1 + x)^p = 1 + px + \frac{p(p-1)}{2!}x^2 + \dots,$$

o termo  $\frac{1}{1 - \phi_1 B} = (1 - \phi_1 B)^{-1} = 1 + \phi_1 B + \phi_1^2 B^2 + \dots$ , onde  $-1 < \phi_1 < 1$ .

Dessa forma, escreve-se o modelo AR(1):  $Y_t = \varepsilon_t + \overset{\text{coef. 1}}{\phi_1} \varepsilon_{t-1} + \overset{\text{coef. 2}}{\phi_1^2} \varepsilon_{t-2} + \dots \overset{\text{...coef. n}}{\rightarrow}$ , ou seja, ele é expresso como um processo MA de ordem infinita (valores dos coeficientes  $\phi_1^n$  decrescentes exponencialmente).



De forma similar,  $MA(1) \Rightarrow Y_t = (1 - \theta_1 B) \varepsilon_t$  pode ser visto como AR de ordem infinita, pela mesma propriedade da dualidade ou melhor, se as condições de invertibilidade e estacionariedade forem obedecidas (serão apresentados mais à frente).

Os processos de maior ordem MA, AR e ARMA podem ser representados por:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t,$$

onde  $Y_t$  é a série observada ajustada à média amostral.

$$\text{Assim, } Y_t = \frac{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)}{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)} \varepsilon_t$$

sendo:

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q ;$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \text{ e}$$

$$H(B) = \frac{\theta_q(B)}{\phi_p(B)} \text{ pode ser denominada função de transferência.}$$

Para representar o modelo de ARIMA (p,d,q),  $Y_t$  nos modelos acima é substituído por  $w_t$ , sendo  $w_t$  a série estacionária:

$$w_t = (1 - B) Y_t \quad \text{1ª diferença}$$

ou, de forma geral,  $w_t = (1 - B)^d Y_t$  ou seja, a série estacionária de ordem d.

Assim, o modelo ARIMA (p,d,q) escreve-se:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) \varepsilon_t$$

Por exemplo, o modelo ARIMA (1,1,1) assume a forma compactada:

$$(1 - \phi_1 B)(1 - B) Y_t = (1 - \theta_1 B) \varepsilon_t$$

O modelo operacional do sistema gerador da série original  $Y_t$  deve obedecer às exigências 1 e 2:

- 1) Unicidade de representação
- 2) Equilíbrio estatístico em torno da média

O que garante:

- 1) Unicidade da representação, são as condições de invertibilidade de  $w_t$ .

- 2) Equilíbrio em torno do valor de uma média fixa, são as condições de estacionariedade de  $w_t$ .

## 7. As Condições de Estacionariedade e de Invertibilidade

Seja o modelo autoregressivo puro AR(p). O modelo autoregressivo é válido para todas as observações da série. Assim, pode ser escrito para a média das observações da série estacionária  $Z_t$ :

$$\bar{Z}_t = \alpha + \phi_1 \bar{Z}_{t-1} + \phi_2 \bar{Z}_{t-2} + \dots + \phi_p \bar{Z}_{t-p}$$

Supondo-se conhecido o valor teórico  $\mu$  ou a média populacional, sendo que essa verdadeira média é constante em qualquer defasagem em que seja calculada, pode-se escrever:

$$\mu = \alpha + \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu,$$

$$\text{ou seja: } \mu = \frac{\alpha}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)}.$$

A análise da função de autocorrelação de processos estacionários leva ao estabelecimento de conjuntos de condições necessárias e suficientes para os valores  $\phi_k$  que garantem a estacionariedade, isto é, média finita de variância invariante do tempo que se somam à condição necessária  $(\phi_1 + \phi_2 + \dots + \phi_p) < 1$ .

Para o conjunto de modelos abaixo, as condições de estacionariedade são as seguintes:

<u>Modelo:</u>	<u>Condições de estacionariedade</u>
$Z_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1}$	Nenhuma
$Z_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$	Nenhuma
$Z_t = \alpha + \phi_1 Z_{t-1} + \varepsilon_t$	$ \phi_1  < 1$
$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \varepsilon_t$	$\left\{ \begin{array}{l} \phi_1 + \phi_2 < 1, \quad  \phi_2  < 1 \\ \phi_2 - \phi_1 < 1 \end{array} \right.$
$Z_t = \alpha + \phi_1 Z_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$	

No processo ARIMA (ou ARMA) a condição de invertibilidade significa que sua parte MA pode ser invertida em AR (puro). Os pesos aplicados aos valores passados de  $Z_t$  são decrescentes, na medida em que a defasagem aumenta. Para o conjunto de modelos abaixo, são condições de invertibilidade:

<u>Modelo</u>	<u>Condições de invertibilidade</u>
$Z_t = \alpha + \varepsilon_t - \theta_1\varepsilon_{t-1}$	$ \theta_1  < 1$
$Z_t = \alpha + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$	$\theta_1 + \theta_2 < 1,  \theta_2  < 1$ $\theta_2 - \theta_1 < 1$
$Z_t = \alpha + \phi_1 Z_{t-1} + \varepsilon_t$	Nenhuma
$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \varepsilon_t$	Nenhuma
$Z_t = \alpha + \phi_1 Z_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$	$ \theta_1  < 1$

As condições de invertibilidade e estacionariedade estabelecem restrições aos valores das estimativas dos parâmetros dos modelos. Como será visto, para a estimação desses parâmetros são obtidas inicialmente suas estimativas preliminares. As estimativas preliminares dos parâmetros dos modelos de séries temporais também devem satisfazer às condições de estacionariedade e invertibilidade. Por consequência, a estimativa preliminar do termo constante  $\alpha$  deve ser feita com base nas estimativas dos  $\theta$ 's e  $\phi$ 's que satisfazem as condições de estacionariedade e invertibilidade.

## 8. O Termo Constante

Considere-se a introdução do termo constante  $\alpha$  no modelo genérico:

ARMA (p, q)  $\rightarrow$  ARMA ( $\alpha, \phi$ 's,  $\theta$ 's), ou seja:

$$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$\varepsilon_1, \varepsilon_2, \dots$  são assumidos estatisticamente independentes

$$E(\varepsilon_i) = 0; E(\varepsilon_i^2) = \sigma_\varepsilon^2; \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

Com o operador B, pode-se escrever o modelo na forma multiplicativa (à exceção da inclusão aditiva do termo constante  $\alpha$ ):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = \\ = \alpha + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

A análise do processo autoregressivo puro permite concluir que  $\alpha = \mu \phi_p(B)$

onde

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p)$$

$\mu$  = média populacional de todas as realizações do processo estacionário sob estudo

O termo constante  $\alpha$  pode ser obtido nas três situações a seguir:

1)  $\phi_p(B)$  não incluído no modelo ( $\phi_p(B) = 1$ ) e  $\theta_q(B)$  pode ser incluído, logo:

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

Assim,

$$\phi_p(B) Z_t = \alpha + \theta_q(B) \varepsilon_t$$

sendo  $Z_t = \alpha + \theta_q(B) \varepsilon_t$  ou seja, o modelo de médias móveis,

$$\text{onde } \alpha = \mu \phi_p(B) = \mu$$

2)  $\phi_p(B)$  pode ser incluído e  $\theta_q(B)$  não pode ser incluído ( $\theta_q(B) = 1$ )

Assim,

$$\phi_p(B) Z_t = \alpha + \theta_q(B) \varepsilon_t$$

$\phi_p(B) Z_t = \alpha + \varepsilon_t$  ou seja, o modelo autoregressivo,

onde

$$\alpha = \mu \phi_p(B)$$

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

$$= \mu - \phi_1 B \mu - \phi_2 B^2 \mu - \dots - \phi_p B^p \mu$$

$$= \mu - \phi_1 \mu - \phi_2 \mu - \dots - \phi_p \mu$$

$$= \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$$

3) Ambos  $\phi_p(B)$  e  $\theta_q(B)$  podem ser incluídos

$$\text{Assim, } \phi_p(B)Z_t = \alpha + \theta_q(B) \varepsilon_t$$

$$\alpha = \mu\phi_p(B) = \mu(1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p) = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$$

Para obter-se a estimativa amostral de  $\mu$ , calcula-se a média amostral da série de observações  $Z_b, Z_{b+1}, \dots, Z_n$ , tal que:

$$\bar{Z} = \frac{\sum_{t=b}^n Z_t}{n - b + 1} \Rightarrow \text{estimativa de } \mu$$

Assim, se  $\bar{Z}$  for estatisticamente diferente de zero,

$\Rightarrow \mu \neq 0$  é hipótese razoável.

Como regra prática, obtém-se a estatística t (Students):

$$t = \frac{\bar{Z}}{S_z / (n - b + 1)^{1/2}}, \text{ onde } S_z = \left[ \frac{\sum_{t=b}^n (Z_t - \bar{Z})^2}{(n - b + 1) - 1} \right]^{1/2}, \text{ como a qual realiza-se o}$$

teste de hipóteses.

Quando  $Z_t$  é a série de diferenças em relação à série original  $Y_t$ , o termo constante  $\alpha$ , ao ser incluído no modelo, incorpora-se como medida da tendência determinística existente, sendo que a tendência estocástica é medida pelo modelo ARMA em seus parâmetros  $\phi_p$  e  $\theta_q$ .

Quando a série modelada é a série original  $Y_t$ , no caso da média amostral não ser estatisticamente diferente de zero, considera-se que o termo constante  $\alpha$  não deve ser incluído na formação do modelo e que  $Y_t$  está flutuando em torno de uma média zero.

## 9. Guia para Escolha do Modelo (AR(1), AR(2), MA(1), MA(2), ARMA(1,1))

No estágio de identificação, as especificações funcionais dos modelos são escolhidas com base na avaliação dos padrões dos correlogramas, de forma que:

A) Modelos: MA(1) e MA(2)

O correlograma tem picos nas defasagens 1, 2, ..., q e truncamento após, e o correlograma parcial apresenta padrão de decaimento.

Nesse caso, se:

(1) Se o correlograma tem pico na defasagem 1 e truncamento após e o correlograma parcial decai com uma forma exponencial o modelo tentativo pode ser  $Y_t = \alpha - \theta_1 \varepsilon_{t-1} + \varepsilon_t$ .

(2) Se o correlograma tem pico nas defasagens 1 e 2 e truncamento após e o correlograma parcial decai num misto de padrão exponencial / onda senoidal, o modelo tentativo parece ser  $Y_t = \alpha - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \varepsilon_t$ .

#### B) Modelos: AR(1) e AR(2)

O correlograma apresenta padrão de decaimento e o correlograma parcial tem picos nas defasagens 1, 2, ..., p e truncamento após.

Nesse caso, se:

(1) Se o correlograma parcial tem picos na defasagem 1 e o correlograma apresenta decaimento num padrão exponencial, o modelo tentativo pode ser:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \varepsilon_t.$$

(2) Se o correlograma parcial tem picos nas defasagens 1 e 2 e o correlograma apresenta decaimento num padrão misto exponencial / onda senoidal o modelo tentativo pode ser  $Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ .

#### C) Modelos: MA(q) e AR(p)

Se o correlograma tem picos nas defasagens 1, 2, ..., q e truncamento após e o correlograma parcial tem picos nas defasagens 1, 2, ..., p e truncamento após, devem ser avaliadas as situações:

1. Se o correlograma trunca-se mais abruptamente, o modelo é MA(q).
2. Se o correlograma parcial trunca-se mais abruptamente, o modelo é AR(p).

3. Se ambos os correlogramas parecem truncar-se igualmente abruptamente, os modelos tentativos podem ser:

1) MA(q) e não AR(p)

2) AR(p) e não MA(q)

Como regra prática, em geral MA(q) leva ao melhor modelo e deve ser tentado em primeiro lugar.

D) ARMA (p,q)

O correlograma apresenta padrão de decaimento e o correlograma parcial, também. Nesse caso, o modelo tentativo deve ter uma parte AR(p) e uma parte MA(q).

Deve-se observar o princípio da parcimônia ou seja, por exemplo: se o correlograma apresenta padrão de decaimento exponencial e o correlograma parcial, também, o modelo tentativo deve ser, inicialmente, o mais simples:

$$Y_t = \alpha + \phi_1 Y_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t.$$

E) Nem MA(q), nem AR(p).

O correlograma tem pequenos valores (sem picos) e o correlograma parcial, idem. O modelo tentativo deve ser  $Y_t = \alpha + \varepsilon_t$ .

## 10. A Estimação dos Parâmetros do Modelo

Considere que o modelo escolhido tentativo seja ARMA(1,1), que em forma de desvios em relação à média tem a forma matemática  $X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ .

A estimação dos parâmetros desse modelo deve se feita através de métodos de estimação não-linear. Assim, os passos seguintes são observados:

1º) valores iniciais das estimativas dos parâmetros  $\phi_1$  e  $\theta_1$  (e  $\alpha$ , se for o caso) são definidos.

2º) modificações nesses valores são obtidas, de forma a minimizar a soma dos quadrados dos erros.

Para avaliar a significância estatística dos valores finais dos parâmetros, devem ser construídos intervalos de confiança, realizados testes de hipóteses e verificadas as condições de estacionariedade e invertibilidade (para a unicidade de representação).

Este tópico será retomado em detalhes mais adiante.

## 11. Verificação da Adequação do Modelo

Supondo-se que  $\phi_1 = 0,35$  e  $\theta_1 = 0,20$  sejam os valores finais estimados para os parâmetros do modelo ARMA(1,1), escreve-se:

$$\hat{X}_t = 0,35 X_{t-1} + \varepsilon_t - 0,20 \varepsilon_{t-1}.$$

Com os valores de  $\hat{X}_t$  gerados pelo modelo, obtém-se os resíduos  $e_t = X_t - \hat{X}_t$ . Se ajuste feito do modelo à série temporal de observações for bom, a série de  $e_t$  conforma-se como o processo gerador do ruído branco.

Nesse momento, deve ser feita a análise de autocorrelação dos resíduos, obtendo-se estimativa da estatística  $\chi^2$  para o teste de Box-Pierce (ou Ljung-Box).

Caso a série de  $e_t$  não se comporte como o ruído branco, deve-se escolher novo modelo. A análise do correlograma e do correlograma parcial pode sugerir qual o padrão do processo gerador da série original que deixou de ser modelado e está presente na série de resíduos.

### 11.1 Testes da Precisão do Modelo

Estes testes correspondem ao exame das propriedades estatísticas dos resíduos, pois qualquer dependência entre seus valores ou existência de não-aleatoriedade poderá estar indicando a presença de informação que deveria ser explorada para melhorar a precisão do modelo ao representar o padrão gerador da série temporal.

São eles:

1) Teste do sobre ajustamento ou teste da significância do parâmetro adicional ou da estimativa da variância dos resíduos  $\hat{\sigma}_\varepsilon^2$ .



Obtém-se a estimativa amostral da variância dos resíduos  $\hat{\sigma}_\varepsilon^2$  ajustada para os graus de liberdade. Se  $\hat{\sigma}_\varepsilon^2$  decrescer, significa que a adição do parâmetro adicional melhorou o ajuste.

Para isso, obtém-se a estimativa de  $\sigma_\varepsilon^2$  por  $\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\alpha}, \hat{\theta}, \hat{\phi})}{n - p - q}$ ,

onde  $S(\alpha, \hat{\theta}, \hat{\phi}) \equiv$  soma do quadrado dos resíduos.

Observa-se que, para amostras de tamanho grande, usa-se indistintamente (n-p-q) ou (n) no denominador.

2) Teste da adequação do modelo ou teste da estatística Qui-Quadrado de Box-Pierce modificado por Ljung e Box.

Um grande valor da estatística Q aponta para a inadequação do modelo, sendo:

$$Q = n(n+2) \sum_{k=1}^m (n-k)^{-1} r_k^2,$$

$\chi^2$  crítico  $\Rightarrow$  tabela Qui-Quadrado.

$r_k =$  autocorrelações residuais,  $k = 1, \dots, m$ .

$n =$  número de observações utilizadas para ajustar o modelo (após método das diferenças)

$m = 6, 12, 18, 24$  e  $30$  ou  $m$  é tal que  $(m-p-q) = 20$

$m-p-q \equiv$  graus de liberdade

Para verificar a adequação do modelo, além do teste da estatística Qui-Quadrado, deve-se proceder ao exame da significância estatística dos valores individuais das autocorrelações e autocorrelações parciais nos correlogramas dos resíduos.

Caso o teste tenha indicado bom modelo, é essencial apenas o exame complementar do correlograma das autocorrelações dos resíduos.

### 3) Análise do periodograma

Essa análise permite observar os efeitos não-aleatórios periódicos. Como trata-se de análise no domínio da frequência, o teste correspondente, ou teste de Kolmogorov-Smirnov (Box e Jenkins (1976), “Time-Series Analysis: Forecasting and Control”, Holden-Day, cap. 8.) é visto em capítulo à frente.

4) Comparação dos correlogramas da série de dados originais ou transformados, se for o caso, com os correlogramas da série obtida com o modelo. As respectivas funções de autocorrelação devem ter gráficos similares.

## 12. Uso do Modelo para a Previsão

O uso do modelo ARMA (ou ARIMA) para a previsão será apresentado por meio de um exemplo (Bowerman & O’Connel, 1987).

Seja: Uma empresa fabrica e vende toalhas de papel absorvente. Na Figura 17 é apresentado o gráfico da série de vendas ( $Y_t$ ) de 120 semanas de dados de vendas semanais em unidades de 10.000 rolos. É objetivo da gerência da empresa prever as vendas semanais para planejar e programar a produção, o orçamento e estimar as necessidades de recursos e a capacidade das instalações para a produção e a estocagem do produto.

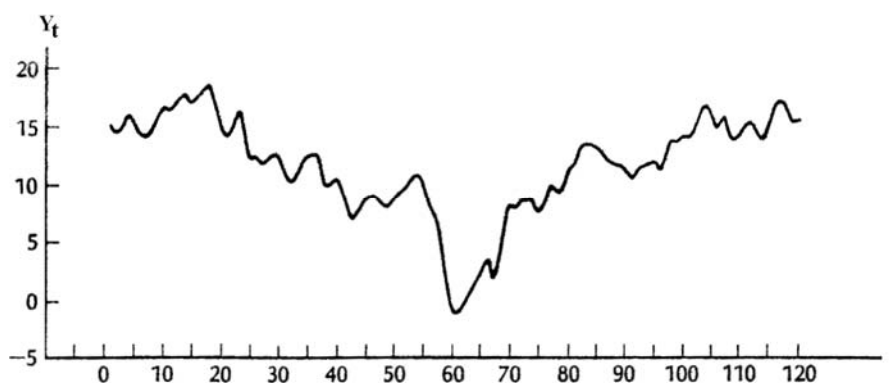


Figura 17 – Gráfico da série original de vendas de toalhas de papel absorvente (unidades de 10.000 rolos)

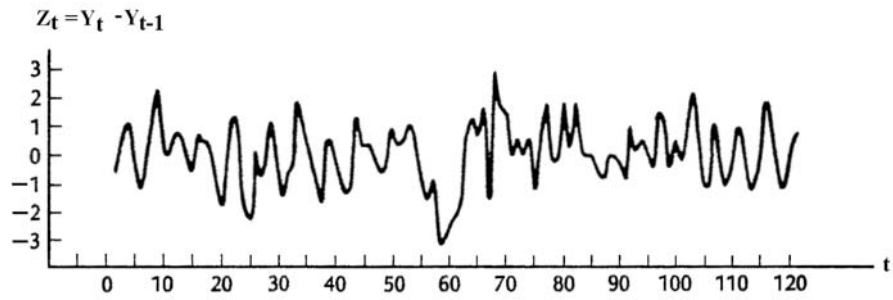


Figura 18 – Gráfico da série de primeiras diferenças

A análise da Figura 17 mostra claramente que a série  $Y_t$  não é estacionária. Dessa forma na Figura 18 é apresentado o gráfico das primeiras diferenças  $Z_t = \Delta Y_t$ . As Figuras 19 e 20, respectivamente, mostram os correlogramas de  $Y_t$  e  $Z_t$ .

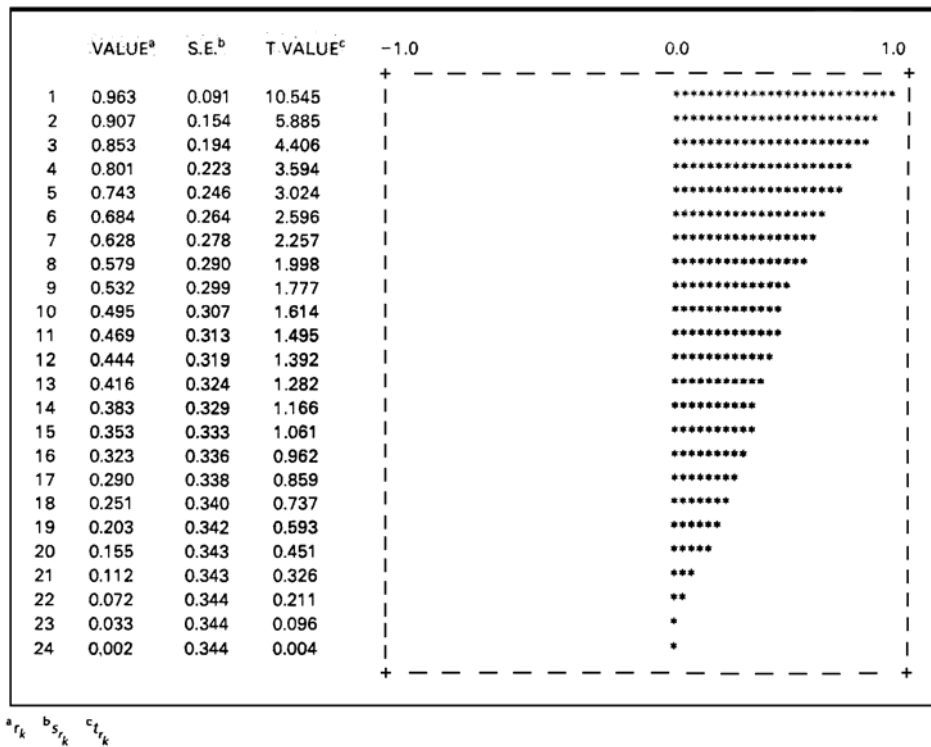


Figura 19 – Correlograma da série original

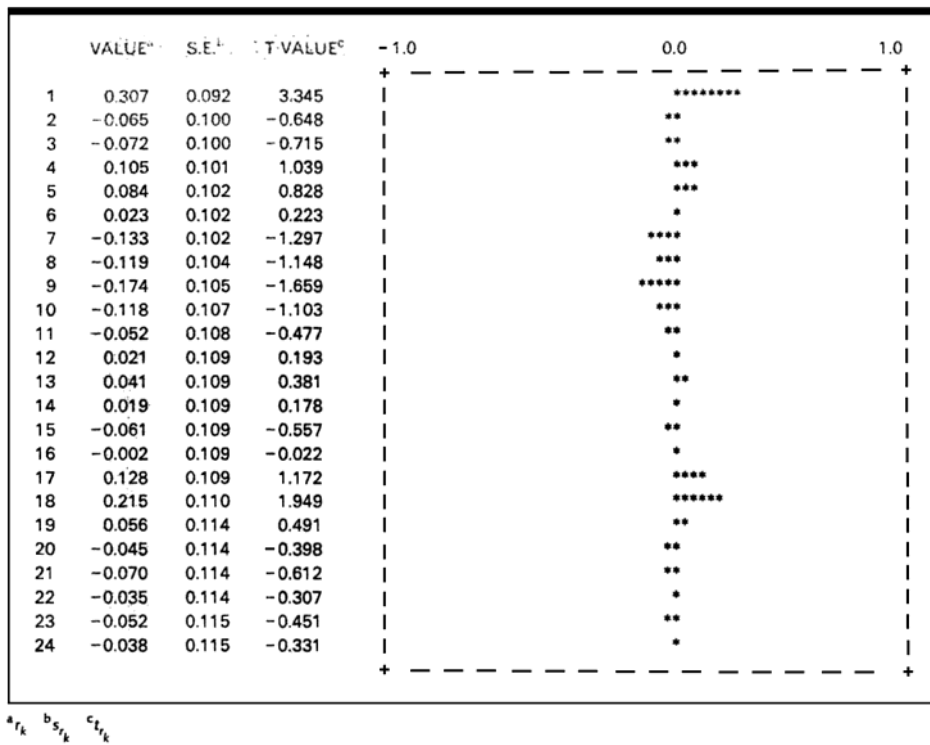


Figura 20 – Correlograma da série de primeiras diferenças

Enquanto na Figura 19 confirma-se a não-estacionariedade nos dados originais, na Figura 20 apenas o primeiro valor de autocorrelação é significativo ou seja:  $|t_{r_1}| > 1,6$  com truncamento após essa defasagem (pode ser observado que os valores sucessivos das autocorrelações são não significativos e próximos de zero com  $|t_{r_k}| < 1,6$  para  $k = 2$  ou  $3$  e  $|t_{r_k}| < 2$ , para  $k > 3$ ).

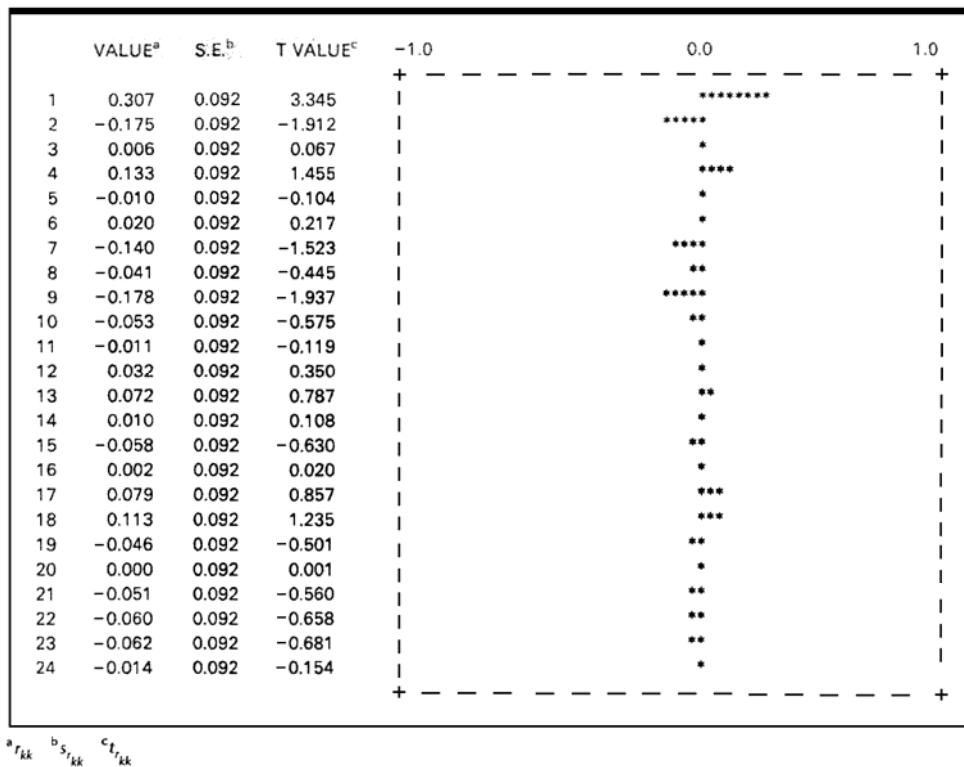


Figura 21 – Correlograma parcial da série de primeiras diferenças

Na Figura 21 é apresentado o correlograma parcial da série de primeiras diferenças. Observa-se que apenas o primeiro valor das autocorrelações parciais é significativo, com  $|t_{r_{11}}| = 3,345 > 2$ . Há um truncamento de valores, pois  $|t_{r_{kk}}| < 2$  para  $k > 1$ , após essa defasagem. Entretanto os valores em módulo da estatística  $t_{r_{kk}}$  para  $k \geq 2$  são 1,912; 1,455; 1,523; ... indicando que o truncamento no correlograma parcial não é tão abrupto quanto no correlograma (nesse, os valores em módulo da estatística  $|t_{r_k}|$  são 0,648; 0,715; 1,039; 0,828; ...).

Dessa forma, o modelo tentativo para representar a série é o modelo de médias móveis  $Z_t = \theta_1(B) \varepsilon_t$ . Nesse exemplo, deve-se avaliar a inclusão ou não do termo constante  $\alpha$ . Ou seja:  $Z_t = \alpha + \theta_1(B) \varepsilon_t$ , se  $\alpha$  for incluído. Observa-se que, no modelo de médias móveis,  $\alpha = \mu$ .

Assim, obtém-se a estimativa amostral  $\bar{Z}$  para  $\mu$ , de forma que:

$$\bar{Z} = \frac{\sum_{t=2}^{120} Z_t}{120 - 2 + 1} = 0,005423$$

$$S_Z = \left[ \frac{\sum_{t=2}^{120} (Z_t - \bar{Z})^2}{(120 - 2 + 1) - 1} \right]^{1/2} = 1,104$$

$$t = \frac{0,005423}{1,104 / (120 - 2 + 1)^{1/2}}, \text{ logo: } |t| = 0,0536 < 2$$

O valor da estatística  $t$  permite concluir que o termo constante  $\alpha$  não deve ser incluído no modelo, o que significa  $Z_t$  está flutuando em torno de uma média zero. Como a série original é  $Y_t$  ( $Z_t$  é a série de diferenças) não há tendência determinística nos valores observados. Qualquer tendência será estocástica.

Assim, o modelo tentativo para representar a série é do tipo MA(1) e não contém termo constante, sendo escrito:  $Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ .

Como  $Z_t = Y_t - Y_{t-1}$ , o modelo é escrito:  $Y_t = Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ .

Considerando que se conhece a estimativa final do parâmetro  $\theta_1$  que obedece às condições de estacionariedade e invertibilidade,  $\hat{\theta}_1 = -0,3545$ , o modelo assume a forma:  $Y_t = Y_{t-1} + \varepsilon_t + 0,3545 \varepsilon_{t-1}$ .

Com esse modelo tentativo pode-se obter as estimativas dos valores das vendas para, então, prever-se vendas futuras. Assim, obtém-se a estimativa  $\hat{Y}_1$  ou seja:

$$\hat{Y}_1 = Y_0 + \hat{\varepsilon}_1 + 0,3545 \hat{\varepsilon}_0.$$

Entretanto,  $Y_0$  não é conhecido (corresponde a período anterior ao que se tem dados), o que mostra necessário, como orientação, que seja observado que:

1. A previsão pontual dos futuros desvios aleatórios é zero. Assim:  $\hat{\varepsilon}_1 = 0$ .
2. A previsão pontual dos desvios aleatórios passados é obtida com informação do passado ou seja, equivale a  $(Y_{t-k} - \hat{Y}_{t-k})$  se  $\hat{Y}_{t-k}$  for conhecido, e 0, se  $\hat{Y}_{t-k}$  não for conhecido (ou se  $\hat{Y}_{t-k}$  não tiver podido ser calculado).

Nesse caso,  $Y_0 - \hat{Y}_0$  não pode ser obtida pois  $\hat{Y}_0$  não pode ser calculado, assim como  $\hat{Y}_1$ , pois  $Y_0$  não é conhecido.

Os valores seguintes são obtidos da mesma forma:  $\hat{Y}_2 = Y_1 + \hat{\varepsilon}_2 + 0,3545\hat{\varepsilon}_1$ ,  $\hat{\varepsilon}_2 = 0$  e  $\hat{\varepsilon}_1 = Y_1 - \hat{Y}_1 = 0$  pois  $\hat{Y}_1$  não foi obtido, logo  $\hat{Y}_2 = 15 + 0 + 0 = 15$ .

Por outro lado,

$\hat{Y}_3 = Y_2 + \hat{\varepsilon}_3 + 0,3545\hat{\varepsilon}_2$ , sendo  $\hat{\varepsilon}_3 = 0$  e  $\hat{\varepsilon}_2 = Y_2 - \hat{Y}_2 = 14,4064 - 15 = -0,5936$ , logo  $\hat{Y}_3 = 14,4064 + 0 + 0,3545(-0,5936) = 14,1960$ .

Sucessivamente, obtém-se as estimativas até  $\hat{Y}_{120} = 14,9553$ , de forma que  $Y_{120} - \hat{Y}_{120} = 15,6453 - 14,9553 = 0,6900$ . A partir desse ponto, pode-se utilizar o modelo para previsão de valores futuros.

Sendo  $\hat{Y}_t = Y_{t-1} + \hat{\varepsilon}_t - \hat{\theta}_1\hat{\varepsilon}_{t-1}$ , pode-se escrever:

$\hat{Y}_{t+\tau}(t) = Y_{t+\tau-1} + \hat{\varepsilon}_{t+\tau} - \hat{\theta}_1\hat{\varepsilon}_{t+\tau-1}$ , sendo  $\hat{Y}_{t-\tau}(t)$  a previsão feita no período t para  $\tau$  períodos à frente de t.

A previsão das vendas de toalhas de papel absorvente para o período 121 é:  $\hat{Y}_{121}(120) = Y_{120} + \hat{\varepsilon}_{121} - (-0,3545)(Y_{120} - \hat{Y}_{120}) = 15,6453 + 0,3545(0,6900) = 15,8899$ .

Utilizando-se um pacote estatístico pode-se obter o desvio padrão do erro de previsão para cada período a partir do limite t do tamanho de amostra. Assim, o intervalo de confiança com 95% de probabilidade de conter o verdadeiro valor  $Y_{121}$  é  $[15,8899 - 2,0362; 17,9261]$  obtido por pacote estatístico computacional (Bowerman e O'Connell, 1987).

Pode-se afirmar que, uma vez conhecida, se a observação das vendas  $Y_{121}$  pertencer ao intervalo, a previsão foi bem sucedida. De estudos anteriores, sabe-se que a metodologia de Box-Jenkins é precisa para o curto-prazo.

O uso do modelo a partir desse ponto permite obter:  $\hat{Y}_{122} = Y_{121} + \hat{\varepsilon}_{122} - \hat{\theta}_1\hat{\varepsilon}_{121}$ , sendo usado no lugar do valor observado  $Y_{121}$  (supondo-se não conhecido) a sua estimativa conhecida ou seja:  $\hat{Y}_{121}(120)$ .

Da mesma forma, o valor  $\hat{Y}_{122}$  corresponde na verdade a  $\hat{Y}_{122}(120) = \hat{Y}_{121}(120) + \hat{\varepsilon}_{122} - \hat{\theta}_1 \hat{\varepsilon}_{121} = 15,8899 + 0 + 0$ , sendo  $\hat{\varepsilon}_{121} = 0$  pois  $Y_{121}$  não foi observado. Assim,  $\hat{Y}_{122}(120) = 15,8899$ .

Observa-se que os sucessivos valores pontuais previstos são iguais ao valor previsto para o período 121, equivalente a  $\tau = 1$ . Isso permite dizer que a memória do modelo MA(1) é de um período.

Da mesma forma à anterior, obtendo-se o desvio padrão do erro de previsão por pacote estatístico, o intervalo de confiança é  $[15,8899 - 3,4278, 19,3177]$ . Observa-se que, quanto maior  $\tau$ , maior é o desvio padrão do erro de previsão e o intervalo da previsão, o que significa mais imprecisa ela é.

De forma geral,  $\hat{Y}_{120+\tau}(120) = Y_{120+\tau-1} + \hat{\varepsilon}_{120+\tau} - \hat{\theta}_1 \hat{\varepsilon}_{120+\tau-1}$ , o que significa que, para  $\tau \geq 2$ ,  $\hat{\varepsilon}_{120+\tau} = \hat{\varepsilon}_{120+\tau-1} = 0$  e  $\hat{Y}_{120+\tau}(120) = \hat{Y}_{120+\tau-1}(120) + \hat{\varepsilon}_{120+\tau} - \hat{\theta}_1 \hat{\varepsilon}_{120+\tau-1} = 15,8899$ .

Na medida em que novos valores forem conhecidos, por exemplo,  $Y_{121} = 16,1099$ , pode-se calcular  $\hat{\varepsilon}_{121} = Y_{121} - \hat{Y}_{121}(120) = 0,2200$ .

Assim,  $\hat{Y}_{122}(121) = Y_{121} + \hat{\varepsilon}_{122} - \hat{\theta}_1 \hat{\varepsilon}_{121} = 16,1879$  e os valores sucessivos são recalculados, como é exemplo  $\hat{Y}_{123}(121) = 16,1879$ .

### 13. Estimação do Modelo

Seja:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t, \text{ onde } w_t = (1-B)^d y_t.$$

A equação pode ser reescrita como:

$$\varepsilon_t = \theta^{-1}(B)\phi(B)(1-B)^d y_t, \text{ ou seja:}$$

$$\varepsilon_t = \theta^{-1}(B)\phi(B)w_t$$

O objetivo da estimação é encontrar um conjunto de parâmetros ( $\theta$  e  $\phi$ ) que minimizam a soma do quadrado dos erros, de forma que:

$$S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_t \varepsilon_t^2$$



A estimativa amostral dos parâmetros dá origem à minimização do somatório do quadrado dos resíduos do modelo:

$$S(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = \sum_t \hat{\varepsilon}_t^2$$

Na presença dos termos de médias móveis, há não-linearidade. Logo, usa-se um procedimento iterativo de estimação não-linear para obter os valores das estimativas dos parâmetros do modelo  $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$  que minimizam  $\sum_t \hat{\varepsilon}_t^2$ .

Uma vez que os parâmetros tenham sido estimados, é preciso testar se os resíduos têm distribuição aleatória de média zero (pelo teste da estatística Q de Box-Pierce, por exemplo).

O procedimento iterativo de estimação não-linear dos parâmetros permite que se teste a significância desses estimadores por meio da estatística t. Além disso, é preciso verificar se esses parâmetros obedecem às condições de estacionariedade e invertibilidade.

Após o exame da aleatoriedade dos resíduos da estimação e da significância dos estimadores, pode ser preciso redefinir o modelo, repetir a estimação e realizar novos exames para teste da adequação do modelo. Uma vez obtida satisfação com o modelo, ele pode ser usado para previsão.

Algumas vezes, mais de um modelo tentativo é considerado adequado após o exame da aleatoriedade dos resíduos e das propriedades estatísticas dos estimadores. Nesse caso, pode-se proceder ao teste da simulação ou seja, à comparação entre os valores gerados pelos modelos e os valores amostrais da observação  $y_t$ , obtendo-se a soma do quadrado dos desvios caso a caso. Aquele modelo que corresponder ao menor somatório deve ser considerado o mais adequado.

Entretanto, sugere-se também reter as especificações alternativas e utilizá-las para a previsão e, com base na avaliação do desempenho das previsões, escolher a especificação que pareça ser mais pertinente com a situação analisada.

### 13.1 A Estimação Não-Linear dos Parâmetros do Modelo

Assume-se que existem  $T + d$  observações disponíveis da série homogênea não-estacionária  $y_t$  de ordem  $d$  ou seja:  $y_{-d+1}, \dots, y_0, y_1, \dots, y_T$ . Após diferenciar a série  $d$  vezes, obtém-se a série estacionária  $w_t$  ( $w_1, \dots, w_T$ ).

Para estimar os parâmetros do modelo ARMA  $(p,d)$  associado à série  $w_t$  baseia-se no pressuposto da distribuição normal e na independência dos termos do erro  $\varepsilon_1, \dots, \varepsilon_T$  ou seja  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ .

Dessa forma, pode-se definir a função de verossimilhança logarítmica condicional associada aos  $(p+q+1)$  parâmetros  $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)$ :

$$L = -T \log \sigma_\varepsilon - \frac{S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)}{2 \sigma_\varepsilon^2}$$

A função é dita condicional pois depende dos valores do passado e de valores não observados  $w_0, w_{-1}, \dots, w_{-p+1}$  e  $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1}$ .

Max  $L$  corresponde à minimização de  $S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ .

Para a estimação dos parâmetros é preciso escolher os valores não-observados passados de  $w_t$  ( $w_0, w_{-1}, \dots$ , etc.). Como solução ao problema, Pindyck & Rubinfeld (1991) propõem que esses valores sejam os valores esperados incondicionais. Assim,  $\varepsilon_0 = \dots = \varepsilon_{-q+1} = 0$  e, se o nível médio da série  $\alpha = 0$ ,  $w_0 = \dots = w_{-p+1} = 0$ .

Dessa forma, o problema é determinar  $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$  que minimizam:

$$S = \sum_{t=1}^T [\varepsilon_t | \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q]^2$$

A estimação não-linear pode ser obtida pelo método iterativo que se baseia na aproximação linear dos dois primeiros termos da expansão da função  $[\varepsilon_t | \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q]$  em série de Taylor em torno de um ponto correspondente ao conjunto de valores iniciais dos parâmetros  $\phi$  e  $\theta$ , cujo vetor pode ser representado por  $\beta_0$ . Representa-se  $w$  o vetor dos valores da variável  $w_t$ ,  $t = 1, \dots, T$ . Assim, a aproximação linear corresponde aos dois primeiros termos de:

$$[\varepsilon_t] = [\varepsilon_t | w, \beta_0] + \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0}) \frac{\partial [\varepsilon_t]}{\partial \beta_i} \Big|_{\beta=\beta_0} +$$

$$+ \frac{1}{2} \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0})^2 \frac{\partial^2 [\varepsilon_t]}{\partial \beta_i^2} \Big|_{\beta=\beta_0} + \dots$$

onde  $\beta_{i,0}$  é o valor da estimativa preliminar do parâmetro  $\beta_i$ .

Fazendo-se  $x_{i,t} = -\frac{\partial [\varepsilon_t]}{\partial \beta_i} \Big|_{\beta=\beta_0}$  e considerando que  $x_{i,t}$  e  $[\varepsilon_t | \mathbf{w}, \boldsymbol{\beta}_0]$  correspondem

a valores resultantes da aplicação das respectivas funções em um determinado ponto (relativos às estimativas iniciais dos parâmetros + os valores de  $w_t$ ), pode-se escrever:

$$[\varepsilon_t | \mathbf{w}, \boldsymbol{\beta}_0] + \sum_{i=1}^{p+q} \beta_{i,0} x_{i,t} = \sum_{i=1}^{p+q} \beta_i x_{i,t} + [\varepsilon_t], \text{ onde será feito } \varepsilon_{t,0} = [\varepsilon_t | \mathbf{w}, \boldsymbol{\beta}_0], \text{ para}$$

simplicificação.

O primeiro lado de igualdade corresponde aos diferentes valores da variável dependente para  $t = 1, \dots, T$  obtidos com um conjunto de estimativas iniciais de parâmetros e valores observados da série estacionária  $w_t$ . O segundo lado da igualdade corresponde à função linear dos parâmetros a estimar multiplicados pelos valores correspondentes das variáveis independentes obtidas em torno das estimativas iniciais dos parâmetros aos quais se soma o termo do erro aleatório.

Novas estimativas dos  $\beta_i$  podem, então, ser obtidas por regressão linear.

À medida que novos valores das estimativas são obtidos, esses são tratados como valores iniciais e o procedimento se repete, com o recálculo dos valores das variáveis independentes e dependentes, em torno de um novo ponto (correspondente a essas novas estimativas).

A regressão linear é realizada até que os valores estimados por ela se aproximem dos valores anteriormente obtidos para os parâmetros (critério de convergência). Nesse ponto, se aplicam os procedimentos para avaliação da significância dos estimadores e da qualidade do ajustamento.

Se for impossível obter a convergência, diferentes estimativas preliminares devem ser definidas e, caso a impossibilidade continue, novas especificações funcionais devem ser tentadas.

Observa-se que, mesmo quando a convergência ocorrer na primeira tentativa de valores iniciais, é aconselhado testar a existência de múltiplas soluções por meio da

variação desse conjunto inicial de estimadores. A estimativa final será aquela que corresponder à menor soma do quadrado dos erros (mínimo global).

### 13.2 Exemplo do Método de Estimação Não-Linear

Como exemplo do primeiro passo do método iterativo de estimação não-linear (Pindyck & Rubinfeld, 1991), seja o modelo ARMA(1,1) escrito como:

$$\varepsilon_t = \frac{1 - \phi_1 B}{1 - \theta_1 B} w_t, \text{ onde } w_t \text{ é a série estacionária.}$$

Esse modelo tem dois parâmetros ( $\phi_1$  e  $\theta_1$ ). Assim:

$$x_{1,t} = -\left. \frac{\partial \varepsilon_t}{\partial \phi_1} \right|_{\phi_{1,0}, \theta_{1,0}} = \frac{B}{1 - \theta_{1,0} B} w_t$$

$$x_{2,t} = -\left. \frac{\partial \varepsilon_t}{\partial \theta_1} \right|_{\phi_{1,0}, \theta_{1,0}} = -\frac{B - \phi_{1,0} B^2}{(1 - \theta_{1,0} B)^2} w_t$$

$x_{1,t}$  e  $x_{2,t}$  correspondem a valores numéricos calculados para  $t = 1, \dots, T$ . Pela expansão dessas expressões:

$$x_{1,t} = \theta_{1,0} x_{1,t-1} + w_{t-1}$$

$$x_{2,t} = 2\theta_{1,0} x_{2,t-1} - \theta_{1,0}^2 x_{2,t-2} - w_{t-1} + \phi_{1,0} w_{t-2}.$$

Esses valores são função dos valores defasados:  $x_{1,t-1}$ ,  $w_{t-1}$ ,  $x_{2,t-1}$ ,  $x_{2,t-2}$  e  $w_{t-2}$ , assim é preciso estimar valores defasados preliminares. Logo, considerando que o nível médio da série original é  $\alpha = 0$ ,

$$x_{2,1} = x_{2,0} = w_0 = w_{-1} = x_{1,0} = 0.$$

Dessa forma, gera-se a série de valores  $x_{2,t}$ , onde

$$x_{2,0} = x_{2,1} = 0 \quad \text{e} \quad x_{2,2} = 2\theta_{1,0} x_{2,1} - \theta_{1,0}^2 x_{2,0} - w_1 + \phi_{1,0} w_0$$

$$\Rightarrow x_{2,2} = -w_1$$

$$x_{2,3} = 2\theta_{1,0} x_{2,2} - \theta_{1,0}^2 x_{2,1} - w_2 + \phi_{1,0} w_1$$

$$\Rightarrow x_{2,3} = -2\theta_{1,0} w_1 - w_2 + \phi_{1,0} w_1, \text{ e assim por diante.}$$

O mesmo procedimento se aplica à série de valores  $x_{1,t}$ .

Uma série de valores deve ser gerada para  $\varepsilon_{t,0} = -\frac{1 - \phi_{1,0}B}{1 - \theta_{1,0}B} w_t$ ,  $t = 1, \dots, T$ ,

sendo:

$$\varepsilon_{t,0} = w_t - \phi_{1,0} w_{t-1} + \theta_{1,0} \varepsilon_{t-1,0}, \text{ onde } \varepsilon_{0,0} = 0.$$

Com essas três series de valores ( $x_{1,t}$ ,  $x_{2,t}$  e  $\varepsilon_{t,0}$ ) pode-se estimar novos parâmetros do seguinte modelo de regressão linear:

$$\varepsilon_{t,0} + \phi_{1,0} x_{1,t} + \theta_{1,0} x_{2,t} = \phi_1 x_{1,t} + \theta_1 x_{2,t} + \varepsilon_t$$

#### 14. Guia para a Identificação de Modelos Sazonais

Quando as séries têm comportamento em que é remarcada a presença de sazonalidade, é preciso a identificação de uma forma particular do modelo de Box-Jenkins de ordem (p, P, q, Q) que descreve a série temporal estacionária  $Z_b, Z_{b+1}, \dots, Z_n$ . Nesse modelo, P e Q dizem respeito às características de médias móveis e autoregressivas do processo de geração das séries associadas ao fator sazonal. Seja o modelo ARMA multiplicativo (na forma compactada) geral (ou seja, sazonal):

$$\phi_p(B) \phi_P(B^L) Z_t = \theta_q(B) \theta_Q(B^L) \varepsilon_t.$$

De forma análoga aos modelos de séries não-sazonais (ou desazonalizadas), deve-se verificar se:

1. O termo constante  $\alpha$  deve ser incluído no modelo;
2. Quais dos operadores  $\phi_p(B)$ ,  $\phi_P(B^L)$ ,  $\theta_q(B)$  e  $\theta_Q(B^L)$  devem ser incluídos no modelo, e
3. A ordem de cada operador que será incluído no modelo.

Considerando que

$$\bar{Z} = \frac{\sum_{t=b}^n Z_t}{n - b + 1} \quad \text{e} \quad S_Z = \left[ \frac{\sum_{t=b}^n (Z_t - \bar{Z})^2}{(n - b + 1) - 1} \right]^{1/2}$$

são a média e o desvio padrão da série temporal  $Z_b, Z_{b+1}, \dots, Z_n$ , o procedimento aproximado para decidir se  $Z$  é estatisticamente diferente de zero, e decidir pela inclusão do termo constante:  $\alpha = \mu \phi_p(B) \phi_p(B^L)$  no modelo geral de Box-Jenkins, é verificar se o valor absoluto do termo:  $\frac{\bar{Z}}{S_z / (n - b + 1)^{1/2}}$  é maior que 2, sendo  $\bar{Z} \equiv$  a média amostral e  $S_z \equiv$  o desvio padrão amostral.

Da mesma forma que para séries não-sazonais, se a série de valores  $Z_b, \dots, Z_n$  é a série original, a média ser zero significa que os valores da série original flutuam em torno de uma média zero. Se, por outro lado,  $Z_b, \dots, Z_n$  representam a série diferenciada, afirmar que a média da população é zero equivale a dizer que a série original não tem tendência determinística. Ao contrário, diz-se que existe essa tendência nos valores da série. Se a série não exibe tendência determinística, qualquer outra tendência será estocástica.

Além disso, para determinar quais dos operadores  $\phi_p(B), \phi_p(B^L), \theta_q(B)$  e  $\theta_q(B^L)$  devem ser incluídos no modelo de Box-Jenkins geral:

(a) É analisado o comportamento do correlograma e do correlograma parcial para determinar se algum operador não sazonal de ordem  $q$ ,  $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ , e se algum operador não sazonal autoregressivo de ordem  $p$ ,  $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ , devem ser utilizados, conforme a Tabela 1.

## TABELA 1

Em geral, é preciso determinar se uma transformação particular do tipo:

$$Z_t = (1 - B^L)^D (1 - B)^d Y_t$$

transformou a série temporal original  $Y_1, Y_2, \dots, Y_n$ , que apresentou variação sazonal, numa série estacionária  $Z_b, Z_{b+1}, \dots, Z_n$ .

Para isso, devem ser examinados os valores do correlograma de  $Z_t$  no nível não sazonal e no nível sazonal.

Define-se (muitas vezes arbitrariamente) o comportamento do correlograma (ou do correlograma parcial) no nível não sazonal como sendo o comportamento da função de autocorrelação nas defasagens 1 até  $(L - 3)$ . Para informação mensal ( $L = 12$ ), esse é o comportamento nas defasagens 1 até 9. Para informação trimestral ( $L = 4$ ), esse é o comportamento na defasagem 1. Entretanto, nesse caso, freqüentemente considera-se também a defasagem 2 e possivelmente a defasagem 3 como tendo valores representativos do nível não sazonal.

Os comportamentos apresentados no correlograma e no correlograma parcial no nível não sazonal são similares àqueles analisados anteriormente. Mais especificamente, define-se as defasagens 1, 2, e possivelmente 3 como sendo as defasagens baixas não sazonais no correlograma e diz-se que um pico existe numa defasagem baixa não sazonal  $k$  no correlograma se:  $|t_{rk}| > 1,6$ .

Além disso, define-se as defasagens 4, 5, ... ,  $L - 3$  como as defasagens não sazonais altas no correlograma, e diz-se que existe um pico numa defasagem alta não sazonal  $k$  no correlograma se:  $|t_{rk}| > 2$ .

Análise similar é feita no correlograma parcial. Diz-se que existe um pico numa defasagem não sazonal  $k$  ( $k = 1, 2, \dots, L - 3$ ) no correlograma parcial se:  $|t_{kk}| > 2$ .

(b) Define-se o comportamento do correlograma (ou do correlograma parcial) no nível sazonal como o comportamento da função de autocorrelação nas defasagens iguais (ou próximas) a  $L, 2L, 3L$  e  $4L$ . Além disso:

b1. Define-se as defasagens  $L$ ,  $2L$ ,  $3L$  e  $4L$  como defasagens sazonais exatas. Um pico existe numa defasagem sazonal exata  $k$  do correlograma se:  $|t_{r_k}| > 1,25$  e um pico existe numa defasagem exata  $k$  do correlograma parcial se:  $|t_{r_{kk}}| > 2$ .

b2. Define-se as defasagens  $L-2$ ,  $L-1$ ,  $L+1$ ,  $L+2$ ,  $2L-2$ ,  $2L-1$ ,  $2L+1$ ,  $2L+2$ ,  $3L-2$ ,  $3L-1$ ,  $3L+1$ ,  $3L+2$ ,  $4L-2$ ,  $4L-1$ ,  $4L+1$  e  $4L+2$  como defasagens sazonais aproximadas. Um pico existe numa defasagem sazonal aproximada  $k$  do correlograma se:  $|t_{r_k}| > 1,6$ . Existe um pico numa defasagem sazonal aproximada  $k$  do correlograma parcial se:  $|t_{r_{kk}}| > 2$ .

O valor da função de autocorrelação no correlograma (ou correlograma parcial) trunca após a defasagem  $k$  no nível sazonal se não houver picos em defasagem sazonais exatas ou aproximadas maiores que essa defasagem  $k$ .

Além disso, o correlograma (ou correlograma parcial) apresenta padrão de decaimento no nível sazonal se a função de autocorrelação não é truncada, mas, ao invés disso, decresce numa forma de padrão de decaimento definida nas defasagens do nível sazonal.

Em geral, pode ser mostrado que, se os valores do correlograma da série  $Z_b, Z_{b+1}, \dots, Z_n$ :

- (1) truncam razoavelmente rápido ou decaem razoavelmente rápido no nível não sazonal, e
- (2) truncam razoavelmente rápido ou decaem razoavelmente rápido no nível sazonal,

então essa série tem valores que podem ser considerados estacionários.

Caso o oposto ocorra, a série é considerada não estacionária e novas transformações devem ser feitas na série de observações.

Quando o correlograma (ou correlograma parcial) trunca razoavelmente rápido no nível não sazonal, isso ocorre frequentemente após a defasagem  $k \leq 2$ . No nível sazonal, isso ocorre frequentemente após a defasagem  $k \leq L + 2$ .



## 15. Estimativas Preliminares dos Parâmetros do Modelo Geral

Os pacotes de “software” computacional de séries temporais começam com estimativas pontuais preliminares dos parâmetros a serem estimados e então aplicam uma técnica de busca interativa para obter as estimativas pontuais de mínimos quadrados dos parâmetros. Alguns pacotes iniciam o procedimento com estimativas geradas internamente mas outros requerem que o usuário as forneça. Essas estimativas devem obedecer às condições de invertibilidade e estacionariedade.

As estimativas pontuais preliminares podem ser obtidas:

1) Estabelecendo-se as estimativas preliminares dos operadores  $\phi_p(B)$ ,  $\phi_p(B^L)$ ,  $\theta_q(B)$ ,  $\theta_q(B^L)$  iguais a 0,1.

$$\begin{aligned} 2) \quad & \text{Obtendo-se a estimativa preliminar de } \alpha = \mu \phi_p(B) \phi_p(B^L) = \\ & = \mu(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \dots - \phi_{p,L} B^{pL}) = \\ & = \mu(1 - \phi_1 - \dots - \phi_p)(1 - \phi_{1,L} - \phi_{2,L} - \dots - \phi_{p,L}) = \\ & \cong \bar{Z}(1 - 0,1 - 0,1 - \dots - 0,1)(1 - 0,1 - 0,1 - \dots - 0,1) \end{aligned}$$

onde  $\bar{Z}$  é a média amostral da série temporal estacionária  $Z_b, Z_{b+1}, \dots, Z_n$ .

Quando se obtém as estimativas finais de mínimos quadrados dos parâmetros do modelo, deve-se verificar se essas estimativas satisfazem às condições de estacionariedade e invertibilidade (Tabela 2). Se não as satisfizerem, isto sugere que o modelo pode não ser adequado.

Tabela 2 - Condições de estacionariedade e de invertibilidade sobre os parâmetros de primeira e segunda ordem dos operadores  $\phi_p(B)$ ,  $\phi_p(B^L)$ ,  $\theta_q(B)$  e  $\theta_q(B^L)$ .

<b>Operador</b>	<b>Condições de estacionariedade</b>	<b>Condições de invertibilidade</b>
$\theta_1(B) = (1 - \theta_1 B)$	Nenhuma	$ \theta_1  < 1$
$\theta_2(B) = (1 - \theta_1 B - \theta_2 B^2)$	Nenhuma	$\theta_1 + \theta_2 < 1$ $\theta_2 - \theta_1 < 1$ $ \theta_2  < 1$
$\theta_1(B^L) = (1 - \theta_{1,L} B^L)$	Nenhuma	$ \theta_{1,L}  < 1$
$\theta_2(B^L) = (1 - \theta_{1,L} B^L - \theta_{2,L} B^{2L})$	Nenhuma	$\theta_{1,L} + \theta_{2,L} < 1$ $\theta_{2,L} - \theta_{1,L} < 1$ $ \theta_{2,L}  < 1$
$\phi_1(B) = (1 - \phi_1 B)$	$ \phi_1  < 1$	Nenhuma
$\phi_2(B) = (1 - \phi_1 B - \phi_2 B^2)$	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $ \phi_2  < 1$	Nenhuma
$\phi_1(B^L) = (1 - \phi_{1,L} B^L)$	$ \phi_{1,L}  < 1$	Nenhuma
$\phi_2(B^L) = (1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L})$	$\phi_{1,L} + \phi_{2,L} < 1$ $\phi_{2,L} - \phi_{1,L} < 1$ $ \phi_{2,L}  < 1$	Nenhuma

## 16. Melhorias no Modelo Sazonal Tentativo

Em geral, se as estatísticas  $Q'$  (ou  $Q$ ) indicam que a adequação de um modelo Box-Jenkins deve ser rejeitada, deve-se investigar o correlograma e o correlograma parcial da série de resíduos para identificar (tentativamente) um modelo descrevendo os resíduos. Esse deve ser combinado ao modelo da série de forma a gerar um novo modelo tentativo.

A Tabela 3 apresenta de forma sucinta os valores críticos da estatística  $t$  para identificar picos no correlograma e no correlograma da série de resíduos.

Tabela 3 – Valores absolutos críticos da estatística t para identificar picos no correlograma da série e no correlograma da série de resíduos

Defasagens	Valores Críticos no Correlograma	Valores Críticos no Correlograma dos Resíduos
<i>Não-sazonais</i>		
• Baixas (1,2 e talvez 3)	1,6	1,25
• Outras defasagens não-sazonais	2,0*	1,6
<i>Sazonais</i>		
• Exatas (L, 2L, 3L, 4L)	1,25	1,25
• Aproximadas (L-2, L-1, L+2, L+1, ...)	1,6	1,25
• Semi-periódicas sazonais (0,5L, 1,5L, 2,5L, 3,5L)	1,6	1,25
<i>Outras</i>	2,0*	1,6

(Bowerman & O'Connell, 1993, pág. 582)

A Tabela 3 apresentou um sumário dos valores críticos da estatística t para identificar picos no correlograma dos resíduos, onde os valores são aqueles das escolhas de Pankratz (1983), excetuando os marcados com (\*). No caso do correlograma parcial dos resíduos, os valores absolutos da estatística t devem ser comparados com o valor crítico 2.

Por outro lado, se  $Q'$  (ou  $Q$ ) indicam que não se deve rejeitar a adequação do modelo, todos os valores individuais da estatística t em módulo devem ser comparados com o valor crítico 2, como uma avaliação complementar ao da estatística  $Q'$  (ou  $Q$ ).

#### 17. Lista de Verificações sobre os Modelos ARIMA (Levenbach e Cleary, 1984, pág. 457-458)

Conforme discutido por Levenbach e Cleary (1984, págs. 457-458), devem ser feitas as seguintes verificações nas etapas de ajustamento de modelos ARIMA à uma série temporal:

1. Verificar se a série é estacionária.
2. Exame da matriz de correlações. Não pode haver um alto grau de correlação entre as estimativas dos parâmetros dos modelos (por exemplo, acima de 0,9).
3. Verificar se todos os termos do modelo foram devidamente incluídos.

4. As estimativas dos parâmetros e seus desvios padrão devem ser examinados. Testes de hipótese devem ser feitos para avaliar a inclusão do parâmetro no modelo.
5. Se um termo autoregressivo regular ou sazonal é incluído, as estimativas dos parâmetros não devem estar próximas de 1,0 (por exemplo, acima de 0,90 ou 0,95). Se uma delas é próxima de 1,0, uma diferenciação regular ou de ordem sazonal deve ser tentada e um teste de previsão feito para determinar qual modelo é melhor: o diferenciado ou o autoregressivo.
6. Verificar se a soma do quadrado dos erros e o desvio padrão dos resíduos diminuem enquanto o ajuste do modelo melhora.
7. A estatística  $\chi^2_{\text{calculada}}$  é menor  $\chi^2_{\text{crítico}}$  para os graus correspondentes de liberdade?
8. Observar se há padrão significativo no correlograma dos resíduos. Rever o correlograma e o correlograma parcial para verificar padrão restante nos resíduos. Atenção maior deve ser dada ao correlograma (padrões de decaimento dos picos das autocorrelações defasadas com início em um pico de valor próximo de 1,0 indicam processo autoregressivo regular. Padrão de decaimento dos picos originando-se num valor mais baixo indica padrão misto ARMA).
9. Uma constante de tendência determinística foi estimada? Sugere-se que a constante seja incluída no final do procedimento. Se a média dos resíduos do modelo final for maior que um desvio padrão em relação a zero, deve ser colocada uma constante no modelo e avaliada a sua significância estatística (Há momentos em que não se deseja colocar uma tendência determinística no modelo de previsão).
10. Tenha atenção para a "super-diferenciação". Uma verificação do correlograma pode ajudar nessa questão. Essa é aparente quando a estimativa do parâmetro regular ou sazonal de médias móveis assume valor próximo de 1,0.
11. Algumas combinações dos parâmetros são improváveis. Se elas estiverem presentes, reavaliar a análise feita. São elas:

- Diferenças sazonais e parâmetros autoregressivos sazonais.
- Parâmetros de médias móveis sazonais e autoregressivos sazonais.
- Parâmetros de médias móveis sazonais em defasagens outras que 4 e 12 (e seus múltiplos) para séries trimestrais e mensais.

## 18. Um Tópico Avançado: Estimação Não Linear

Seja a equação:

$$Y = f(X_1, X_2, \dots, X_k, \beta_1, \beta_2, \dots, \beta_p) + \varepsilon$$

onde:

$f \equiv$  função não linear de  $k$  variáveis independentes e  $p$  parâmetros.

Como critério de estimação dos parâmetros, busca-se a minimização da soma dos desvios ao quadrado. Para isso, sejam as  $T$  observações de  $Y, X_1, X_2, \dots, X_k$ , de forma que:

$$S = \sum_{t=1}^T [Y_t - f(X_{1t}, \dots, X_{kt}, \beta_1, \dots, \beta_p)]^2 \text{ é a soma dos desvios ao quadrado e}$$

$\hat{\beta}_1, \dots, \hat{\beta}_p$  são os estimadores não-lineares de  $\beta_1, \dots, \beta_p$ .

Se  $\varepsilon \sim N(0, \sigma^2) \Rightarrow \hat{\beta}_1, \dots, \hat{\beta}_p$  são os estimadores de máxima verossimilhança ou os estimadores dos  $\beta$ 's que tem a maior probabilidade de ter gerado as observações amostrais dos  $Y$ 's.

A função de verossimilhança é  $L = (2\pi\sigma_\varepsilon^2)^{-T/2} e^{-S/2\sigma_\varepsilon^2}$ , para a qual se quer obter a máxima verossimilhança ou seja, maximizar  $L$ . Deve ser observado que a minimização de  $S$  em relação aos  $\beta$ 's equivale à maximização de  $L$  em relação aos  $\beta$ 's.

Para obter as estimativas dos  $\beta$ 's que minimizam  $S$ , pode-se utilizar, por exemplo, três métodos computacionais de estimação não linear. Nos dois primeiros, a especificação da função  $S$  somada à eficiência computacional define o melhor método:

### 1) Busca Direta

Obtém-se os valores de S associados a diferentes conjuntos de parâmetros, de forma a definir aquele conjunto associado ao S mínimo, que é o conjunto ótimo de parâmetros.

Esse método é válido apenas em casos de um ou dois parâmetros, pelo grande número de combinações de parâmetros para se realizar a busca direta.

### 2) Otimização Direta

Define-se o sistema de equações normais:  $\frac{\partial S}{\partial \beta_i} = 0, i = 1, \dots, p$  ou seja:

$$\sum_{t=1}^T \left[ 2[Y_t - f(X_{1t} \dots X_{kt}, \beta_1, \dots, \beta_p)] \frac{\partial f}{\partial \beta_i} \right] = 0, i = 1, \dots, p,$$

ao qual aplica-se métodos computacionais eficientes.

### 3) Método de Linearização Iterativa

O método desenvolve-se segundo três passos:

- (a) A equação é linearizada através de expansão de séries de Taylor em torno de um conjunto inicial de parâmetros  $(\beta_{1,0} \dots \beta_{p,0})$ . Deve haver uma variação de conjuntos iniciais para observar se o ótimo local ou global foi obtido.
- (b) Aplicação de Mínimos Quadrados Ordinários (MQO) para a geração de parâmetros da equação linearizada que minimizem a soma do quadrado dos desvios.
- (c) Equação re-linearizada em torno do novo conjunto de parâmetros (obtidos em (b)). Volta-se ao passo (b).

Na medida que se tem convergência, há condição de parada na aplicação do método. Esse método possibilita a aplicação de testes estatísticos ( $R^2$ , estatística t, ...), por exemplo, avaliando o ajustamento a cada iteração e ao final, quando se verificou convergência.

Para definir as variáveis independentes e dependentes para a aplicação de MQO, seja a equação para Y na forma de série expandida em torno de um conjunto inicial de parâmetros:

$$Y = f(X_1, \dots, X_k, \beta_{1,0}, \dots, \beta_{p,0}) + \sum_{i=1}^p \left[ \frac{\partial f}{\partial \beta_i} \right]_0 \overbrace{(\beta_i - \beta_{i,0})}^{\Delta_i} + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \left[ \frac{\partial^2 f}{\partial \beta_i \partial \beta_j} \right]_0 (\Delta_i \cdot \Delta_j) + \dots + \varepsilon$$

Nessa série, considera-se a aproximação linear como composta apenas pelos dois primeiros termos após o sinal de igualdade mais o termo do erro aleatório. A equação pode ser reescrita de forma a definir as variáveis independentes e a variável dependente ou seja:

$$\begin{aligned} & \text{Nova variável independente} \\ & \underbrace{Y - f(X_1, \dots, X_k, \beta_{1,0}, \dots, \beta_{p,0}) + \sum_{i=1}^p \beta_{i,0} \left[ \frac{\partial f}{\partial \beta_i} \right]_0}_{\text{p variáveis independentes}} = \\ & = \sum_{i=1}^p \beta_i \underbrace{\left[ \frac{\partial f}{\partial \beta_i} \right]_0}_{\text{p variáveis independentes}} + \varepsilon \end{aligned}$$

Aplicando os Mínimos Quadrados Ordinários (MQO) com base nesse conjunto de variáveis, obtém-se o conjunto de parâmetros  $\beta_{1,1}, \beta_{2,1}, \dots, \beta_{p,1}$  e, em decorrência, são obtidas as novas variáveis dependente e independentes, respectivamente:

$$Y - f(X_1, \dots, X_k, \beta_{1,1}, \dots, \beta_{p,1}) + \sum_{i=1}^p \beta_{i,1} \left[ \frac{\partial f}{\partial \beta_i} \right]_1 \text{ e } \left[ \frac{\partial f}{\partial \beta_i} \right]_1, i = 1, \dots, p.$$

De acordo com o método, aplica-se os MQO na equação linearizada equivalente a esse novo conjunto de variáveis e assim sucessivamente até observar-se o

critério de parada, por exemplo, quando  $\left| \frac{\beta_{i,j+1} - \beta_{i,j}}{\beta_{i,j}} \right| < \delta, i = 1, \dots, p$ , sendo  $\delta$  estabelecido pequeno arbitrariamente.

## 19. Exemplo de Aplicação da Metodologia a Série Temporal Sazonal

A gerência de um grupo de hotéis necessita da previsão de curto-prazo do número de quartos ocupados em seus estabelecimentos, de forma a se programar e contratar pessoal-extra para atender a demanda dos meses de verão, estabelecer contratos de fornecimento de itens que tem um grande tempo de ressurgimento, programar gastos com propaganda etc..

Para isto, foi analisada a série histórica do número de quartos ocupados por dia durante 15 anos, entre 1973 e 1987. Como o planejamento necessário é mensal, o conjunto de dados da série histórica foi reduzido a médias mensais pela divisão do total mensal pelo número de dias do mês. A Figura 22 apresenta o gráfico dessa série temporal.

Foi decidido realizar a análise da série temporal com os dados entre 1973 e 1986, de forma que o ano de 1987 fosse usado para verificar a adequação do modelo.

A Figura 22 revela forte tendência e padrão sazonal. A amplitude do padrão sazonal parece aumentar com o aumento do nível da série.

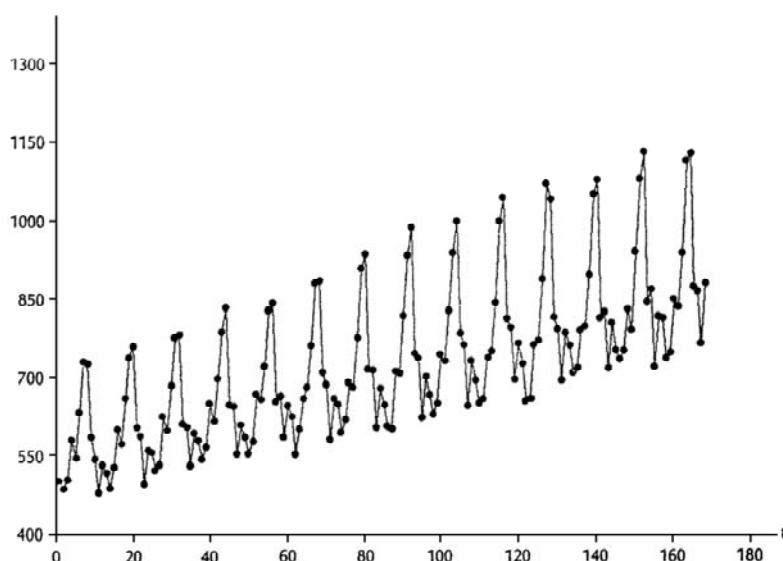


Figura 22 - Gráfico da série de ocupação de quartos do hotel ( $y_t$ )

Dessa forma, é feita uma transformação na base de dados obtendo-se o log (natural) de  $y_t$ , de forma que  $y_t^* = \ln y_t$ , cujo gráfico é mostrado na Figura 23. Observa-se que a amplitude da variação sazonal foi equalizada, mas a série continua não-estacionária em relação à tendência (assim como tem padrão sazonal). A análise dos



correlogramas à Figura 24 (a) e (b) confirma a não-estacionariedade da série transformada.

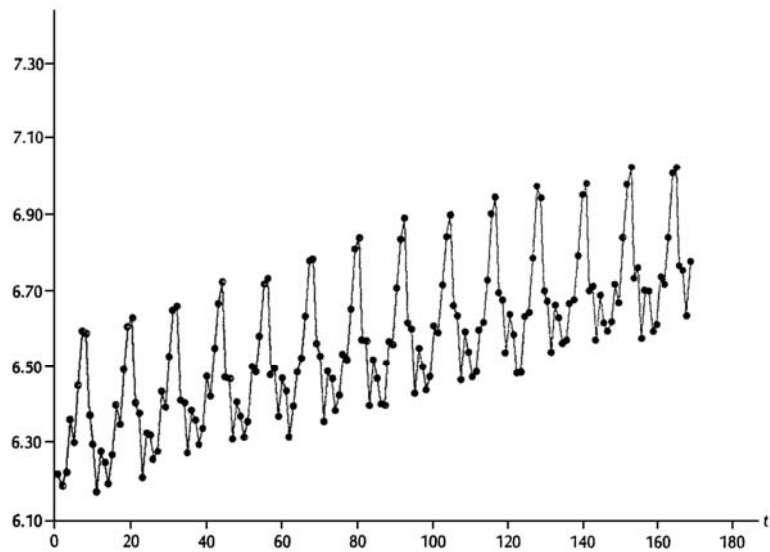


Figura 23 – Gráfico da série transformada  $y_t^* = \ln y_t$

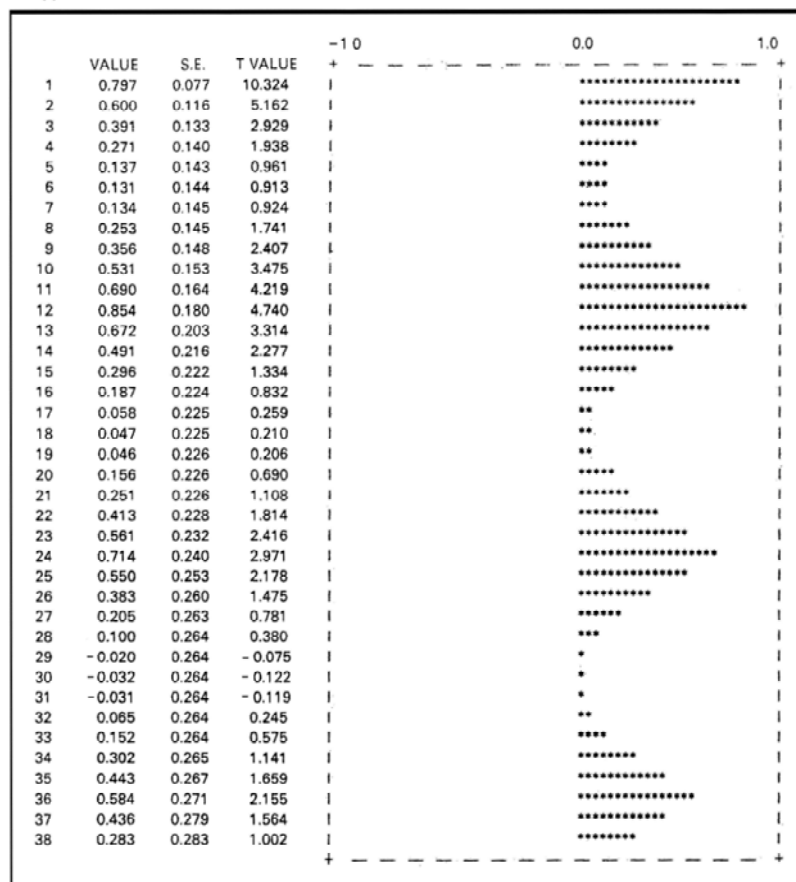


Figura 24 (a) – Correlograma da série transformada  $y_t^* = \ln y_t$

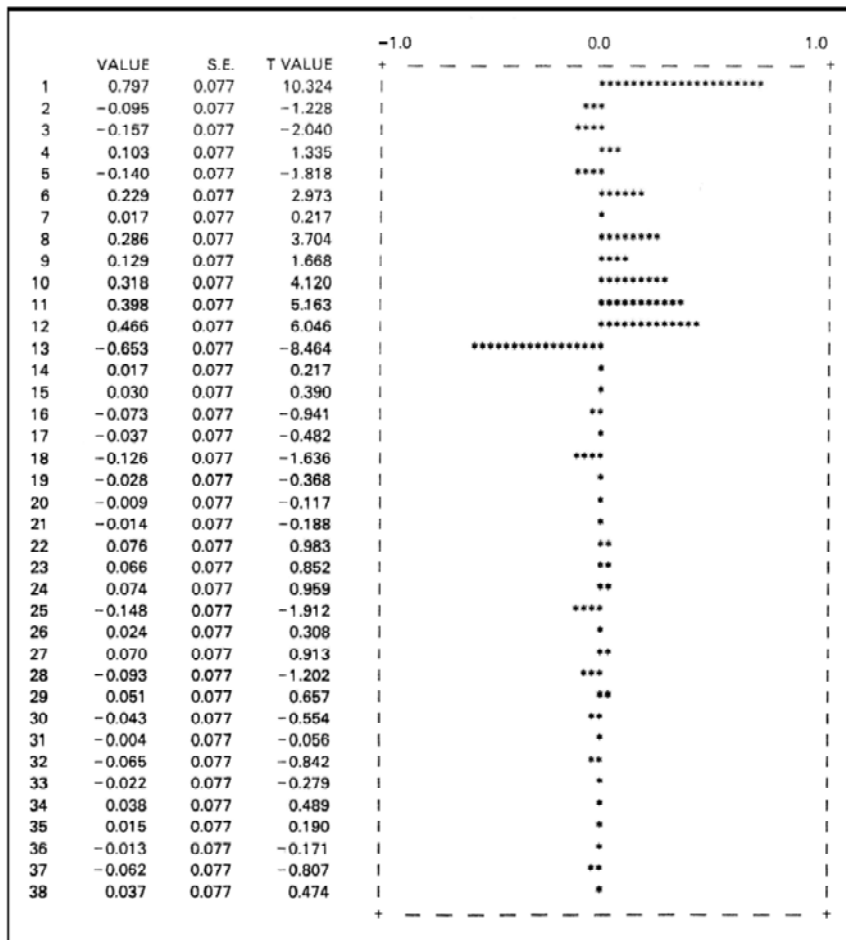


Figura 24 (b) – Correlograma parcial da série transformada  $y_t^* = \ln y_t$

É feita a transformação  $z_t = y_t^* - y_{t-1}^*$ , e os correlogramas simples e parcial relativos à série resultante são mostrados na Figura 25 (a) e (b). No correlograma simples, observa-se que no nível não-sazonal as autocorrelações decaem vagarosamente e que no nível sazonal elas decaem extremamente devagar. Logo  $z_t$  pode ser considerada não-estacionária.

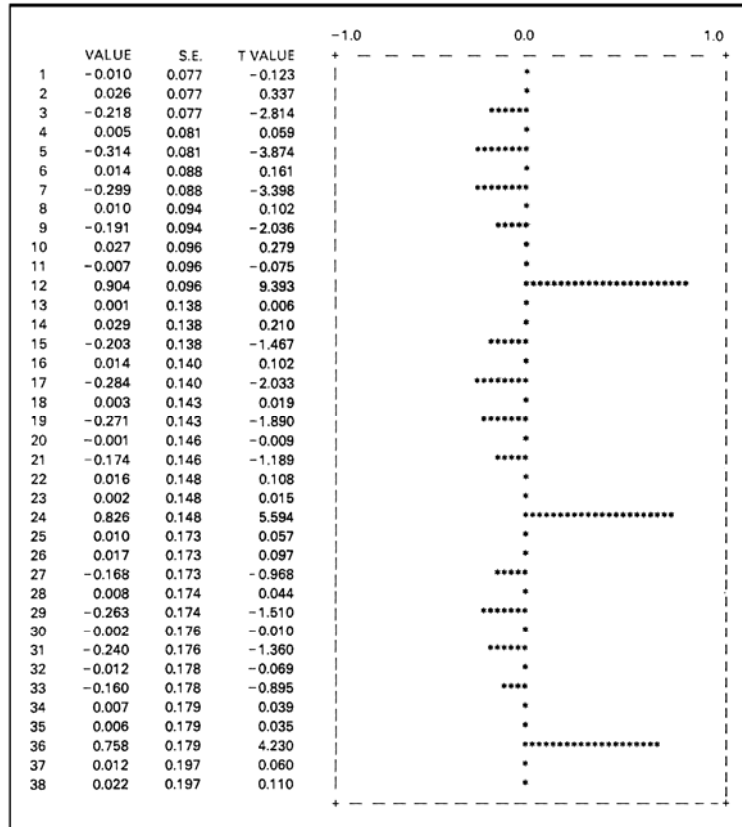


Figura 25 (a) – Correlograma da série transformada  $z_t = y_t^* - y_{t-1}^*$

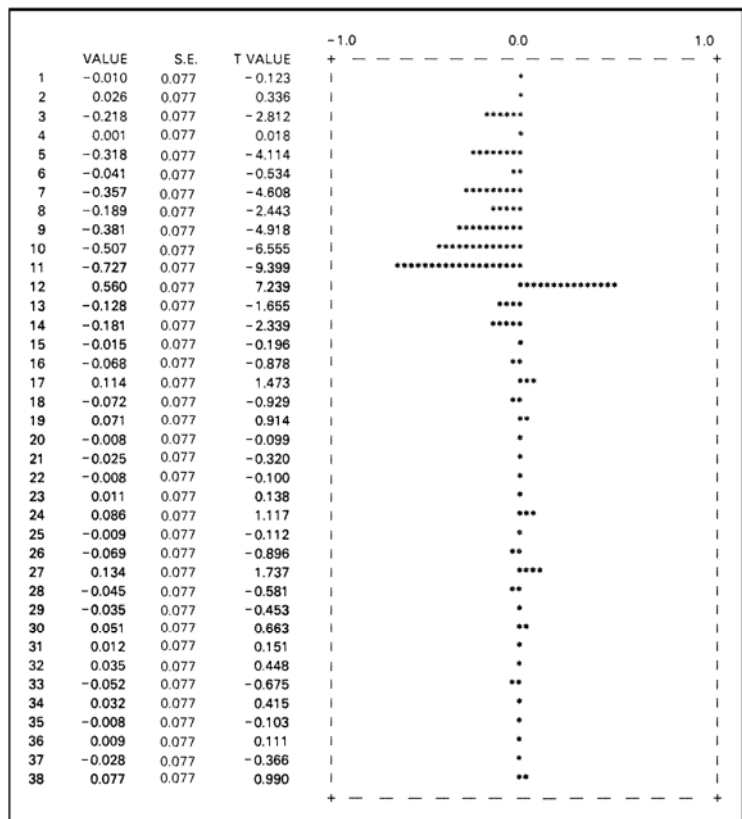


Figura 25 (b) – Correlograma parcial da série transformada  $z_t = y_t^* - y_{t-1}^*$

É feita a transformação  $z_t = y_t^* - y_{t-12}^*$ . A Figura 26 (a) e (b) apresenta os correlogramas relativos à essa transformação. No nível não-sazonal, no correlograma simples, observa-se picos nos períodos 1, 3 e 5, mas que parecem fazer parte de um padrão de decaimento rápido. No correlograma parcial observa-se picos nos períodos 1, 3 e 5 e há indicação de truncamento após o período 5.

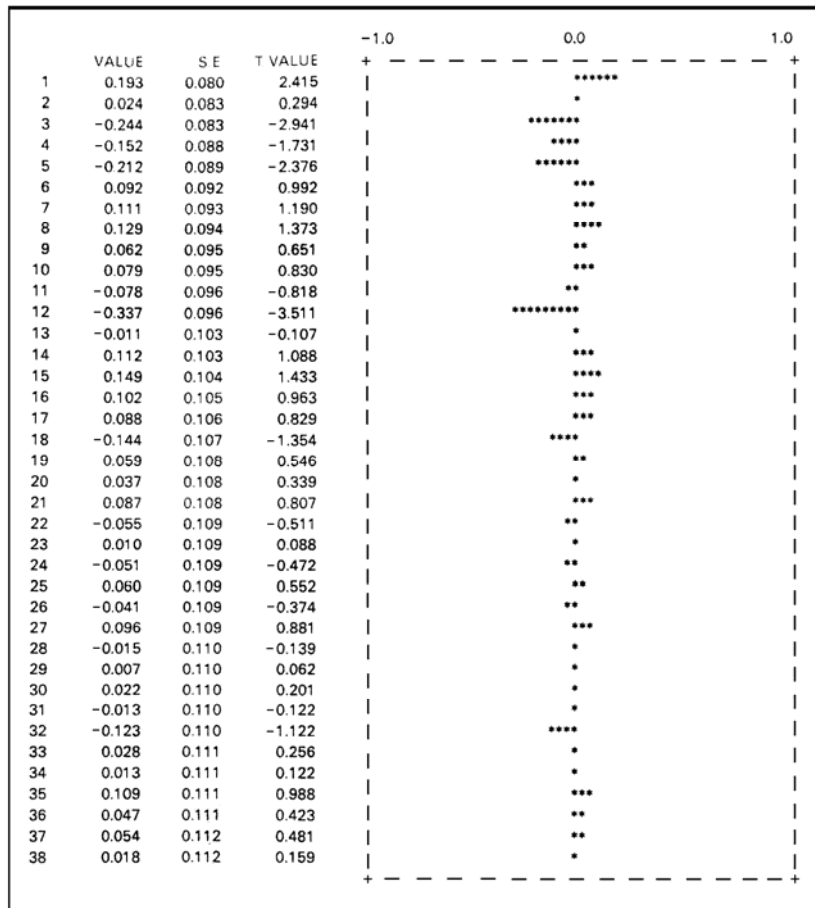


Figura 26 (a) – Correlograma da série transformada  $z_t = y_t^* - y_{t-12}^*$

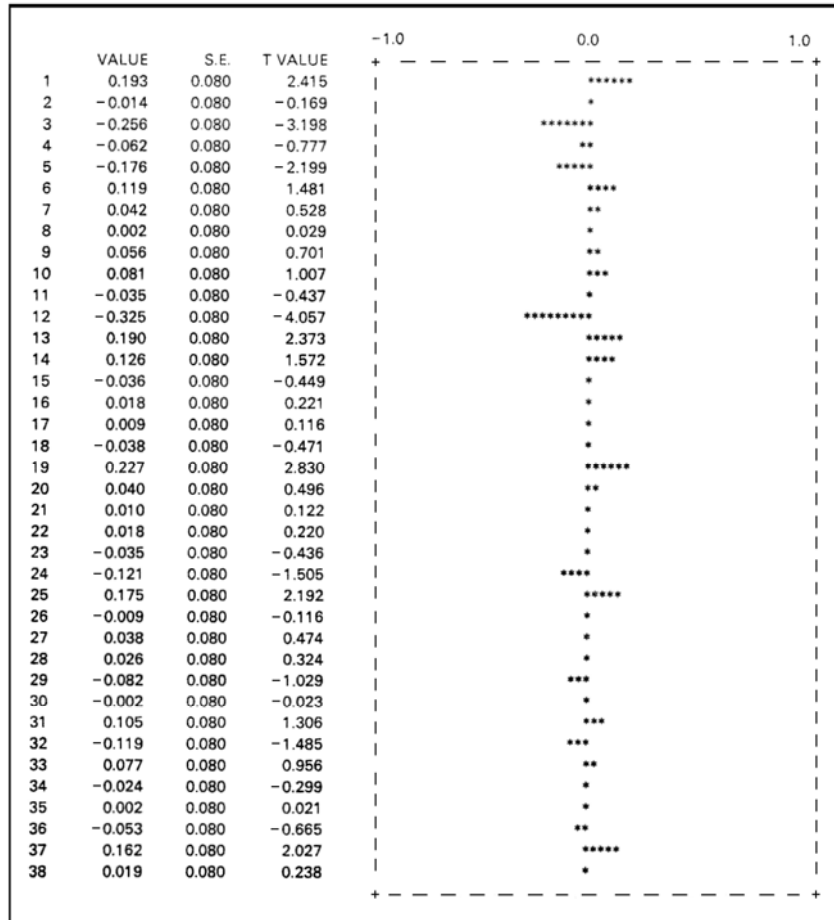


Figura 26 (b) – Correlograma parcial da série transformada  $z_t = y_t^* - y_{t-12}^*$

No nível sazonal, observa-se que o correlograma tem pico e trunca após a defasagem de período 12 ( $|t_{r_{12}}| = 3,511$ ). Por outro lado, no correlograma parcial a função decai razoavelmente rápido, com picos decrescentes nas defasagens sazonais exatas e aproximadas ( $|t_{r_{12,12}}| = 4,057$ ;  $|t_{r_{25,12}}| = 2,192$ ;  $|t_{r_{37,37}}| = 2,027$ ).

A análise do correlograma tanto no nível não-sazonal quanto no sazonal permite concluir que a transformação  $z_t = y_t^* - y_{t-12}^*$  gerou uma série temporal estacionária.

Entretanto, é feita a transformação de primeiras diferenças na série de diferenças de ordem 12, gerando a série relativa a  $z_t = y_t^* - y_{t-1}^* - y_{t-12}^* + y_{t-13}^*$ .

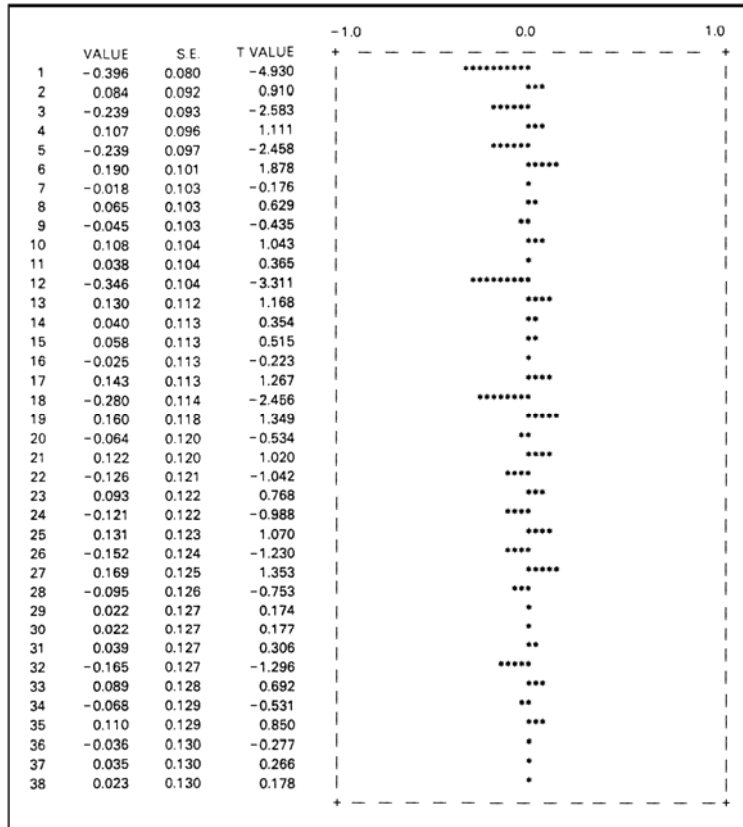


Figura 27 (a) – Correlograma da série  $z_t = y_t^* - y_{t-1}^* - y_{t-2}^* + y_{t-3}^*$

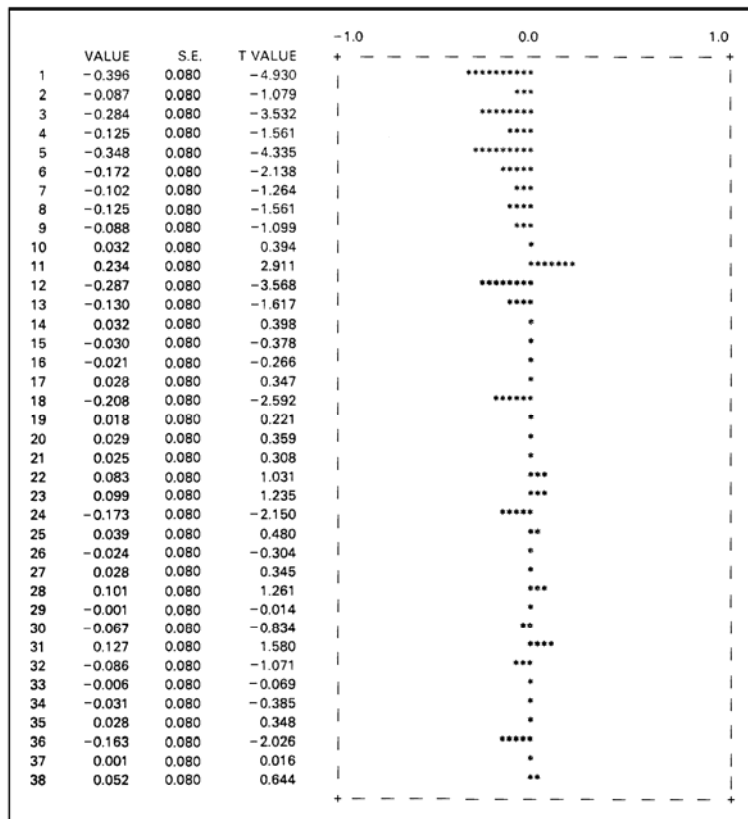


Figura 27 (b) – Correlograma parcial da série  $z_t = y_t^* - y_{t-1}^* - y_{t-2}^* + y_{t-3}^*$

A análise do correlograma (Figura 27(a)) dessa série é feita. Observa-se truncamento no nível sazonal (após defasagem de período 18) e decaimento razoavelmente rápido no nível não-sazonal. Entretanto, comparando com os padrões observados no correlograma da Figura 26(a), conclui-se que obter as diferenças de primeira ordem não acelerou o truncamento ou o decaimento, implicando em dizer que a transformação  $z_t = y_t^* - y_{t-12}^*$  parece o bastante para garantir a estacionariedade e a transformação  $y_t^* - y_{t-1}^* - y_{t-12}^* + y_{t-13}^*$  estaria gerando série “menos” estacionária que  $y_t^* - y_{t-12}^*$ . Dessa forma, é desnecessário avaliar o correlograma parcial da Figura 27(b).

Assim, é preciso avaliar o comportamento das autocorrelações nos correlogramas da Figura 26 (a) e (b) para identificar os operadores de médias móveis e autoregressivos que devem compor o(s) modelo(s) tentativo(s). Como resultado dessa avaliação, no nível não-sazonal, o correlograma tem padrão de decaimento razoavelmente rápido e o correlograma parcial tem picos nas defasagens 1, 3 e 5, com truncamento após. Dessa forma, o operador  $\phi_p(B)$  deve ser incluído no modelo, podendo assumir três formas:

- (a)  $(1 - \phi_1 B - \phi_3 B^3 - \phi_5 B^5)$ , que corresponde a um modelo pouco usual.
- (b)  $(1 - \phi_1 B - \phi_3 B^3)$
- (c)  $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)$

No nível sazonal, o correlograma tem pico na defasagem 12 e truncamento após, enquanto o correlograma parcial apresenta padrão de decaimento rápido. Isto indica o uso do operador  $\theta_Q(B^L)$ . Uma vez que o único pico é na defasagem 12,  $L = 12$  e o operador sazonal é:

$$\theta_1(B^{12}) = (1 - \theta_{1,12} B^{12})$$

Os modelos tentativos são:

- (a)  $(1 - \phi_1 B - \phi_3 B^3 - \phi_5 B^5) z_t = \alpha + (1 - \theta_{1,12} B^{12}) \varepsilon_t$
- (b)  $(1 - \phi_1 B - \phi_3 B^3) z_t = \alpha + (1 - \theta_{1,12} B^{12}) \varepsilon_t$
- (c)  $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) z_t = \alpha + (1 - \theta_{1,12} B^{12}) \varepsilon_t$

Para concluir se a constante  $\alpha$  deve ser incluída no modelo (significando que a série varia em torno de média  $\neq 0$ ), deve-se obter o valor amostral da média e avaliar sua significância estatística.

Assim, lembrando que  $z_t = y_t^* - y_{t-12}^*$ ,

Obtém-se:

$$\bar{z} = \frac{\sum_{t=13}^{168} z_t}{168 - 13 + 1} = 0,0333$$

$$S_z = \left[ \frac{\sum_{t=13}^{168} (z_t - \bar{z})^2}{(168 - 13 + 1) - 1} \right]^{1/2} = 0,02319$$

$$t = \frac{\frac{\bar{z}}{S_z}}{\frac{1}{\sqrt{n - b + 1}}} = \frac{0,033}{0,02319} = 17,7951$$

Logo  $|t| > 2$ , levando à conclusão de que:

$\bar{z}$  é estatisticamente  $\neq 0$  e

$$\alpha = \mu \phi_p(B) \phi_p(B^L).$$

A análise da adequação dos três modelos tentativos indica que a forma  $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) z_t = \alpha + (1 - \theta_{1,12} B^{12}) \varepsilon_t$  é a mais apropriada (Bowerman e O'Connell, 1976). Logo:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) (y_t^* - y_{t-12}^*) = \alpha + (1 - \theta_{1,12} B^{12}) \varepsilon_t$$

$$y_t^* - \phi_1 B y_t^* - \phi_2 B^2 y_t^* - \phi_3 B^3 y_t^* - y_{t-12}^* + \phi_1 B y_{t-12}^* + \phi_2 B^2 y_{t-12}^* + \phi_3 B^3 y_{t-12}^* = \alpha + \varepsilon_t - \theta_{1,12} B^{12} \varepsilon_t$$

A equação pode ser escrita:

$$y_t^* = \alpha + y_{t-12}^* + \phi_1 (y_{t-1}^* - y_{t-13}^*) + \phi_2 (y_{t-2}^* - y_{t-14}^*) + \phi_3 (y_{t-3}^* - y_{t-15}^*) - \theta_{1,12} \varepsilon_{t-12} + \varepsilon_t$$



As estimativas preliminares dos parâmetros do modelo obedecem às condições de invertibilidade e estacionariedade:

$$\phi_1 + \phi_2 + \phi_3 < 1 \text{ (necessária, mas não suficiente)}$$

$$|\theta_{1,12}| < 1$$

São elas:  $\hat{\phi}_1 = 0,1$ ;  $\hat{\phi}_2 = 0,1$ ;  $\hat{\phi}_3 = 0,1$  e  $\hat{\theta}_{1,12} = 0,1$ , e

$$\alpha = 0,33 (1 - 0,1 - 0,1 - 0,1) = 0,0231.$$

As estimativas finais de mínimos quadrados são:

$$\hat{\phi}_1 = 0,2922; \hat{\phi}_2 = 0,1674; \hat{\phi}_3 = -0,2408 \text{ e } \hat{\theta}_{1,12} = 0,5917,$$

que satisfazem às condições de invertibilidade e estacionariedade, permitindo que se obtenha  $\alpha = 0,0258$ .

A estatística t para cada um dos parâmetros estimados é maior do que 2, em módulo, permitindo considerá-los significantes estatisticamente.

A avaliação do coeficiente de correlação entre os parâmetros estimados mostra que o valor absoluto das correlações é inferior a 0,9, o que permite concluir que as estimativas finais dos parâmetros não são altamente correlacionadas. Correlações muito altas sugerem que as estimativas são de má qualidade. Nesse caso, se deverá investigar os correlogramas para encontrar um modelo alternativo que possa ter melhores resultados nas verificações finais da estimação.

Aplicando o teste da estatística Q de Box-Pierce aos resíduos obtidos quando o modelo anterior é ajustado à série temporal e fazendo  $m - np = 20$ , onde  $np \equiv$  número de parâmetros é igual a 4,  $n' = n - (d + LD) = 156$ , sendo d e D o número de diferenciações de não-sazonal e sazonal, respectivamente, obtém-se:

$$Q = n' \sum_{l=1}^m r_l^2(\hat{\epsilon}) = 26,712$$

Observa-se que o valor tabulado de  $\chi^2_{\text{crítico}}$  é:

$$\chi^2_{(0,05)} [(m - np) = 20 \text{ graus de liberdade}] = 31,4104$$

Dessa forma, conclui-se que:

$Q = 26,712 < 31,4104 = \chi^2_{(0,05)}(20 \text{ graus de liberdade})$  e, assim, a adequação do modelo não pode ser rejeitada para um nível de significância de 0,05.

Os valores individuais das estatísticas  $t$  das autocorrelações e autocorrelações parciais da série de resíduos são menores do que dois (em módulo) (excetuando-se o caso de uma autocorrelação dos resíduos, na defasagem 10), logo pode-se concluir que o modelo é adequado.

Antes de prosseguir com a previsão com o modelo estimado, alguns comentários devem ser feitos:

- 1) Recomenda-se identificar picos nos correlogramas da série de resíduos de acordo com o guia para identificação de modelos sazonais (seção ? e Figura ?) apenas quando o teste  $Q$  (ou  $Q'$ ) indicar que a adequação do modelo deve ser rejeitada.
- 2) Se  $Q$  (ou  $Q'$ ) indicarem que a adequação do modelo não deve ser rejeitada, deve-se concentrar em examinar os valores individuais das autocorrelações amostrais dos resíduos cujas estatísticas  $t$  forem maiores que 2 (em módulo).
- 3) Se há indicação de que a adequação do modelo deve ser rejeitada, recomenda-se que somente sejam considerados picos significativos no correlograma parcial dos resíduos aqueles cujos valores da estatística  $t$  em módulo sejam maiores que 2.
- 4) No caso da avaliação de modelos alternativos, aquele com menor desvio-

padrão  $S = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - np}}$  indica o melhor ajuste e deve conduzir a previsões mais precisas dos intervalos de predição.

## 19.1 O Uso do Modelo Tentativo para a Previsão

## Anexo 1

### A Abordagem da Máxima Verossimilhança

Supõe-se que um modelo é:

$Y = f(X_1, \dots, X_k, \beta_1, \dots, \beta_p) + \varepsilon$ , onde  $\varepsilon$  é normalmente distribuído e satisfaz os pressupostos do modelo clássico de regressão linear.

A distribuição de probabilidade de  $Y$  pode ser obtida por:

$$p(Y_i) = f(Y_i, X_i, \beta) = \left[ \frac{1}{2\pi \sigma^2} \right]^{1/2} \exp \left[ \left( \frac{-1}{2\sigma^2} \right) (Y_i - f(\dots))^2 \right],$$

onde:

$$f(\dots) = f(X_{1i}, \dots, X_{ki}, \beta_1, \dots, \beta_p)$$

Os estimadores de máxima verossimilhança de um parâmetro  $\beta$  são os estimadores  $\hat{\beta}$  que maximizam:

$$p(Y_1) p(Y_2) \dots p(Y_N)$$

onde:

$$L = \prod_{i=1}^N p(Y_i) \text{ é referida como a função de verossimilhança,}$$

ou seja:  $\hat{\beta}$  são os estimadores que mais provavelmente geram as observações  $Y_i$ .

Aplicando logaritmo na função de verossimilhança obtém-se (Gujarati, 2006, pág. 93):

$$\begin{aligned} \log L &= \sum_{i=1}^N \log[f(Y_i, X_i, \beta)] = \\ &= -\left(\frac{N}{2}\right) \log 2\pi - \left(\frac{N}{2}\right) \log \sigma^2 - \\ &\quad - \left(\frac{1}{2}\right) \sum_{i=1}^N \frac{(Y_i - f(X_{1i}, \dots, X_{ki}, \beta_1, \dots, \beta_p))^2}{\sigma^2} \end{aligned}$$

Dessa forma, maximizar  $\log L$  corresponde à minimização de:

$$S = \sum_{i=1}^N (Y_i - f(X_{i1}, \dots, X_{ki}, \beta_1, \dots, \beta_p))^2$$

Considerando o modelo a duas variáveis  $Y_i = \alpha + \beta X_i + \varepsilon_i$ , diferencia-se a função  $\log L$  parcialmente em relação a  $\alpha$ ,  $\beta$  e  $\sigma^2$ .

Como resultado, obtém-se:

$$\frac{\partial(\log L)}{\partial \alpha} = \frac{1}{\sigma^2} \sum (Y_i - \alpha - \beta X_i) = 0$$

$$\frac{\partial(\log L)}{\partial \beta} = \frac{1}{\sigma^2} \sum [X_i (Y_i - \alpha - \beta X_i)] = 0$$

$$\frac{\partial(\log L)}{\partial \sigma^2} = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \alpha - \beta X_i)^2 = 0$$

Prosseguindo com a solução desse sistema de equações, são obtidos os estimadores de máxima verossimilhança:

$$\alpha' = \bar{Y} - \beta' \bar{X}$$

$$\beta' = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\sigma^{2'} = \frac{\sum (Y_i - \alpha' - \beta' X_i)^2}{N}$$

Os estimadores de máxima verossimilhança de  $\alpha$  e  $\beta$  são identicamente iguais aos estimadores de mínimos quadrados. Entretanto,  $\sigma^{2'}$  é um estimador viesado de  $\sigma^2$  (embora consistente).

## Referências

- Bowerman e O'Connell (1987), *Times Series Forecasting: Unified Concepts and Computer Implementation*, Duxburg Press.
- Dickey, P.A. e Fuller, W.A. (1979), "Distribution of Estimators for Autoregressive Time Series With a Unit Root", *Journal of the American Statistical Association*, Vol. 74, pp. 427-431.
- Granger e Newbold (1977), *Forecasting Economic Time Series*, Academic Press, NY.
- Gujarati, D. (2006), *Econometria Básica*, Elsevier.
- Levenbach e Cleary (1984), *The Modern Forecaster: The Forecasting Process Through Data Analysis*, Lifetime Learning Publications, Califórnia.
- Pankratz, A. (1983), *Forecasting With Univariate Box-Jenkins Models: Concepts and Cases*, Wiley, N.Y.
- Pindyck, R.S. e Rubinfeld, P.L. (1991), *Econometric Models & Economic Forecasts*, McGraw-Hill International Editions.

VERSÃO EM REVISAO