

## 3

### A Modelagem Univariada por Box & Jenkins

#### 3.1

##### Preliminares da modelagem por Box & Jenkins

Serão a seguir apresentados alguns conceitos imprescindíveis para a compreensão da técnica Box & Jenkins (BJ) selecionados a partir de Box, G.E.P., Jenkins, GM (1994)[4] e Souza & Camargo (2004)[13].

- a) *Processo ruído branco*: ocorre quando a série temporal não depende do seu passado. Nesta situação a série não é autoprojetável e todas as autocorrelações são nulas. Um processo desse tipo representa a aleatoriedade total entre os dados. Logo, para uma série “branca”, não se pode usar um modelo autoexplicativo para descrevê-la, não é possível prevê-la por uma modelagem univariada.
- b) *Objetivo da análise de uma série temporal*: dada uma série temporal que não é “branca”, isto é, que apresenta uma estrutura de dependência serial, deve-se formular o melhor modelo matemático possível levando em conta o que está sendo acusado pelo correlograma.
- c) *Correlograma*: gráfico que mostra a estrutura de dependência entre os dados. Com relação ao que é mostrado pelo correlograma, enfatiza-se que uma autocorrelação baixa não significa independência. O gráfico evidencia apenas a autocorrelação linear. Independência é um conceito mais abrangente.
- d) *Processo estocástico*: toda variável aleatória que guarda uma estrutura de dependência no tempo descreve um processo desta natureza.

- e) *Processo estocástico estacionário*: processo que permanece em equilíbrio estatístico. Em outras palavras, com a sua média e a sua variância constantes.
- f) *Processo estacionário de 2ª ordem*: processo em que a média e a variância são constantes. Além disso, neste tipo de processo, a autocovariância não depende do tempo 't' onde se está na série, mas somente do *lag* de defasagem 'k'. Isso significa que se no primeiro ano de dados a autocorrelação é igual a 0.7, por exemplo, no último ano também será igual a 0.7.
- g) *Série não estacionária homogênea*: séries não estacionária na média ou no nível, mas que, ao passar por sucessivas diferenças, geram um processo estacionário.
- h) *Autocovariância de lag k*: medida de dependência linear entre  $Z_t$  e  $Z_{t+k}$ .

$$\gamma_k = E\{[Z_t - \mu_z][Z_{t+k} - \mu_z]\}. \text{ Quando } k=0; \gamma_0 = \sigma_z^2.$$

É bom enfatizar que a autocovariância é uma função par. Para todo inteiro k:

$$\gamma_k = \gamma_{-k}.$$

Sendo assim,  $\text{cov}[Z_t, Z_{t+k}] = \text{cov}[Z_t, Z_{t-k}]$ . Portanto, pode-se definir  $\gamma_k$  somente para  $k \geq 0$ .

- i) *Autocorrelação de lag k* ( $\rho_k$ ): medida padronizada da dependência linear entre  $Z_t$  e  $Z_{t+k}$ . Graficamente a correlação existente é mostrada pelo correlograma que desenha a FAC (função de autocorrelação total).

$$\rho_k = \frac{E\{[Z_t - \mu_z][Z_{t+k} - \mu_z]\}}{[E(Z_t - \mu_z)^2(Z_{t+k} - \mu_z)^2]^{1/2}} = \frac{\gamma_k}{\gamma_0}; \quad \rho_k = 1, \text{ para } k=0 \\ = 0, \text{ para } k>0.$$

As propriedades da função de autocorrelação para um processo estacionário são:

$$\rho_0 = 1$$

$$|\rho_k| \leq 1$$

$$\rho_k = \rho_{-k}$$

- j) *Autocorrelação parcial de lag k* ( $\phi_{kk}$ ): medida de dependência linear somente entre  $Z_t$  e  $Z_{t+k}$  eliminando a dependência entre os valores intermediários. Esta dependência pode ser vista graficamente no correlograma que descreve a FACP (função de autocorrelação parcial). Pelo método de Durbin exposto por Souza & Camargo (2004) [13]:

$$\hat{\phi}_{k-1,k-1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} * \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j}$$

Tem-se também:

$$\hat{\phi}_{k-1,j} = \hat{\phi}_{kj} - (\hat{\phi}_{k-1,k-1}) * (\hat{\phi}_{k,k-1-j}); \text{ onde } j = 1, \dots, k.$$

- k) *Operador de atraso (B)*: é uma notação muito utilizada na modelagem Box & Jenkins. Ao aplicar o operador de retardo, o expoente da variável é atrasado 'k' períodos. 'B<sup>k</sup>' é relativo a 'k' períodos passados.

$$B^k Z_t = Z_{t-k}; \text{ onde 'B' denota períodos passados.}$$

Exemplo:

$$Z_t = \alpha_1 Z_{t-2} + \alpha_2 Z_{t-4} + \beta_1 a_{t-1} + \beta_2 a_t$$

$$= (\alpha_1 B^2 + \alpha_2 B^4) Z_t + (\beta_1 B + \beta_2) a_t.$$

l) *Modelos AR(p)*: modelo autoregressivo de ordem 'p'. Regressão em função de  $Z_t$ . Por isso, o nome Auto-Regressivo (AR).

Exemplo:

AR (1): representa a estrutura de dependência da informação atual com o período imediatamente anterior.

AR (2): estrutura de dependência vai até dois períodos passados.

m) *Modelos MA (q)*: modelo de ordem 'q' que faz a regressão utilizando o erro de previsão 'q' períodos passados. Fornece uma expressão em função dos resíduos. O nome Médias Móveis se dá porque a série é uma função soma algébrica ponderada dos termos de erro que se movem no tempo. Mas o nome não tem haver com o conceito literal de média.

n) *modelos BJ*: refere-se a encontrar a estrutura ARMA, ARIMA ou SARIMA que descreve a série em estudo.

### 3.2

#### Os modelos Box & Jenkins

O fundamento teórico da técnica de previsão por Box & Jenkins é gerar um processo estacionário de 2ª ordem passando um ruído branco por um filtro linear de memória infinita.

$$a_t \rightarrow \boxed{\Psi(B)} \rightarrow Z_t$$

A idéia é que um modelo com infinitos parâmetros não é implementável dado a impossibilidade de estimar tamanha quantidade de parâmetros. Assim, a

modelagem BJ consiste em encontrar a estrutura de autocorrelação presente na série temporal e a autocorrelação existente entre os termos do erro aleatório. Desta forma, infinitos parâmetros são colocados como a razão de dois finitos parâmetros:

$$Z_t = \psi(B) a_t \rightarrow \Psi(B) = \frac{\theta(B)}{\Phi(B)} \rightarrow \Phi(B) Z_t = \theta(B) a_t$$

Onde:

$\Psi(B)$  : infinitos parâmetros.

$\frac{\theta(B)}{\Phi(B)}$ : finitos parâmetros.

$\Phi(B)$ : parâmetro da parte AR da série. Refere-se à modelagem autoregressiva da série.

$\theta(B)$  : parâmetro da parte MA do modelo. Refere-se à modelagem da autocorrelação entre os termos do erro do modelo.

Como o ideal é estimar modelos com economia de parâmetros, quando ocorrer da ordem 'p' aumentar muito, a quantidade de parâmetros deve ser reduzida através de defasagens do ruído branco. Defasar o ruído e não mais a série é equivalente.

Portanto, modelar uma série pelo método BJ consiste em definir a estrutura ARMA mais adequada para descrever os dados em estudo. A condição essencial para a aplicação dessa técnica é que a série seja estacionária de 2ª ordem. Caso contrário, do conjunto original de dados, cria-se uma nova série realizando sucessivas diferenças.

Exemplo:

$X_1, X_2, X_3, X_4, \dots, X_{t-1}$  : série original por suposição não estacionária de 2ª ordem.

Por diferenças:

$$Z_1 = X_2 - X_1 ;$$

$$Z_2 = X_3 - X_2 ;$$

$$Z_3 = X_4 - X_3 \dots\dots\dots$$

Onde:

$Z_1, Z_2, Z_3, \dots, Z_{t-1}$  : série estacionária de 2ª ordem.

Essas diferenças representam o que se chama de derivada para uma variável discreta. Se a série tem um crescimento linear, com a 1ª diferença ficará horizontal. Se o crescimento é quadrático, será preciso fazer a 2ª diferença. Matematicamente essas diferenciações são representadas pelo operador diferença  $\nabla^d$  (nabla). O índice sobrescrito 'd' diz respeito ao número de diferenças que horizontalizam a série tornando-a estacionária de 2ª ordem.

$$Z_t = \nabla^d X_t.$$

Onde:

$X_t$ : série original antes de aplicar as diferenças.

$\nabla^d$ : operador diferença nabla.

$Z_t$ : série estacionária após aplicar as diferenças.

O número de diferenciações aplicadas à série para torná-la estacionária é representado na estrutura de modelagem pela letra 'I'.

$$\text{Para } d = 1 \rightarrow \nabla^1 = X_t - X_{t-1}.$$

$$\text{Para } d = 2 \rightarrow \nabla^2 = X_t - (2 X_{t-1}) + X_{t-2}.$$

Na prática executam-se quantas diferenças forem necessárias para chegar a uma série estacionária. Em geral, quando for preciso aplicar 3 ou 4 diferenças, a série não é adequada para ser modelada por Box & Jenkins.

Portanto, em resumo:

- ARMA (p, q): utilizado para modelar séries originalmente estacionárias de 2ª ordem.

Modelo:  $[\Phi(B) X_t = \theta(B) a_t]$

- ARIMA (p, d, q): utilizado para modelar séries originalmente não estacionárias pela aplicação de sucessivas diferenciações.

Modelo:  $[\Phi(B) \nabla^d X^d = \theta(B) a_t]$

Se mesmo após as diferenciações a série não se estacionar, deve-se fazer algum tipo de transformação nos dados, como a logarítmica por exemplo. Porém, de acordo com Esteves (2003) [7], isso pode gerar estimativas tendenciosas que necessitam de algum fator de correção.

Quando for preciso incluir a componente de sazonalidade no modelo, isso pode ser feito recorrendo à modelagem SARIMA. Esta estrutura, além de modelar a correlação convencional das partes AR e MA do modelo, também observa a correlação presente entre os períodos sazonais.

Portanto, para descrever a sazonalidade existem estimadores específicos para esta finalidade com as suas respectivas partes AR (P) e MA (Q) sazonal. As letras maiúsculas 'P' e 'Q' simbolizam a ordem sazonal do modelo. Esta ordem é identificada observando exclusivamente os *lags* sazonais.

Notação para representar o modelo com um ciclo sazonal:

$$\begin{aligned} \text{SARIMA } (p, d, q) * (P, D, Q)_s \\ = (\varphi, \nabla, \theta) * (\Phi, \nabla_s, \Theta) \end{aligned}$$

Estrutura SARIMA:

$$\varphi(B) \Phi(B^S) \nabla^d \nabla^D X_t = \theta(B) \Theta(B^S) a_t$$

As ordens  $p$  (AR) e  $q$  (MA) podem ser determinadas respectivamente nos gráficos desenhados pela FACP (função de autocorrelação parcial) e pela FAC (função de autocorrelação total). Para tanto, é preciso estar atento a forma de decréscimo da série e em qual *lag* acontece o corte brusco. Este corte se dá no *lag* cuja correlação é mais significativa determinando a ordem do modelo.

O padrão teórico de um AR é um decréscimo exponencial na FAC e um corte brusco na FACP no *lag* correspondente a ordem do modelo. O padrão MA é um corte brusco na FAC e um decréscimo exponencial na FACP. No modelo ARMA verifica-se a presença de um corte brusco tanto na FAC quanto na FACP.

Com respeito a identificação da ordem sazonal do modelo, o conceito descrito no parágrafo anterior permanece o mesmo. Contudo, neste caso, apenas os *lags* sazonais devem ser avaliados. Um detalhamento da estrutura SARIMA é apresentado:

$$\text{Modelo : } \varphi(B) \Phi(B^S) \nabla^d \nabla^D X_t = \theta(B) \Theta(B^S) a_t$$

Onde:

$$\varphi(B) : 1 - \varphi_1 B^1 - \varphi_2 B^2 - \dots - \varphi_p B^p$$

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

$$\Phi(B^S) : 1 - \Phi_1(B^S) - \Phi_2(B^{2S}) - \dots - \Phi_p(B^{pS})$$

$$\Theta(B^S) : 1 - \Theta_1(B^S) - \Theta_2(B^{2S}) - \dots - \Theta_Q(B^{QS})$$

$$\nabla^d : (1 - B)^d$$

$$\nabla_s^D : (1 - B)^D$$

Após identificar a ordem do modelo BJ, é necessário estimar os parâmetros, ou seja, os valores assumidos pelas quantidades desconhecidas:



$$\xi = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_p, \Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_p, \sigma_a^2).$$

A estimação se dá pelo Método de Máxima Verossimilhança que é melhor descrito na próxima seção. O objetivo é minimizar a soma dos erros ao quadrado.

$$S[\xi = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_p, \Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_p, \sigma_a^2)].$$

$$S(\xi) = \sum a_t^2$$

Sabendo que:

$$\Phi(B) \nabla^d Z_t = \theta(B) a_t$$

Então:

$$a_t = \theta(B)^{-1} \Phi(B) \nabla^d Z_t$$

Onde:

$$a_t = Z_t - \hat{Z}_t(\tau)$$

As duas questões envolvidas na estimação pelo Método de Máxima Verossimilhança são:

- 1) determinar os valores iniciais.
- 2) minimizar a soma do quadrado dos resíduos.

### 3.3

#### Estimação paramétrica dos modelos BJ

Existem softwares estatísticos como FPW (Forecast for Pro Windows) que é capaz de estimar automaticamente os valores ótimos para os parâmetros do modelo Box & Jenkins identificado. Entretanto, este programa apenas pode incorporar um único ciclo sazonal. No FPW ainda só é possível modelar séries com até 500 observações. Para o presente estudo, este software mostra-se

inadequado para realizar a modelagem pretendida considerando dois ciclos: diário e semanal.

Desta maneira, a ordem do modelo foi estimada no excel e a otimização dos parâmetros bem como as previsões foram realizadas no MATLAB em suas versões 6.1 e 6.5. Nesta situação, a escolha dos valores otimizados dos parâmetros do modelo não ocorre de maneira automática sendo necessário adotar um procedimento para a estimação.

### 3.3.1

#### **Metodologia para estimar os parâmetros dos modelos Box & Jenkins**

Identificada a ordem do modelo, é preciso obter as estimativas dos seus parâmetros. De acordo com Souza & Camargo (2004) [13], na parte AR, os parâmetros são lineares e um método simples como o de Mínimos Quadrados Ordinários resolveria facilmente o problema. Porém a parte MA do modelo não é linear e exige métodos mais complexos para efetuar a estimação de forma satisfatória. Sendo assim, uma técnica empregada para executar essa tarefa pode ser o Método de Máxima Verossimilhança que produz estimadores com importantes propriedades que os tornam atraentes para gerar as estimativas. O fundamento teórico desta metodologia é que toda a informação populacional absorvida pelos dados encontra-se na função de verossimilhança, desde que o modelo tenha sido identificado corretamente.

A função de verossimilhança é a função de densidade conjunta (fdp) de todas as observações da série temporal e o seu estimador é aquele que maximiza a probabilidade das quantidades observadas, para média, variância, entre outras, na amostra selecionada. Ou seja, o conjunto de parâmetros que se mostra mais consistente com a amostra.

$$L(\theta/\underline{x}) = f(\underline{x}; \theta)$$

Ressalta-se que na fdp o argumento são os dados amostrais, enquanto na função de verossimilhança, o argumento são os parâmetros. (Spanos (1999) [14]).

### 3.3.2

#### Propriedades dos Estimadores de Máxima Verossimilhança (EMV)

a) *pequenas amostras:*

- 1) Invariância: propriedade muito útil quando o espaço paramétrico tem restrições. Pela propriedade, a inversa é única e tem que existir.

Exemplo:

$$aa \ (n) \quad X \sim N(\theta, 1)$$

$$\hat{\theta} = \bar{X} \quad (\text{estimador de máxima verossimilhança de } \theta)$$

Então, pela propriedade da invariância:

$$n = e^{\theta} \rightarrow n = e^{\hat{\theta}} = e^{\hat{x}}$$

A propriedade da invariância permite dizer que se um parâmetro está representado por um estimador, então, qualquer transformação feita na função que contenha o parâmetro, como a apresentada no exemplo anterior, permitirá obter a estimativa do parâmetro também por sua função transformada. (Bolfarini (2001) [3]).

b) *grandes amostras*

No caso de grandes amostras, os EMV apresentam as seguintes propriedades assintóticas:

- 1) consistentes: o estimador converge em probabilidade para o verdadeiro parâmetro populacional.

1.1) *Condições para um estimador ser consistente:*

- $E(\hat{\theta}) = \theta$  (não tendenciosidade)
  - $\lim_{N \rightarrow \infty} \hat{\sigma}^2 = 0$
- 2) eficientes: são os de variância mínima. Isto é, se dois estimadores forem não tendenciosos e  $[\hat{\sigma}^2(\hat{\theta}_1) < \hat{\sigma}^2(\hat{\theta}_2)]$ , então,  $\hat{\theta}_1$  é mais eficiente que  $\hat{\theta}_2$ . A variância mínima é provada pelo limite de Cramer Rao. Recomenda-se consultar Greene (1993) [8] para informações adicionais a respeito do limite de Cramer Rao.
- 3) normalidade: quando ‘n’ aumenta, a distribuição tende para uma normal o que facilita os cálculos para a média e variância.

Pela suposição  $\epsilon_t \sim N(0, \sigma^2)$  chega-se a função de verossimilhança para  $\xi$ . Isto é,  $L(\xi)$ .

Sob a hipótese de normalidade do ruído, as estimativas do Método de Máxima Verossimilhança são assintoticamente equivalentes às estimativas obtidas por Mínimos Quadrados. Nos modelos ARIMA (p, d, q), a maximização de  $L(\xi)$  se equipara a minimizar a soma do quadrado dos resíduos.

A maximização da verossimilhança permite que os estimadores  $\hat{\phi}$  e  $\hat{\theta}$  estejam associados aos seus respectivos erros padrão possibilitando construir intervalos de confiança para os mesmos. Quanto aos estimadores de Mínimos Quadrados, devem ser tal que para o modelo:

$$\text{Modelo: } \hat{\phi}(B)Z_t = \hat{\theta}(B)\epsilon_t$$

Onde:

$$Z_t = \nabla^d X_t$$

O somatório,  $S(\xi) = \sum (at)^2$ , seja mínimo:

Em síntese, para um modelo ARIMA (p, d, q) com (p + q + 1) parâmetros:

Tem-se:

$$\xi = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q, \sigma_a^2).$$

Por Máxima Verossimilhança:  $L(\xi/x_1, \dots, x_n)$  e assumindo a normalidade do ruído:

$$at \sim N(0, \sigma_a^2)$$

Pode-se dizer:  $EMV \cong EMQ$ .

Para estimar os valores do erro desde o instante  $t = 1$ , é preciso ter as informações relativas ao passado da série e aos erros passados. Estes valores não são observáveis, ou seja,  $X^* = (X_0, X_{t-1}, X_{t-2}, \dots)$  e  $a^* = (a_0, a_{t-1}, a_{t-2}, \dots)$ , não estão disponíveis no banco de dados. Isso torna o procedimento de estimação complexo. Assim, salienta-se que estimar estes valores iniciais antecede ao processo de estimação propriamente dito.

Existem 2 procedimentos para calcular os  $a^*$  e  $w^*$  anteriores ao instante  $t = 1$  que são descritos em Souza & Camargo (2004) [13].

- a) Método condicional: valores iniciais desconhecidos são fornecidos pelo usuário.
- b) Método incondicional: sugerido por Box & Jenkins. Os valores iniciais são estimados a partir da própria série por “*Backforecasting*”.

### 3.3.3

#### Método condicional

Pela técnica condicional os valores atribuídos aos dados não observáveis partem de suposições teóricas. Segundo Souza & Camargo (2004) [13], para uma série estacionária, a estimação dos parâmetros  $\hat{\phi}$  (AR) e  $\hat{\theta}$  (MA) do modelo ARIMA (p, d, q) equivale ao problema de estimar  $\hat{\phi}$  e  $\hat{\theta}$  do modelo ARMA (p,q).

Supondo  $Z_t = \nabla^d X_t$  (série estacionária)

Tem-se:

$$a_t = Z_t - \Phi_1 Z_{t-1} - \dots - \Phi_p Z_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

Onde:

$$E(Z_t) = \mu.$$

Considerando 'at' como um processo ruído branco, com  $a_t \sim N(0, \sigma_a^2)$ , para qualquer que seja 'i' em  $a_i$ , a fdp conjunta do ruído pode ser escrita como uma gaussiana mostrada abaixo:

$$P(\underline{a}) = 1 / (2\pi \sigma_a^2)^{n/2} * \exp \left\{ -\sum_{t=1}^n a_t^2 / 2 \sigma_a^2 \right\}$$

Dessa forma, em síntese, a função de verossimilhança para uma série  $Z_t$  estacionária está condicionada aos vetores de dados passados e não observáveis de  $Z_*$  e  $a_*$ :

$Z_* = (Z_0, Z_{-1}, \dots, Z_{-p+1})$ : valores não observáveis da série temporal.

$a_* = (a_0, a_{-1}, \dots, a_{-q+1})$ : valores não observáveis da parte MA do modelo.

O logaritmo da função de verossimilhança condicional é dado por:

$$L(\underline{\xi} / \underline{\Phi}, \underline{\theta}, \sigma a^2) = -n \log \sigma a - S(\underline{\Phi}, \underline{\theta}) / 2 \sigma a^2$$

Pelo método condicional, uma das formas de obter os valores iniciais não observáveis para  $Z_*$  e  $a_*$  consiste em assumir:

$$E\{a_t\} = 0$$

$$E\{Z_t\} = \bar{Z}$$

Pode-se ainda estabelecer os valores esperados condicionais através de um processo iterativo fazendo  $(Z_*, a_*)$  iguais a zero. Então estima-se o modelo ARMA minimizando  $S^*(\underline{\Phi}, \underline{\theta})$  condicional a estes valores iguais a zero e, dessa maneira, obtendo as estimativas para os valores passados  $Z_*$ .

Por conclusão, uma vez que a função de verossimilhança e a função da soma dos quadrados  $S(\underline{\Phi}, \underline{\theta})$  dependem dos valores passados não observáveis, decidindo-se determiná-los pelo método condicional, a maximização da função de verossimilhança é obtida minimizando a função da soma dos quadrados dos resíduos.

$$\text{Logo: } \text{Max } L^*(\underline{\xi}) = \text{Min } S^*(\underline{\xi}).$$

Quanto ao procedimento incondicional, trata-se de um processo bem mais complexo do que este apresentado. Box & Jenkins (1994) [4] chama este método incondicional de “*backforecasting*” que também permite gerar previsões das quantidades desconhecidas desde o início da série.

Ressalta-se que, na prática, quando as raízes da equação,  $\Phi(B) = 0$ , estão próximas do círculo unitário e o número de observações da série temporal não é grande em relação a ordem  $(p, q)$  do modelo, o procedimento de inicialização condicional pode não ser o mais adequado.

Contudo, para um banco de dados relativamente grande a ‘p’ e ‘q’, como a amostra da demanda de carga horária que está sendo utilizada neste estudo, a função condicional soma dos quadrados equivale a função incondicional e adotar essa mecânica condicional para inicializar a modelagem é suficiente. É pertinente

dizer que para séries pequenas quanto às ordens ( $p$ ,  $q$ ), as estimativas pelo procedimento incondicional são mais eficientes que as obtidas a partir do método condicional.

Neste estudo, considerando que a série modelada é bastante grande, optou-se pelo uso do procedimento condicional por ser mais simples e suficiente para produzir boas estimativas

### 3.4

#### **Critérios de avaliação dos modelos Box & Jenkins**

Estimado o modelo é importante verificar a sua adequação para realizar as previsões pretendidas. Para isso existem testes estatísticos baseados, principalmente, na análise dos ruídos que no caso representam os erros de previsão. Se a modelagem é boa, os ruídos devem formar uma série do tipo ruído branco. Também deve ser aplicado o teste de sobrefixação.

##### a) Teste de sobrefixação

Este teste possibilita observar se na etapa de identificação da ordem do modelo foi cometido algum tipo de engano ou ocorreu alguma distorção. O teste consiste em criar um modelo de ordem superior e verificar a significância do parâmetro extra incorporado.

Salienta-se que não se pode aumentar a ordem do modelo nas partes AR e MA ao mesmo tempo. Primeiro aumenta-se a ordem AR e avalia-se a significância deste novo parâmetro. Se for significativo permanecerá na modelagem, senão será retirado. Só após essa averiguação na parte AR do modelo, a parte MA deverá ser aumentada e, então, testa-se a significância do parâmetro acrescentado. Por fim, a significância conjunta dos parâmetros pode ser testada.

O ideal é encontrar o melhor modelo que seja ainda o mais parcimonioso.



b) Teste dos ruídos

Se o modelo estimado for bom, os resíduos devem apresentar as seguintes características:

- $\mu_{\hat{a}t} \sim \text{zero}$
- $\sigma^2_{\hat{a}t} = \text{constante}$
- FAC ( $\hat{a}t$ )  $\sim$  série ruído branco;  $\rho_k = 1$ , para  $k = 0$   
 $= 0$ , para  $k > 0$ .

Uma forma de avaliação estatística muito empregada para análise dos resíduos é o teste Ljung and Box denotado por 'Q'. Este teste averigua simultaneamente as k-primeiras autocorrelações dos ruídos e testa a hipótese nula de que os erros são descorrelacionados. Se o modelo não possuir os ruídos aleatórios significa que alguma autocorrelação importante não foi captada. Provavelmente faltou incorporar mais parâmetros ou é preciso considerar alguma variável explicativa na modelagem.

$H_0 : \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0$  (k-primeiras autocorrelações dos erros são iguais a zero).

A estatística Q segue uma distribuição Qui-quadrado com (k- p+q) graus de liberdade.

$$\begin{array}{ccc} (Q/ H_0 \text{ verdade}) \sim \chi^2 (k- p + q) & & \\ \downarrow & & \downarrow \\ \text{ruído branco} & & \text{graus de liberdade} \end{array}$$

Onde:

$$Q = N(N+2) * \sum_{i=1}^k \frac{\hat{\rho}_i^2(\mathbf{a})}{(N-i)}$$

↓

N = nº de observações da série diferenciada

Pode acontecer de mais de um modelo passar pelos testes de adequação. Nestas circunstâncias, há procedimentos para escolher um único modelo. Os critérios empregados com este objetivo sugerem que o modelo definitivo seja aquele que minimiza a variância dos ruídos e tenha a menor quantidade de parâmetros. Os principais critérios de desempate que serão aqui descritos são: AIC e BIC.

a) AIC (AKAIKE INFORMATION CRITERION):

Segundo este critério, o melhor modelo para uma mesma série é o que apresenta menor AIC. Quanto mais parâmetros, maior o AIC.

$$\begin{array}{ccc} \text{AIC} = N * \ln \hat{\sigma}_a^2 + 2 * m & \rightarrow & \text{n}^\circ \text{ de parâmetros.} \\ \downarrow & & \downarrow \\ \text{tamanho da série} & & \text{n}^\circ \text{ de parâmetros livres} \end{array}$$

b) BIC (BAYESIAN INFORMATION CRITERION):

Este critério é uma mutação do método AIC. O melhor modelo é o de menor BIC, ou seja, aquele modelo que está explicando mais da variância do erro com o menor número de variáveis. Este critério foi sugerido por Shibata em 1976.

$$\begin{array}{ccc} & & \text{variância da série} \\ & & \uparrow \\ \text{BIC} = N * \ln \hat{\sigma}_a^2 - (N - m) \ln (1 - m/n) + m \ln N + m \ln (\hat{\sigma}_z^2 - 1) / m & & \\ & & \frac{\quad}{(\hat{\sigma}_a^2)} \\ & & \downarrow \\ & & \text{variância do resíduo.} \end{array}$$

Lembra-se que estes critérios servem como desempate apenas quando os MAPE dos modelos se igualam. Se os MAPE não são iguais, então, o melhor modelo será aquele com o menor erro.

### 3.5

#### Modelagem estendida Box & Jenkins incorporando dois ciclos: diário e semanal

A forma estendida da modelagem BJ foi apresentada por Taylor (2002) [15] a fim de modelar os dados de uma série de carga de 30 min em 30 min com dois ciclos: diário e semanal. Este mesmo modelo já foi utilizado para prever séries de carga horária por Darbellay and Slama (2000) [5].

Segundo Taylor, o modelo estendido BJ tem sido aplicado para competir com as previsões da modelagem alternativa Holt-Winters, como a que foi mostrada no capítulo 2. Apesar da ênfase deste estudo ser avaliar ciclos duplos, o modelo BJ estendido pode ser adaptado para incorporar três ou mais ciclos.

A expressão geral do modelo estendido BJ com dois ciclos é dada por:

##### Notação geral:

$$\text{SARIMA}(p, d, q)^*(P_1, D_1, Q_1)_{s1}*(P_2, D_2, Q_2)_{s2}$$

##### Estrutura do modelo:

$$\Phi(B) \Phi(B^{S1}) \Omega(B^{S2}) \nabla_s^{D1} \nabla_s^{D2} \nabla^d X_t = \theta(B) \Theta(B^{S1}) \Psi(B^{S2}) \text{ at.}$$

##### Notação geral para dados horários com ciclos diário e semanal:

$$\text{SARIMA}(p, d, q)^*(P_1, D_1, Q_1)_{24}*(P_2, D_2, Q_2)_{168}$$

##### Estrutura do modelo para dados horários com ciclos diário e semanal:

$$\varphi(B) \Phi(B^{24}) \Omega(B^{168}) \nabla_{24}^{D1} \nabla_{168}^{D2} \nabla^d X_t = \theta(B) \Theta(B^{24}) \Psi(B^{168}) \text{ at.}$$

Onde:

$$\varphi(B) = 1 - \varphi_1(B^1) - \varphi_2(B^2) - \varphi_3(B^3) \dots - \varphi_p(B^p)$$

$$\Phi(B^{24}) = 1 - \Phi_1(B^{24}) - \Phi_2(B^{48}) - \Phi_3(B^{72}) - \dots - \Phi_{p1}(B^{p1(24)})$$

$$\Omega(B^{168}) = 1 - \Omega_1(B^{168}) - \Omega_2(B^{336}) - \Omega_3(B^{504}) - \dots - \Omega_{p2}(B^{p2(168)})$$

$$\nabla_{24}^{D1} = (1-B)^{D1}$$

$$\nabla_{168}^{D2} = (1-B)^{D2}$$

$$\nabla^d = (1-B)^d$$

$X_t$  = série temporal original

$$\theta(B) = 1 - \theta_1(B^1) - \theta_2(B^2) - \theta_3(B^3) - \dots - \theta_q(B^q)$$

$$\Theta(B^{24}) = 1 - \Theta_1(B^{24}) - \Theta_2(B^{48}) - \Theta_3(B^{72}) - \dots - \Theta_{Q1}(B^{Q1(24)})$$

$$\Psi(B^{168}) = 1 - \Psi_1(B^{168}) - \Psi_2(B^{336}) - \Psi_3(B^{504}) - \dots - \Psi_{Q2}(B^{Q2(168)})$$

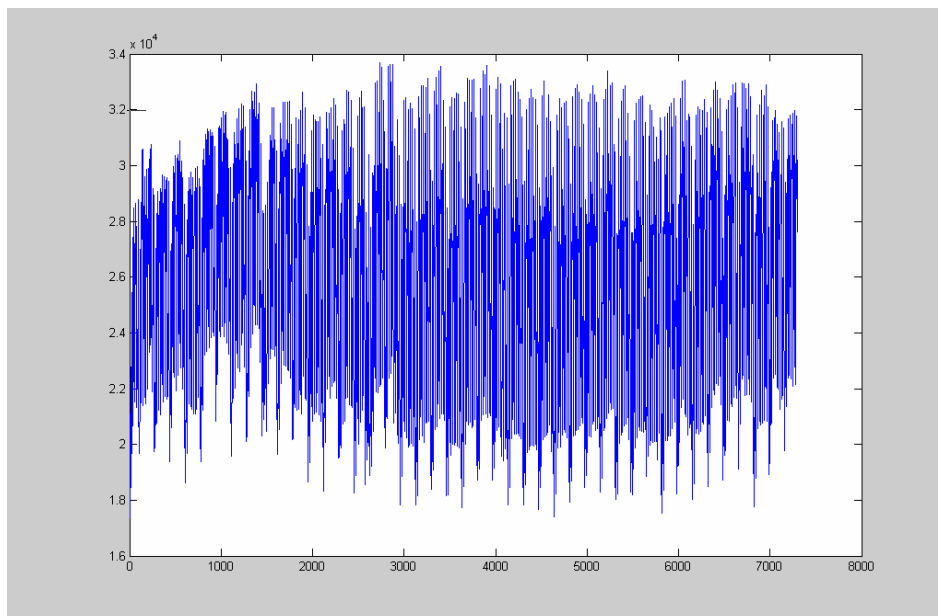
at = erro aleatório

### 3.5.1

#### Preliminares da modelagem empírica

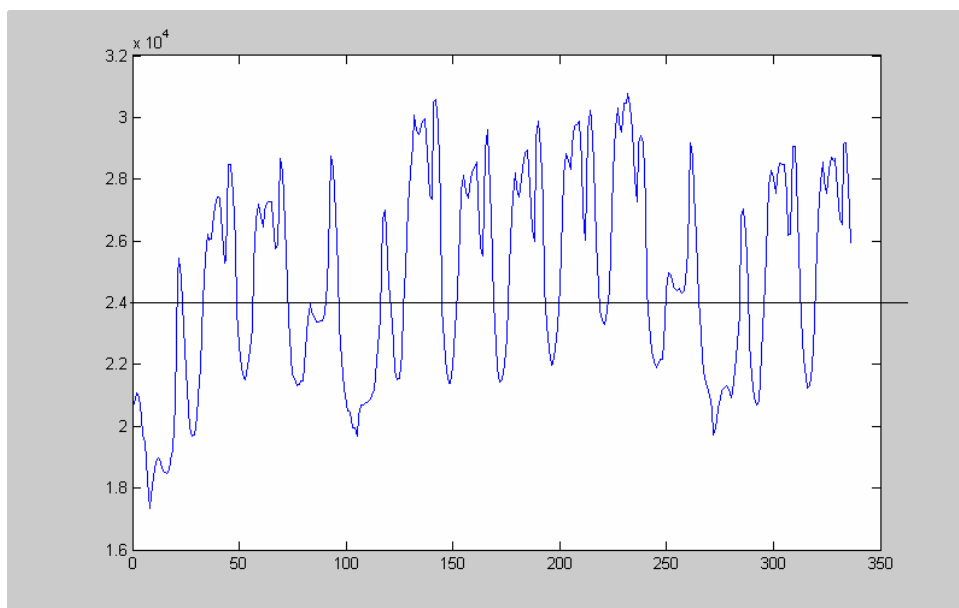
A etapa de modelagem começa verificando se a série está sujeita as condições de estacionariedade de 2º ordem. O gráfico 3.1 dá uma idéia do comportamento dos dados. Recorda-se que, no caso de estacionariedade, graficamente a série deve estar na horizontal, isto significa que o seu nível deve ser constante assim como a sua variância. Aparentemente, os gráficos 3.1 e 3.2 sugerem que a série é estacionária.

Gráfico 3.1: Série de demanda horária\_sudeste (jan-out/2003)



Ao plotar duas semanas de dados ou 336 observações, que correspondem a 14 ciclos diários completos e 2 ciclos semanais, a figura 3.2 e a linha tracejada levam a pensar que o nível está se mantendo constante.

Gráfico 3.2: Série de demanda horária\_sudeste (2 semanas/2003)



Utilizando-se dos recursos do software SPSS, a série de dados foi dividida em 7 grupos e a igualdade entre as médias e a variância dos grupos foi testada. Os resultados são colocados a seguir.

A tabela 3.1 conta que com um sig = 0,000, a hipótese nula de igualdade entre as variâncias dos grupos é rejeitada, o que significa dizer que a série em estudo é heterocedástica.

A tabela 3.2 reporta o teste de igualdade entre as médias. Sob a hipótese nula de médias iguais, os valores de sig inferiores à (0.05) permitem verificar entre quais grupos as médias se diferenciam. Portanto, se há diferença nas médias das categorias, então, o nível não é constante.

Tabela 3.1: Teste de igualdade da variância entre os grupos da série original

## ANOVA

ORIGINAL					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2,62E+09	6	436840958,8	31,233	,000
Within Groups	1,02E+11	7289	13986681,56		
Total	1,05E+11	7295			

Tabela 3.2: Teste de igualdade entre as médias

## Multiple Comparisons

Dependent Variable: ORIGINAL

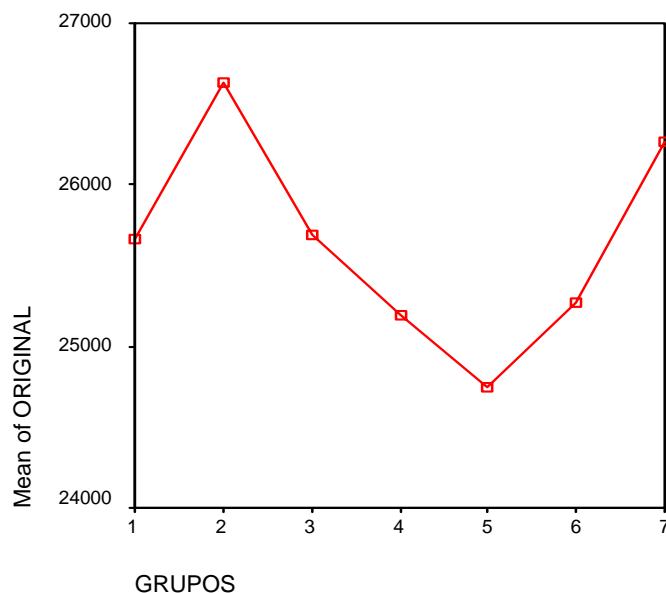
Tamhane

(I) GRUPOS	(J) GRUPOS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-970,59*	154,755	,000	-1440,21	-500,97
	3	-29,68	161,285	1,000	-519,13	459,77
	4	460,00	161,369	,089	-29,71	949,70
	5	909,51*	161,080	,000	420,68	1398,34
	6	390,06	160,991	,279	-98,49	878,62
	7	-613,65*	150,139	,001	-1069,19	-158,10
2	1	970,59*	154,755	,000	500,97	1440,21
	3	940,91*	165,944	,000	437,33	1444,49
	4	1430,58*	166,025	,000	926,76	1934,41
	5	1880,10*	165,745	,000	1377,13	2383,07
	6	1360,65*	165,658	,000	857,94	1863,36
	7	356,94	155,132	,366	-113,76	827,65
3	1	29,68	161,285	1,000	-459,77	519,13
	2	-940,91*	165,944	,000	-1444,49	-437,33
	4	489,67	172,129	,090	-32,67	1012,02
	5	939,19*	171,858	,000	417,67	1460,71
	6	419,74	171,774	,266	-101,52	941,01
	7	-583,97*	161,648	,006	-1074,46	-93,47
4	1	-460,00	161,369	,089	-949,70	29,71
	2	-1430,58*	166,025	,000	-1934,41	-926,76
	3	-489,67	172,129	,090	-1012,02	32,67
	5	449,52	171,937	,173	-72,25	971,28
	6	-69,93	171,853	1,000	-591,44	451,57
	7	-1073,64*	161,731	,000	-1564,39	-582,89
5	1	-909,51*	161,080	,000	-1398,34	-420,68
	2	-1880,10*	165,745	,000	-2383,07	-1377,13
	3	-939,19*	171,858	,000	-1460,71	-417,67
	4	-449,52	171,937	,173	-971,28	72,25
	6	-519,45	171,582	,051	-1040,13	1,24
	7	-1523,16*	161,443	,000	-2013,03	-1033,28
6	1	-390,06	160,991	,279	-878,62	98,49
	2	-1360,65*	165,658	,000	-1863,36	-857,94
	3	-419,74	171,774	,266	-941,01	101,52
	4	69,93	171,853	1,000	-451,57	591,44
	5	519,45	171,582	,051	-1,24	1040,13
	7	-1003,71*	161,354	,000	-1493,31	-514,11
7	1	613,65*	150,139	,001	158,10	1069,19
	2	-356,94	155,132	,366	-827,65	113,76
	3	583,97*	161,648	,006	93,47	1074,46
	4	1073,64*	161,731	,000	582,89	1564,39
	5	1523,16*	161,443	,000	1033,28	2013,03
	6	1003,71*	161,354	,000	514,11	1493,31

\*. The mean difference is significant at the .05 level.

Também o gráfico 3.3 que desenha as médias dos sete grupos permite observar a discrepância retrata pela tabela 3.2.

Gráfico 3.3: Média dos grupos da série original



Neste caso, para aplicar a técnica Box & Jenkins, a série deverá ser diferenciada.

A questão da estacionariedade da série pode também ser avaliada através da função de autocorrelação total (FAC). Quando a dependência entre os *lags* está demorando a decrescer, esse decréscimo lento é um indicativo de que a série provavelmente não é estacionária.

A função de autocorrelação total (FAC) da série original retratada pelos gráficos 3.4, 3.5 e 3.6 a seguir, apresenta-se da seguinte forma: para os *lags* iniciais há um decréscimo rápido. Contudo, para os *lags* sazonais do ciclo 1 e para os *lags* sazonais do ciclo 2, o amortecimento está se dando de modo muito lento. Esta situação evidencia a necessidade de aplicar na série uma diferença sazonal 1 (diário) e uma diferença sazonal 2 (semanal).

Gráfico 3.4: FAC dos 23 primeiros lags da série original

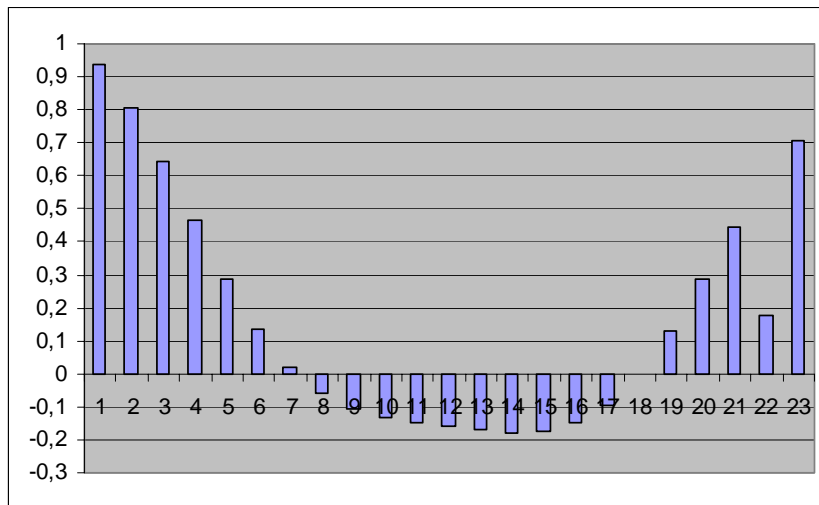


Gráfico 3.5: FAC dos lags sazonais do ciclo1 (lag 24 até lag 480 em intervalos de 24 totalizando 20 lags plotados)

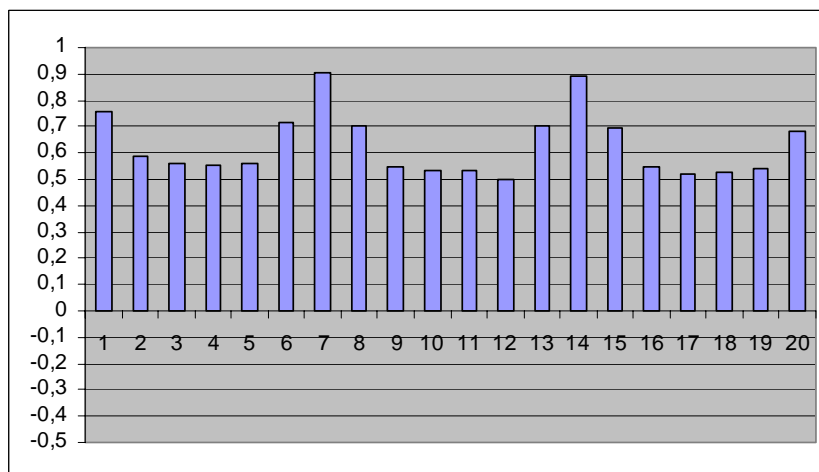
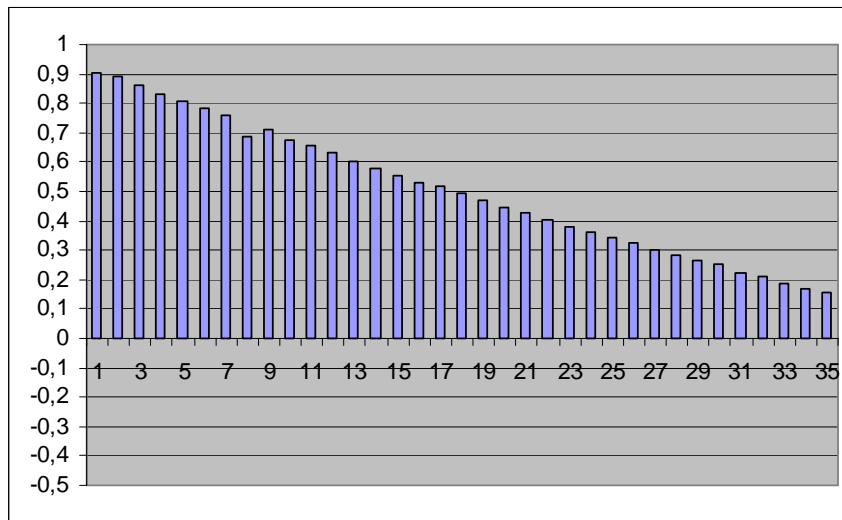




Gráfico 3.6: FAC dos *lags* sazonais do ciclo1 (*lag* 168 até *lag* 5880 em intervalos de 168 totalizando 35 *lags* plotados)



Pela aplicação de uma diferenciação sazonal 1 e sazonal 2 tem-se:

Gráfico 3.7: FAC de 23 *lags* da série após diferença sazonal 1 e 2

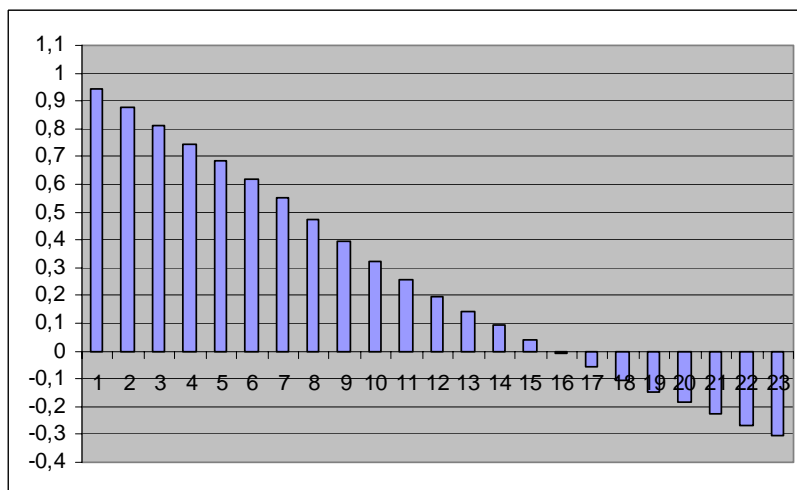


Gráfico 3.8: FAC dos *lags* sazonais do ciclo1 (*lag* 24 até *lag* 480 em intervalos de 24 totalizando 20 *lags* após diferença sazonal 1 e 2).

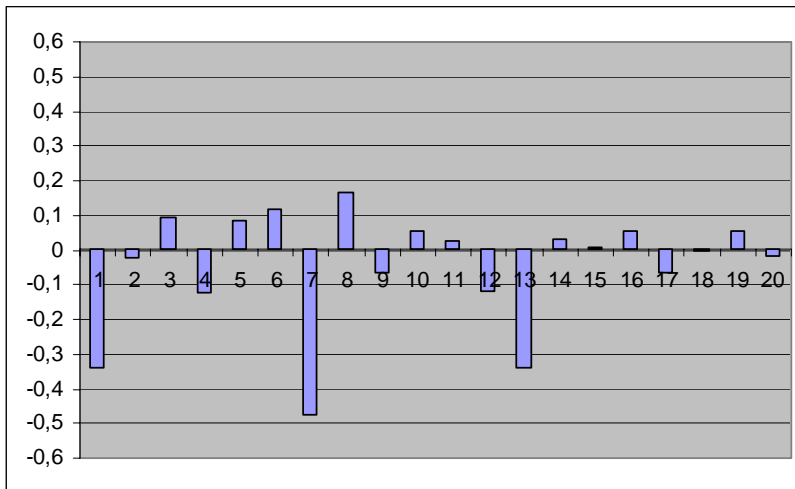
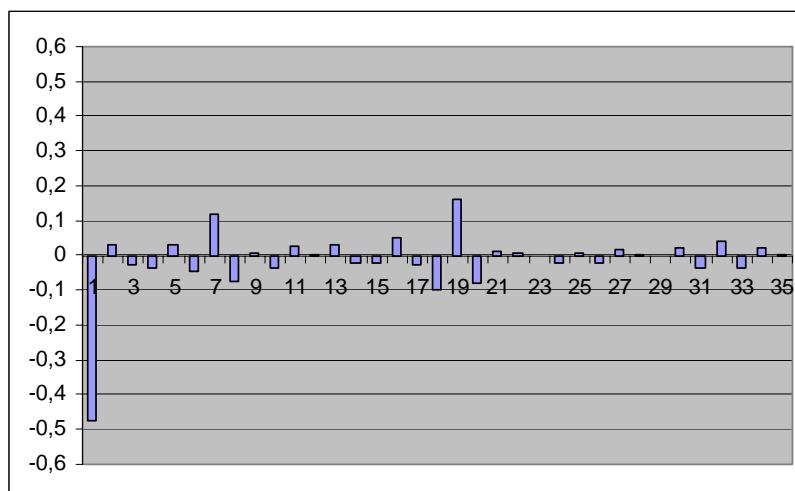


Gráfico 3.9: FAC dos *lags* sazonais do ciclo2 (*lag* 168 até *lag* 5880 em intervalos de 168 totalizando 35 *lags* após diferença sazonal 1 e 2).



Aplicando-se as diferenciações sazonais 1 e 2 nos dados originais, nota-se que os *lags* iniciais (*lag* 1 até *lag* 23) passam a ter um decréscimo lento sendo preciso efetuar ainda uma diferenciação simples na tentativa de estacionar a série.

A seguir a FAC da série após uma diferenciação simples, uma diferenciação sazonal 1 e uma diferenciação sazonal 2.

Gráfico 3.10: FAC de 23 *lags* da série após diferença simples e sazonal 1 e 2.

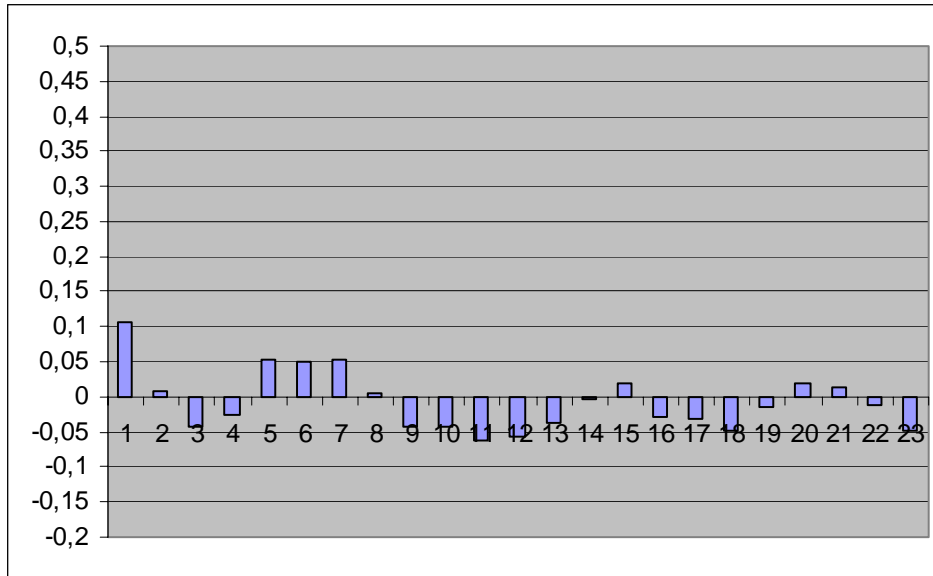


Gráfico 3.11: FAC dos *lags* sazonais do ciclo1 (*lag* 24 até *lag* 480 em intervalos de 24 totalizando 20 *lags* após diferença simples e sazonal 1 e 2).

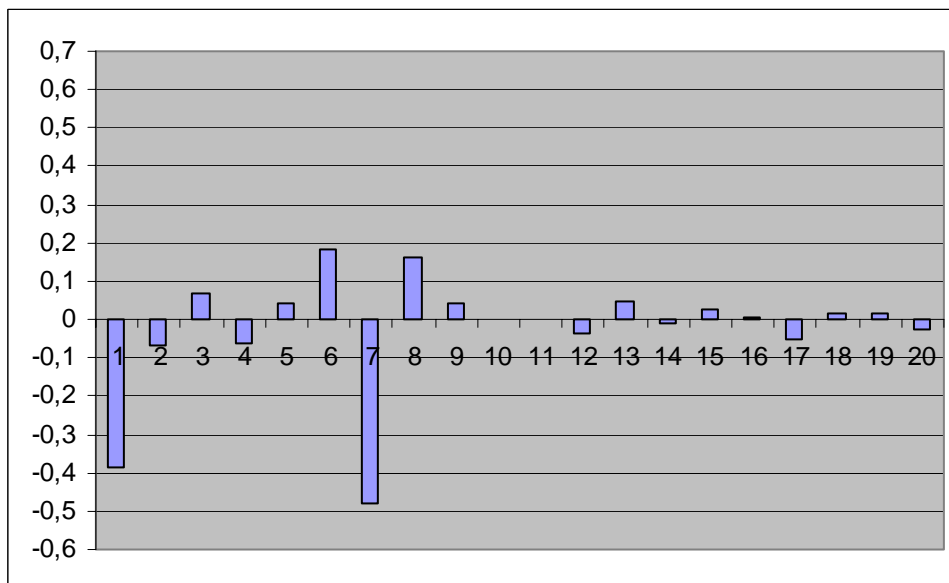
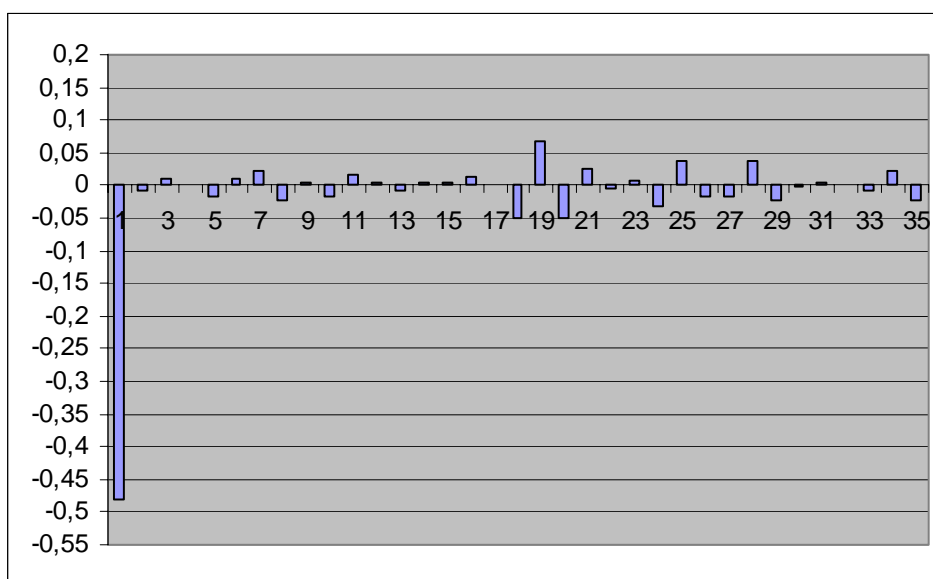


Gráfico 3.12: FAC dos *lags* sazonais do ciclo2 (*lag* 168 até *lag* 5880 em intervalos de 168 totalizando 35 *lags* após diferença simples e sazonal 1 e 2.



Existe ainda uma terceira maneira de investigar a estacionariedade da série além das duas que já foram mostradas: 1) gráfico da série no tempo; 2) FAC; 3) testes de raiz unitária/não estacionariedade. Os testes de raiz unitária são procedimentos mais formais para a averiguação da hipótese de estacionariedade de uma série temporal. Os outros dois procedimentos 1 e 2, apesar de válidos podem gerar resultados enganosos. Há uma variedade de testes disponíveis no Eviews 4.1 que servem para esta finalidade. Um dos mais conhecidos que será aqui implementado é o teste ADF (Augmented Dickey Fuller) no qual a hipótese nula é de que a série possui raiz unitária, ou seja, não é estacionária. A hipótese alternativa é a estacionariedade da série.

Os resultados para a série de dados original é colocado na tabela 3.3 que segue:

Tabela 3.3: Teste de Raiz unitária (ADF) para a série de dados original

Estatística do teste	-1.7897
Valor crítico 1%	- 3.4782
Valor crítico 5%	-2.67828
Valor crítico 10%	-2.1478
Probabilidade	0.5519

De acordo com o quadro 3.3, com um sig=0.5519 aceita-se a hipótese nula do teste de raiz unitária/não estacionariedade dos dados.

Para averiguar se estas diferenciações feitas possibilitaram tornar a série estacionária, o mesmo teste anteriormente executado no SPSS para a série original e o testes ADF pode ser feito para a série diferenciada a fim de observar como se comporta a sua média. No caso dos testes de média e ANOVA do SPSS, a série foi novamente dividida em 7 grupos.

Segundo o resultado da tabela 3.4, com um sig = 1.000 > 0.05, aceita-se a hipótese nula de igualdade das variâncias. Em outras palavras, a série agora é homocedástica. As informações contidas na tabela 3.5 mostram que as médias entre os grupos são iguais. O gráfico 3.13 corrobora com essa conclusão.

Tabela 3.4: Teste da igualdade da variância para a série diferenciada

**ANOVA**

DEMANDA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3497,227	6	582,871	,002	1,000
Within Groups	1,68E+09	7096	237395,520		
Total	1,68E+09	7102			

Tabela 3.5: Teste de igualdade entre as médias para a série diferenciada

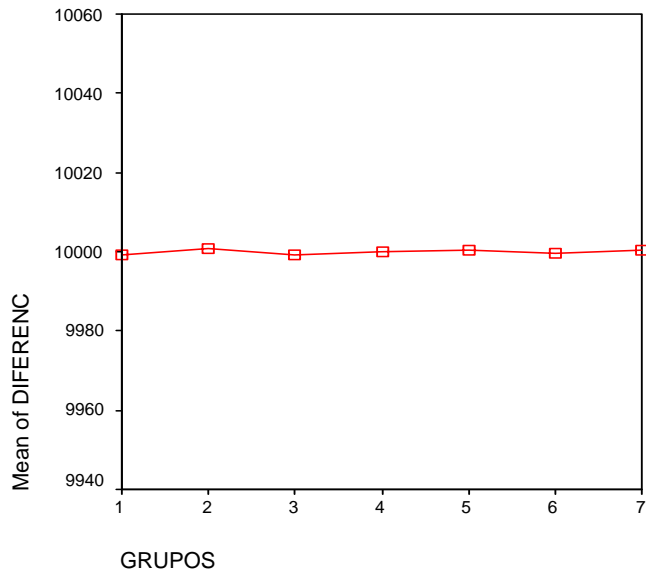
## Multiple Comparisons

Dependent Variable: DIFERENC

Bonferroni

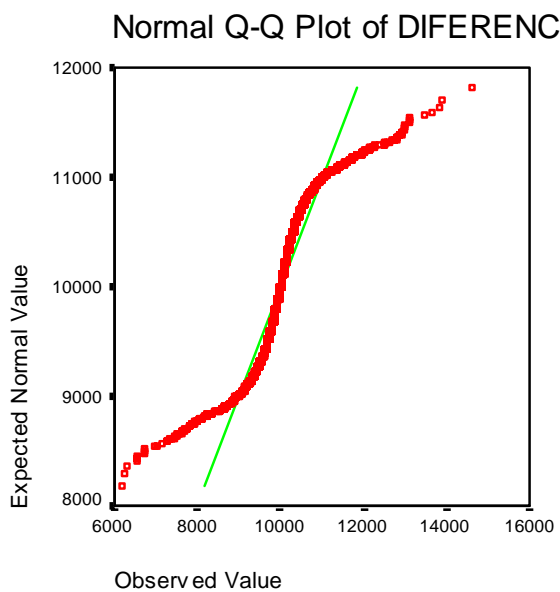
(I) GRUPOS	(J) GRUPOS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-1,80	21,790	1,000	-68,02	64,42
	3	,08	21,790	1,000	-66,14	66,30
	4	-,95	21,790	1,000	-67,17	65,27
	5	-1,24	21,790	1,000	-67,46	64,98
	6	-,53	21,790	1,000	-66,75	65,69
	7	-1,16	21,275	1,000	-65,82	63,50
2	1	1,80	21,790	1,000	-64,42	68,02
	3	1,88	21,790	1,000	-64,34	68,10
	4	,85	21,790	1,000	-65,37	67,08
	5	,56	21,790	1,000	-65,66	66,78
	6	1,27	21,790	1,000	-64,95	67,50
	7	,64	21,275	1,000	-64,02	65,30
3	1	-,08	21,790	1,000	-66,30	66,14
	2	-1,88	21,790	1,000	-68,10	64,34
	4	-1,03	21,790	1,000	-67,25	65,20
	5	-1,32	21,790	1,000	-67,54	64,90
	6	-,61	21,790	1,000	-66,83	65,62
	7	-1,24	21,275	1,000	-65,90	63,42
4	1	,95	21,790	1,000	-65,27	67,17
	2	-,85	21,790	1,000	-67,08	65,37
	3	1,03	21,790	1,000	-65,20	67,25
	5	-,29	21,790	1,000	-66,52	65,93
	6	,42	21,790	1,000	-65,80	66,64
	7	-,21	21,275	1,000	-64,87	64,44
5	1	1,24	21,790	1,000	-64,98	67,46
	2	-,56	21,790	1,000	-66,78	65,66
	3	1,32	21,790	1,000	-64,90	67,54
	4	,29	21,790	1,000	-65,93	66,52
	6	,71	21,790	1,000	-65,51	66,94
	7	,08	21,275	1,000	-64,58	64,74
6	1	,53	21,790	1,000	-65,69	66,75
	2	-1,27	21,790	1,000	-67,50	64,95
	3	,61	21,790	1,000	-65,62	66,83
	4	-,42	21,790	1,000	-66,64	65,80
	5	-,71	21,790	1,000	-66,94	65,51
	7	-,63	21,275	1,000	-65,29	64,02
7	1	1,16	21,275	1,000	-63,50	65,82
	2	-,64	21,275	1,000	-65,30	64,02
	3	1,24	21,275	1,000	-63,42	65,90
	4	,21	21,275	1,000	-64,44	64,87
	5	-,08	21,275	1,000	-64,74	64,58
	6	,63	21,275	1,000	-64,02	65,29

Gráfico 3.13: Médias dos grupos da série diferenciada



A série diferenciada não apresenta uma distribuição normal como pode ser visto na figura 3.14. A normalidade dos dados é um ponto essencial para usar o método de modelagem BJ. Apesar da contatação feita no quadro 3.14, como a série é bastante grande a não normalidade torna-se um fator irrelevante.

Gráfico 3.14: Normalidade da série em estudo



Os resultados da avaliação da estacionariedade pelo teste de raiz unitária ADF são mostrados na tabela 3.6.

Tabela 3.6: Teste de Raiz unitária (ADF) para a série de dados diferenciada

Estatística do teste	-10.421
Valore crítico 1%	- 3.4782
Valore crítico 5%	-2.67822
Valore crítico 10%	-2.1478
Probabilidade	0.0000

Portanto, é possível concluir que, após as diferenciações realizadas, a série em estudo tornou-se estacionária. O valor do sig =0.000 leva a rejeitar a hipótese de não estacionariedade.

Pelos gráficos da FAC explica-se a ordem especificada para a parte MA do modelo.

Quanto a parte simples, pelo gráfico 3.10, todos os *lags* estão abaixo do limite de 0.2 (limite definido no FPW para aceitar a hipótese de decorrelação entre os *lags*), logo, a parte MA simples foi considerada como zero. Para a parte MA do ciclo 1, segundo o gráfico 3.11, como o primeiro *lag* ultrapassa o limite de 0.2 chegando a 0.4 e, em seguida, há um corte brusco do *lag* 1 para o *lag* 2, então MA sazonal 1 foi considerado com ordem 1. Em relação a parte MA referente ao ciclo 2, conforme mostra o gráfico 3.12, o primeiro *lag* apresenta um coeficiente de correlação próximo de 0.5, enquanto os *lags* 2 e 3 estão praticamente zerados. Esse fato leva a concluir que a parte MA do ciclo 2 também é um MA de ordem 1.

Uma vez que, pela análise da FAC, a série foi considerada adequada para ser modelada por BJ deve-se prosseguir o estudo observando a FACP (função de autocorrelação parcial) que possibilita definir a ordem da parte AR da modelagem.

A seguir coloca-se a tabela 3.7 que traz os valores dos  $\hat{\rho}$ s finais estimados para o correlograma da ACF que servirão para calcular e plotar a correlação na FACP.



Tabela 3.7: Rôs finais estimados para a FAC

ro_1	0,106364	ro_96	-0,059802	ro_1680	-0,016258
ro_2	0,007048	ro_120	0,044744	ro_1848	0,017222
ro_3	-0,043439	ro_144	0,180617	ro_2016	0,004813
ro_4	-0,025228	ro_168	-0,479676	ro_2184	-0,006724
ro_5	0,051758	ro_192	0,164686	ro_2352	0,004306
ro_6	0,051366	ro_216	0,041426	ro_2520	0,00312
ro_7	0,052268	ro_240	0,001118	ro_2688	0,012654
ro_8	0,004754	ro_264	0,000113	ro_2856	0,002196
ro_9	-0,043676	ro_288	-0,03634	ro_3024	-0,050924
ro_10	-0,042001	ro_312	0,045708	ro_3192	0,068275
ro_11	-0,061957	ro_336	-0,008545	ro_3360	-0,0496
ro_12	-0,055669	ro_360	0,028106	ro_3528	0,026
ro_13	-0,038066	ro_384	0,006887	ro_3696	-0,003751
ro_14	-0,004273	ro_408	-0,049458	ro_3864	0,005793
ro_15	0,018515	ro_432	0,018611	ro_4032	-0,031386
ro_16	-0,02759	ro_456	0,016812	ro_4200	0,036182
ro_17	-0,030582	ro_480	-0,026593	ro_4368	-0,017252
ro_18	-0,048514	ro_168	-0,479676	ro_4536	-0,015612
ro_19	-0,014226	ro_336	-0,008545	ro_4704	0,037196
ro_20	0,019233	ro_504	0,010354	ro_4872	-0,022339
ro_21	0,015052	ro_672	0,001473	ro_5040	-0,002783
ro_22	-0,012937	ro_840	-0,01564	ro_5208	0,005655
ro_23	-0,046923	ro_1008	0,009284	ro_5376	0,000683
ro_24	-0,38846	ro_1176	0,021998	ro_5544	-0,009244
ro_48	-0,067935	ro_1344	-0,024329	ro_5712	0,02087
ro_72	0,068994	ro_1512	0,003587	ro_5880	-0,023087

O raciocínio para o cálculo da FACP está baseado nas equações abaixo sugeridas por Durbin (Souza & Camargo (2004) [13]):

$$\hat{\phi}_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{22} = \hat{\rho}_2 - \hat{\rho}_1^2$$

$$1 - \hat{\rho}_1^2$$

$$\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22} * \hat{\phi}_{11}$$

$$\hat{\phi}_{33} = \hat{\rho}_3 - \hat{\phi}_{21} * \hat{\rho}_2 - \hat{\phi}_{22} * \hat{\rho}_1$$

$$1 - \hat{\phi}_{21} * \hat{\rho}_1 - \hat{\phi}_{22} * \hat{\rho}_2$$

E, assim sucessivamente para o cálculo de todos os  $\hat{\phi}_{kk}$ .

Os correlogramas da FACP estão a seguir:

Gráfico 3.15: FACP dos 23 lags simples

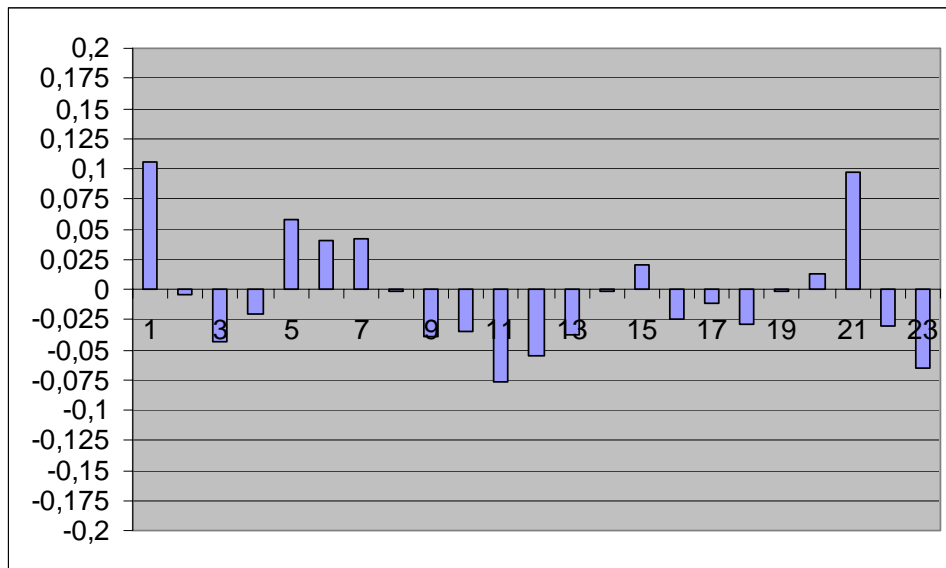


Gráfico 3.16: FACP dos lags sazonais do ciclo 1 (lag 24 até lag 480 em intervalos de 24 lags totalizando 20 lags sazonais 1)

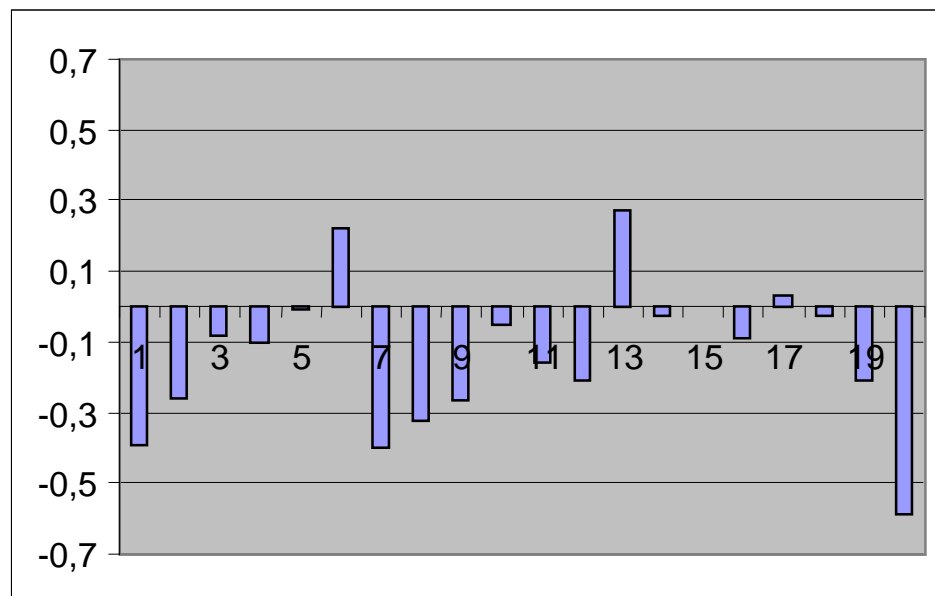
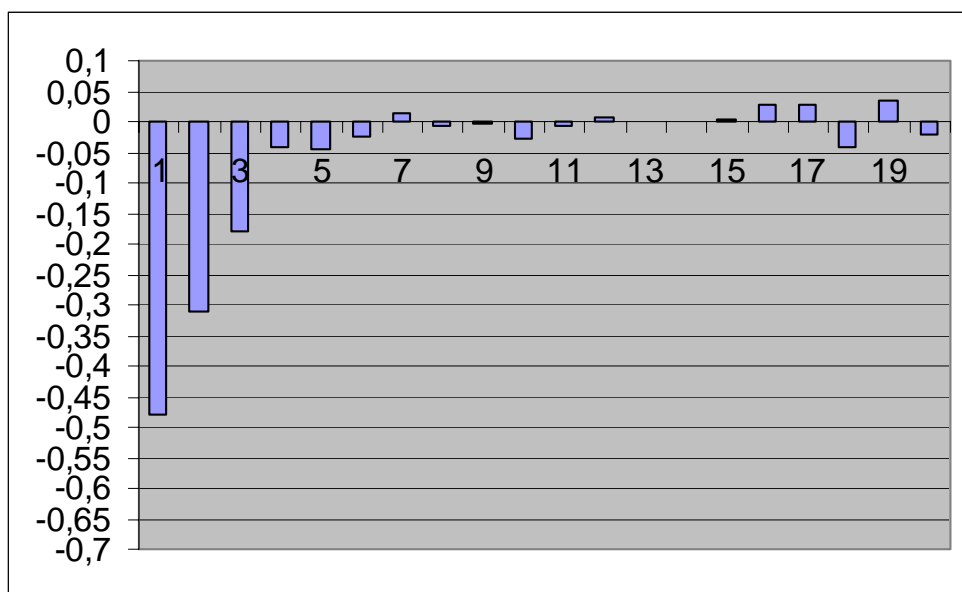


Gráfico 3.17: FACP dos lags sazonais do ciclo 2 (lag 168 até lag 3360 em intervalos de 168 lags totalizando 20 lags sazonais 2)



Pelos correlogramas da FACP, para os *lags* iniciais, o gráfico 3.15 relata que todos os *lags* estão abaixo do limite de 0,2, o que faz a ordem da parte AR simples ser zero. O gráfico 3.16 acusa que o *lag* sazonal 1 tem um coeficiente de correlação próximo de 0,4 havendo em seguida um decréscimo. Portanto, o ciclo 1 tem ordem AR (1)<sub>24</sub>. O gráfico 3.17 mostra que os *lags* 1e 2 do ciclo 2 registram uma correlação significativa, notando-se do *lag* 2 para o *lag* 3 um corte brusco. Assim, para o ciclo 2 define-se um AR (2)<sub>168</sub>.

Diante dos gráficos da FAC, da FACP e das diferenciações feitas, a ordem estimada para modelo BJ em avaliação é:

$$\text{SARIMA } (0,1,0)*(1,1,1)_{24}*(2,1,1)_{168}$$

Como comentário, menciona-se que Taylor & Menezes & Mcsharry (2004) [18] em trabalho exposto em Sidney\_Austrália encontraram para:

**Caso1:** Rio de Janeiro (dados horários)

$$\text{SARIMA } (3,0,3)*(3,0,3)_{24}*(3,0,3)_{168}$$

**Caso2:** Inglaterra (dados de 30 min em 30 min)

SARIMA (2,0,1)\*(2,0,1)<sub>48</sub>\*(1,0,2)<sub>336</sub>

Abrindo o modelo SARIMA (0,1,0)\*(1,1,1)<sub>24</sub>\*(2,1,1)<sub>168</sub>, definido para a série do sudeste em estudo, o detalhamento dessa ordem definida podem ser melhor conferidos.

Para um modelo SARIMA (0,1,0)\*(1,1,1)<sub>24</sub>\*(2,1,1)<sub>168</sub>

$$\nabla^d = (1-B)^1$$

$$\nabla_{24}^D = (1-B^{24})^1$$

$$\nabla_{168}^D = (1-B^{168})^1$$

$$\Phi'(B) = (1 - \Phi_1' B^{24}) : \text{parâmetro da parte AR do ciclo1}$$

$$\Phi''(B) = (1 - \Phi_1'' B^{168} - \Phi_2'' B^{336}) : \text{parâmetro da parte AR do ciclo2}$$

$$\Theta'(B) = (1 - \Theta_1' B^{24}) : \text{parâmetro da parte MA do ciclo1}$$

$$\Theta''(B) = (1 - \Theta_1'' B^{168})$$

$$= \nabla^1 \nabla_{24}^1 \nabla_{168}^1 \Phi'(B^{24}) \Phi''(B^{168}) Z_t = \Theta'(B^{24}) \Theta''(B^{168}) \text{ at}$$

Por partes:

AR

$$= (1-B)^1 (1-B^{24})^1 (1-B^{168})^1 (1 - \Phi_1' B^{24}) (1 - \Phi_1'' B^{168} - \Phi_2'' B^{336}) Z_t$$

$$= (1 - B^{24} - B + B^{25}) (1 - B^{168}) = (1 - B^{168} - B^{24} - B^{192} - B + B^{169} + B^{25} - B^{193})$$

$$= (1 - \Phi_1' B^{24}) (1 - \Phi_1'' B^{168} - \Phi_2'' B^{336}) =$$

$$= (1 - \Phi_1'' B^{168} - \Phi_2'' B^{336} - \Phi_1' B^{24} + \Phi_1' \Phi_1'' B^{192} + \Phi_1' \Phi_1'' B^{360}) Z_t$$

MA

$$= (1 - \Theta' B^{24}) (1 - \Theta'' B^{168}) =$$

$$= (1 - \Theta'' B^{168} - \Theta' B^{24} + \Theta' \Theta'' B^{192}) \text{ at}$$

MODELO

$$(1 - B^{168} - B^{24} + B^{192} - B + B^{169} + B^{25} - B^{193}) (1 - \Phi_1'' B^{168} - \Phi_2'' B^{336} - \Phi_1' B^{24} + \Phi_1' \Phi_1'' B^{192} + \Phi_1' \Phi_1'' B^{360}) Z_t = (1 - \Theta'' B^{168} - \Theta' B^{24} + \Theta' \Theta'' B^{192}) \text{ at.}$$

$$= (1 - \Phi_1'' B^{168} - \Phi_2'' B^{336} - \Phi_1' B^{24} + \Phi_1' \Phi_1'' B^{192} + \Phi_1' \Phi_1'' B^{360} - B^{168} \dots$$

$$+ \Phi_1'' B^{336} + \Phi_2'' B^{504} + \Phi_1'' B^{192} - \Phi_1' \Phi_1'' B^{360} - \Phi_1 \Phi_2'' B^{528} - B^{24} + \Phi_1'' B^{192} \dots$$

$$+ \Phi_2'' B^{360} + \Phi_1'' B^{48} - \Phi_1' \Phi_1'' B^{216} - \Phi_1 \Phi_2'' B^{384} + B^{192} - \Phi_1'' B^{360} - \Phi_2'' B^{528} \dots$$

$$- \Phi_1' B^{216} + \Phi_1' \Phi_1'' B^{384} + \Phi_1 \Phi_2'' B^{552} - B + \Phi_1'' B^{169} + \Phi_2'' B^{337} + \Phi_1' B^{25} \dots$$

$$- \Phi_1' \Phi_1'' B^{193} - \Phi_1' \Phi_2'' B^{361} + B^{169} - \Phi_1'' B^{337} - \Phi_2'' B^{505} - \Phi_1'' B^{193} \dots$$

$$+ \Phi_1' \Phi_1'' B^{361} + \Phi_1 \Phi_2'' B^{529} + B^{25} - \Phi_1'' B^{193} - \Phi_2'' B^{361} - \Phi_1'' B^{49} \dots$$

$$+ \Phi_1' \Phi_1'' B^{217} + \Phi_1' \Phi_2'' B^{385} - B^{193} + \Phi_1'' B^{361} + \Phi_2'' B^{529} + \Phi_1' B^{217} \dots$$

$$- \Phi_1' \Phi_1'' B^{385} - \Phi_1' \Phi_2'' B^{553}) Z_t = (1 - \Theta'' B^{168} - \Theta' B^{24} + \Theta' \Theta'' B^{192}) \text{ at.}$$

$$= Z_t - \Phi_1'' Z_{t-168} - \Phi_2'' Z_{t-336} - \Phi_1' Z_{t-24} + \Phi_1' \Phi_1'' Z_{t-192} + \Phi_1' \Phi_1'' Z_{t-360} \dots$$

$$- Z_{t-168} + \Phi_1'' Z_{t-336} + \Phi_2'' Z_{t-504} + \Phi_1'' Z_{t-192} - \Phi_1' \Phi_1'' Z_{t-360} \dots$$

$$- \Phi_1 \Phi_2'' Z_{t-528} - Z_{t-24} + \Phi_1'' Z_{t-192} + \Phi_2'' Z_{t-360} + \Phi_1'' Z_{t-48} \dots$$

$$- \Phi_1' \Phi_1'' Z_{t-216} - \Phi_1 \Phi_2'' Z_{t-384} + Z_{t-192} - \Phi_1'' Z_{t-360} - \Phi_2'' Z_{t-528} \dots$$

$$- \Phi_1' Z_{t-216} + \Phi_1' \Phi_1'' Z_{t-384} + \Phi_1 \Phi_2'' Z_{t-552} - Z_{t-1} + \Phi_1'' Z_{t-169} \dots$$

$$+ \Phi_2'' Z_{t-337} + \Phi_1' Z_{t-25} - \Phi_1' \Phi_1'' Z_{t-193} - \Phi_1' \Phi_2'' Z_{t-361} + Z_{t-169} \dots$$

$$- \Phi_1'' Z_{t-337} - \Phi_2'' Z_{t-505} - \Phi_1'' Z_{t-193} + \Phi_1' \Phi_1'' Z_{t-361} + \Phi_1 \Phi_2'' Z_{t-529} \dots$$

$$+ Z_{t-25} - \Phi_1'' Z_{t-193} - \Phi_2'' Z_{t-361} - \Phi_1'' Z_{t-49} + \Phi_1' \Phi_1'' Z_{t-217} + \Phi_1' \Phi_2'' Z_{t-385} \dots$$

$$- Z_{t-193} + \Phi_1'' Z_{t-361} + \Phi_2'' Z_{t-529} + \Phi_1' Z_{t-217} - \Phi_1' \Phi_1'' Z_{t-385} - \Phi_1' \Phi_2'' Z_{t-553}$$

$$= \text{at} - \Theta'' \text{at-168} - \Theta' \text{at-24} + \Theta' \Theta'' \text{at-192.}$$

Colocando o modelo em ordem crescente quanto a idade da observação:

$$\begin{aligned}
&= Z(t)-Z(t-1)-Z(t-24)+Z(t-25)-Z(t-168)+Z(t-169)+Z(t-192)-Z(t-193)... \\
&\quad - \Phi'_1 * Z(t-24) + \Phi'_1 * Z(t-25) + \Phi'_1 * Z(t-48) - \Phi'_1 * Z(t-49) + \Phi'_1 * Z(t-192)... \\
&\quad - \Phi'_1 * Z(t-193) - \Phi'_1 * Z(t-216) + \Phi'_1 * Z(t-217) - \Phi_1'' * Z(t-168) + \Phi_1'' * Z(t-169)... \\
&\quad + \Phi_1'' * Z(t-192) - \Phi_1'' * Z(t-193) + \Phi_1'' * Z(t-336) - \Phi_1'' * Z(t-337) - \Phi_1'' * Z(t-360)... \\
&\quad + \Phi_1'' * Z(t-361) - \Phi_2'' * Z(336) + \Phi_2'' * Z(t-337) + \Phi_2'' * Z*(t-360) - \Phi_2'' * Z*(t-361)... \\
&\quad + \Phi_2'' * Z(t-504) - \Phi_2'' * Z(t-505) - \Phi_2'' * Z(t-528) + \Phi_2'' * Z(t-529) + \Phi'_1 * \Phi_1'' * Z(t-192)... \\
&\quad - \Phi'_1 * \Phi_1'' * Z(t-193) - \Phi'_1 * \Phi_1'' * Z(t-216) + \Phi'_1 * \Phi_1'' * Z(t-217) - \Phi'_1 * \Phi_1'' * Z(t-360)... \\
&\quad + \Phi'_1 * \Phi_1'' * Z(t-361) + \Phi'_1 * \Phi_1'' * Z(t-384) - \Phi'_1 * \Phi_1'' * Z(t-385) + \Phi'_1 * \Phi_2'' * Z(t-360)... \\
&\quad - \Phi'_1 * \Phi_2'' * Z(t-361) - \Phi'_1 * \Phi_2'' * Z(t-384) + \Phi'_1 * \Phi_2'' * Z(t-385) - \Phi'_1 * \Phi_2'' * Z(t-528)... \\
&\quad + \Phi'_1 * \Phi_2'' * Z(t-529) + \Phi'_1 * \Phi_2'' * Z(t-552) - \Phi'_1 * \Phi_2'' * Z(t-553)... \\
&= at - \Theta' * a(t-24) - \Theta'' * a(t-168) + \Theta' \Theta'' * a(t-192).
\end{aligned}$$

Sobre os parâmetros é útil esclarecer:

$\Phi'_1$  : parâmetro 1 da parte AR sazonal ciclo 1;

$\Phi_1''$  : parâmetro 1 da parte AR sazonal ciclo2;

$\Phi_2''$  : parâmetro 2 da parte AR sazonal ciclo2;

$\Theta'$  : parâmetro 1 da parte MA sazonal ciclo1;

$\Theta''$  : parâmetro 1 da parte sazonal ciclo2;

Estabelecida a ordem do modelo BJ estendido, é preciso estimar os valores ótimos dos cinco parâmetros. Essa parte do estudo bem como as previsões a serem feitas foram realizadas no software MATLAB 6.1. Para a utilização do MATLAB foi válido consultar Bastos (1997) [2].

O primeiro passo para a estimação dos parâmetros foi jogar 'at' como variável dependente.

$$\begin{aligned}
at &= Z(t)-Z(t-1)-Z(t-24)+Z(t-25)-Z(t-168)+Z(t-169)-Z(t-192)-Z(t-193)... \\
&\quad - \Phi'_1 * Z(t-24) + \Phi'_1 * Z(t-25) + \Phi'_1 * Z(t-48) - \Phi'_1 * Z(t-49) + \Phi'_1 * Z(t-192)... \\
&\quad - \Phi'_1 * Z(t-193) - \Phi'_1 * Z(t-216) + \Phi'_1 * Z(t-217) - \Phi_1'' * Z(t-168) + \Phi_1'' * Z(t-169)... \\
&\quad + \Phi_1'' * Z(t-192) - \Phi_1'' * Z(t-193) + \Phi_1'' * Z(t-336) - \Phi_1'' * Z(t-337) - \Phi_1'' * Z(t-360)... \\
&\quad + \Phi_1'' * Z(t-361) - \Phi_2'' * Z(336) + \Phi_2'' * Z(t-337) + \Phi_2'' * Z*(t-360) - \Phi_2'' * Z*(t-361)...
\end{aligned}$$

$$\begin{aligned}
& +\Phi_2''*Z(t-504)-\Phi_2''*Z(t-505)-\Phi_2''*Z(t-528)+\Phi_2''*Z(t-529)+\Phi_1'*\Phi_1''*Z(t-192)... \\
& -\Phi_1'*\Phi_1''*Z(t-193)-\Phi_1'*\Phi_1''*Z(t-216)+\Phi_1'*\Phi_1''*Z(t-217)-\Phi_1'*\Phi_1''*Z(t-360)... \\
& +\Phi_1'*\Phi_1''*Z(t-361)+\Phi_1'*\Phi_1''*Z(t-384)-\Phi_1'*\Phi_1''*Z(t-385)+\Phi_1'*\Phi_2''*Z(t-360)... \\
& -\Phi_1'*\Phi_2''*Z(t-361) - \Phi_1'*\Phi_2''*Z(t-384)+\Phi_1'*\Phi_2''*Z(t-385)-\Phi_1'*\Phi_2''*Z(t-528)... \\
& +\Phi_1'*\Phi_2''*Z(t-529)+\Phi_1'*\Phi_2''*Z(t-552)-\Phi_1'*\Phi_2''*Z(t-553) +\Theta'*a(t-24) \\
& +\Theta''*a(t-168) - \Theta' \Theta'' *a(t-192).
\end{aligned}$$

Pode ser visto que, para o cálculo de 'at', é necessário conhecer os valores passados da série e do erro. Como a série diferenciada apresenta 7103 observações, ou seja, trata-se de uma série bastante grande, o cálculo de 'at' começou a partir do instante  $t = 554$ , sem prejuízos analíticos e, assim, se tem todos os valores da defasagem de  $Z^*$  (série diferenciada) que vai até (t-553). Embora tenha sido mostrado que a série  $Z$  (diferenciada) não segue uma distribuição normal, a sua distribuição pode ser aproximada para uma normal devido ao grande número de observações. De acordo com Judge (1988) [10], como 'at' está linearmente relacionada com os termos de  $Z$ , então, por suposição 'at' segue uma distribuição normal.

Logo:  $EMV \cong EMQ$ .

Ainda considerando que a parte do erro na modelagem constitui-se num MA (1) para o ciclo 1 e para o ciclo 2, os valores passados do termo aleatório, de acordo com o método condicional, foram assumidos como zero até que os respectivos valores da defasagem de 'at' fossem conhecidos pelo próprio processo de estimação realizado.

Quanto ao intervalo de busca para a otimização dos parâmetros, os limites inferior e superior foram estabelecidos, como explicado em Box & Jenkins (1994) [4]:

- Para um processo AR (1) estacionário:  $\Phi_1 \in ]-1,1[$
- Para um processo AR (2) estacionário:  $\Phi_1 \in ]-2,2[$  e  $\Phi_2 \in ]-1,1[$
- Para um processo MA (1) estacionário e inversível:  $\Theta \in ]1,1[$ .

É bom esclarecer que estas condições garantem que se  $k \rightarrow \infty$ ,  $\hat{\rho}$  e  $\phi_{kk}$  devem decrescer.

A busca dentro dos respectivos limites para cada um dos cinco parâmetros inicialmente aconteceu em intervalos de 0.1. Depois intervalos menores foram sendo construídos em torno do último valor otimizado até que a procura se deu em intervalos de 0.01. A combinação de valores para os parâmetros que gerou a menor soma dos erros quadráticos foi o critério considerado para estabelecer os valores otimizados.

Com base nestas explicações, os valores ótimos encontrados para os parâmetros do modelo foram:

$$\Phi'_1 = -0.46$$

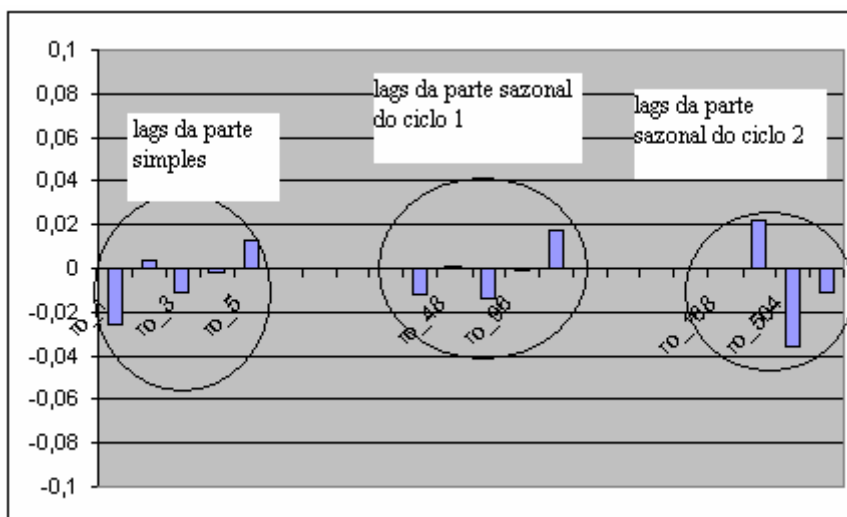
$$\Phi''_1 = -0.60$$

$$\Phi''_2 = -0.19$$

$$\Theta' = 0.85$$

$$\Theta'' = 0.83$$

Gráfico 3.18: Correlograma dos resíduos do modelo. SARIMA  $(0,1,0)*(1,1,1)_{24}*(2,1,1)_{168}$





Observa-se pela figura 3.18 que os resíduos do modelo identificado formam uma série do tipo ruído branco, pois o coeficiente de correlação está abaixo do limite de 0.2 (critério do FPW). Quando os erros não são aleatórios, normalmente, tenta-se uma melhor identificação da ordem do modelo através do teste de sobrefixação ou a inclusão de alguma variável causal.

Dando continuidade a modelagem empírica pretendida, conhecidos os valores “ótimos” dos parâmetros, as previsões 01 passo-à-frente e 24 passos-à-frente podem ser realizadas.

Equação de Z no período (t + l):

l = horizonte de previsão.

- Para 01 passo-à-frente: l = 1.
- Para 24 passos-à-frente: l = 24.

$$\begin{aligned}
 Z(7104+l) = & Z((7103)+1-1)+Z((7103)+1-24)-Z((7103)+1-25)+Z((7103)+1-168)... \\
 & -Z((7103)+1-169)-Z((7103)+1-192)+Z((7103)+1-193)+(-0.46)*Z((7103)+1-24)... \\
 & -(-0.46)*Z((7103)+1-25)-(-0.46)*Z((7103)+1-48)+(-0.46)*Z((7103)+1-49)... \\
 & -(-0.46)*Z((7103)+1-192)+(-0.46)*Z((7103)+1-193)+(-0.46)*Z((7103)+1-216)... \\
 & -(-0.46)*Z((7103)+1-217)+(-0.60)*2*Z((7103)+1-168)-(-0.60)*((7103)+1-169)... \\
 & -(-0.60)*Z((7103)+1-192)+(-0.60)*Z((7103)+1-193)-(-0.60)*Z((7103)+1-336)... \\
 & +(-0.60)*Z((7103)+1-337)+(-0.60)*Z((7103)+1-360)-(-0.60)*Z((7103)+1-361)... \\
 & -(-0.19)*Z((7103)+1-504)+(-0.19)*Z((7103)+1-505)+(-0.19)*Z((7103)+1-528)... \\
 & -(-0.19)*Z((7103)+1-529)-(-0.46)*(-0.60)*Z((7103)+1-192)... \\
 & +(-0.46)*(-0.60)*Z((7103)+1-193)+(-0.46)*(-0.60)*Z((7103)+1-216)... \\
 & -(-0.46)*(-0.60)*Z((7103)+1-217)+(-0.46)*(-0.60)*Z((7103)+1-360)... \\
 & -(-0.46)*(-0.60)*Z((7103)+1-361)-(-0.46)*(-0.60)*Z((7103)+1-384)... \\
 & +(-0.46)*(-0.60)*Z((7103)+1-385)-(-0.46)*(-0.19)*Z((7103)+1-360)... \\
 & +(-0.46)*(-0.19)*Z((7103)+1-361)+(-0.46)*(-0.19)*Z((7103)+1-384)... \\
 & -(-0.46)*(-0.19)*Z((7103)+1-385)+(-0.46)*(-0.19)*Z((7103)+1-528)... \\
 & -(-0.46)*(-0.19)*Z((7103)+1-529)-(-0.46)*(-0.19)*Z((7103)+1-552)...
 \end{aligned}$$

$$+(-0.46)*(-0.19)*Z((7103)+1-553)-(0.85)*a((7103)+1-24)...$$

$$-(0.83)*a((7103)+1-168)+(0.85)*(0.83)*a((7103)+1-192);$$

Para chegar a equação de previsão, basta tirar as expectâncias de cada termo. Lembra-se que o valor esperado do erro futuro é sempre nulo.

Como o banco de dados 'Z' representa a série com 3 diferenças: uma diferença simples, uma diferença sazonal 1 e uma diferença sazonal 2. Para facilitar a programação feita no MATLAB 6.1 e a avaliação do modelo, os dados *out-of-sample* passaram pelas mesmas diferenciações executadas na série *in-sample*. Os resultados desta modelagem e seus respectivos MAPE podem ser conferidos no capítulo V deste trabalho quando tanto os resultados dos modelos BJ quanto os da modelagem HW são melhor explicados.

### 3.6

#### Considerações finais

Este capítulo pretendeu apresentar a metodologia univariada por Box & Jenkins em sua formulação originalmente proposta e também em sua forma estendida a fim de adaptar a técnica às características específicas da série em estudo. Para os dados referentes à demanda energética horária do sudeste em 2003 foi preciso incorporar dois ciclos ao modelo. Porém, como anteriormente mencionado, conforme a necessidade, mais ciclos podem ser comportados.

Os principais aspectos da técnica BJ foram abordados de maneira sintetizada de forma a dar condições para o leitor ter uma perfeita compreensão da proposta aqui levantada. Para isso, a teoria alternativa foi exemplificada a partir de uma modelagem empírica cujos resultados da previsão podem ser vistos no capítulo 5.

O alicerce teórico fornecido nesta etapa do estudo é uma ferramenta que, certamente, tornará possível desenvolver modelos mais acurados para a representação dos mais diversos problemas e, dessa maneira, obter uma modelagem sensível para capturar as particularidades de cada série, seja na área financeira, econômica, energética, entre outras.