

Econometria III

Professor Thiago Morello

Nota de aula 2

(referências: Bueno, capítulos 2 e 3; Enders, cap.2)

(1) Processos estocásticos

Definição adaptada de Spanos (2019) e Bueno (2008):

Um processo estocástico é uma sequência ordenada de variáveis aleatórias. Geralmente, tal ordenação é função do tempo. Ou seja, para os fins desta disciplina, um processo estocástico é uma sequência de variáveis aleatórias ordenadas no tempo.

Um requisito técnico é o de que todas as variáveis processo estocástico sejam definidas no mesmo espaço de probabilidade – espaço de probabilidade é um conjunto de três elementos, sendo eles: (i) o espaço amostral ou conjunto de todos os eventos possíveis, (ii) a sigma-álgebra, ou uma categoria específica de subconjuntos do espaço amostral e (iii) a função distribuição de probabilidades.

Um processo estocástico geralmente tem dois componentes: (i) componente não-explicitamente aleatório, o qual especifica com uma fórmula, (ao menos parte da) dependência temporal, e (ii) componente explicitamente aleatório ou referente às perturbações. Este último, geralmente (mas não sempre), não é sujeito à dependência temporal. Um exemplo é o processo auto-regressivo de ordem 1 (AR(1)), dado por: $y_t = a_0 + a_1 y_{t-1} + e_t$, em que $a_0 + a_1 y_{t-1}$ é o componente determinista e e_t o aleatório.

(2) Estacionaridade

Por estacionaridade se entende, em linhas gerais, a conservação, ao longo do tempo, sem qualquer alteração, da função distribuição de probabilidade. Não sendo este o caso, a análise estatística se torna impossível, pois, por um lado, observa-se apenas um valor por instante de tempo, mas, por outro, múltiplos valores são necessários para estimar os parâmetros de uma função de distribuição de probabilidade. Segue, abaixo, uma definição de estacionaridade:

Um processo estocástico é fracamente estacionário (ou estacionário em termos de covariância) se:

$$(1) E[y_t] = E[y_{t-1}] = \mu$$

$$(2) V[y_t] = V[y_t] = \sigma^2$$

$$(3) \text{cov}[y_t, y_{t-j}] = \gamma_j$$

A terceira condição é a que mais explicitamente se refere ao grau de dependência temporal inerente a séries estacionárias: tal dependência, medida pela autocovariância, é função exclusivamente da distância temporal, ou defasagem, separando duas realizações da série. Ou seja, a autocovariância não varia ao longo do tempo.

É condição necessária para a estacionaridade fraca que as raízes características da equação a diferenças tenham todas módulo inferior à unidade.

A definição mais restritiva, e, que portanto, garante estacionaridade “forte”, ou estrita, é a de que a função de distribuição de probabilidade conjunta seja igual em quais dois instantes de tempo (não importa a distância entre eles). Trata-se da definição 2.4 do livro de Bueno.

Aproveitando a menção à covariância, cabe definir, assim como feito por Enders, a autocorrelação como $\text{corr}[y_t, y_{t-j}] = \gamma_j/\gamma_0$, em que $\gamma_0 = \text{cov}[y_t, y_t] = E[(y_t - E[y_t])(y_t - E[y_t])] = E[(y_t - E[y_t])^2] = V[y_t]$.

Um exemplo de um processo estocástico estacionário é o ruído branco, um processo básico, o qual serve de fundamento para a especificação de outros processos estocásticos. O ruído branco é puramente aleatório, sendo denotado como uma perturbação, e_t , cujas propriedades são:

1. Expectativa nula, $E[e_t] = 0$, $t = 1, 2, \dots$
2. Variância constante, $V[e_t] = E[(e_t - E[e_t])^2] = \sigma^2$
3. Autocovariância nula: $\text{cov}[e_t, e_{t-j}] = 0$, qualquer que sejam t e j

3 Processos estacionários

3.1 Objetivo

O objetivo da análise econométrica de séries de tempo, no que tange a séries unidimensionais, é descobrir qual processo estocástico teórico melhor reproduz uma série temporal observada. Tal descoberta se dá por meio de três ações, (i) a interpretação de gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP), (ii) testes de hipótese de (in)significância conjunta para defasagens superiores (teste de Ljung-Box) e (iii) minimização de estatísticas conhecidas como critérios de informação (de Akaike, Schwarz e Hannan-Quinn).

É importante assinalar que o procedimento que se acaba de descrever é empregado para o caso de séries estacionárias, a quais constituem o objeto desta nota de aula – séries não estacionárias serão tratadas na próxima nota de aula. O principal a assinalar é que as próximas seções desta nota de aula tratarão de séries estacionárias, apenas.

Esta seção tem como objeto a primeira ação. O objetivo é ensinar a identificar o número de defasagens autorregressivas e de média móvel necessárias para mais bem reproduzir uma série empírica. Para tanto, serão (i) definidas FAC e FACP, (ii) descrito o procedimento de computação da FAC e da FACP (manual e automático), (iii) exibidos gráficos da FAC e FACP para alguns processos estocásticos teóricos básicos e (iv) resumidas as características gerais dos gráficos de FAC e FACP de processos AR(p), MA(q) e ARMA(p,q).

3.1 Função de autocorrelação (FAC)

A autocorrelação populacional entre dois valores de uma série de tempo, o primeiro ocorrendo no instante t e o segundo no instante $t - j$ é:

$$\text{corr}(y_t, y_{t-j}) = \frac{\text{cov}(y_t, y_{t-j})}{dp(y_t)dp(y_{t-j})} = \frac{E[y_t - E[y_t]]E[y_{t-j} - E[y_{t-j}]]}{dp(y_t)dp(y_{t-j})}$$

Sendo y_t estacionária, é possível simplificar a expressão anterior, pois (i) expectativa e variâncias são constantes no tempo e (ii) a covariância é função exclusivamente da defasagem captada pelo termo “ j ”. Com isso, a expressão anterior passa a:

$$\begin{aligned} \text{corr}(y_t, y_{t-j}) &= \frac{E[y_t - E[y]]E[y_{t-j} - E[y]]}{dp(y)^2} = \frac{E[y_t - E[y]]E[y_{t-j} - E[y]]}{E[y - E[y]]E[y - E[y]]} = \frac{\gamma_j}{\gamma_0} \\ &= \rho_j. \end{aligned}$$

Ou seja, no numerador a expectativa interna é constante e, no denominador, há a variância da série. Com isso, retoma-se a notação sintética para autocovariância da seção anterior com a letra grega “gama”.

Já a autocorrelação amostral de ordem j para uma série estacionária é:

$$\hat{\rho}_j = \text{corr}(\widehat{y_t}, \widehat{y_{t-j}}) = \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})(y_{t-j} - \bar{y})}{\widehat{dp(y_t)^2}} = \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})(y_{t-j} - \bar{y})}{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2}$$

Esta estatística, é, pois, uma função de j , o tamanho da defasagem. É por isso que a sequência de valores da estatística para níveis de j progressivamente maiores, i.e., para $j = 1, 2, 3, \dots$, é chamada de função de autocorrelação (FAC) amostral.

3.2 Obtendo a função de autocorrelação (FAC)

4.2.1 Definição e premissas

Fórmulas para variância, autocovariâncias e autocorrelações, e, portanto, para a FACP podem ser obtidas a partir do sistema de equações de Yule-Walker, o que será feito nessa seção para alguns processos estocásticos básicos.

4.2.2 Processo AR(1)

4.2.2.1 Cálculo da expectativa do AR(1)

O AR(1) pode ser escrito, denotando-se a constante por “c” e o coeficiente por ϕ_1 , como:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \varepsilon_t \text{ é um ruído branco}$$

A solução de tal processo é:

$$y_t = a_0 \sum_{i=0}^{t-1} \phi_1^i + \phi_1^t y_0 + \sum_{i=0}^{t-1} \phi_1^i \varepsilon_{t-i}$$

Tomando o caso limite em que $t \rightarrow \infty$, ou seja, considerado a solução de “longo prazo”, tem-se:

$$y_t = c \sum_{i=0}^{\infty} \phi_1^i + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

Uma vez que $0 < \phi_1 < 1$, é condição para que o processo seja estacionário (algo necessário para que ele seja útil para análise econométrica).

Cabe notar que $\sum_{i=0}^{\infty} \phi_1^i$ é a soma de uma progressão geométrica infinita com razão $q = \phi_1$ e primeiro termo $a_1 = \phi_1^0 = 1$. A fórmula de tal soma é: $S = a_1/(1-q) = \frac{1}{1-\phi_1}$. Com isso:

$$y_t = c \frac{1}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}$$

Do que decorre:

$$E[y_t] = E\left[\frac{c}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}\right] = \frac{c}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i E[\varepsilon_{t-i}] = \frac{c}{1-\phi_1}$$

A última passagem decorrendo da nulidade da média de um ruído branco.

4.2.2.2 Equações de Yule-Walker para o AR(1)

Cabe retomar a definição do processo AR(1):

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \varepsilon_t \text{ é um ruído branco}$$

Da mesma maneira, pode-se especificar um processo AR(1) para uma transformação da série original, a transformação de interesse sendo a subtração da série de sua expectativa. O processo transformado, $y_t - E[y_t] = y_t - \mu = y_t^*$, pode ser especificado da seguinte forma:

$$y_t^* = \phi_1 y_{t-1}^* + \varepsilon_t(1)$$

(note a omissão da constante)

Agora o processo (1) será manipulado algebricamente para com isso obter as equações de Yule-Walker (ver Enders, 2015, p.60):

Multiplique a equação (1) por y_{t-j}^* , para $j = 0, 1, 2, \dots$

$$\begin{aligned} y_t^* y_t^* &= c y_t^* + \phi_1 y_{t-1}^* y_t^* + \varepsilon_t y_t^* \\ y_t^* y_{t-1}^* &= c y_{t-1}^* + \phi_1 y_{t-1}^* y_{t-1}^* + \varepsilon_t y_{t-1}^* \\ y_t^* y_{t-2}^* &= c y_{t-2}^* + \phi_1 y_{t-1}^* y_{t-2}^* + \varepsilon_t y_{t-2}^* \\ &\dots \\ y_t^* y_{t-j}^* &= c y_{t-j}^* + \phi_1 y_{t-1}^* y_{t-j}^* + \varepsilon_t y_{t-j}^* \end{aligned}$$

Tome as expectativas dos dois lados do sistema anterior, obtendo o sistema abaixo o qual será referido como “sistema A”:

$$\begin{aligned} E[y_t^* y_t^*] &= c E[y_t^*] + \phi_1 E[y_{t-1}^* y_t^*] + E[\varepsilon_t y_t^*] \\ E[y_t^* y_{t-1}^*] &= c E[y_{t-1}^*] + \phi_1 E[y_{t-1}^* y_{t-1}^*] + E[\varepsilon_t y_{t-1}^*] \\ E[y_t^* y_{t-2}^*] &= c E[y_{t-2}^*] + \phi_1 E[y_{t-1}^* y_{t-2}^*] + E[\varepsilon_t y_{t-2}^*] \\ &\dots \\ E[y_t^* y_{t-j}^*] &= c E[y_{t-j}^*] + \phi_1 E[y_{t-1}^* y_{t-j}^*] + E[\varepsilon_t y_{t-j}^*] \end{aligned}$$

É possível simplificar o sistema anterior com base nos fatos a seguir listados.

Fato 1: $E[y_t^* y_t^*] = E[y_t^{*2}] = E[(y_t - E[y_t])^2] = V[y_t] = \gamma_0$

Fato 2: *para $s > 1$* , $E[y_t^* y_{t-s}^*] = E[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])] = cov[y_t, y_{t-s}] = \gamma_s$

Fato 3: $E[y_t^*] = E[(y_t - E[y_t])] = E[y_t] - E[y_t] = 0$

Fato 4:

$$\begin{aligned} E[\varepsilon_t y_t^*] &= E[\varepsilon_t (y_t - \mu)] = \{1\} = E \left[\varepsilon_t \left(c \sum_{i=0}^{\infty} \phi_1^i + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} - \mu \right) \right] = \{2\} \\ &= E \left[\varepsilon_t \left(\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \right) \right] = E \left[\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \varepsilon_t \right] = E \left[\phi_1^0 \varepsilon_t \varepsilon_t + \sum_{i=1}^{\infty} \phi_1^i \varepsilon_{t-i} \varepsilon_t \right] \\ &= E \left[\phi_1^0 E[\varepsilon_t^2] + \sum_{i=1}^{\infty} \phi_1^i E[\varepsilon_{t-i} \varepsilon_t] \right] = E \left[\sigma^2 + \sum_{i=1}^{\infty} \phi_1^i 0 \right] = \sigma^2 \end{aligned}$$

{1} Foi usada a solução do processo AR(1);

{2} Foi usado o fato de que a média do AR(1) é $c \sum_{i=0}^{\infty} \phi_1^i = \frac{c}{1-\phi_1}$.

Ou seja:

$$E[\varepsilon_t y_t^*] = \sigma^2$$

Fato 5: para $s > 0$

$$\begin{aligned} E[\varepsilon_t y_{t-s}^*] &= E[\varepsilon_t (y_{t-s} - \mu)] = \{1\} = E\left[\varepsilon_t \left(\sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-s-i}\right)\right] = \sum_{i=0}^{\infty} \phi_1^i E[\varepsilon_{t-s-i} \varepsilon_t] \\ &= \sum_{i=0}^{\infty} \phi_1^i 0 = 0 \end{aligned}$$

{1} Passagens análogas às utilizadas no fato 4.

Ou seja:

$$E[\varepsilon_t y_{t-s}^*] = 0, \text{ para } s > 1$$

Incorporando os fatos 1 a 5 no sistema A, chega-se a:

$$\gamma_0 = \phi_1 \gamma_1 + \sigma^2$$

$$\gamma_1 = \phi_1 \gamma_0$$

$$\gamma_2 = \phi_1 \gamma_1$$

...

$$\gamma_j = \phi_1 \gamma_{j-1}$$

Com base nas primeiras duas equações, calcula-se γ_0 , a variância de y_t :

$$\gamma_0 = \phi_1 \phi_1 \gamma_0 + \sigma^2 = \phi_1^2 \gamma_0 + \sigma^2 \rightarrow \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}$$

Com base nas demais equações calculam-se as covariâncias de ordem $j = 1, 2, \dots$, da seguinte forma:

$$\phi_j = \phi_1^{1+j} \frac{\sigma^2}{1 - \phi_1^2}$$

Já as autocorrelações podem ser calculadas dividindo-se todas as equações do sistema, exceto pela primeira, pela variância γ_0 (notar que dividir pela variância é dividir pelo produto do desvio-padrão por ele mesmo e, de fato, como tal desvio é o mesmo para todas as defasagens da variável, tem-se o mesmo denominador da definição geral da correlação):

$$\frac{\gamma_1}{\gamma_0} = \phi_1 \frac{\gamma_0}{\gamma_0}$$

$$\frac{\gamma_2}{\gamma_0} = \phi_1 \frac{\gamma_1}{\gamma_0}$$

...

$$\frac{\gamma_j}{\gamma_0} = \phi_1 \frac{\gamma_{j-1}}{\gamma_0}$$

Ou, alternativamente:

$$\rho_1 = \phi_1$$

$$\rho_2 = \phi_1 \rho_1$$

...

$$\rho_j = \phi_1 \rho_{j-1}$$

Com isso, a função de autocorrelação é $\rho_j = \phi_1^j, j = 1, 2, \dots$

Resumo

Uma síntese do obtido com o sistema de Yule-Walker é, se $y_t \sim \text{AR}(1)$, então:

$$E[y_t] = \mu, V[y_t] = \frac{\sigma^2}{1-\phi_1^2}, \text{cov}(y_t, y_{t-j}) = \phi_1^j, j = 1, 2, \dots$$

É notório, pois, que expectativa, variância e covariâncias são invariantes no tempo, e, conclusivamente, o processo $\text{AR}(1)$ é estacionário. Cabe também assinalar que isso apenas é verdade pois o termo de perturbação é um ruído branco.

4.2.3 Processo AR(p)

Trata-se do processo:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

É possível demonstrar que a solução deste processo é dada por (Enders, 2015, equação 2.20):

$$y_t = \frac{c}{1 - \sum_{j=1}^p \phi_j} + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

Em que $c_1, c_2, \dots$, são coeficientes indeterminados.

A expectativa do $\text{AR}(p)$ é:

$$E[y_t] = E\left[\frac{c}{1 - \sum_{j=1}^p \phi_j} + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}\right] = \frac{c}{1 - \sum_{j=1}^p \phi_j} + \sum_{i=0}^{\infty} c_i E[\varepsilon_{t-i}] = \frac{c}{1 - \sum_{j=1}^p \phi_j}$$

Com base neste e em outros resultados, é possível utilizar a abordagem de Yule-Walker para obter a variância e as correlações do processo AR(p). Este exercício não será feito nesta disciplina. Basta observar que a função de autocorrelação terá o seguinte formato (Bueno, 2008, p.41):

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, j = 1, 2, \dots$$

O principal é que tanto expectativa, como variância e covariâncias são invariantes no tempo, e, consequentemente o processo AR(p) é estacionário.

4.2.4 Processo MA(1)

$$E[y_t] = E[\mu + \theta \varepsilon_{t-1} + \varepsilon_t] = \mu$$

Aplicando o procedimento de Yule-Walker, o primeiro passo é transformar o processo subtraindo a série da própria expectativa:

$$y_t^* = y_t - \mu = \theta \varepsilon_{t-1} + \varepsilon_t$$

No segundo passo, multiplica-se a equação do processo transformado por y_{t-j} , $j = 1, 2, 3, \dots$

$$y_t^* y_t^* = \theta \varepsilon_{t-1} y_t^* + \varepsilon_t y_t^*$$

$$y_t^* y_{t-1}^* = \theta \varepsilon_{t-1} y_{t-1}^* + \varepsilon_t y_{t-1}^*$$

$$y_t^* y_{t-2}^* = \theta \varepsilon_{t-1} y_{t-2}^* + \varepsilon_t y_{t-2}^*$$

$$y_t^* y_{t-3}^* = \theta \varepsilon_{t-1} y_{t-3}^* + \varepsilon_t y_{t-3}^*$$

...

$$y_t^* y_{t-j}^* = \theta \varepsilon_{t-1} y_{t-j}^* + \varepsilon_t y_{t-j}^*$$

Tomando as expectativas:

$$E[y_t^* y_t^*] = \theta E[\varepsilon_{t-1} y_t^*] + E[\varepsilon_t y_t^*]$$

$$E[y_t^* y_{t-1}^*] = \theta E[\varepsilon_{t-1} y_{t-1}^*] + E[\varepsilon_t y_{t-1}^*]$$

$$E[y_t^* y_{t-2}^*] = \theta E[\varepsilon_{t-1} y_{t-2}^*] + E[\varepsilon_t y_{t-2}^*]$$

$$E[y_t^* y_{t-3}^*] = \theta E[\varepsilon_{t-1} y_{t-3}^*] + E[\varepsilon_t y_{t-3}^*]$$

...

$$E[y_t^* y_{t-j}^*] = \theta E[\varepsilon_{t-1} y_{t-j}^*] + E[\varepsilon_t y_{t-j}^*]$$

As expectativas do lado direito têm de ser simplificadas com base na fórmula do processo MA(1).

(A) Primeiras expectativas à direita ($\theta E[\varepsilon_{t-1}y_t^*], E[\varepsilon_{t-1}y_{t-1}^*], \theta E[\varepsilon_{t-1}y_{t-2}^*], \dots, \theta E[\varepsilon_{t-1}y_{t-j}^*]$)

$$(A.1) \quad \text{Expectativa 1: } \theta E[\varepsilon_{t-1}y_t^*] = \theta E[\varepsilon_{t-1}(\theta\varepsilon_{t-1} + \varepsilon_t)] = \theta^2 E[\varepsilon_{t-1}^2] + \theta E[\varepsilon_t \varepsilon_{t-1}].$$

Neste ponto recorda-se a natureza de ruído branco da perturbação, ou seja, os fatos de que $E[\varepsilon_t] = 0$, $\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = E[\varepsilon_t \varepsilon_{t-j}] = 0$, $j = 1, 2, \dots$, $V[\varepsilon_t] = E[\varepsilon_t^2] = \sigma^2$, $t = 1, 2, \dots$. Do que decorre: $E[\varepsilon_{t-1}^2] = \sigma^2$, $E[\varepsilon_{t-1}\varepsilon_t] = 0$. E, pois:

$$\theta E[\varepsilon_{t-1}y_t^*] = \theta^2 E[\varepsilon_{t-1}^2] + \theta E[\varepsilon_t \varepsilon_{t-1}] = \theta^2 \sigma^2 + \theta 0 \leftrightarrow \theta E[\varepsilon_{t-1}y_t^*] = \theta^2 \sigma^2$$

$$(A.2) \quad \text{Expectativa 2: } \theta E[\varepsilon_{t-1}y_{t-1}^*] = \theta E[\varepsilon_{t-1}(\theta\varepsilon_{t-2} + \varepsilon_{t-1})] = E[\theta^2 \varepsilon_{t-1}\varepsilon_{t-2} + \theta \varepsilon_{t-1}^2] = \theta^2 E[\varepsilon_{t-1}\varepsilon_{t-2}] + \theta E[\varepsilon_{t-1}^2] = \theta^2 0 + \theta \sigma^2 \leftrightarrow \theta E[\varepsilon_{t-1}y_{t-1}^*] = \theta \sigma^2.$$

$$(A.3) \quad \text{Expectativa 3: } \theta E[\varepsilon_{t-1}y_{t-2}^*] = \theta E[\varepsilon_{t-1}(\theta\varepsilon_{t-3} + \varepsilon_{t-2})] = E[\theta^2 \varepsilon_{t-1}\varepsilon_{t-3} + \theta \varepsilon_{t-1}\varepsilon_{t-2}] = \theta^2 E[\varepsilon_{t-1}\varepsilon_{t-3}] + \theta E[\varepsilon_{t-1}\varepsilon_{t-2}] = \theta^2 0 + \theta 0 = 0 \leftrightarrow \theta E[\varepsilon_{t-1}y_{t-2}^*] = 0. \text{ Essa nulidade se estende para todas as demais expectativas, pois nelas o índice do termo de perturbação, } t-1, \text{ é sempre maior em mais do que uma unidade do que o índice da série transformada (e com isso tem-se expectativas de produtos de perturbações referentes à instantes distintos).}$$

(B) Segundas expectativas à direita ($E[\varepsilon_t y_t^*], E[\varepsilon_t y_{t-1}^*], E[\varepsilon_t y_{t-2}^*], \dots$)

$$(B.1) \quad \text{Expectativa 1: } E[\varepsilon_t y_t^*] = E[\varepsilon_t(\theta\varepsilon_{t-1} + \varepsilon_t)] = \theta E[\varepsilon_t \varepsilon_{t-1}] + E[\varepsilon_t^2] = \sigma^2;$$

$$(B.2) \quad \text{Expectativa 2: } E[\varepsilon_t y_{t-1}^*] = E[\varepsilon_t(\theta\varepsilon_{t-2} + \varepsilon_{t-1})] = \theta E[\varepsilon_t \varepsilon_{t-2}] + E[\varepsilon_t \varepsilon_{t-1}] = 0. \text{ Tal nulidade se estende para as demais expectativas, nas quais os índices temporais da perturbação e da série diferem em magnitude ainda maior.}$$

Incorporando os resultados obtidos para os dois conjuntos de expectativas do lado direito da equação, tem-se:

$$\gamma_0 = \theta^2 \sigma^2 + \sigma^2 \leftrightarrow \gamma_0 = (1 + \theta^2) \sigma^2$$

$$\gamma_1 = \theta \sigma^2$$

$$\gamma_2 = 0$$

$$\gamma_3 = 0$$

...

$$\gamma_j = 0$$

Dividindo o sistema por γ_0 para obter as autocorrelações:

$$\begin{aligned}
1 &= 1 \\
\rho_1 &= \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1 + \theta^2} \\
\rho_2 &= \frac{\gamma_2}{\gamma_0} = 0 \\
\rho_3 &= 0 \\
&\dots \\
\rho_j &= 0
\end{aligned}$$

Ou seja, a FAC é positiva apenas para a defasagem $j = 1$, reduzindo-se imediatamente à zero a partir de $j = 2$.

4.2.5 Processo MA(1): estacionaridade e invertibilidade

Os resultados obtidos na sub-seção anterior podem ser resumidos da seguinte maneira. Sendo $y_t \sim \text{MA}(1)$:

$$E[y_t] = \mu ; V[y_t] = \sigma^2; \text{cov}(y_t, y_{t-j}) = 0, \text{ para } j = 1, \dots, J.$$

Ou seja, expectativa, variância e covariância são invariantes no tempo, e, portanto, o processo MA(1) é estacionário. Mas, deve-se assinalar, estacionariedade não é condição suficiente para que seja possível utilizar o processo MA(1) como um instrumento para reproduzir uma série temporal. Ao menos não por meio da metodologia que será utilizada nesta disciplina, a qual foi proposta por Box e Jenkins, e será apresentada na próxima seção (quanto a isso ver Enders, 2015, p.77). A condição de invertibilidade também tem de ser atendida. Esta condição exige que seja possível escrever um processo MA(q) como um processo AR(∞). Uma maneira de verifica-la é escrever o processo MA(1) com base no operador de defasagem.

[Operador defasagem] Antes de continuar é preciso explicar em que consiste tal operador. Ele, quando aplicado a uma série de tempo, “atrasa” a série em um instante de tempo. O valor que ocorre em $t = 1$ passa a ocorrer em $t = 2$, o valor que ocorre em $t = 2$ passa a ocorrer em $t = 3$ (passa a demorar mais para que um dado valor ocorra). Formalmente o operador é denotado por “L” de “lag”, e é tal que $Ly_t = y_{t-1}$, $L^2y_t = y_{t-2}$ e $L^jy_t = y_{t-j}$, $j = 1, 2, \dots$. Tal operador simplifica a notação de algumas operações, como, por exemplo, a primeira diferença da variável, i.e., $y_t - y_{t-1} = y_t - Ly_t = (1-L)y_t$, e a própria fórmula do processo MA(1), como segue.

$Y_t \sim \text{MA}(1) \rightarrow y_t = c + \theta e_{t-1} + e_t$, $e_t \sim \text{RB}$. Utilizando o operador defasagem:

$$y_t - c = \theta L e_t + e_t \rightarrow y_t - c = e_t(1 + \theta L) \rightarrow (y_t - c)(1 + \theta L)^{-1} = e_t(1)$$

Agora retomando a apresentação da propriedade de invertibilidade, tal condição surge exatamente na equação (1). Se $|\theta| < 1$, então $1 + \theta L$ é a soma de uma progressão geométrica infinita convergente. Basta recordar que tal soma, denotando-se o primeiro

termo da progressão por “ a_1 ” e a razão por “ q ”, é igual a $a_1/(1-q)$. Como as progressões que vimos até agora todas têm $a_1 = 1$, então, $1 - q = 1 + \theta L$, e pois, $q = -\theta L$. A progressão portanto é $1, -\theta L, \theta^2 L^2, -\theta^3 L^3, \theta^4 L^4, \dots$

Com isso, a expressão (1) pode ser escrita como:

$$(y_t - c) (1 + \theta L)^{-1} = e_t \rightarrow (y_t - c) (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \theta^4 L^4 + \dots) = e_t$$

Usando a propriedade distributiva nos operadores de defasagem, os quais por sua vez ficam aplicados à y , tem-se:

$$y_t^* - \theta y_{t-1}^* + \theta^2 y_{t-2}^* - \theta^3 y_{t-3}^* + \theta^4 y_{t-4}^* + \dots = e_t$$

Ou seja, um processo $AR(\infty)$.

Essa inversão é útil para o cálculo da FAC e da FACP de processos $MA(1)$. É também possível inverter processos $MA(q)$, $q > 1$, mas a condição necessária para tanto é a de que as raízes do polinômio característico do $MA(q)$ estejam todas fora do círculo unitário do plano complexo.

4.2.4 Processo $MA(q)$

No apêndice, o sistema de Yule-Walker do processo $MA(q)$ é montado. O principal resultado a apontar é que expectativa, variância e covariância são constantes no tempo, e há, pois, estacionariedade.

4.2.5 Processo $ARMA(1,1)$

Trata-se do processo:

$$y_t = c + \phi y_{t-1} + \theta e_{t-1} + e_t, \text{ em que } e_t \sim RB(0, \sigma^2).$$

Escrevendo como desvios da média, multiplicando por $y_t, y_{t-1}, y_{t-2}, \dots$, e calculando as expectativas:

$$\begin{aligned} E[y_t^* y_t^*] &= \phi E[y_{t-1}^* y_t^*] + \theta E[e_{t-1}^* y_t^*] + E[e_t^* y_t^*] \\ E[y_t^* y_{t-1}^*] &= \phi E[y_{t-1}^* y_{t-1}^*] + \theta E[e_{t-1}^* y_{t-1}^*] + E[e_t^* y_{t-1}^*] \\ E[y_t^* y_{t-2}^*] &= \phi E[y_{t-1}^* y_{t-2}^*] + \theta E[e_{t-1}^* y_{t-2}^*] + E[e_t^* y_{t-2}^*] \\ &\dots \\ E[y_t^* y_{t-j}^*] &= \phi E[y_{t-1}^* y_{t-j}^*] + \theta E[e_{t-1}^* y_{t-j}^*] + E[e_t^* y_{t-j}^*] \quad (A) \end{aligned}$$

Os termos desconhecidos requerendo derivação algébrica são, primeiramente:

$$E[e_{t-1}^* y_{t-j}^*] = E[e_{t-1}^* (\phi y_{t-j-1}^* + \theta e_{t-j-1} + e_{t-j})] = \phi E[e_{t-1}^* y_{t-j-1}^*] + \theta E[e_{t-1} e_{t-j-1}] + E[e_{t-1} e_{t-j}]$$

Cabe considerar três casos, o primeiro, $j = 0$:

$$\begin{aligned} E[e_{t-1}^* y_t^*] &= \phi E[e_{t-1}^* y_{t-1}^*] + \theta E[e_{t-1} e_{t-1}] + E[e_{t-1} e_t] = \phi E[e_{t-1}^* (\phi y_{t-2}^* + \theta e_{t-2} + e_{t-1})] + \theta \sigma^2 + 0 \\ &= \phi \sigma^2 + \theta \sigma^2 = (\phi + \theta) \sigma^2. \end{aligned}$$

O segundo caso é $j = 1$:

$$E[e_{t-1} y_{t-1}^*] = \phi E[e_{t-1} y_{t-2}^*] + \theta E[e_{t-1} e_{t-2}] + E[e_{t-1} e_{t-1}] = \sigma^2.$$

O terceiro caso é $j > 1$:

$$E[e_{t-1} y_{t-j}^*] = \phi E[e_{t-1} y_{t-j-1}^*] + \theta E[e_{t-1} e_{t-j-1}] + E[e_{t-1} e_{t-j}] = 0$$

Já o segundo termo desconhecido é:

$$E[e_t y_t^*] = E[e_t (\phi y_{t-1}^* + \theta e_{t-1} + e_t)] = \phi E[e_t y_{t-1}^*] + \theta E[e_t e_{t-1}] + E[e_t^2] = 0 + 0 + \sigma^2$$

E o terceiro termo desconhecido:

$E[e_t y_{t-1}^*]$, $j > 1$, tal termo sendo nulo pois:

$$E[e_t (\phi y_{t-j-1}^* + \theta e_{t-j-1} + e_{t-j})] = \phi E[e_t y_{t-j-1}^*] + \theta E[e_t e_{t-j-1}] + E[e_t e_{t-j}] = 0.$$

Com isso o sistema (A) passa a:

$$E[y_t^* y_t^*] = \phi E[y_{t-1}^* y_t^*] + \theta(\phi + \theta) \sigma^2 + \sigma^2$$

$$E[y_t^* y_{t-1}^*] = \phi E[y_{t-1}^* y_{t-1}^*] + \theta \sigma^2$$

$$E[y_t^* y_{t-2}^*] = \phi E[y_{t-1}^* y_{t-2}^*]$$

$$E[y_t^* y_{t-j}^*] = \phi E[y_{t-1}^* y_{t-j}^*]$$

...

$$E[y_t^* y_{t-j}^*] = \phi E[y_{t-1}^* y_{t-j}^*] + \theta E[e_{t-1} y_{t-j}^*] + E[e_t y_{t-j}^*]$$

Utilizando a notação sintética:

$$\gamma_0 = \phi \gamma_1 + \theta(\phi + \theta) \sigma^2 + \sigma^2$$

$$\gamma_1 = \phi \gamma_0 + \theta \sigma^2$$

$$\gamma_2 = \phi \gamma_1$$

...

$$\gamma_j = \phi \gamma_{j-1}$$

Substituindo a segunda equação na primeira, obtém-se:

$$\gamma_0 = \phi(\phi \gamma_0 + \theta \sigma^2) + \theta(\phi + \theta) \sigma^2 + \sigma^2$$

$$\gamma_0 = \phi^2 \gamma_0 + \sigma^2[\phi \theta + \theta(\phi + \theta) + 1]$$

$$\gamma_0 = \sigma^2[\phi \theta + \theta(\phi + \theta) + 1]/(1-\phi^2)$$

$$\gamma_0 = \sigma^2[2\phi\theta + \theta^2 + 1]/(1-\phi^2)$$

Esta sendo a variância do processo.

Utilizando este fato obtém-se γ_1 :

$$\gamma_1 = \phi \sigma^2[2\phi\theta + \theta^2 + 1]/(1-\phi^2) + \theta \sigma^2$$

$$\gamma_1 = \{\phi [2\phi\theta + \theta^2 + 1]/(1-\phi^2) + \theta\} \sigma^2$$

Denotando o termo $\{\phi [2\phi\theta + \theta^2 + 1]/(1-\phi^2) + \theta\}$ por “B”, tem-se:

$$\gamma_1 = B\sigma^2$$

E utilizando-se as outras equações, obtém-se as autocovariâncias de ordem superior:

$$\gamma_2 = \phi B\sigma^2$$

$$\gamma_3 = \phi \gamma_2 = \phi^2 B\sigma^2$$

$$\gamma_j = \phi^j B\sigma^2$$

Com isso as autocorrelações são $\rho_j = \phi^j B$.

Do que se conclui que expectativa, variância e autocorrelações são invariantes no tempo, e, portanto, o processo é estacionário. Mas isso depende, conforme afirma Bueno (2008, p.31) das raízes características dos processos AR e MA estarem fora do círculo unitário do plano complexo.

3.3 Função de autocorrelação parcial (FACP)

A FAC, ao calcular a correlação entre duas observações separadas por j instantes de tempo, p.ex., y_1 e y_3 , acaba captando também, potencialmente, a influência de observações intermediárias, no caso do exemplo, y_2 . E isso pois y_2 é correlacionada com y_1 ou com y_3 (ou com ambas) nos dados. As correlações captadas pela FAC, não são, portanto, puras. Ou seja, não refletem apenas a relação existente entre y_1 e y_3 . Ou, de modo geral, e para uma defasagem de j instantes de tempo, não refletem apenas duas observações separadas por j instantes de tempo.

É exatamente para obter correlações puras que se utiliza a função de autocorrelação parcial (FACP). Esta é composta por correlações calculadas de maneira a eliminar a influência de observações intermediárias. Ou seja, a correlação entre y_t e y_{t-j} é obtida após serem eliminadas as influências das defasagens de ordem inferior à j, ou seja, das defasagens iguais a $y_{t-1}, y_{t-2}, \dots, y_{t-j-1}$.

O processo de purificação inerente à geração da FACP é o mesmo implícito à regressão múltipla. Na regressão múltipla, deseja-se obter o coeficiente medindo a influência, sobre a variável dependente, de uma variável independente específica, x_k . O objetivo é que esta influência seja pura, i.e., seja livre da influência das outras variáveis independentes, x_s , $s \neq k$. Tal objetivo é alcançado incluindo-se todas as outras variáveis independentes na

regressão múltipla, como variáveis de controle, juntamente com a variável independente principal. Ao fazer isso, o coeficiente da última será puro.

O mesmo procedimento é adotado para gerar a FACP, mas, porém, de maneira gradual. Mais precisamente, trata-se de regredir y_t contra y_{t-j} , incluindo todas as defasagens intermediárias, e, portanto, $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-j-1}$. As defasagens intermediárias são, portanto, variáveis de controle, cuja influência é, pois, captada separadamente da influência y_{t-j} , esta última contida no coeficiente de y_{t-j} . Mas há um detalhe importante: deve-se estimar uma regressão linear para cada defasagem. Isso pode ser enunciado de maneira mais clara por meio da descrição do procedimento completo, tal como segue:

1. Obter a correlação pura entre y_t e y_{t-1} como estimativa pontual para $\phi_{1,1}$, obtida esta a partir da estimação da seguinte regressão linear: $y_t = \phi_{1,1}y_{t-1} + e_{1t}$;
2. Obter a correlação pura entre y_t e y_{t-2} como estimativa pontual para $\phi_{2,2}$, obtida esta a partir da estimação da seguinte regressão linear: $y_t = \phi_{2,1}y_{t-1} + \phi_{2,2}y_{t-2} + e_{2t}$;
3. Repetir o procedimento para $\phi_{3,3}$, estimado a partir de $y_t = \phi_{3,1}y_{t-1} + \phi_{3,2}y_{t-2} + \phi_{3,3}y_{t-3} + e_{3t}$
4. E assim por diante até a última defasagem desejada, esta denotada como J .

Organizando os resultados em uma tabela com duas colunas, a primeira o tamanho da defasagem, j , e a segunda a estimativa pontual $\hat{\phi}_{j,j}$, tem-se a FACP de ordem J . A qual também pode ser representada graficamente no plano $j \times \hat{\phi}_{j,j}$.

Uma propriedade importante da FACP é, conforme assinalado por Bueno (2008), a nulidade da mesma, ou seja, da estimativa pontual $\hat{\phi}_{j,j}$ para uma defasagem, j , superior à ordem do processo estocástico autorregressivo subjacente aos dados. Ou seja, um processo $AR(p)$ tem FACP não nula para uma defasagem inferior ou igual a p , apresentando FACP nula para defasagens superiores a p . Esta nulidade, porém, é fato apenas para processos autorregressivos. Já os outros dois processos apresentam padrões diferentes para a FACP. O processo $MA(q)$ exibe um decaimento exponencial contínuo, não relacionado à ordem q . Já o processo $ARMA(p,q)$ apresenta decaimento exponencial a partir da p -ésima defasagem – e isso pois o componente autorregressivo (AR) se torna nulo a partir da p -ésima defasagem, sobrando apenas o padrão do componente MA.

4 Metodologia de Box-Jenkins: uma implementação possível

4.1 Teorema de Wold

Este teorema, conforme assinala Bueno (2008), é fundamental, pois garante com que a metodologia que será apresentada a seguir seja válida para ampla gama de séries de tempo. Uma versão simplificada e suficiente para esta disciplina do enunciado do teorema de Wold é:

Todo processo estocástico, linear ou não, pode ser adequadamente aproximado por um processo estocástico linear.

Como a metodologia de Box-Jenkins toma por base processos estocásticos lineares, o teorema de Wold garante a aplicabilidade de tal metodologia para todas as séries de tempo empíricas possíveis.

Mas é preciso ressaltar que tal teorema não garante uma performance satisfatória estatística desejável para o processo linear, seja em termos de reprodução da estrutura de autocorrelação (parcial) da mesma, e, menos ainda, em termos de acurácia de previsão de valores futuros da série. O teorema apenas garante que a metodologia de Box Jenkins não tende, caso implementada rigorosamente, a cometer erro sistematicamente grande no que tange à reprodução da série de tempo original, dentro da amostra observada.

4.2 Objetivos

Neste momento é preciso lembrar o objetivo da análise: reproduzir uma série temporal empírica a partir de um processo estocástico teórico. No estágio atual da apresentação do conteúdo da disciplina, estamos considerando um subconjunto específico de processos estocásticos teóricos, a assim-chamada família ARMA, ou seja, trata-se de todas as combinações possíveis de processos autoregressivos (AR) e de processos de médias móveis (MA). Essas combinações incluem processos estocásticos contendo apenas um tipo de componente, AR ou MA, o que se denota, respectivamente, como $ARMA(p,0)$ e $MA(0,q)$.

Portanto, o objetivo é: reproduzir uma série temporal empírica a partir de um processo estocástico $ARMA(p,q)$.

Este objetivo suscita a seguinte pergunta: mas qual processo $ARMA(p,q)$ exatamente? Ou seja, quais são as ordens “p” e “q”?

A resposta está exatamente nas funções de autocorrelação (FAC) e de autocorrelação parcial (FACP). É possível, por meio de inspeção gráfica, detectar, aproximadamente, as ordens “p” e “q” que melhor reproduzem a série de tempo empírica analisada.

De fato, esta detecção aproximada é o primeiro estágio do procedimento proposto por Box e Jenkins em obra originalmente publicada em 1976, o qual será chamado de “metodologia de Box-Jenkins” doravante. Tal procedimento tem quatro estágios: (i) identificação, (ii) estimação, (iii) diagnóstico, (iv) previsão (Gujarati, 2000, seção 22.3, Enders (2015, capítulo 2, seção 8). Todos esses estágios se referem ao processo ARMA, ou seja, trata-se de:

- (i) Identificar especificações candidatas do processo ARMA que melhor reproduzem a série de tempo;
- (ii) Estimar as especificações candidatas do ARMA;
- (iii) Diagnosticar as especificações candidatas estimadas, selecionando a melhor delas;
- (iv) Utilizar a especificação ARMA selecionada para previsão.

A rigor, esta é a primeira versão da metodologia de Box-Jenkins que será apresentada nesta disciplina. E isso pois é possível, e algumas vezes adequado, incluir a etapa de

previsão como parte do diagnóstico. A segunda versão será objeto da próxima nota de aula, a qual tratará de previsão.

Há um desafio na implementação prática, ou computacional, da metodologia de Box-Jenkins. Ele é referente ao estágio de identificação. Trata-se do desafio de elaborar uma lista suficientemente, mas não excessivamente, completa de especificações candidatas do ARMA. O cerne deste desafio está em um trade-off importante, entre custo e benefício da otimização. Uma lista muito grande representa um custo alto, pois todas as demais etapas terão de ser aplicadas a ela, ou seja, um grande número de especificações terá de ser estimado e comparado no estágio de diagnóstico. Uma lista pequena representa um benefício pequeno. Para entender é útil relembrar do conceito de otimização com restrições, utilizado em microeconomia. Quanto menor a lista, mais restrições o analista impõe, artificialmente, ao conjunto de especificações possíveis, e, portanto, maior é a probabilidade de que “a melhor especificação” seja, na verdade, um ótimo local e não global. Acaba-se, pois, ignorando especificações que seriam ainda melhores.

4.2 Um possível algoritmo inicial

Há diversas maneiras de enfrentar o desafio de implementação da metodologia de Box-Jenkins, equilibrando o trade-off intrínseco. Será aqui proposto um algoritmo que representa apenas um dos diversos algoritmos possíveis. Tal algoritmo procura evitar uma lista excessivamente grande de especificações candidatas. Mas, porém, ao custo de eventualmente omitir especificações com boa performance (bom ajuste aos dados). O algoritmo proposto será referido como “algoritmo 1”, e segue abaixo. Ao descrevê-lo utiliza-se o termo “correlograma” para referir-se aos gráficos da FAC e FACP.

1. Identificação: dois critérios são sugeridos:

- a. Inspeccionar os gráficos da FAC e FACP referentes à série de tempo, i.e., à variável dependente (Y), procurando observar sinais que apontem para ordens “p” e “q” específicas. Para isso, usar a seguinte tabela, reproduzida de Bueno (2008, p.43), a qual é uma versão simplificada das tabelas 2.1 de Enders (2015) e 6.1 de Box, Jenkins e Reinsel (2008):

Especificação	FAC	FACP
$AR(p) = ARMA(p,0)$	Decaimento exponencial	Truncada na defasagem p
$MA(q) = ARMA(0,q)$	Truncada na defasagem q	Decaimento exponencial
$ARMA(p,q)$	Decaimento exponencial a partir da defasagem q + 1	Decaimento exponencial a partir da defasagem p + 1

- b. Ordem máxima de significância: trata-se da maior defasagem para a qual há autocorrelação estatisticamente significativa nos gráficos da FAC e da FACP. Este nível é, adicionalmente à tabela anterior, também um indicador das ordens “p” e “q”, da seguinte forma: (i) uma ordem máxima $q_{\max}$ na FAC, ou

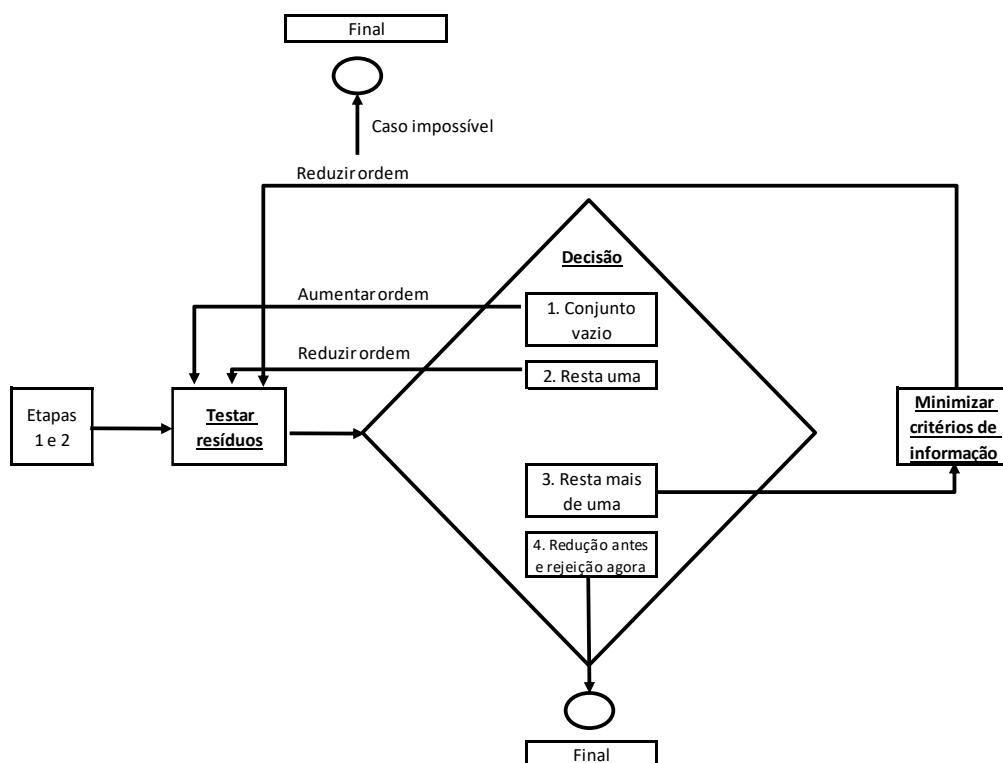
- seja, truncamento a partir de $q_{\max}+1$, sugere um processo $MA(q_{\max}+1)$ e, (ii) uma ordem máxima $p_{\max}$ na FACP, i.e., truncamento a partir de $p_{\max}+1$, sugere um processo $AR(p_{\max}+1)$. É recomendável, além de especificações AR e MA isoladas, também tomar por base a especificação $ARMA(p_{\max}+1, q_{\max}+1)$;
- c. Também é útil aplicar um teste de significância conjunta às autocorrelações referente às defasagens de ordem um até ordem j^* . Para isso utiliza-se a estatística de Ljung-Box:
 - i. A hipótese nula do teste é: (H_0) as autocorrelações (parciais) de ordem $j = 1, \dots, j^*$, são todas nulas. A hipótese alternativa (H_1) do teste é a negação lógica da hipótese nula, i.e., pelo menos uma ordem não-nula;
 - ii. A estatística do teste é, para o teste referente às “s” primeiras defasagens, $Q = T(T+2) \sum_{k=1}^s \frac{\hat{\rho}_k^2}{T-k} \sim \chi^2(s)$ (Enders, 2015, p.68) a distribuição aproximada (assintótica), pois, é a qui-quadrado com número de graus de liberdade igual à s (igual à maior defasagem considerada); A sugestão de Enders (2015, p.68) é tomar $j^* = T/4$, em que T = tamanho da série empírica;
 - d. Uma vez que a inspeção visual nunca é exata, e é comum ficar em dúvida quanto a mais de uma possibilidade, recomenda-se uma combinação dos dois critérios acima (tabela e ordem máxima de significância). A combinação recomendada é: tomar como lista inicial todas as especificações sugeridas pelos dois critérios, conjuntamente;
2. Estimação: esta etapa é automaticamente feita pelos softwares estatísticos disponíveis, bastando conhecer o comando necessário para tanto. Basta afirmar que, apesar do método de mínimos quadrados ordinários ser consistente, ele não é eficiente, propriedade esta possuída pelo método de máxima verossimilhança, o qual é geralmente utilizado¹;
 3. Diagnóstico, parte 1, lista inicial de especificações. Diagnosticar todas as especificações da lista inicial da seguinte maneira:
 - a. Estimar a especificação e obter o correlograma dos resíduos (atenção: trata-se do correlograma dos resíduos e não da série original/variável Y);
 - b. Como um processo estocástico ARMA é estacionário, a perturbação do ARMA é obrigatoriamente um ruído branco. Um ruído branco tem FAC e FACP nulas para todas as ordens de defasagem. Eis a primeira operação de diagnóstico. Trata-se de verificar o critério de correlograma nulo para todas as ordens de defasagem do ARMA estimado. Este critério é verificado a partir de um teste de significância individual para a autocorrelação referente a uma dada ordem. O teste é reportado automaticamente para todas as ordens de $j = 1, \dots, j^*$, e também para a autocorrelação parcial, por comandos autocorrelograma de softwares estatísticos;

¹ Mais detalhes sobre estimação podem ser encontrados nos capítulos 7 de Hamilton e nas seções 3.4 e 3.5 de Bueno (2008).

- c. A segunda operação de diagnóstico é a realização do teste de significância conjunta de Ljung-Box referente a todas as autocorrelações (parciais) de ordem 1,..., j*:
 - i. A estatística do teste é a estatística de Ljung-Box, a qual, para as primeiras “s” defasagens, é $Q = T(T + 2) \sum_{k=1}^s \frac{\hat{\rho}_k^2}{T-k} \sim \chi^2(s - (p + q + 1))$ (Enders, 2015, p.68). Atenção: uma vez que o teste é aplicado aos resíduos, o número de graus de liberdade da distribuição qui-quadrado tem de ser descontado pelo número de parâmetros – e isso pois é necessário, para calcular os resíduos, estimar primeiramente os parâmetros da especificação candidata (Enders, 2015, p. 68). Esses parâmetros incluem o intercepto (ou “constante”), e os coeficientes dos termos AR e os coeficientes dos termos MA.
 - ii. Há uma implicação prática relevante: não é possível calcular o p-valor do teste para $s \leq (p + q + 1)$, pois o número de graus de liberdade é obrigatoriamente uma grandeza positiva;
 - d. A terceira operação de diagnóstico é o cálculo dos critérios de informação de Akaike (AIC), Bayesian Information Criterion (BIC) e Schwartz Bayesian Criterion (SBC);
 - e. Redução da lista a partir de dois procedimentos: eliminação e classificação:
 - i. Eliminação: eliminar da lista inicial as especificações que (i) rejeitem a hipótese nula do teste de significância conjunta para pelo menos uma defasagem e (ii) rejeitem, para pelo menos um tipo de teste de significância individual (com base na FAC ou na FACP), a hipótese nula para pelo menos uma defasagem;
 - ii. Classificação: este procedimento depende do resultado do procedimento anterior:
 1. Se todas as especificações da lista inicial forem eliminadas, tem-se o caso “conjunto vazio”. Seguir para o próximo passo;
 2. Se restar apenas uma especificação não-eliminada, esta é a especificação selecionada pela classificação;
 3. Se restarem pelo menos duas especificações não eliminadas, classifica-las com base nos três critérios de informação. Ou seja, selecionar apenas aquela com menores níveis de pelo menos dois dos três critérios de informação;
4. Diagnóstico, parte 2, segunda lista de especificações. A segunda lista depende do resultado obtido no passo anterior:
- a. Se não ocorreu “conjunto vazio”:
 - i. Deve-se reduzir em uma unidade a ordem da especificação selecionada, i.e., para MA(q-1), AR(p-1), ou para ARMA(p-1, q) e ARMA(p, q-1).
 - ii. Estimar tais processos e aplicar, aos resíduos de cada um, os testes de significância individual e conjunta para FAC e FACP. Também calcular os três critérios de informação;

- iii. Eliminar da lista as candidatas em que pelo menos um teste rejeite a hipótese nula;
 - iv. Repetir os passos i a iii anteriores, reduzindo sempre em uma unidade a ordem da especificação, até obter, em pelo menos um testes de significância, a rejeição da hipótese nula. Ou seja, até descartar a primeira candidata;
- b. Se ocorreu “conjunto vazio”:
 - i. Deve-se aumentar em uma unidade as ordens de todas as especificações da lista inicial, ou seja, para $AR(p+1)$, $MA(q+1)$, $ARMA(p+1, q)$ e $ARMA(p, q+1)$;
 - ii. Estimar as especificações aumentadas e aplicar, aos resíduos, os testes de significância individual e conjunta para FAC e FACP. Também calcular os três critérios de informação;
 - iii. Descartar da lista candidatas que rejeitem a hipótese nula em pelo menos um teste;
 - iv. Repetir os passos i a iii anteriores até pelo menos duas candidatas não rejeitarem a hipótese nula.
- c. Diagnóstico, parte 3, seleção da melhor candidata. O critério de classificação é simples: a melhor candidata é aquela que, de todas as candidatas não eliminadas, atinge menor nível em pelo menos dois dos três critérios de informação.

Figura Fluxograma das etapas 3 e 4 do algoritmo



4.3 Esclarecimentos

Quanto ao algoritmo anterior, há dois esclarecimentos a fazer:

1. O primeiro esclarecimento é uma advertência: o algoritmo anterior é sujeito a erro. Conforme afirmado de início, ele pode ignorar especificações com boa performance, ou seja, que resultam em resíduos com comportamento de ruído branco e atingem níveis relativamente baixos dos critérios de informação. Um algoritmo com menor probabilidade de erro será apresentado na nota de aula 3;
2. De maneira relacionada ao primeiro esclarecimento, conforme afirmam Gujarati, Enders e Bueno, a metodologia de Box-Jenkins é muitas vezes entendida como uma arte e nunca como um procedimento exato;
3. Como Bueno (2008) realisticamente afirma, reconhecer, a partir da FAC e FACP, sinais indicando um processo ARMA é mais difícil do que sinais indicando processos isolados AR ou MA. E isso pois observar a ordem de defasagem a partir da qual FAC e FACP começam a decair é difícil. Apesar da recomendação de sempre considerar, na lista inicial, uma combinação ARMA das ordens significativas máximas, ainda assim processos ARMA de boa performance podem ser ignorados;

O terceiro esclarecimento diz respeito às fórmulas dos critérios de informação. Elas têm a seguinte estrutura geral, $SQR + P$, em que “SQR” é a soma dos quadrados dos resíduos do processo ARMA estimado e “P” é uma penalização. Conforme bem explica Bueno (2008, p.47), a estrutura geral da fórmula equilibra benefício e custo de aumento da ordem do processo ARMA em uma unidade. No caso, o benefício é o aumento da parcela da série de tempo explicada pelo processo, este benefício sendo, tal como na regressão linear, inversamente proporcional à parcela não-explicada, ou seja, à SQR. Já o custo é a perda de graus de liberdade, o que reduz a precisão (aumenta a volatilidade) do estimador. Daí porque se deseja minimizar os critérios de informação, uma vez que eles são a soma de dois fatores indesejáveis, parcela não explicada da série de tempo e perda de graus de liberdade. Os critérios específicos de informação, BIC, AIC e SBC, diferem na forma matemática da penalização, a qual é uma função do número de parâmetros estimados. O que se dá da seguinte maneira:

1. $AIC = T \cdot \log(SQR) + 2 \cdot K$
2. $BIC = T \cdot \log(SQR) + \log(T) \cdot K$
3. $HQ = T \cdot \log(SQR) + 2 \cdot K \cdot \log(\log(T))$

Apêndice: obtenção da FAC para o processo MA(q)

(Observação: este apêndice não foi ainda finalizado)

O processo de interesse é:

$$y_t = \mu + \sum_{s=1}^q \theta_s \varepsilon_{t-s} + \varepsilon_t$$

A expectativa dele é:

$$E[y_t] = E \left[\mu + \sum_{s=1}^q \theta_s \varepsilon_{t-s} + \varepsilon_t \right] = \mu$$

Agora inicia-se o procedimento de Yule-Walker, transformando com subtração da expectativa:

$$y_t^* = y_t - \mu = \sum_{s=1}^q \theta_s \varepsilon_{t-s} + \varepsilon_t$$

O segundo passo, o qual multiplicação a equação do MA(q) por y_{t-s} , $s = 0, 1, 2, \dots$, conduz ao sistema abaixo:

$$y_t^* y_t^* = \theta \varepsilon_{t-1} y_t^* + \varepsilon_t y_t^*$$

$$y_t^* y_t^* = \sum_{s=1}^q (\theta_s \varepsilon_{t-s} y_t^*) + \varepsilon_t y_t^*$$

$$y_t^* y_{t-1}^* = \sum_{s=1}^q (\theta_s \varepsilon_{t-s} y_{t-1}^*) + \varepsilon_t y_{t-1}^*$$

$$y_t^* y_{t-2}^* = \sum_{s=1}^q (\theta_s \varepsilon_{t-s} y_{t-2}^*) + \varepsilon_t y_{t-2}^*$$

$$y_t^* y_{t-3}^* = \sum_{s=1}^q (\theta_s \varepsilon_{t-s} y_{t-3}^*) + \varepsilon_t y_{t-3}^*$$

...

$$y_t^* y_{t-j}^* = \sum_{s=1}^q (\theta_s \varepsilon_{t-s} y_{t-j}^*) + \varepsilon_t y_{t-j}^*$$

Tomando as expectativas:

$$E[y_t^* y_t^*] = \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} y_t^*] + E[\varepsilon_t y_t^*]$$

$$y_t^* y_{t-1}^* = \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} y_{t-1}^*] + E[\varepsilon_t y_{t-1}^*]$$

$$y_t^* y_{t-2}^* = \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} y_{t-2}^*] + E[\varepsilon_t y_{t-2}^*]$$

$$y_t^* y_{t-3}^* = \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} y_{t-3}^*] + E[\varepsilon_t y_{t-3}^*]$$

...

$$y_t^* y_{t-j}^* = \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} y_{t-j}^*] + E[\varepsilon_t y_{t-j}^*]$$

Com isso:

$$E[y_t^* y_t^*] = \sum_{s=1}^q \theta_s E \left[\varepsilon_{t-s} \left(\sum_{s=1}^q \theta_s \varepsilon_{t-s} + \varepsilon_t \right) \right] + E[\varepsilon_t y_t^*]$$

$$y_t^* y_{t-1}^* = \sum_{s=1}^q \theta_s E \left[\varepsilon_{t-s} \left(\sum_{s=1}^q \theta_s \varepsilon_{t-1-s} + \varepsilon_{t-1} \right) \right] + E[\varepsilon_t y_{t-1}^*]$$

$$y_t^* y_{t-2}^* = \sum_{s=1}^q \theta_s E \left[\varepsilon_{t-s} \left(\sum_{s=1}^q \theta_s \varepsilon_{t-2-s} + \varepsilon_{t-2} \right) \right] + E[\varepsilon_t y_{t-2}^*]$$

$$y_t^* y_{t-3}^* = \sum_{s=1}^q \theta_s E \left[\varepsilon_{t-s} \left(\sum_{s=1}^q \theta_s \varepsilon_{t-3-s} + \varepsilon_{t-3} \right) \right] + E[\varepsilon_t y_{t-3}^*]$$

...

$$y_t^* y_{t-j}^* = \sum_{s=1}^q \theta_s E \left[\varepsilon_{t-s} \left(\sum_{s=1}^q \theta_s \varepsilon_{t-j-s} + \varepsilon_{t-j} \right) \right] + E[\varepsilon_t y_{t-j}^*]$$

Ou seja:

$$E[y_t^* y_t^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-i} \varepsilon_{t-j}] + \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} \varepsilon_t] + E[\varepsilon_t y_t^*]$$

$$E[y_t^* y_{t-1}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-1-i} \varepsilon_{t-j}] + \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} \varepsilon_{t-1}] + E[\varepsilon_t y_{t-1}^*]$$

$$E[y_t^* y_{t-2}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-2-i} \varepsilon_{t-j}] + \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} \varepsilon_{t-2}] + E[\varepsilon_t y_{t-2}^*]$$

$$E[y_t^* y_{t-3}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-3-s} \varepsilon_{t-s}] + \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} \varepsilon_{t-3}] + E[\varepsilon_t y_{t-3}^*]$$

...

$$E[y_t^* y_{t-j}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-j-s} \varepsilon_{t-s}] + \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} \varepsilon_{t-j}] + E[\varepsilon_t y_{t-j}^*]$$

Uma das principais características do sistema é a zeragem do lado direito para uma defasagem k maior do que a ordem do processo, i.e., para $k > q$.

Basta notar que, a k -ésima equação tem a seguinte forma:

$$E[y_t^* y_{t-k}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-k-s} \varepsilon_{t-s}] + \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} \varepsilon_{t-k}] + E[\varepsilon_t y_{t-k}^*]$$

$E[\varepsilon_{t-k-s} \varepsilon_{t-s}] = 0$, para $k > q$, qualquer que seja $s \leq q$.

$E[\varepsilon_{t-s} \varepsilon_{t-k}] = 0$, para $k > q$, qualquer que seja $s \leq q$.

E é sempre verdade que $E[\varepsilon_t y_{t-k}^*] = 0$, para $k > 0$.

Com isso é adequado reescrever o sistema como:

$$E[y_t^* y_t^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-i} \varepsilon_{t-j}] + \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} \varepsilon_t] + E[\varepsilon_t y_t^*]$$

$$E[y_t^* y_{t-1}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-1-i} \varepsilon_{t-j}] + \sum_{s=1}^q \theta_s E[\varepsilon_{t-s} \varepsilon_{t-1}] + E[\varepsilon_t y_{t-1}^*]$$

$$E[y_t^* y_{t-2}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-2-i} \varepsilon_{t-j}] + \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} \varepsilon_{t-2}] + E[\varepsilon_t y_{t-2}^*]$$

$$E[y_t^* y_{t-3}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-3-s} \varepsilon_{t-s}] + \sum_{i=1}^q \theta_i E[\varepsilon_{t-i} \varepsilon_{t-3}] + E[\varepsilon_t y_{t-3}^*]$$

...

$$E[y_t^* y_{t-q}^*] = \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-q-s} \varepsilon_{t-s}] + \sum_{i=1}^q \theta_s E[\varepsilon_{t-s} \varepsilon_{t-q}] + E[\varepsilon_t y_{t-q}^*]$$

$$E[y_t^* y_{t-(q+1)}^*] = 0$$

...

$$E[y_t^* y_{t-j}^*] = 0$$

Para continuar a simplificar o sistema, é necessário utilizar alguns fatos quanto aos somatórios. Já sabemos que $E[\varepsilon_{t-s} y_{t-l}^*] = \sigma^2$, se $s = l$, e $E[\varepsilon_{t-s} y_{t-l}^*] = 0$, se $s \neq l$.

Com isso, os somatórios duplos são:

Primeira equação:

$$\begin{aligned} \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-i} \varepsilon_{t-j}] &= \left(\sum_{i=1}^q \theta_i^2 E[\varepsilon_{t-i} \varepsilon_{t-i}] + \sum_{\substack{j=1 \\ j \neq 1}}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-i} \varepsilon_{t-j}] \right) \\ &= \left(\sum_{i=1}^q \theta_i^2 \sigma^2 + \sum_{\substack{j=1 \\ j \neq 1}}^q \sum_{i=1}^q 0 \right) = \sigma^2 \sum_{i=1}^q \theta_i^2 \end{aligned}$$

Segunda equação:

$$\begin{aligned} \sum_{j=1}^q \sum_{i=1}^q \theta_j \theta_i E[\varepsilon_{t-1-i} \varepsilon_{t-j}] &= \left(\sum_{j=1}^q \theta_j \theta_1 E[\varepsilon_{t-1-i} \varepsilon_{t-1}] + \sum_{j=1}^q \sum_{i=2}^q \theta_j \theta_i E[\varepsilon_{t-1-i} \varepsilon_{t-j}] \right) \\ &= \left(\sum_{j=1}^q \theta_j \theta_1 E[\varepsilon_{t-1-i} \varepsilon_{t-1}] + \sum_{l=2}^{q+1} \sum_{i=2}^q \theta_{l-1} \theta_i E[\varepsilon_{t-l} \varepsilon_{t-i}] \right) \\ &= \left(\sum_{j=1}^q \theta_j \theta_1 E[\varepsilon_{t-1-i} \varepsilon_{t-1}] + \sum_{i=2}^q \theta_q \theta_i E[\varepsilon_{t-q-1} \varepsilon_{t-i}] \right. \\ &\quad \left. + \sum_{l=2}^{q+1} \sum_{i=2}^q \theta_{l-1} \theta_i E[\varepsilon_{t-l} \varepsilon_{t-i}] \right) \\ &= \left(\sum_{j=1}^q 0 + \sum_{j=2}^q 0 + \sum_{l=2}^q \sum_{i=2}^q \theta_{l-1} \theta_i E[\varepsilon_{t-l} \varepsilon_{t-j}] \right) \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{l=2}^q \theta_{l-1} \theta_l [\varepsilon_{t-l} \varepsilon_{t-l}] + \sum_{\substack{l=2 \\ l \neq i}}^q \sum_{i=2}^q \theta_{l-1} \theta_i E[\varepsilon_{t-l} \varepsilon_{t-j}] \right) \\
&= \left(\sum_{l=2}^q \theta_{l-1} \theta_l \sigma^2 \right) = \sigma^2 \left(\sum_{l=2}^q \theta_{l-1} \theta_l \right)
\end{aligned}$$

E assim por diante, ou seja, no sentido descendente do sistema, cada somatório duplo contém uma unidade a menos de $\theta_s^2 \sigma^2$.

Já os somatórios simples são sempre:

$$\sum_{s=1}^q \theta_s E[\varepsilon_{t-s} \varepsilon_{t-k}] = \theta_s \left(E[\varepsilon_{t-k} \varepsilon_{t-k}] + \sum_{\substack{s=1 \\ s \neq k}}^q \varepsilon_{t-s} \varepsilon_{t-k} \right) = \theta_s (\sigma^2)$$

Com isso o sistema reduz-se a:

$$E[y_t^* y_t^*] = \sigma^2 \sum_{i=1}^q \theta_i^2 + \sigma^2 = \sigma^2 \left(1 + \sum_{i=1}^q \theta_i^2 \right)$$

$$E[y_t^* y_{t-1}^*] = \sigma^2 \left(\sum_{l=2}^q \theta_{l-1} \theta_l \right) + \theta_1 \sigma^2 = \sigma^2 \left(\sum_{l=2}^q \theta_{l-1} \theta_l + \theta_1 \right)$$

$$E[y_t^* y_{t-2}^*] = \sigma^2 \left(\sum_{l=2}^q \theta_{l-2} \theta_l \right) + \theta_2 \sigma^2 = \sigma^2 \left(\sum_{l=2}^q \theta_{l-2} \theta_l + \theta_2 \right)$$

$$E[y_t^* y_{t-3}^*] = \sigma^2 \left(\sum_{l=2}^q \theta_{l-3} \theta_l \right) + \theta_3 \sigma^2 = \sigma^2 \left(\sum_{l=2}^q \theta_{l-3} \theta_l + \theta_3 \right)$$

...

$$E[y_t^* y_{t-q}^*] = \sigma^2 \left(\sum_{l=2}^q \theta_{l-q} \theta_l \right) + \theta_q \sigma^2 = \sigma^2 \left(\sum_{l=2}^q \theta_{l-q} \theta_l + \theta_q \right)$$

$$E[y_t^* y_{t-(q+1)}^*] = 0$$

...

$$E[y_t^* y_{t-j}^*] = 0$$