

2. Séries Temporais

2.1. Definição

Um processo estocástico é uma função aleatória Y_t que evolui no tempo (e/ou no espaço), definida sob um espaço de probabilidades. Mais precisamente, um processo estocástico é uma família $Y = \{Y(t), t \in T\}$ tal que para cada $t \in \mathbb{R}$, $Y(t)$ é uma variável aleatória. Se $T \equiv \mathbb{Z} = \{1, \dots, t\}$, diz-se que o processo é de parâmetro discreto e é denotado por Y_t . Se $T \in \mathbb{R}$, diz-se que o processo é de parâmetro contínuo e é denotado por $Y(t)$.

Suponha que para um dado valor de t , Y_t se origine de um experimento que pode ser repetido em condições idênticas. Então a cada repetição y_t obtém-se um registro de valores de Y_t no tempo que é uma realização de um processo estocástico. Sendo assim, uma série temporal, que é um conjunto de observações de uma determinada variável Y geradas sequencialmente no tempo, também pode ser definida como sendo uma parte da trajetória, ou de uma realização parcial, uma amostra finita, de um processo estocástico. Uma série temporal $\{y_t\}_{t=1}^T$ é uma realização amostral dentre todas as séries possíveis de tamanho T que poderiam ter sido geradas por um mesmo processo estocástico, SOUZA & CAMARGO (2004).

Considere um processo estocástico $\{Y_t, t \in \mathcal{T}\}$, onde \mathcal{T} é o conjunto índice que controla a evolução do tempo. A média de Y_t é definida por $\mu_t = E(Y_t)$, onde $E(\cdot)$ denota o operador ‘esperança’. A covariância entre Y_t e Y_s é dada por $\gamma(t, s) = Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)]$, e $\gamma(t, t) = Cov(Y_t, Y_t)$ é a variância de Y_t .

Um processo estocástico está estatisticamente determinado quando se conhecem as funções de distribuição conjunta até a T -ésima ordem. Na prática ocorrem duas situações problemáticas: não se conhecem todas as funções de distribuição até a T -ésima ordem e, comumente tem-se apenas uma realização do processo estocástico em

questão, a partir da qual se deseja inferir todas as características do mecanismo gerador da série. Para superar estas dificuldades, assumem-se duas restrições: Estacionariedade e Ergodicidade. Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio. A definição mais precisa para a estacionariedade é dada a seguir.

Quando a covariância entre dois membros de um processo estocástico depender unicamente da distância temporal entre eles, e este processo tem média μ e variância σ^2 constantes, este processo estocástico é dito ser um processo **(fracamente) estacionário, ou estacionário de segunda ordem**, ou seja, é um processo que preserva suas características ao longo do tempo, $Y_t \sim \text{Dist}(\mu, \sigma^2)$ para todo t . Mais formalmente, uma série temporal $\{y_t, t \in \mathcal{T}\}$ é estacionária se:

- a) $E(y_t) = \mu_t, \forall t \in \mathcal{T}$,
- b) $\text{Var}(y_t) = \sigma^2, \forall t \in \mathcal{T}$.
- c) $\text{Cov}(y_t, y_{t+k}) = \gamma(k)$, ou seja, depende só de $k, \forall t, t+k \in \mathcal{T}$ e $k \neq 0$.

Existe uma definição mais forte, a de Estacionariedade Estrita. Uma série temporal é dita ser **estritamente (ou fortemente) estacionária** se as funções de distribuição conjuntas de $\{y_{t_1}, \dots, y_{t_m}\}$ e $\{y_{t_1+k}, \dots, y_{t_m+k}\}$ são idênticas para todo inteiro positivo m e para todos $t_1, \dots, t_m, t_1+k, \dots, t_m+k \in \mathcal{T}$. Ou seja, se as funções de distribuição conjuntas de $\{y_{t_1}, \dots, y_{t_m}\}$ e $\{y_{t_1+k}, \dots, y_{t_m+k}\}$ são iguais pra períodos de tempo de mesma duração. Neste caso suas estatísticas não são afetadas por variações devido à escolha da origem dos tempos, pois y_t e y_{t+k} são identicamente distribuídas para qualquer k . Todavia em geral, não é possível testar se este tipo de estacionariedade se verifica, pois na maioria das vezes a função de distribuição é desconhecida.

E uma série é **estritamente estacionária de ordem finita**, quando para um determinado valor i , a estacionariedade estrita do processo não é válida para todo $t_j \in T$, mas apenas para $j \leq i$.

Por **ergodicidade** entende-se a condição em que apenas uma realização do processo é suficiente para obter todas as estatísticas do mesmo. Algumas propriedades dos processos estacionários que também se aplicam aos processos ergódicos (ambos no

sentido amplo) são: média e variância constantes; função de autocorrelação e autocovariância independentes da origem dos tempos.

Uma série temporal pode ainda ser caracterizada quanto à presença de componentes de tendência, componentes sazonais, componentes cíclicas e componentes de ruído.

O movimento sistemático de aumento ou decréscimo, linear ou não linear, de uma série temporal ao longo do tempo é chamado de **tendência**. A tendência representa o efeito de longo prazo ao redor da média, ou seja, o movimento dominante em uma série temporal.

Entende-se por **sazonalidade**, por sua vez, flutuações periódicas recorrentes que podem aparecer em períodos específicos do ano, quando as observações são intra- anuais, ou seja, registradas mensalmente, trimestralmente ou semanalmente, por exemplo. MORETTIN & TOLOI (2006, p.64) afirmam que “é difícil definir, tanto do ponto de vista conceitual como estatístico, o que seja sazonalidade” por isso consideram como sazonais os fenômenos que ocorrem regularmente de ano para ano é uma abordagem empírica. Estes autores afirmam também que as séries sazonais são caracterizadas por apresentarem correlação alta em “lags sazonais”, isto é, intervalos múltiplos de doze meses no caso das séries de valores mensais. Portanto, a sazonalidade deve ser visualizada em gráficos de autocorrelação parcial das séries onde os picos significantes em intervalos regulares durante o ano confirmam a presença da componente sazonal.

Componentes **cíclicos** são variações com comportamento similar à componente sazonal, caracterizadas por movimentos variáveis periódicos ou oscilatórios, mas que ao se relacionam diretamente a uma medida temporal, e está suscetível a influências de fatores externos. Por este motivo, os ciclos não possuem duração uniforme e assim, tipicamente, o movimento cíclico possui um comprimento maior do que o observado na variação sazonal, períodos superiores a um ano.

Componentes de **ruído** são constituídas por movimentos ascendentes ou descendentes, de grande instabilidade e aleatórios não explicados por variações cíclicas ou pela tendência, devidos ao acaso ou fatores aleatórios.

Na análise de séries temporais, é comum realizar a identificação destas componentes, através da decomposição das componentes da série investigada. A maneira clássica de decomposição fundamenta-se em um modelo aditivo ou multiplicativo envolvendo os componentes da série. A decomposição aditiva pode ser representada através da expressão $Y_t = m_t + S_t + \varepsilon_t$, e caso o modelo seja multiplicativo, pode-se representar a decomposição como $Y_t = m_t S_t \varepsilon_t$, em que m_t e S_t representam as componentes de tendência e sazonalidade (incluindo a variação cíclica), respectivamente e ε_t retrata o ruído ou erro aleatório. A SSA aplicada à série temporal visa fazer uma decomposição aditiva da série através da análise espectral singular da série.

Assim como ocorre nos procedimentos desenvolvidos para observações independentes, as séries temporais podem ser univariadas ou multivariadas. A abordagem desta tese é toda dedicada a séries univariadas, em tempo discreto com observações no conjunto contínuo e em intervalos equidistantes no tempo.

2.2. O Operador de Diferenças

Dada uma série temporal $\{y_t\}_{t=1}^T$, a primeira diferença da série é definida por $\nabla y_t = (1 - B)y_t = y_t - B(y_t)$, onde $B(y_t) = y_{t-1}$. $\nabla y_t = y_t - y_{t-1}$.

A segunda diferença da série é dada por $\nabla^2 y_t = \nabla(\nabla y_t) = \nabla(y_t - y_{t-1}) = \nabla y_t - \nabla y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$.

E recursivamente, a n -ésima diferença da série é dada por:

$$\nabla^n y_t = (1 - B)^n y_t = \sum_{r=0}^n (-1)^r \binom{n}{r} y_{t-r}, \text{ onde } \binom{n}{r} = \frac{n!}{r!(n-r)!}. \quad (1)$$

Convém ressaltar que, ao se tomar n diferenças da série $\{y_t\}_{t=1}^T$, são perdidas n observações.

2.3. Processo Ruído Branco

Um processo estocástico $\{\varepsilon_t, t \in \mathcal{T}\}$ de variáveis aleatórias não correlacionadas com média zero e variância finita e constante é chamado de ruído branco e é denotado por $\varepsilon_t \sim RB(0, \sigma^2)$.

2.4. Processo Passeio Aleatório

Seja $\{y_t\}_{t=1}^T$ uma série dada por $y_t = y_{t-1} + \varepsilon_t$, onde $\varepsilon_t \sim RB(0, \sigma^2)$. Este é um processo não estacionário conhecido como passeio aleatório (“random walk”). Logo, em um passeio aleatório, o valor da série no instante t é o valor da série no instante anterior mais um erro de média zero, com variância constante.

Sendo assim, recursivamente o passeio aleatório pode ser escrito como:

$$y_t = y_0 + \sum_{j=1}^t \varepsilon_j, \quad (2)$$

de forma que o valor esperado, ou seja, a média de y_t em qualquer instante é igual ao valor y_0 , uma constante. E a variância y_t é dada por

$$Var(y_t) = Var\left(\sum_{j=1}^t \varepsilon_j\right) = t\sigma^2 \cdot \sum_{j=1}^t \varepsilon_j, \quad (3)$$

ou seja, à medida que t aumenta, a variância de y_t cresce indefinidamente, o que viola uma das condições de estacionariedade. No entanto, tirando a primeira diferença do passeio aleatório, a nova série passa ser um ruído branco, que é uma série estacionária.

Uma das características do passeio aleatório é a persistência dos choques aleatórios, isto é, o efeito de cada termo de erro ε_t não se “dissipa” ao longo do tempo. Assim, podemos dizer que o passeio aleatório tem memória infinita, pois “guarda” a informação de todos os choques sofridos até o período corrente.

Um passeio aleatório com deslocamento, ou *drift*, δ é dado por

$$\begin{aligned} y_t &= y_{t-1} + \delta + \varepsilon_t \\ &= y_0 + \delta t + \sum_{j=1}^t \varepsilon_j \end{aligned} \quad (4)$$

onde $\varepsilon_t \sim RB(0, \sigma^2)$. Com a inserção do deslocamento δ , y_t se desloca para cima ou para baixo dependendo de δ ser positivo ou negativo; δ especifica assim a rapidez com que a componente de tendência determinística cresce.

A média e variância de um passeio aleatório com deslocamento δ são dadas por $E(y_t) = t\delta$ e $Var(y_t) = t\sigma^2$. Ou seja, nem a média nem a variância do passeio aleatório com deslocamento são constantes, e assim ele é um processo claramente não estacionário. Porém, escrevendo a primeira equação em termos da primeira diferença tem-se $y_t - y_{t-1} = \delta + \varepsilon_t$, a primeira diferença do passeio aleatório com deslocamento é um processo estacionário, como se verifica também para o passeio aleatório sem deslocamento.

2.5. Modelos ARIMA

A análise de séries temporais, segundo o enfoque de BOX & JENKINS (1976), atualizado em BOX, JENKINS & REINSEL (1994), tem como objetivo principal a realização de previsão. Essa metodologia permite que valores futuros de uma série sejam previstos tomando por base apenas seus valores presentes e passados. Isso é feito explorando a correlação temporal que existe geralmente entre os valores exibidos pela série. A relação temporal considerada pelo enfoque de Box & Jenkins é representada formalmente por um conjunto de processos estocásticos genericamente denominados modelos ARIMA (termo oriundo da expressão *Auto Regressive Integrate Moving Average*). Por envolverem apenas uma série de tempo, eles são classificados como modelos univariados.

Os modelos ARIMA resultam da combinação de três componentes também denominados “filtros”: o componente Auto Regressivo (AR), o filtro de Integração (I) e o componente de Médias Móveis (MA). Uma série de tempo pode conter os três filtros ou apenas um subconjunto deles, resultando daí várias alternativas de modelos passíveis de análise pela metodologia Box & Jenkins.

Seja $\{y_t\}_{t=1}^T$ uma realização do processo estocástico $\{Y_t\}_{t=1}^T$, isto é, uma série temporal estacionária de segunda ordem que não apresenta movimentos sazonais. Um modelo que pode ser usado para modelar o comportamento dinâmico desta série é o modelo auto-regressivo de ordem p , $AR(p)$ definido por:

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (5)$$

onde $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros auto-regressivos; c é uma constante (a serem estimados) e $\varepsilon_t \sim RB(0, \sigma^2)$.

Outro modelo comumente usado é o modelo de Médias Móveis de ordem q , $MA(q)$, dado por:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (6)$$

onde $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros de médias móveis; c é uma constante (a serem estimados) e $\varepsilon_t \sim RB(0, \sigma^2)$.

Para obter maior flexibilidade no ajuste de uma série temporal, às vezes é vantajoso incluir tanto termos auto-regressivos como termos de médias móveis. Esta composição caracteriza o modelo auto-regressivo e de médias móveis $ARMA(p, q)$, onde p é a ordem autorregressiva e q é a ordem de média móveis, dado por:

$$y_t = C + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (7)$$

onde C é uma constante, $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros auto regressivos; $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros de médias móveis, a serem estimados e $\varepsilon_t \sim RB(0, \sigma^2)$.

O modelo em (7) pode ser representado em termos de dois polinômios, os quais são obtidos com o defasamento $\nabla = (1 - B)$, onde B é definido por $B^d(y_t) = y_{t-d}$. Assim uma nova representação para (7) seria:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(y_t - \mu) = (1 - \theta_1 B - \dots - \theta_q B^q)\varepsilon_t, \quad (8)$$

onde, $\mu, \phi_k \in \mathbb{R}$ e $\theta_j \in \mathbb{R}$ denotam os parâmetros do modelo de Box & Jenkins a serem estimados, e ε_t o ruído em t .

A verificação da estacionariedade de segunda ordem em uma série temporal pode ser realizada através da análise de perfil do gráfico da função de autocorrelação simples ρ_k , a qual é definida em (9).

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}. \quad (9)$$

onde \bar{y} é a média da série temporal $\{y_t\}_{t=1}^T$; e k , a defasagem da autocorrelação. A determinação das ordens p e q pode ser realizada por meio de uma análise do perfil dos gráficos das funções de autocorrelação (ρ_k) - ACF (*autocorrelation function*) - e de autocorrelação parcial (φ_{kk}) - PACF (*partial autocorrelation function*) (HAMILTON, 1994; MORETTIN & TOLOI, 2006).

Existem testes específicos para estacionariedade, os famosos “testes da raiz unitária”, cujo detalhamento teórico foge ao objetivo deste trabalho. O mais simples deles, baseado na função de autocorrelação, requer o cálculo da estatística de Ljung-Box. Outro procedimento muito adotado é o teste Aumentado de Dickey-Fuller (ADF). GUJARATI (2000) explicita que ele pode ser aplicado a processos com termo de erro autocorrelacionado. Detalhes sobre o teste ADF e sua implementação no software EViews 5.0 são apresentados por SEILER (2004, p. 279). Outro teste da raiz unitária comumente usado, disponível nos softwares R e Eviews é o teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS).

Caso a série em questão não seja estacionária, a metodologia de Box & Jenkins aplica-se a também a um subgrupo específico de séries não estacionárias: séries que se tornam estacionárias após a aplicação de diferenças. O número de diferenças necessárias para tornar uma série estacionária é chamado ordem de integração da série. Um processo é dito ser integrado de ordem d se $\{y_t\}_{t=1}^T$ é não estacionária, mas a série resultante após d diferenças em $\{y_t\}_{t=1}^T$ é estacionária. Se esta série resultante das d diferenças pode ser modelada por um modelo ARMA(p,q) tem-se a seguinte formulação:

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)[(1 - B)^d y_t - \mu] = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t, \quad (10)$$

e dizemos que $\{y_t\}_{t=1}^T$ é um processo autoregressivo integrado de médias móveis, um ARIMA (p,d,q), onde o parâmetro de diferenças d , por definição, assume valores inteiros positivos.

Quando no modelo ARIMA (p, d, q) o parâmetro d assume valores fracionários, então o processo pode ter um comportamento de longa dependência (HOSKING, 1981) sendo de longa dependência se $d \in (0; 0,5)$ e de curta dependência se $d \in (-0,5; 0)$. Modelos com esta característica são conhecidos como modelos autorregressivos integrados fracionalmente e de médias móveis ARFIMA (p, d, q) .

Os modelos de Box & Jenkins podem ser utilizados para a modelagem de séries temporais que apresentam sazonalidade (HIPEL & McLEOD, 1994). Supondo que $\{y_t\}_{t=1}^T$ apresente sazonalidade, o modelo Box & Jenkins para $\{y_t\}_{t=1}^T$ é dado por

$$\varphi(B)(1 - \Phi_1 B - \dots - \Phi_p B^{PS}) \nabla^d (1 - B^S)^D y_t = \theta(B)(1 - \theta_1 B - \dots - \theta_q B^{QS}) \varepsilon_t, \quad (11)$$

onde: $\varphi(B) = (1 - \varphi_1 B - \dots - \varphi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, d são os polinômios das diferenças simples; D é a ordem das diferenças sazonais; S é o período sazonal; $\varphi_k \in \mathbb{R}$, e $\theta_j \in \mathbb{R}$ são, respectivamente, os coeficientes dos polinômios não sazonais; e $\Phi_m \in \mathbb{R}$ e $\Theta_n \in \mathbb{R}$ são, respectivamente, os coeficientes dos polinômios sazonais (HAMILTON, 1994; MORETTIN, 1997). Os modelos sazonais são descritos como modelos sazonais autorregressivos integrados e de médias móveis SARIMA $(p, d, q) \times (P, D, Q)_S$.

Assim como no modelo ARIMA (p, d, q) , se pelo menos um dos dois parâmetros d ou D assume valores fracionários, temos um modelo sazonal autoregressivo fracionalmente integrado e de médias móveis SARFIMA $(p, d, q) \times (P, D, Q)_S$.

2.6. Medidas de Qualidade do Ajuste

Seja a série temporal $\{y_t\}_{t=1}^T$, modelada por uma abordagem que resultou uma série $\{\hat{y}_t\}_{t=1}^T$ de valores ajustados por esta abordagem. A acuidade do método de previsão utilizada nesta abordagem é comumente mensurada pelas medidas (métricas) citadas a seguir:

a) Erro absoluto médio (*Mean Absolute Deviation-MAD*):

$$MAD = \frac{\sum_{i=1}^T |y_i - \hat{y}_i|}{T} \quad (12)$$

b) Raiz do erro quadrado médio (*Root Mean Square Error – RMSE*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^T (y_i - \hat{y}_i)^2}{T}} \quad (13)$$

c) Raiz do erro quadrado médio Padronizado (*RMSEP*) :

$$RMSEP = \sqrt{\frac{\sum_{i=1}^T (y_i - \hat{y}_i)^2}{\sum_{i=1}^T (y_i - \bar{y})^2}} \quad (14)$$

d) Erro de previsão (*Forecast Error*):

$$FE = \sqrt{\frac{\sum_{i=1}^T (y_i - \hat{y}_i)^2}{T-k}} \quad (15)$$

onde k é o número de parâmetros do modelo.

e) Erro percentual absoluto médio (*Mean Absolute Percentual Error - MAPE*):

$$MAPE = \frac{\sum_{i=1}^T \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{T} \times 100 \quad (16)$$

f) Coeficiente de explicação R^2 : mede o quanto a variação total dos dados é explicada pelo modelo. É calculado através da comparação do erro do modelo e a variação dos dados da série em torno da sua média:

$$R^2 = \left(1 - \frac{\sum_{i=1}^T (y_i - \hat{y}_i)^2}{\sum_{i=1}^T (y_i - \bar{y})^2} \right) \times 100 \quad (17)$$

g) Critério de Informação Bayesiano (*Bayesian Information Criterion – BIC*):

$$BIC = -2\ln L(\psi) + [2k\ln(T)]/T \quad (18)$$

onde $L(\psi)$ é uma função de verossimilhança maximizada.

h) *Geometric Mean Realtive Absolute Error – GMRAE*: Compara o erro do modelo selecionado com o método ingênuo de previsão, que considera como melhor preditor

de y no instante i , o seu valor no instante $i - 1$. Deseja-se que o GMRAE seja menor que 1. O cálculo é feito através da fórmula:

$$GMRAE = \sqrt[T]{\prod \left| \frac{y_i - \hat{y}_i}{y_i - y_{i-1}} \right|} \quad (19)$$

Estas medidas são automaticamente calculadas nos principais softwares dedicados a análises de séries temporais. O modelo que minimiza as métricas e o critério *BIC*, bem como apresenta maior R^2 é o que tem o melhor desempenho nas previsões.