

APÊNDICE C

FUNDAMENTOS ESTATÍSTICOS – SÉRIES TEMPORAIS¹

Neste Apêndice são apresentados alguns conceitos de estatística úteis para validar os modelos de previsão de demanda de energia, sobretudo os que envolvem método de regressão, econometria e modelos de séries temporais. Duas referências foram utilizadas para a preparação deste Apêndice: a segunda edição do livro de William W. S. Wei, intitulado *Time Series Analysis (Univariate and Multivariate Methods)*, publicada em 2006, e outro mais recente, publicado em 2012, por R. L. S. Bueno sob o título *Econometria de Séries Temporais*.

VARIÁVEL ALEATÓRIA E DISCRETA

Quando uma variável assume resultados diversos entre uma observação e outra em razão de fatores relacionados à chance, ela é chamada de variável aleatória.

a) Variável aleatória discreta

Definição: seja x uma variável aleatória. Se o número de valores possíveis de x for finito ou infinito enumerável, então x é denominada de variável aleatória discreta. A cada resultado possível x_i está associada uma probabilidade de ocorrência $p(x_i)$ com as seguintes condições para todo i :

$$p(x_i) \geq 0 \quad \Leftrightarrow \quad \sum_{i=1}^{\infty} p(x_i) = 1 \quad (\text{C.1})$$

a) Variável aleatória contínua

Definição: diz-se que x é uma variável aleatória contínua se existir uma função f , denominada de função densidade de probabilidade f_{dp} de x , que satisfaça as seguintes condições:

$$f(x) \geq 0 \quad \Leftrightarrow \quad \int_{-\infty}^{+\infty} f(x) dx = 1 \quad (\text{C.2})$$

Para quaisquer a e b , com $-\infty < a < b < +\infty$, tem-se que:

$$p(a \leq x \leq b) = \int_a^b f(x) dx \quad (\text{C.3})$$

b) Função de distribuição acumulada

Definição: seja x uma variável aleatória discreta. Define-se a função F como a função de distribuição acumulada da variável aleatória x como:

$$F(x) = \sum_j p(x_j) \quad (\text{C.4})$$

¹ Este apêndice foi escrito por Dr. João B. Marques (jbdmarques@gmail.com).

Onde o termo de soma é estendido a todos os índices j que satisfaçam a condição $x_j \leq x$. Se x for uma variável aleatória contínua, então:

$$F(x) = \int_{-\infty}^x f(s)ds \quad (\text{C.5})$$

Onde s é uma variável muda de integração. Observe que se $F(x)$ é a função de distribuição acumulada da variável aleatória contínua x , com *fdp* $f(x)$, então:

$$\frac{dF(x)}{dx} = f(x) \quad (\text{C.6})$$

Para todo x no qual F seja derivável.

VALOR ESPERADO, ESPERANÇA CONDICIONAL E INCONDICIONAL

O valor esperado de uma variável aleatória discreta, ou simplesmente esperança, é a soma das probabilidades de ocorrência de cada evento multiplicada pelo seu valor. É o valor médio *esperado* de um experimento se ele for repetido muitas vezes, portanto:

$$E(x) = \sum_{i=1}^n x_i p(x_i) \quad (\text{C.7})$$

Se a variável for contínua com *fdp* igual a $f(x)$, então a esperança é dada por:

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx \quad (\text{C.8})$$

Duas propriedades são básicas em relação ao valor esperado, a primeira é:

$$E(ax + by) = aE(x) + bE(y) \quad (\text{C.9})$$

Onde a e b são constantes, e a segunda é:

$$E(xy) \neq E(x)E(y) \quad (\text{C.10})$$

Contudo, se as variáveis x e y são independentes, então:

$$E(xy) = E(x)E(y) \quad (\text{C.11})$$

a) Esperança condicional e não condicional

Esses conceitos são fundamentais em econometria de séries temporais, uma vez que a ordem dos dados do conjunto da população é uma condicionante. Considere o espaço amostral Φ a esperança não condicional (ou incondicional) de uma variável aleatória x , que é definida por:

$$E(x | \Phi) = E(x) \quad (\text{C.12})$$

Onde o conjunto a que pertence a esperança não está definida claramente no espaço amostral. A lei das expectativas totais afirma que:

$$E[E(x | \Phi)] = E(x | \Phi) = E(x) \quad (\text{C.13})$$

Seja Ω todos os subconjuntos do espaço amostral Φ sobre o qual a variável aleatória x está inserida. A lei das expectativas iteradas, aplicada a dois conjuntos $A, B \in \Omega$, é definida por:

$$E[E(x | A, B) | A] = E(x | A) \quad (\text{C.14})$$

O que implica que é sobre o menor conjunto de informação que se determina a esperança condicional.

VARIÂNCIA

Definição: seja x uma variável aleatória. Define-se a variância de x , denotada por $\text{Var}(x)$ ou σ_x^2 , como:

$$\text{Var}(x) = E[x - E(x)]^2 \quad (\text{C.15})$$

A raiz quadrada positiva de $\text{Var}(x)$ é denominada de desvio padrão. Outra forma de expressar a $\text{Var}(x)$ é:

$$\text{Var}(x) = E(x^2) - [E(x)]^2 \quad (\text{C.16})$$

a) Propriedades da variância

Se a e b são constantes, então:

$$\text{Var}(x + a) = \text{Var}(x) \quad (\text{C.17})$$

E:

$$\text{Var}(ax) = a^2 \text{Var}(x) \Leftrightarrow \text{Var}(ax + b) = a^2 \text{Var}(x) \quad (\text{C.18})$$

Também:

$$\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x, y) \quad (\text{C.19})$$

Onde $\text{Cov}(x, y)$ representa a covariância entre as variáveis aleatórias x e y :

$$\text{Cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] \quad (\text{C.20})$$

Onde μ_x e μ_y são as médias das distribuições das variáveis x e y , cuja função de correlação é:

$$\rho(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\sigma_x^2} \sqrt{\sigma_y^2}} \quad (\text{C.21})$$

DEFINIÇÃO DE PROCESSO ESTOCÁSTICO

Um processo estocástico é uma família de variáveis aleatórias $Z(\omega, t)$ indexada no tempo, onde ω pertence ao espaço amostral e t refere-se ao tempo. Para um determinado t , $Z(., t)$ é uma variável aleatória e para um dado espaço amostral ω , $Z(\omega, .)$ é uma realização e $Z(\omega, t)$ é um número real. Uma população que consista de todas as possíveis realizações é chamada de processo estocástico

gerador de dados. Portanto, uma série temporal é uma realização ordenada de um processo estocástico qualquer.

Suponha infinitas medições da temperatura do convés de n sondas de petróleo durante o dia. Assim, seria possível montar um conjunto com as n sequências seguintes:

$$\left(\left\{ Z_t^{(1)} \right\}_{t=-\infty}^{\infty}, \left\{ Z_t^{(2)} \right\}_{t=-\infty}^{\infty}, \dots, \left\{ Z_t^{(n)} \right\}_{t=-\infty}^{\infty} \right) \quad (\text{C.22})$$

Mais exatamente, em cada instante t existiriam n temperaturas ou n observações relacionadas ao tempo t , ou seja:

$$\left\{ Z_t^{(1)}, Z_t^{(2)}, \dots, Z_t^{(n)} \right\} \quad (\text{C.23})$$

Esse é um conjunto de dados cuja distribuição é, possivelmente, normal. Assim, vários momentos dessa série podem ser estimados; por exemplo, o primeiro e o segundo momentos (esperança e variância, respectivamente). Convém suprimir o espaço amostral e simplificar $Z(\omega, t)$ para $Z(t)$ ou Z_t ($t=0, \pm 1, \pm 2, \dots$). Dessa forma, a esperança não condicional da variável aleatória $Z(t)$, ou simplesmente Z_t é:

$$\mu_t = E(Z_t) = \int_{-\infty}^{+\infty} Z_t f(Z_t) dZ_t \quad (\text{C.24})$$

Onde $f(Z_t)$ é a função densidade de probabilidade – ver Eq. (C.8). Note que a esperança é a média esperada de uma distribuição de frequência. A função de variância de um processo estocástico é:

$$\sigma_t^2 = E(Z_t - \mu_t)^2 \quad (\text{C.25})$$

E a função de covariância entre duas variáveis aleatórias Z_{t_1} e Z_{t_2} é:

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}) \quad (\text{C.26})$$

Assim, por analogia, a função de correlação fica definida como:

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}} \quad (\text{C.27})$$

AUTOCOVARIÂNCIA, AUTOCORRELAÇÃO

a) Autocovariância

A autocovariância representa a covariância da variável aleatória com ela mesma defasada de k passos, onde as séries temporais são tomadas do mesmo processo estocástico. Em um processo estritamente estacionário, a função de distribuição é a mesma para todo t , portanto $\mu_t = \mu$ é uma constante, desde que $E(Z_t) < \infty$ e a variância $\sigma_t^2 = \sigma^2$ para todo t seja também uma constante. Assim, define-se a autocovariância como:

$$\gamma_k = \text{Cov}(Z_t, Z_{t+k}) = E[(Z_t - \mu_t)(Z_{t+k} - \mu_{t+k})] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \quad (\text{C.28})$$

Note que, por definição, γ_0 é a própria variância. Observe também que, de acordo com definição em (C.28), as variâncias não condicionais de $Z_t = \mu + \varepsilon_t$ e $Z_{t+k} = \mu + \varepsilon_{t+k}$ são idênticas.

b) Autocorrelação

A autocorrelação é definida por:

$$\rho_k = \frac{\gamma_k}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (\text{C.29})$$

Onde $\text{Var}(Z_t) = \text{Var}(Z_{t+k}) = \gamma_0$.

c) Função de autocovariância e função de autocorrelação (FAC)

Uma vez que γ_k e ρ_k são funções de k (defasagem), γ_k e ρ_k são chamados de função de autocovariância e de autocorrelação, respectivamente. Esses conceitos são importantes na determinação do melhor modelo para representar uma série temporal específica.

EXEMPLO C.1 – Suponha $Z_t = \mu + \varepsilon_t$ um processo estocástico com $\varepsilon_t \sim i.i.d. (0, \sigma^2)$. Determine a função de autocovariância e a função de autocorrelação.

Solução

De acordo com a Eq. (C.28) tem-se que:

$$\gamma_k = E[(Z_t - \mu)(Z_{t+k} - \mu)] = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (\text{C.30})$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma^2} = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

Note que o ruído ε_t é *i.i.d.* (idêntica e independentemente distribuído), portanto, de acordo com a Eq. (C.11), quando $k \neq 0$ resulta que:

$$E(\varepsilon_t \varepsilon_{t+k}) = E(\varepsilon_t)E(\varepsilon_{t+k}) = 0 \quad (\text{C.31})$$

d) Função de autocorrelação parcial (FACP)

Além da autocorrelação entre Z_t e Z_{t+k} , pode-se investigar a correlação entre Z_t e Z_{t+k} após remoção da dependência linear destas variáveis com as variáveis intervenientes $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$. Mais exatamente, podem-se investigar os erros destas variáveis obtidas de suas melhores estimativas por método de regressão linear.

Considere $\{Z_t\}$ um processo estacionário e, sem perda de generalidades, assumamos que $E(Z_t) = 0$. Considere a expressão a seguir o melhor estimador de Z_{t+k} :

$$\hat{Z}_{t+k} = \alpha_1 Z_{t+k-1} + \alpha_2 Z_{t+k-2} + \dots + \alpha_{k-1} Z_{t+1} \quad (\text{C.32})$$

Onde α_i ($1 \leq i \leq k-1$) são os coeficientes obtidos por método de regressão minimizando a esperança do quadrado dos resíduos, ou seja:

$$E(Z_{t+k} - \hat{Z}_{t+k})^2 \quad (\text{C.33})$$

De forma similar, considere a seguinte expressão o melhor estimador de Z_t , ou seja:

$$\hat{Z}_t = \beta_1 Z_{t+1} + \beta_2 Z_{t+2} + \dots + \beta_{k-1} Z_{t+k-1} \quad (C.34)$$

Onde β_i ($1 \leq i \leq k-1$) são os coeficientes obtidos por método de regressão minimizando a esperança do quadrado dos resíduos:

$$E(Z_t - \hat{Z}_t)^2 \quad (C.35)$$

Assim, a função de autocorrelação parcial ϕ_{kk} entre Z_t e Z_{t+k} é definida como a correlação ordinária entre $(Z_t - \hat{Z}_t)$ e $(Z_{t+k} - \hat{Z}_{t+k})$, ou seja:

$$\phi_{kk} = \frac{\text{Cov}[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{\text{Var}(Z_t - \hat{Z}_t)} \sqrt{\text{Var}(Z_{t+k} - \hat{Z}_{t+k})}} \quad (C.36)$$

Utilizando as propriedades da esperança e da variância, pode-se chegar ao seguinte método geral:

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \quad (C.37)$$

Ou de forma genérica:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad (C.38)$$

Onde ρ_k é a função de autocorrelação entre Z_t e Z_{t+k} .

e) Média amostral

Quando se tem apenas uma realização de um processo estacionário, o estimador natural da média $\mu = E(Z_t)$ é definido por:

$$\bar{Z}_t = \frac{1}{n} \sum_{i=1}^n Z_t \quad (C.39)$$

Onde \bar{Z}_t é a média temporal de n observações. Note que:

$$E(\bar{Z}_t) = \frac{1}{n} \sum_{i=1}^n E(Z_i) = \frac{1}{n} \cdot n\mu = \mu \quad (\text{C.40})$$

De onde se conclui que \bar{Z}_t é o estimador não enviesado de μ .

f) Função de autocovariância amostral

De forma similar, quando se tem apenas uma realização, o estimador da autocovariância é:

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}), \quad \text{ou} \quad \hat{\gamma}_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) \quad (\text{C.41})$$

Ambos os estimadores da Eq. (C.41) são não enviesados e pode-se demonstrar facilmente que:

$$E(\hat{\gamma}_k) \approx \gamma_k - \frac{k}{n} \gamma_k - \left(\frac{n-k}{n}\right) \text{Var}(\bar{Z}) \quad \text{e} \quad E(\hat{\gamma}_k) = \gamma_k - \text{Var}(\bar{Z}) \quad (\text{C.42})$$

g) Função de autocorrelação amostral (FAC)

Para uma dada realização de uma série temporal, a função de autocorrelação amostral é definida como:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2}, \quad k = 0,1,2,\dots \quad (\text{C.43})$$

EXEMPLO C.2 – Considere os dez valores de uma série temporal conforme Tabela C.1:

Tabela C.1 – Amostras de uma série temporal (10 observações)

t	Z_t	Z_{t+1}	Z_{t+2}	Z_{t+3}	...	Z_{t-1}	Z_{t-1}
1	13	8	15	4			
2	8	15	4	4		13	
3	15	4	4	12		8	13
4	4	4	12	11		15	8
5	4	12	11	7		4	15
6	12	11	7	14		4	4
7	11	7	14	12		12	4
8	7	14	12			11	12
9	14	12				7	11
10	12					14	7

Solução

Note que a média é $\bar{Z} = 10$ e, portanto:

$$\begin{aligned} \hat{\rho}_1 &= \frac{(13-10)(8-10) + (8-10)(15-10) + \dots + (7-10)(14-10) + (14-10)(12-10)}{(13-10)^2 + (8-10)^2 + \dots + (14-10)^2 + (12-10)^2} = \frac{-27}{144} = -0,118 \\ \hat{\rho}_2 &= \frac{(13-10)(15-10) + (8-10)(4-10) + \dots + (11-10)(14-10) + (7-10)(12-10)}{144} = -0,201 \\ \hat{\rho}_3 &= \frac{(13-10)(4-10) + (8-10)(4-10) + \dots + (12-10)(14-10) + (11-10)(12-10)}{144} = 0,181 \\ &\dots \end{aligned} \tag{C.44}$$

De modo geral, $\hat{\rho}_k = \hat{\rho}_{-k}$, ou seja, a função de autocorrelação amostral é simétrica em relação à origem $k=0$. Podem-se utilizar os dados obtidos e montar um gráfico da função de autocorrelação contra a defasagem. Ver Figura C.1.

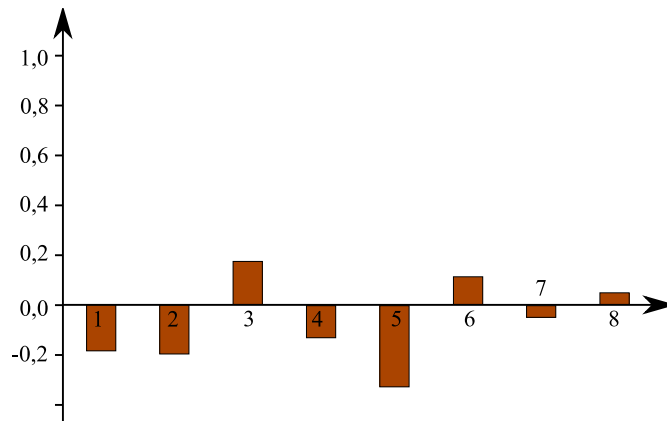


Figura C.1 – Função de autocorrelação amostral – FAC

h) Função de autocorrelação parcial amostral (FACP)

A função de autocorrelação parcial amostral $\hat{\phi}_{kk}$ é obtida substituindo ρ_i por $\hat{\rho}_i$ na Eq. (C.38). No entanto, em vez de apresentá-la em forma de determinantes, é melhor utilizar a forma recursiva a seguir:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad \text{e} \quad \hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \cdot \hat{\phi}_{k,k+1-j}, \quad j = 1, 2, \dots, k \tag{C.45}$$

EXEMPLO C.3 – Usando os dados do exemplo anterior, determine $\hat{\phi}_{11}$, $\hat{\phi}_{22}$, $\hat{\phi}_{21}$, e $\hat{\phi}_{33}$.

Solução

Partindo das Eqs. (C.37) e (C.38) tem-se que:

$$\begin{aligned}
\hat{\phi}_{11} &= \hat{\rho}_1 = -0,188 \\
\hat{\phi}_{22} &= \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{-0,201 - (-0,188)}{1 - (-0,188)^2} \\
\hat{\phi}_{21} &= \hat{\phi}_{11} - \hat{\phi}_{22} \cdot \hat{\phi}_{11} = (-0,188) - (-0,245)(-0,188) = -0,234
\end{aligned} \tag{C.46}$$

A partir da Eq. (C.45), tem-se que:

$$\hat{\phi}_{3,3} = \frac{\hat{\rho}_3 - \hat{\phi}_{21}\hat{\rho}_2 - \hat{\phi}_{22}\hat{\rho}_1}{1 - \hat{\phi}_{21}\hat{\rho}_1 - \hat{\phi}_{22}\hat{\rho}_2} = \frac{0,088}{0,907} = 0,097 \tag{C.47}$$

ESTACIONARIDADE, ERGOCIDADE E RUÍDO BRANCO

a) Estacionaridade

Por definição, se o processo estocástico Z_t tem esperança e autocovariância independentes do tempo, então se trata de uma série fracamente estacionária. Para tanto é necessário que se observe as seguintes condições:

$$\begin{aligned}
E(Z_t)^2 &< \infty \\
E(Z_t) &= \mu, \text{ para todo } t \in \{0, \pm 1, \pm 2, \dots, \pm t\} \\
E[(Z_t - \mu)(Z_{t+k} - \mu)] &= \lambda_k
\end{aligned} \tag{C.48}$$

A primeira condição assegura que o segundo momento não centrado é finito (podendo ser diferente para diferentes períodos). A segunda condição exige igualdade entre as médias (as distribuições podem sofrer alterações com o tempo). A terceira condição é que a variância seja igual para todo o período e independentemente do tempo.

b) Ergocidade

Para estimar uma série temporal é necessário atender à propriedade de ergocidade, além da estacionaridade. Suponha que uma realização $Z(\omega, \cdot)$ é uma realização de um processo estocástico. A média temporal é definida por:

$$\bar{Z}^{(\omega)} = \frac{1}{n} \sum_{t=1}^n Z_t^{(\omega)} \tag{C.49}$$

Se $\bar{Z}^{(\omega)}$ convergir para $E(\bar{Z}^{(\omega)})$, então existe ergocidade. Como as médias são todas iguais para diferentes intervalos de tempo, basta uma realização para se ter a média. O que se pretende é que a esperança de cada observação seja igual (estacionaridade) e se possa estimar essa esperança tomando-se a média temporal das observações (ergocidade). Pode-se provar que, se a soma das covariâncias for finita, Z_t é ergódico para o primeiro momento.

c) Ruído branco

Uma sequência de realizações $\{\varepsilon_t\}$ é um ruído branco se a média for zero em cada realização e se a variância for constante e não for correlacionada a qualquer outra realização da própria série. Se uma sequência de realizações $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ atende às condições em (C.50):

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ E(\varepsilon_t^2) &= \sigma^2 \text{ para todo } t \\ E(\varepsilon_t \varepsilon_{t-j}) &= 0, \text{ para todo } j \neq 0 \end{aligned} \quad (\text{C.50})$$

Então o processo $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ é um ruído branco.

MODELOS DE SÉRIES TEMPORAIS ESTACIONÁRIAS

A seguir, são indicados os processos estocásticos, com suas respectivas funções de autocovariância e autocorrelação, mais utilizados no tratamento de variáveis que interessam aos modelos de previsão de demanda de energia. Esses processos são também largamente utilizados em outras áreas de pesquisa, sobretudo em econometria e finanças. George Box e Gwilym Jenkins (1976) desenvolveram uma metodologia de tratamento de séries temporais estacionárias e popularizaram uma família de modelos denominada ARIMA (*autoregressive integrated moving average*).

a) Processo autorregressivo AR(p)

Para entender esta metodologia de George Box e Gwilym Jenkins (1970), é conveniente iniciar com a definição do que seja uma série temporal autorregressiva AR(p), que é determinada por:

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t = \mu + \sum_{j=1}^p \phi_j Z_{t-j} + \varepsilon_t \quad (\text{C.51})$$

Onde Z_t é a observação da série no tempo t ; ϕ_p é o parâmetro de ordem p do modelo; μ é um número real e ε_t é o resíduo ou erro (ruído branco²). Outra forma de apresentar uma série temporal AR(p) é:

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \dots + \phi_p \dot{Z}_{t-p} + \varepsilon_t = \sum_{j=1}^p \phi_j \dot{Z}_{t-j} + \varepsilon_t \quad (\text{C.52})$$

Onde $\dot{Z}_t = Z_t - \mu$. É comum também usar operadores que defasam as variáveis no tempo para representação de séries temporais de modo mais compacto. Por exemplo, pode-se reescrever a Eq. (C.51) com operadores, assim tem-se que:

$$Z_t = \mu + \phi_1 B Z_t + \phi_2 B^2 Z_t + \dots + \phi_p B^p Z_t + \varepsilon_t \quad (\text{C.53})$$

Onde $Z_{t-1} = B Z_t$ e $B(B Z_t) = B^2 Z_t$. De maneira geral, $B^p Z_t = Z_{t-p}$. Assim, a Eq. (C.53) pode ser resumida para:

² Uma sequência ε_t é um ruído branco se a média (por conveniência) for zero e a variância constante e não correlacionada a qualquer outra simulação da própria série.

$$\phi_p(B)Z_t = \mu + \varepsilon_t \quad (\text{C.54})$$

Onde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$. Assim, um processo autorregressivo de primeira ordem AR(1) é dado por:

$$(1 - \phi_1 B)\dot{Z}_t = \varepsilon_t, \quad \text{ou} \quad \dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \varepsilon_t \quad (\text{C.55})$$

A função de autocovariância γ_k de um modelo AR(1) é obtida conforme a seguir:

$$\gamma_k = E(\dot{Z}_t \dot{Z}_{t-k}) = E(\phi_1 \dot{Z}_{t-k} \dot{Z}_{t-1}) + E(\dot{Z}_{t-k} \varepsilon_t) = \phi_1 E(\dot{Z}_{t-k} \dot{Z}_{t-1}) + 0 \quad (\text{C.56})$$

Ou seja:

$$\gamma_k = \phi_1 \gamma_{k-1}, \quad k \geq 1 \quad (\text{C.57})$$

A função de autocorrelação se resume a:

$$\rho_k = \phi_1 \rho_{k-1} \quad k \geq 1 \quad (\text{C.58})$$

Adotando $\rho_0 = 0$. Nesse modelo, quando o módulo de $\phi_1 < 1$ e se o processo é estacionário, a função de autocorrelação decai exponencialmente de duas formas: se $0 < \phi_1 < 1$, o valor da função decai e é sempre positiva; e se $0 < \phi_1 < 1$, a autocorrelação também decai, iniciando com valor negativo e trocando de sinal a cada passo k . Já a função de autocorrelação parcial (FACP) é:

$$\phi_{kk} \begin{cases} \rho_1 = \phi_1, & k = 1 \\ 0, & k \geq 2 \end{cases} \quad (\text{C.59})$$

O processo autorregressivo de segunda ordem AR(2) é definido como:

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \varepsilon_t, \quad \text{ou} \quad (1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = \varepsilon_t \quad (\text{C.60})$$

As funções de autocovariância, autocorrelação (FAC) e autocorrelação parcial (FACP) para o processo AR(2) são, respectivamente:

$$\begin{aligned} \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad k \geq 1, & \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 1 \\ \phi_{kk} &= \begin{cases} \rho_1 = \frac{\phi_1}{1 - \phi_2}, & k = 1 \\ \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, & k = 2 \\ 0, & k = 3 \end{cases} \end{aligned} \quad (\text{C.61})$$

EXEMPLO C.4 – Simule 250 valores a partir de um processo estocástico AR(1), conforme modelo indicado na Eq. (C.62) a seguir, e elabore dois gráficos indicando os valores das funções de autocorrelação amostral (FAC) e autocorrelação parcial amostral (FACP) para 10 passos de defasagem k .

$$(1 - \phi_1 B)(Z_t - 10) = \varepsilon_t, \quad \text{com} \quad \phi_1 = \pm 0,9 \quad (\text{C.62})$$

Solução

O modelo descrito na Eq. (C.62) pode ser reescrito como:

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \varepsilon_t, \quad \text{onde } \dot{Z}_t = Z_t - 10 = Z_t - \mu \quad (C.63)$$

Note que, de acordo com a Eq. (C.58), a forma de decaimento da FAC depende do valor de ϕ_1 , conforme comentado. Para cada simulação realizada, há um resultado diferente para os valores da função de autocorrelação (FAC).

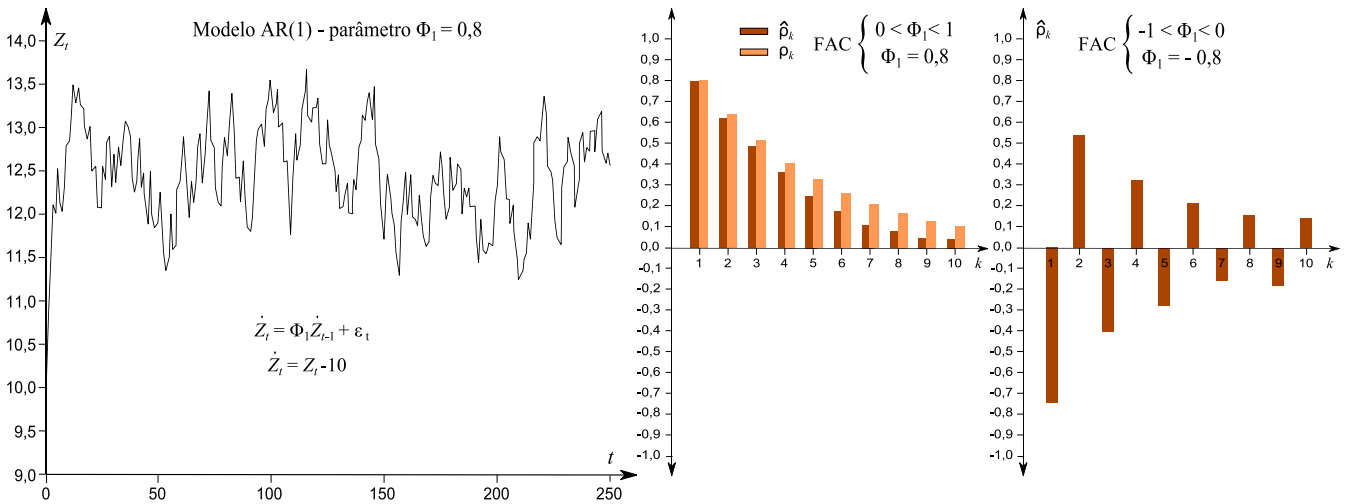


Figura C.2 – Modelo AR(1)

O modelo AR(1) pode ser útil para representar diversas séries temporais de variáveis econômicas ou variáveis de interesse nas análises de produção e preço de energia. Por exemplo, a inflação medida pelo IPCA entre janeiro de 1995 e dezembro de 2009 segue um modelo AR(1) em que $\phi_1=0,55$. Veja na Figura C.3 o índice mensal de inflação (IPCA) durante o período de 1995 a 2009 e uma simulação AR(1), onde $Z_t = \phi_1 Z_{t-1} + \varepsilon_t$ (onde $Z_{t-1} = 1,7\%$ em janeiro de 1995).

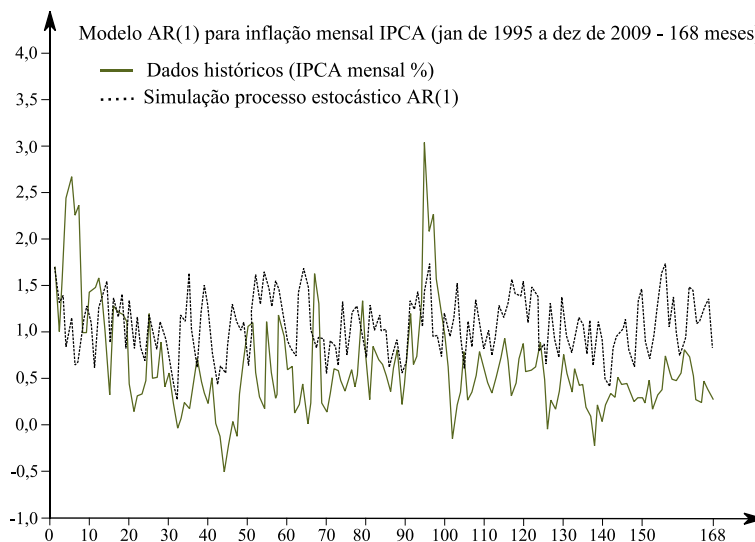


Figura C.3 – Inflação mensal IPCA – jan. 1995 a dez. 2009
 Fonte: IBGE

b) Processo de médias móveis MA(q)

O processo geral AR(p) é definido como:

$$\dot{Z}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad \text{ou} \quad \dot{Z}_t = \theta_q(B) \varepsilon_t \quad (\text{C.64})$$

Onde $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$. Assim, o processo de médias móveis de primeira ordem MA(1) fica reduzido a:

$$\dot{Z}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta_1 B) \varepsilon_t \quad (\text{C.65})$$

As funções de autocovariância, autocorrelação (FAC) e autocorrelação parcial (FACP) para o processo MA(1) são, respectivamente:

$$\gamma_k = \begin{cases} (1 + \theta_1^2) \sigma_\varepsilon^2, & k = 0 \\ -\theta_1 \sigma_\varepsilon^2, & k = 1 \\ 0, & k > 1 \end{cases}, \quad \rho_k = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2}, & k = 1 \\ 0, & k > 1 \end{cases} \quad (\text{C.66})$$

E:

$$\phi_{kk} = \frac{-\theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \quad k \geq 1 \quad (\text{C.67})$$

O processo de médias móveis de segunda ordem MA(2) é definido como:

$$\dot{Z}_t = (1 - \theta_1 B - \theta_2 B^2) \varepsilon_t \quad (\text{C.68})$$

As funções de autocovariância, autocorrelação (FAC) e autocorrelação parcial (FACP) para o processo MA(2) são, respectivamente:

$$\gamma_k = \begin{cases} (1 + \theta_1^2 + \theta_2^2) \sigma_\varepsilon^2, & k = 0 \\ -\theta_1 (1 - \theta_2) \sigma_\varepsilon^2, & k = 1 \\ -\theta_2 \sigma_\varepsilon^2, & k = 2 \\ 0, & k > 2 \end{cases}, \quad \rho_k = \begin{cases} \frac{-\theta_1 (1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, & k = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases} \quad (\text{C.69})$$

E:

$$\phi_{kk} = \begin{cases} \rho_1, & k = 1, \\ \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, & k = 2 \\ \frac{\rho_1^3 - \rho_1 \rho_2 (2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2 (1 - \rho_2)}, & k = 3 \end{cases} \quad (\text{C.70})$$

c) Processo autorregressivo de médias móveis ARMA(p, q)

Uma extensão natural é a composição de um processo puro AR(p) com outro MA(q), formando, assim, um processo estocástico ARMA(p, q). O processo geral ARMA(p, q) é definido como:

$$\phi_p(B)\dot{Z}_t = \theta_q(B)\varepsilon_t \quad (\text{C.71})$$

Onde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ e $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$. Assim, o processo ARMA(1,1) fica reduzido a:

$$(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)\varepsilon_t \quad (\text{C.72})$$

As funções de autocovariância, autocorrelação (FAC) e autocorrelação parcial (FACP) para o processo ARMA(1) são, respectivamente:

$$\gamma_k = \left\{ \begin{array}{ll} \frac{\phi_1(1 + \theta_1^2 - 2\phi_1\theta_1)\sigma_\varepsilon^2}{(1 - \phi_1^2)}, & k = 0 \\ \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)\sigma_\varepsilon^2}{(1 - \phi_1^2)}, & k = 1 \\ \phi_1\gamma_{k-1}, & k \geq 2 \end{array} \right\}, \quad \rho_k = \left\{ \begin{array}{ll} 1, & k = 0 \\ \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}, & k = 1 \\ \phi_1\rho_{k-1}, & k \geq 2 \end{array} \right\} \quad (\text{C.73})$$

A função de autocorrelação do processo ARMA(1,1) e demais, além de complicada não é necessária.

MODELOS DE SÉRIES TEMPORAIS NÃO ESTACIONÁRIAS

Muitas variáveis econômicas são séries temporais não estacionárias, particularmente aquelas relacionadas ao crescimento das atividades produtivas. Em geral, a média e a variância das séries temporais não estacionárias não são constantes e dependem do tempo t . Por exemplo, o PIB do Brasil entre os anos de 1950 e 2000 forma uma série temporal não estacionária, no entanto, possivelmente a taxa de crescimento no PIB pode ser uma série estacionária. De um modo geral, algumas séries não estacionárias homogêneas tornam-se estacionárias após alguns filtros ou derivações.

a) Filtros nos processos estocásticos

É comum realizar algumas transformações em séries temporais. A essas transformações dá-se o nome de filtragem. A metodologia de Box e Jenkins (1970) é para séries temporais estacionárias, no entanto, se uma em especial não for estacionária, ela deve então ser diferenciada até que se torne, para que se aplique a metodologia. A diferenciação é um tipo de filtro. Se uma variável Z_t não é estacionária, pode-se definir uma nova variável que corresponde à primeira diferença ΔZ_t . Se a estacionaridade não for atingida com a primeira diferenciação, fazem-se novas operações sucessivas até obter a estacionaridade. Assim, define-se a primeira diferenciação da variável Z_t com a seguinte notação:

$$\Delta Z_t = (1 - B)Z_t = Z_t - Z_{t-1} \quad (\text{C.74})$$

A segunda diferenciação seria:

$$\Delta^2 Z_t = \Delta \Delta Z_t = \Delta(Z_t - Z_{t-1}) = Z_t - Z_{t-1} - (Z_{t-1} - Z_{t-2}) = Z_t - 2Z_{t-1} + Z_{t-2} \quad (\text{C.75})$$

Ou:

$$\Delta^2 Z_t = (1-B)^2 Z_t = (1-2B+B^2)Z_t \Leftrightarrow \Delta^d Z_t = (1-B)^d Z_t \quad (\text{C.76})$$

Com isso, pode-se definir adequadamente um modelo ARIMA.

b) Invertibilidade

Outra forma útil para descrever um processo estocástico autorregressivo AR é regredir o valor de Z com suas variáveis passadas somadas a um ruído branco (choque), ou seja:

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \dots + \varepsilon_t = \varepsilon_t + \sum_{j=1}^{\infty} \pi_j \dot{Z}_{t-j} \quad (\text{C.77})$$

Ou:

$$\pi(B)\dot{Z}_t = \varepsilon_t \quad (\text{C.78})$$

Onde:

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j, \quad \text{com} \quad 1 + \sum_{j=1}^{\infty} |\pi_j| < \infty \quad (\text{C.79})$$

De acordo com Box e Jenkins (1976), um processo é invertível se for possível reescrevê-lo na forma da Eq. (C.77), conforme condição estabelecida na Eq.(C.79). Note que, para um processo linear $Z_t = \psi_t(B)\varepsilon_t$ ser invertível é necessário reescrevê-lo na forma de um processo AR. Isso se faz substituindo Z_t na Eq. (C.78), ou seja:

$$\pi(B)\psi(B)\varepsilon_t = \varepsilon_t \quad (\text{C.80})$$

Portanto:

$$\pi(B)\psi(B) = 1 \Leftrightarrow \pi(B) = \frac{1}{\psi(B)} = \psi^{-1}(B) \quad (\text{C.81})$$

A condição de invertibilidade garante que os pesos dos valores passados (π_j) podem ser obtidos a partir dos pesos dos choques passados ψ_j , o ruído branco (*white noise*). Além da estacionaridade, essa condição garante que os pesos π_j decaem à medida que a série é deslocada para trás. Mais exatamente, os maiores pesos devem ser atribuídos às observações mais recentes. Portanto, conforme argumenta Box e Jenkins (1976), se um processo é não invertível, a previsão por meio dele não faz sentido.

c) Processo ARIMA (p, d, q)

Se Z_t torna-se estacionária após d diferenciações e se a série resultante for um modelo ARMA(p, q), então diz-se que Z_t é uma variável descrita por um modelo ARIMA(p, d, q). Mais exatamente, fazendo-se d diferenciações sobre a variável Z_t resulta num modelo ARMA (p, q) semelhante ao modelo apresentado na Eq. (C.71). Ou seja, um modelo ARIMA(p, d, q) é definido como:

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)\varepsilon_t \quad (\text{C.82})$$

Onde o parâmetro θ_0 assume funções distintas dependendo do grau de diferenciação d . Quando $d=0$, o processo original é estacionário θ_0 e representa a média do processo, ou seja: $\theta_0 = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$. Se $d \geq 1$, θ_0 é chamado de termo de tendência determinística. Um modelo ARIMA(p, d, q) refere-se, respectivamente, às ordens de autorregressão, integração e média móvel.

d) Processo SAR(p), SMA(q), SARMA(p, q) e SARIMA(p, d, q)

Existem variáveis que apresentam comportamento oscilatório recorrente e com periodicidade homogênea que podem ser representadas por séries temporais. Essas séries são ditas sazonais. A diferenciação sazonal é nada mais do que a diferenciação com defasagem de um período maior que unitário, um período s , por exemplo. Assim, tem-se que a d -ésima diferenciação sazonal s é definida como:

$$\Delta_s^d Z_t = (1 - B_s)^d Z_{t-s} \quad (\text{C.83})$$

Onde $B_s^p Z_t = Z_{t-ps}$. Um modelo estocástico autorregressivo é dito sazonal de ordem p , SAR(p), quando seus valores são regredidos de seus valores anteriores defasados de s ou múltiplos de s . Assim, tem-se que:

$$Z_t = \mu + \phi_1 Z_{t-s} + \phi_2 Z_{t-2s} + \dots + \phi_p Z_{t-ps} + \varepsilon_t = \mu + \sum_{j=1}^p \phi_j Z_{t-js} + \varepsilon_t \quad (\text{C.84})$$

De forma similar, um modelo SMA(q) é dito de médias móveis sazonal de ordem q , quando representado por:

$$Z_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-s} - \theta_2 \varepsilon_{t-2s} - \dots - \theta_q \varepsilon_{t-qs} = \mu + \sum_{j=0}^q \theta_j \varepsilon_{t-js}, \quad \theta_0 = 1 \quad (\text{C.85})$$

Por analogia, um modelo SARMA(p, q) nada mais é do que uma combinação de dois termos, um SAR(p) e um SMA(q), de modo que:

$$Z_t = \mu + \overbrace{\sum_{i=1}^p \phi_i Z_{t-is}}^{\text{Autorregressivo}} + \overbrace{\sum_{j=0}^q \theta_j \varepsilon_{t-js}}^{\text{Média móvel}}, \quad \theta_0 = 1 \quad (\text{C.86})$$

Do mesmo modo, se a d -ésima diferenciação de uma série reproduz uma série estacionária do tipo SARMA(p, q), diz-se que esta série é integrada com ordem d , ou seja, um modelo SARIMA(p, d, q).

PREVISÃO (ESTIMATIVAS FUTURAS)

Como já mencionado, as funções FAC e FACP são importantes na determinação do melhor modelo para representar uma série temporal específica. Para esse item, considere um modelo geral ARIMA (p, d, q):

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)\varepsilon_t \quad (\text{C.87})$$

Onde $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ e $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$ e o $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. O parâmetro θ_0 foi omitido, sem perda de generalidades, por questão de simplificação.

a) Estimativa do erro em previsões com base num modelo ARMA

Para determinar o erro de dados previstos pelo método dos mínimos quadrados para um modelo ARMA, deve-se inicialmente assumir um caso onde $d=0$ e $\mu=0$, ou seja:

$$\phi(B)Z_t = \theta(B)\varepsilon_t \quad (\text{C.88})$$

Uma vez que o modelo é estacionário, pode-se representá-lo a partir de um processo de médias móveis, ou seja:

$$Z_t = \psi(B)\varepsilon_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_1\varepsilon_{t-2} + \dots = \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j}, \quad \text{com } \psi_0 = 1 \quad (\text{C.89})$$

Onde:

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j \quad (\text{C.90})$$

Para $t=n+l$, tem-se que:

$$Z_{n+l} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{n+l-j} \quad (\text{C.91})$$

Suponha agora que se conheçam as observações até $t=n$ e seja necessário estimar l -passo a frente, ou seja, o valor de Z_{n+l} como uma combinação linear das observações conhecidas, $Z_n, Z_{n-1}, Z_{n-2}, \dots$, uma vez que Z_t pode ser escrito conforme Eq. (C.89). Assim, conforme notação, o valor estimado $\hat{Z}_n(l)$ de Z_{n+l} é escrito como:

$$\hat{Z}_n(l) = \psi_l^* \varepsilon_n + \psi_{l+1}^* \varepsilon_{n-1} + \psi_{l+2}^* \varepsilon_{n-2} + \dots = \sum_{j=0}^{\infty} \psi_{l+j}^* \varepsilon_{n-j} \quad (\text{C.92})$$

Onde os coeficientes podem ser determinados. Pode-se demonstrar facilmente que o erro é minimizado quando:

$$\psi_{l+j}^* = \psi_{l+j} \quad (\text{C.93})$$

Mais exatamente, a melhor estimativa da variável l -passo a frente é dada pela esperança condicionada ao conjunto de dados disponíveis, ou seja:

$$\hat{Z}_n(l) = E(Z_{n+l} | Z_n, Z_{n-1}, Z_{n-2}, \dots) \quad (\text{C.94})$$

Cujo erro:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j} \Leftrightarrow \hat{Z}_n(l) = Z_{n+l} - e_n(l) \quad (\text{C.95})$$

Com isso, pode-se estimar a variável para qualquer passo além dos n dados disponíveis. Em um processo normal, caso se deseje estabelecer os desvios, limitando a variável para uma faixa de $(1-\alpha)100\%$ de possíveis resultados, calcula-se com a expressão a seguir:

$$\hat{Z}_n(l) \pm N_{\alpha/2} \left[1 + \sum_{j=1}^{l-1} \psi_j^2 \right]^{1/2} \sigma_\varepsilon \quad (\text{C.96})$$

Onde $N_{\alpha/2}$ é o desvio padrão normal tal que $P(N > N_{\alpha/2}) = \alpha/2$.

b) Estimativa do erro em previsões com base num modelo ARIMA

Partindo do modelo geral, Eq. (C.87), pode-se reescrevê-lo para prever Z_{t+l} conforme processo AR, uma vez que o processo ARIMA é invertível. Ou seja:

$$\pi(B)Z_{t+l} = \varepsilon_{t+l} \quad (\text{C.97})$$

Onde:

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)} \quad (\text{C.98})$$

Ou, de forma equivalente:

$$Z_{t+l} = \varepsilon_{t+l} + \sum_{j=1}^{\infty} \pi_j Z_{t+l-j} \quad (\text{C.99})$$

Pode-se demonstrar³ que:

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{n+l-j} \quad (\text{C.100})$$

Onde, nesse caso:

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-1-i} \psi_i, \text{ com } j = 1, 2, \dots, l-1. \quad (\text{C.101})$$

Note que a Eq. (C.100) é idêntica à Eq. (C.95). Assim, a estimativa para um processo estocástico ARIMA é descrita com uma média ponderada das estimativas prévias. Ou seja:

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty} \pi_j \hat{Z}_n(l-j), \text{ com } l \geq 1 \quad (\text{C.102})$$

Pode-se perceber para $t \leq n$, por processo recursivo, que a estimativa acaba sendo expressa como a soma ponderada dos valores correntes. Por exemplo:

$$\hat{Z}_n(1) = \pi_1 Z_n + \pi_2 Z_{n-1} + \pi_3 Z_{n-2} + \dots = \sum_{j=1}^{\infty} \pi_j Z_{n+1-j} \quad (\text{C.103})$$

E:

$$\hat{Z}_n(2) = \pi_1 \hat{Z}_n(1) + \pi_2 Z_n + \pi_3 Z_{n-1} + \dots = \pi_1 \sum_{j=1}^{\infty} \pi_j Z_{n+1-j} + \sum_{j=1}^{\infty} \pi_{j+1} Z_{n+1-j} = \sum_{j=1}^{\infty} \pi_j^{(2)} Z_{n+1-j} \quad (\text{C.104})$$

³ Consultar demonstração em Wei (2006).

Onde:

$$\pi_j^{(2)} = \pi_1 \pi_j + \pi_{j+1} \tag{C.105}$$

Portanto, de maneira geral:

$$\hat{Z}_n(2) = \sum_{j=1}^{\infty} \pi_j^{(l)} Z_{n+1-j} \tag{C.106}$$

Onde:

$$\pi_j^{(l)} = \sum_{i=1}^{l-1} \pi_1 \pi^{(l-i)} + \pi_{j+l-1}, \quad \text{para } l > 1, \quad \text{e } \pi_j^{(1)} = \pi_j \tag{C.107}$$

IDENTIFICAÇÃO DE MODELOS

Em análise de séries temporais, um dos passos mais importante é a determinação do modelo que represente a série temporal disponível. Para tanto, é necessário conhecer com propriedade as funções FAC e FACP, que, na prática, são desconhecidas. Note que as equações dessas funções citadas até aqui só foram possíveis com o conhecimento *a priori* do modelo. Os padrões de comportamento das funções FAC e FACP podem indicar os modelos geradores do processo estocástico que os definem. Por exemplo, sabe-se que a função de autocorrelação do processo MA(1) apresenta um decaimento exponencial para a FAC e um truncamento “cut off” após o passo p para a função de autocorrelação parcial. Portanto, a representação dessas funções construída com base em uma realização indica que a realização seja possivelmente gerada por um processo estocástico MA(1).

Tabela C.2 –Características das funções FAC E FACP

Processo	FAC	FACP
AR(p)	Declínio exponencial ou onda amortecida	Truncamento após o passo p
MA(q)	Truncamento após o passo p	Declínio exponencial ou onda amortecida
ARMA(p,q)	Declínio exponencial	Declínio exponencial

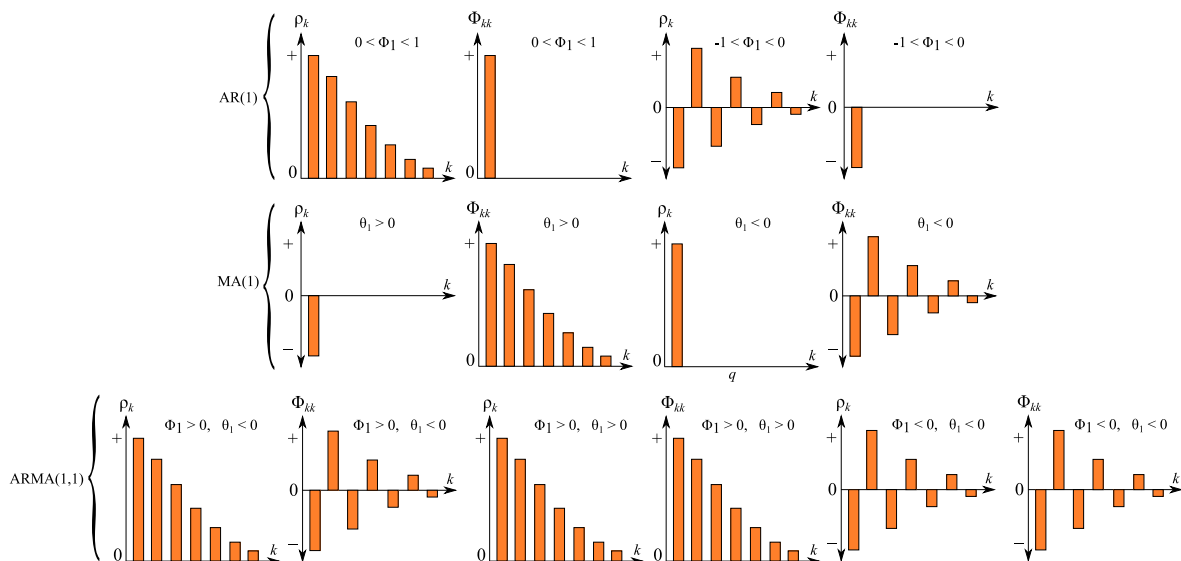


Figura C.4 – Padrões das funções FAC E FACP para os processos AR(1), MA(1) E ARMA(1,1)

A Tabela C.2 indica as características teóricas gerais das funções FAC e FACP para os processos $AR(p)$, $MA(q)$ e $ARMA(p,q)$ e a Figura C.4 exemplifica os padrões para os casos específicos $AR(1)$, $MA(1)$ e $ARMA(1,1)$.

PROCESSOS ESTOCÁSTICOS DE WIENER

Uma classe de processo não estacionário estocástico é chamada de processo de Wiener, também conhecida como movimento browniano. É um processo estocástico de tempo contínuo com três propriedades importantes:

- É um processo de Markov – o que implica dizer que o valor corrente tem informação suficiente para a melhor previsão do valor futuro;
- O incremento é independente – isso significa que as distribuições de probabilidade das alterações da variável em qualquer intervalo de tempo são independentes;
- As mudanças no processo em qualquer intervalo finito de tempo são normalmente distribuídas, com variância crescendo linearmente com o tempo.

Essa classe de processo é bastante difundida na previsão de variáveis econômicas e pode ser utilizada em modelos de demanda de energia agregada com modelos de econometria. Todos os processos de passeio aleatório, tanto de estado contínuo como discreto, com ou sem deriva, satisfazem à propriedade de um processo de Markov e são chamados, conseqüentemente, de processos Markovianos.

a) Passeio aleatório

O passeio aleatório é um caso particular do modelo $ARIMA(p,d,q)$. Se $p=0$, $d=1$ e $q=0$, a Eq. (C.82) se reduz a:

$$(1 - B)Z_t = \varepsilon_t \Leftrightarrow Z_t = Z_{t-1} + \varepsilon_t \quad (C.108)$$

Note que o passeio aleatório é um caso particular, também, do modelo $AR(1)$ quando $\phi_1=1$ – ver Eq. (C.55).

b) Movimento browniano com *drift*

O processo mais simples de Wiener é o movimento browniano com deriva, ou seja, *drift*:

$$dZ = \alpha dt + \sigma dz \quad (C.109)$$

Onde dz é o incremento de Wiener. Na Eq. (C.109), α é chamado de parâmetro do *drift*, e σ é o parâmetro da variância. Note que para qualquer intervalo Δt , a variação de x , denotado por Δx , é normalmente distribuída, com valor esperado $E(\Delta x) = \alpha \Delta t$ e variância $\text{Var}(\Delta x) = \sigma^2 \Delta t$. A Figura C.5 mostra quatro simulações da Eq. (C.109) com deriva $\alpha=0,2$ por ano e desvio padrão $\sigma=1$ por ano. Realizando uma projeção de 50 anos (por exemplo, de 1950 até 2000) com $x(0)=0$ e tomando amostras mensais, a trajetória para a variável $x(t)$ resulta na seguinte equação:

$$Z_t = Z_{t-1} + 0.01667 + 0.2887\varepsilon_t \quad (C.110)$$

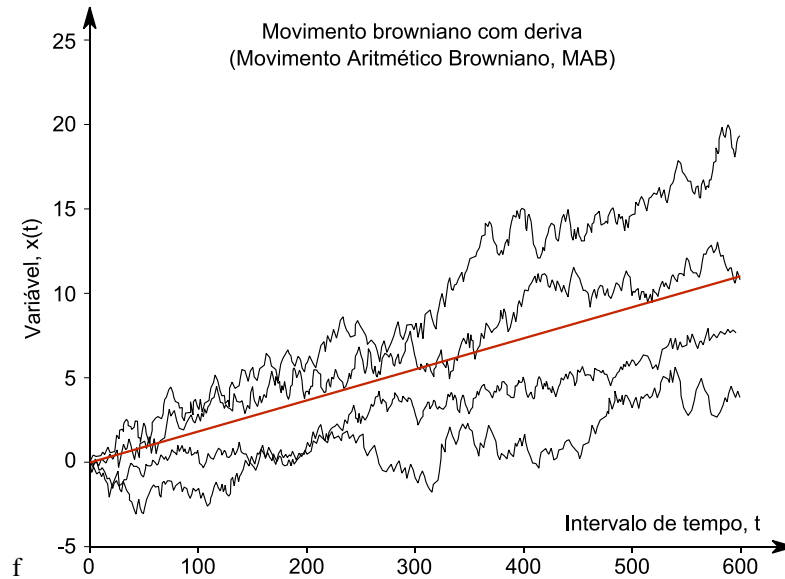


Figura C.5 – Exemplo de MB com tendência (período 1950 - 2000, amostragem mensal)

A Figura C.6 mostra a melhor estimativa de um processo estocástico do tipo MB com *drift*. Utilizando-se da Eq. (C.110) para representar as primeiras 288 amostras (24 anos de 12 meses), constroem-se curvas de previsão para as demais amostras (600–288=312). Por se tratar de um processo de Markov, é necessária apenas a última informação (dez/1974) para se construir uma curva de previsão. A previsão dos valores de x para um tempo T é dada pela seguinte equação:

$$\hat{x}_{1974+T} = x_{1974} + 0.01667T \tag{C.111}$$

O gráfico indica uma curva tipo MAB com três outras curvas de intervalos de confiança, ou seja, as trajetórias dos valores previstos para $x(t)$ são obtidas com valor indicado na Eq. (C.111) adicionado ou subtraído de um, dois ou três desvios-padrão.

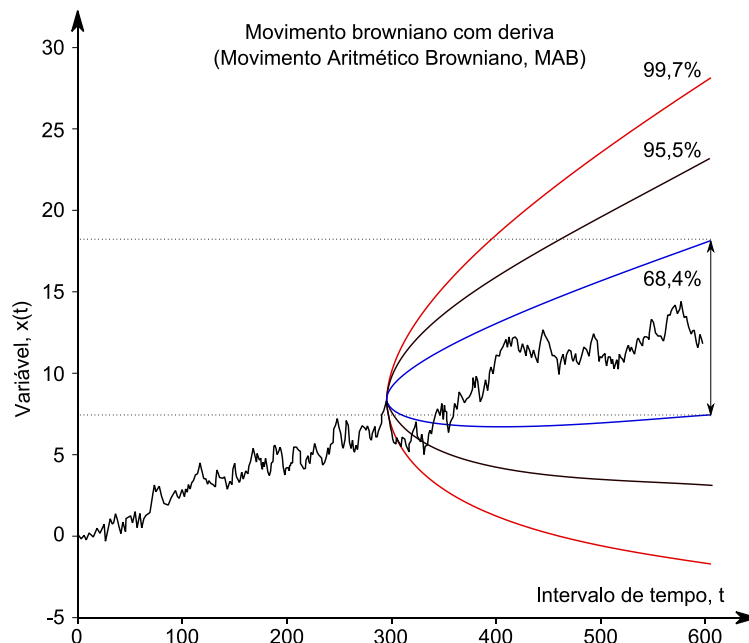


Figura C.6 – Melhor estimativa para um MB com *drift* (intervalo de confiança: 68,4%, 95,5% e 99,7%)

A variância do processo de Wiener cresce linearmente com o tempo e o desvio padrão com a raiz quadrada do tempo. Portanto, para uma trajetória com intervalo de confiança de 68,4% (um desvio-padrão) a previsão de T meses é:

$$x_{1974} + 0,01667T \pm 0,2887\sqrt{T} \quad (\text{C.112})$$

c) Movimento browniano generalizado – Processos de Itô

O processo de Wiener serve para representar uma gama vasta de processos aleatórios. Os processos indicados aqui são casos especiais do movimento Browniano generalizado com *drift*, cuja equação é:

$$dx = a(x,t)dt + b(x,t)dz \quad (\text{C.113})$$

Onde dz , novamente, é o incremento do processo de Wiener e $a(x,t)$ e $b(x,t)$ são funções não aleatórias conhecidas. Note que o *drift* e o coeficiente da variância são funções do tempo. O processo contínuo definido na Eq. (C.113) é chamado de processo de Itô.

Considere a média e a variância desse processo. Uma vez que $E(dz)=0$, $E(dx)=a(x,t)dt$, a variância de dx é igual a $E[dx^2]-(E[dx])^2$, o qual contém termos dt , $(dt)^2$ e termos em $(dt)(dz)$ com ordem $(dt)^{3/2}$. Para um dt infinitesimal, os termos em $(dt)^2$ e $(dt)^{3/2}$ podem ser ignorados e a variância resulta em:

$$\text{Var}[dx] = b^2(x,t)dt \quad (\text{C.114})$$

O termo $a(x,t)$ refere-se à taxa instantânea esperada do processo de Itô e o termo $b^2(x,t)$ como a taxa de variância instantânea.

d) Movimento Browniano geométrico

Um caso especial da Eq. (C.113) é o movimento geométrico Browniano com deriva, *drift*, em que $a(x,t)=\alpha x$ e $b(x,t)=\sigma t$, onde α e σ são constantes. Nesse caso, a Eq. (C.113) se torna:

$$dx = \alpha x dt + \sigma x dz \quad (\text{C.115})$$

Conforme já discutido, o percentual de variação de x , $\Delta x/x$, é normalmente distribuído. Já as variações expressas em logaritmo natural de x , variações absolutas de x e Δx têm distribuição lognormal. A relação entre x e seu logaritmo tem detalhes mais complicados nesse conceito. Nesta seção é mostrado que se $x(t)$ é dado pela Eq. (C.115), então $F(x)=\log x$ é um movimento Browniano simples com *drift*:

$$dF = \left(\alpha - \frac{1}{2}\sigma^2 \right) dt + \sigma dz \quad (\text{C.116})$$

Portanto, sobre um intervalo finito de tempo t , as variações de x em termos de logaritmo são normalmente distribuídas com média e variância, respectivamente:

$$\alpha - \frac{1}{2}\sigma^2, \quad \sigma^2 \quad (\text{C.117})$$

Para a variável $x(t)$ em si, pode ser demonstrado que se o valor corrente é $x(0)=x_0$, então o valor esperado de $x(t)$ é dado por:

$$E[x(t)] = x_0 e^{\alpha t} \tag{C.118}$$

E a variância $x(t)$ é dada por:

$$Var[x(t)] = x_0^2 e^{2\alpha t} (e^{\sigma^2 t} - 1) \tag{C.119}$$

O resultado da esperança do MGB pode ser usado para o cálculo do valor presente esperado descontado sobre um período de tempo. Por exemplo, note que:

$$E\left[\int_0^\infty x(t)e^{-rt} dt\right] = \int_0^\infty x_0 e^{-(r-\alpha)t} dt = x_0 / (r - \alpha) \tag{C.120}$$

fornece a taxa de desconto r que excede a taxa de crescimento α . Esse procedimento será usado quando for necessário calcular o valor presente descontado de um fluxo de caixa de rendimentos que segue um movimento geométrico Browniano.

O MGB é frequentemente usado para modelar preços de ativos financeiros, taxas de atratividade ou de juros, salários, preço de produtos e outras variáveis econômicas e financeiras. A Figura C.7 mostra três simulações da Eq. (C.115) considerando um *drift* $\alpha=0,09$ (9% ao ano) e $\sigma=0,2$ (20% ao ano). Os valores adotados referem-se à taxa de crescimento anual esperada e ao desvio padrão do índice da bolsa NYMEX, respectivamente. Nesse exemplo, o tempo refere-se ao período entre 1950 e 2000, tomando amostras mensais. Assim, tem-se a seguinte equação para representar a variável aleatória $x(t)$:

$$x_t = 1,0075x_{t-1} + 0,0577x_{t-1}\varepsilon_t \tag{C.121}$$

Com $x_{1950}=100$. Note que 9% *a.a.* equivale a 0,75% *a.m.* (1/12 de 9%) e 0,2 de desvio padrão ao ano equivale a 0,13, ou seja, raiz quadrada de 0,2/12 e ε_t é tomado de uma distribuição normal com média zero e desvio padrão unitário. Assim, o valor 0,0577 representa α/σ .

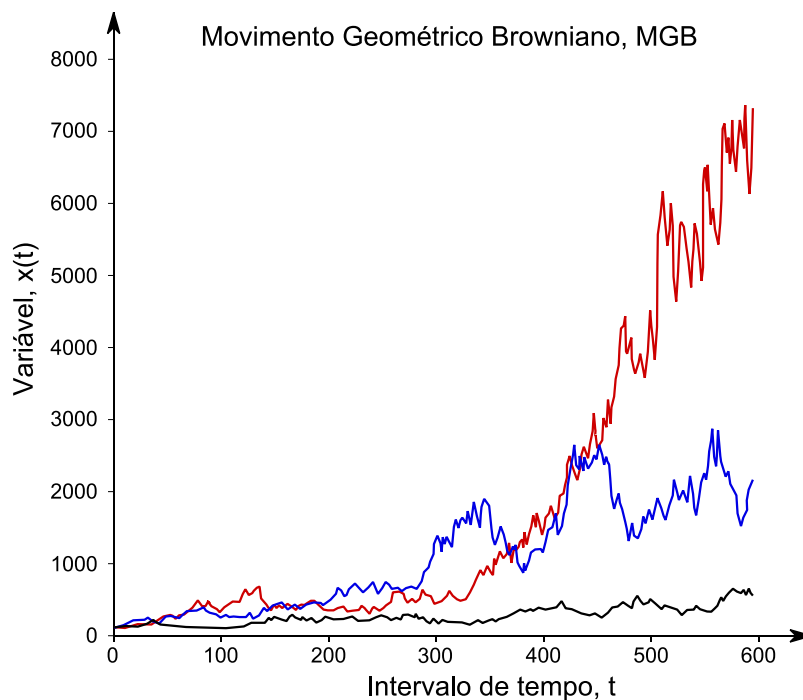


Figura C.7 – Exemplos de MGB (período 1950 - 2000, amostragem mensal)

Na Figura C.8 estão indicadas a linha de tendência média e as linhas de previsões com intervalo de confiança de 66% (dados amostrados a partir de 1970 – 300 amostras mensais), sendo que as curvas superior e inferior foram feitas a partir de 1974 (48 amostras iniciais). Como o MGB segue um processo de Markov, apenas o valor $x(t)$ de dezembro de 1974 é necessário para construir as curvas de previsão. A linha de tendência média é dada por:

$$\hat{x}_{1974+T} = (1.0075)^T x_{1974} \quad (\text{C.122})$$

Onde T é dado em meses e as curvas limites do intervalo de confiança seguem as equações indicadas em (C.123).

$$(1.0075)^T (1.0577)^{\sqrt{T}} x_{1974}, \quad \text{e} \quad (1.0075)^T (1.0577)^{-\sqrt{T}} x_{1974} \quad (\text{C.123})$$

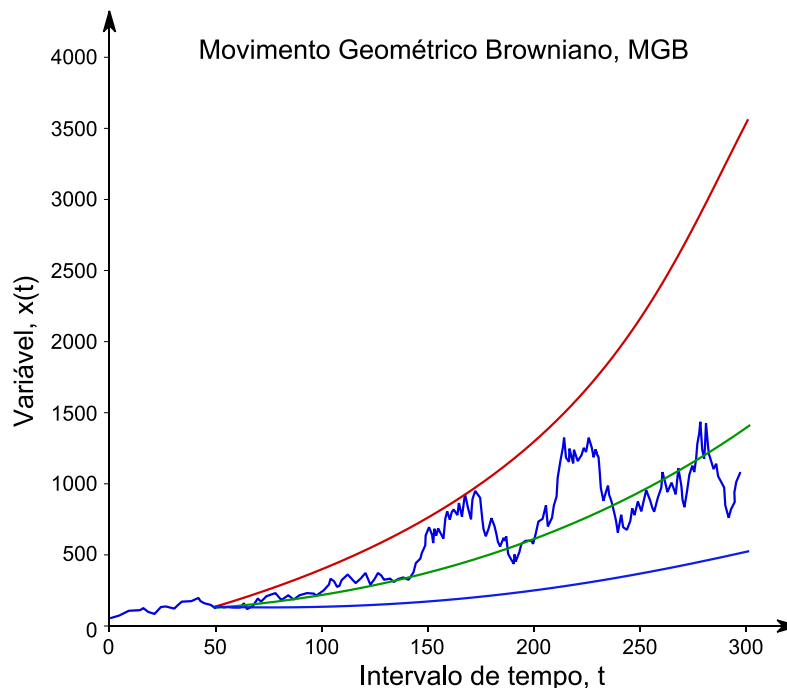


Figura C.8 – Melhor estimativa para um MGB (intervalo de confiança de 66%, período de 1970 a 1995, amostragem mensal)

e) Processo estocástico com reversão à média

A tendência do MB é afastar-se do ponto inicial. De fato, isso é compatível com algumas variáveis econômicas – por exemplo, com preços de ativos especulativos – mas não com outras. Considere por exemplo, o preço de uma lista de commodities tal como o aço e o petróleo. Embora se possam modelar esses preços com MB, pode-se argumentar que tais preços estão relacionados no longo prazo ao custo marginal de produção. Em outras palavras, embora o preço do óleo, por exemplo, possa sofrer altas e quedas no mercado internacional⁴, no longo prazo o preço dessa commodity tende ao seu custo marginal. Essa condição leva ao modelo estocástico com reversão à

⁴ Por exemplo, o preço do óleo pode flutuar em função de guerras e/ou revoluções envolvendo países produtores de petróleo ou mesmo devido a determinações dos países da OPEP (Organização dos Países Exportadores de Petróleo). Apesar dessa modelagem ser utilizada para essa commodity, é importante observar que os preços marginais em cada país produtor diferem substancialmente da cotação internacional, uma vez que os custos de exploração são diversificados em função das condições geológicas de cada jazida.

média. O mais simples modelo de processo estocástico de reversão à média – também conhecido como processo de Ornstein-Uhlenbeck – segue a Eq. (C.124):

$$dx = \eta(\bar{x} - x)dt + \sigma dt \quad (\text{C.124})$$

Onde η é a velocidade da reversão e \bar{x} é o nível normal da variável x , isto é, o nível para o qual x tende a reverter. Se a variável x representa o preço de uma commodity, então \bar{x} pode ser o seu custo marginal de longo prazo. Note que a variação esperada de x depende da diferença entre x e \bar{x} . Se x é maior (ou menor) que \bar{x} , é mais provável que o preço caia (ou suba) no próximo intervalo de tempo. Portanto, embora esse processo satisfaça a propriedade de um processo de Markov, os incrementos não são independentes. Se o valor corrente $x(0)$ é x_0 e a variável x segue o processo da Eq. (C.124), então o valor esperado em qualquer tempo t é:

$$E[x(t)] = \bar{x} + (x_0 - \bar{x})e^{-\eta t} \quad (\text{C.125})$$

E a variância de $(x_t - \bar{x})$ é:

$$\text{Var}[x_t - \bar{x}] = \frac{\sigma^2}{2\eta}(1 - e^{-2\eta t}) \quad (\text{C.126})$$

Note que o valor esperado de x_t converge para \bar{x} à medida que o tempo cresce e a variância converge para $\sigma^2/2\eta$. Se $\eta \rightarrow \infty$, $\text{Var}[x_t] \rightarrow 0$, o que implica que a variável x nunca se desvia do valor \bar{x} , mesmo que momentaneamente, e se $\eta \rightarrow 0$, $\text{Var}[x_t] \rightarrow \sigma^2/t$, ou seja, comporta-se como um processo de movimento Browniano simples.

A Eq. (C.124) é uma versão de um processo autorregressivo de primeira ordem de tempo discreto, AR(1). Especificamente, a Eq. (C.124) é o caso limite quanto $\Delta t \rightarrow 0$ do seguinte processo AR(1):

$$x_t - x_{t-1} = \bar{x}(1 - e^{-\eta}) + (e^{-\eta} - 1)x_{t-1} + \varepsilon_t \quad (\text{C.127})$$

Onde ε_t é distribuído normalmente com média zero e desvio padrão σ_ε , e:

$$\sigma_\varepsilon^2 = \frac{\sigma^2}{2\eta}(1 - e^{-2\eta}) \quad (\text{C.128})$$

O parâmetro da Eq. (C.124) pode ser estimado usando dados de tempo discreto disponíveis na seguinte regressão:

$$x_t - x_{t-1} = a + bx_{t-1} + \varepsilon_t \quad (\text{C.129})$$

E a partir da equação $\bar{x} = -a/b$, $\hat{\eta} = -\log(1 + \hat{b})$, e:

$$\hat{\sigma} = \hat{\sigma}_\varepsilon \sqrt{\frac{\log(1 + \hat{b})}{(1 + \hat{b})^2 - 1}} \quad (\text{C.130})$$

Onde $\hat{\sigma}_\varepsilon$ é o erro padrão da regressão.

É fácil generalizar a Eq. (C.124). Por exemplo, se pode esperar que $x(t)$ convirja para \bar{x} conforme Eq. (C.124), mas a taxa de variância cresce com x . Então, pode-se usar o seguinte processo:

$$dx = \eta(\bar{x} - x)dt + \sigma x dt \quad (\text{C.131})$$

Alternativamente, modificações proporcionais na variável podem ser modeladas como um simples processo de reversão à média. Isso equivale a descrever $x(t)$ com o seguinte processo:

$$dx = \eta x(\bar{x} - x)dt + \sigma x dt \quad (\text{C.132})$$

Os diferentes processos de reversão à média têm implicações nas decisões de investimento. A seguinte questão pode ser levantada: os preços das commodities e outros bens são modelados adequadamente como um processo de MB ou por um processo de reversão à média? Observando dados dos preços de petróleo de um longo tempo (120 anos, por exemplo), é fácil sugerir que esses preços seguem um processo de reversão à média com baixa taxa de velocidade. Utilizando testes de raiz unitários (ao invés do teste t) com 120 anos de dados, é fácil rejeitar a hipótese de passeio aleatório. No entanto, para dados de 30 ou 40 anos, o teste de rejeição falha. Esse parece ser o caso de muitas outras variáveis econômicas, pois, usando poucos dados, é difícil distinguir estatisticamente os processos entre passeio aleatório e processo de reversão à média.

LEMA DE ITÔ

O processo indicado na Eq. (C.113) é de tempo contínuo e não diferenciável. No entanto, em muitas situações se deseja trabalhar com funções de processo estocástico com suas derivadas. Por exemplo, quando se deseja obter o valor da opção de investimento de uma plataforma como função do preço do óleo quando esta função é representada por um processo estocástico tipo MGB. Para isso, haverá a necessidade de diferenciar e/ou integrar funções gerais do processo de Itô. Isso é feito utilizando-se do Lema de Itô.

O Lema de Itô é representado facilmente por uma série de expansão de Taylor. Suponha que $x(t)$ segue o processo indicado na Eq. (C.113) e que uma função $F(x,t)$ seja duas vezes diferenciável em relação a x e uma em relação a t . A regra do cálculo elementar define a diferencial de primeira ordem de F como:

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial t} dt \quad (\text{C.133})$$

Incluindo diferenciais de ordem superior em relação à x , resulta que:

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial t} dt + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (dx)^2 + \frac{1}{6} \frac{\partial^3 F}{\partial x^3} (dx)^3 + \dots \quad (\text{C.134})$$

Os termos de ordem superior (terceira, quarta etc) podem ser desprezados. Assim, o Lema de Itô fica definido pela seguinte equação:

$$dF = \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial x} dx + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (dx)^2 \quad (\text{C.135})$$

Substituindo dx da equação que define o processo de Itô – Eq. (C.113) –, resulta em:

$$dF = \left[\frac{\partial F}{\partial t} + a(x,t) \frac{\partial F}{\partial x} + \frac{1}{2} b^2(x,t) \frac{\partial^2 F}{\partial x^2} \right] dt + b(x,t) \frac{\partial F}{\partial x} dz \quad (\text{C.136})$$

Pode-se facilmente estender a expansão da série de Taylor para funções de vários processos de Itô. Por exemplo, supondo que $F=F(x_1, x_2, \dots, x_m, t)$ seja uma função temporal de m processos de Itô x_1, x_2, \dots, x_m , onde:

$$dx_i = a_i(x_1, x_2, \dots, x_m, t)dt + b_i(x_1, x_2, \dots, x_m, t)dz_i, \quad i = 1, 2, \dots, m \quad (C.137)$$

Com $E(dz_i, dz_j) = \rho_{ij}dt$. Então, aplicando-se o Lema de Itô, resulta a seguinte diferencial dF :

$$dF = \frac{\partial F}{\partial t} dt + \sum_{i=1}^m \frac{\partial F}{\partial x_i} dx_i + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 F}{\partial x_i \partial x_j} dx_i dx_j \quad (C.138)$$

Assim, substituindo dx_i da Eq. (C.137), resulta:

$$dF = \left[\frac{\partial F}{\partial t} + \sum_{i=1}^m a_i(x_1, x_2, \dots, x_m, t) \frac{\partial F}{\partial x_i} + \frac{1}{2} \sum_{i=1}^m b_i^2(x_1, x_2, \dots, x_m, t) \frac{\partial^2 F}{\partial x_i^2} + \frac{1}{2} \sum_{i \neq j}^m \rho_{ij} b_i(x_1, x_2, \dots, x_m, t) \frac{\partial^2 F}{\partial x_i \partial x_j} \right] dt + \sum_{i=1}^m b_i(x_1, x_2, \dots, x_m, t) \frac{\partial F}{\partial x_i} dz_i \quad (C.139)$$

EXEMPLO C.5 – Considere a função $F(x) = x^\beta$, onde x segue um comportamento tipo MGB de acordo com a Eq. (C.115). Indique como calcular a esperança do valor presente descontado de uma taxa r definida por:

$$E \left[\int_0^\infty F(x(t)) e^{-rt} \right] \quad (C.140)$$

Solução

O Lema de Itô pode ser aplicado na função F , resultando em:

$$dF = \beta x^{\beta-1} [\alpha x dt + \sigma x dz] + \frac{1}{2} \beta(\beta-1) x^{\beta-2} \sigma^2 x^2 dt \quad (C.141)$$

Simplificando, chega-se a:

$$dF = \left[\beta\alpha + \frac{1}{2} \beta(\beta-1)\sigma^2 \right] F dt + \beta\sigma F dz \quad (C.142)$$

Note que a Eq. (C.142) indica que F segue um MGB. Assim, pode-se usar a Eq. (C.120) para dedução do valor esperado, ou seja:

$$dF = \frac{x_o^\beta}{\left[r - \beta\alpha - \frac{1}{2} \beta(\beta-1)\sigma^2 \right]} \quad (C.143)$$

Desde que o denominador seja positivo.

Para outros procedimentos de séries temporais, incluindo séries temporais multivariadas, o leitor pode consultar Montgomery e Johnson (1976) e Morettin e Toloí (1985). Nessas referências podem-se consultar os passos da metodologia de Box e Jenkins (1976) e os testes estatísticos dos parâmetros estimados e os resíduos.

REFERÊNCIAS E LEITURAS SUGERIDAS

- Box, G. E. P., and G. M. Jenkins. 1976. *Time Series Analysis, Forecasting, and Control*. San Francisco: Holden-Day, Inc.
- BUENO, R. L. S. 2012. *Econometria de Séries Temporais*. 2nd ed. São Paulo: Editora Cengage Learning.
- Montgomery, C.D., and L.A. Johnson. 1976. *Forecasting and Time Series Analysis*. New York: McGraw-Hill.
- Morettin, P.A., and C.M.C. Toloi. 1985. *Previsão de Séries Temporais*. 2nd ed. São Paulo: Atual Editora.
- Wei, W. W. S. 2006. *Time Series Analysis – Univariate and Multivariate Methods*. 2nd ed. Boston: Pearson Education.