

Métodos de Previsão e Séries Temporais

A. Manuela Gonçalves

2021/2022

Conteúdo

1	Introdução	1
1.1	<i>Packages</i> Séries Temporais	3
2	Séries temporais	5
2.1	Conceito de série temporal	5
2.1.1	Objetivos da análise de uma série temporal	6
2.1.2	Abordagens de análise de uma série temporal	7
2.1.3	Componentes de uma série temporal	7
2.2	Processos estocásticos	12
2.2.1	Processos estocásticos estacionários	13
2.2.2	Processos estocásticos não estacionários	19
3	Modelos e métodos de previsão	31
3.1	Modelos SARIMA e a metodologia Box-Jenkins	32
3.1.1	Modelos de processos estacionários	34
3.1.2	Modelos de processos não estacionários	41
3.1.3	Metodologia Box-Jenkins	45
3.1.4	Análise dos resíduos	50
3.1.5	Seleção de modelos	52
3.1.6	Previsão pontual e intervalar	54
3.2	Métodos de alisamento exponencial	55

Conteúdo

3.2.1	Método de alisamento exponencial simples	57
3.2.2	Método linear de Holt	59
3.2.3	Método de Holt-Winters	60
3.2.4	Outros métodos de alisamento exponencial	62
3.2.5	Intervalos de previsão	63
3.3	Medidas de avaliação	64
3.3.1	Medidas de avaliação utilizadas	67

Capítulo 1

Introdução

Os métodos de previsão constituem uma ferramenta fundamental nos processos de tomada de decisão em diversas áreas, como em Economia, Gestão e Finanças, ou em Ambiente, entre outras.

Existem vários métodos de previsão suportados por diferentes metodologias estatísticas, desde os mais complexos, que envolvem diferentes parâmetros, aos métodos mais simples e de fácil compreensão. Estes métodos podem ser baseados em modelos de séries temporais designados por modelos de Box-Jenkins, modelos de decomposição clássica associados a modelos de regressão linear múltipla e métodos de alisamento exponencial, entre outros. Uma descrição mais exhaustiva dos vários métodos de previsão pode ser encontrada em De Gooijer e Hyndman (2006).

No entanto, com a existência de múltiplas abordagens, a opção por apenas uma delas torna-se complicada. As medidas de avaliação, baseadas na análise da precisão das previsões, são, em muitos dos casos, o principal critério de seleção de um método de previsão. Através destas medidas é possível avaliar a eficácia de um modelo ou método de previsão, mostrando o quanto ele é capaz de descrever a série temporal em estudo e de obter previsões que venham a revelar-se o mais precisas possível.

Apesar de existirem, ainda durante o século XIX, algumas tentativas (determinísticas) de estudar séries temporais, é já no século XX que Yule (1927) estabelece a noção de estocasticidade em séries temporais, postulando que cada série pode ser considerada como uma realização de um processo estocástico (De Gooijer e Hyndman, 2006). Em 1970, baseados nos trabalhos de Yule (1926) e Wold (1938), Box e Jenkins (1970) desenvolvem uma abordagem prática e sistemática para a construção de modelos autorregressivos e de médias móveis integrados (*autorregressive integrated moving average*, ARIMA), que não é mais do que um processo de modelação iterativo em três fases: identificação, estimação e validação. Também durante o

século XX, mais precisamente nos anos 50 e 60, surgem, com os trabalhos de Brown (1959, 1963), Holt (1957) e Winters (1960), os modelos de alisamento exponencial, usualmente designados por modelos de Holt-Winters.

Uma série temporal é uma realização de um processo estocástico cujo espaço de parâmetros é um conjunto discreto de índices que representa o tempo. Em geral, numa série temporal, não se pode assumir que as observações são realizações independentes de uma população (ou de diferentes populações). A dependência das observações que constituem a série temporal é um conceito fundamental.

A estrutura probabilística que reflete a estrutura subjacente à série temporal é um processo estocástico. A ideia de seleção de um modelo de série temporal é muito importante, pois é com base nessa estrutura probabilística que se fazem previsões.

Um modelo de série temporal designa-se linear ou não linear conforme o valor atual da série temporal é função linear das observações passadas ou não. Os modelos lineares têm a vantagem de serem mais fáceis de interpretar e de implementar que os não lineares, além da simplicidade e baixo custo que apresentam. Devido a essas características, os métodos lineares são muito populares na previsão de vendas no segmento do retalho e que, se estes forem capazes de capturar o comportamento dos dados, devem ser os escolhidos. No entanto, algumas características de dados reais (e.g., ciclos irregulares, *outliers*, etc.) não são possíveis de modelar usando esses métodos, pelo que se deve optar, nesses casos, por métodos não lineares.

Para avaliar a precisão das previsões existem vários critérios. No entanto, nem todos os critérios são adequados para todas as situações (Hyndman e Koehler, 2006; Hyndman, 2006). Por exemplo, medidas que dependam da escala dos dados (e.g., o erro quadrático médio, EQM, a raiz do erro quadrático médio, REQM, e o erro absoluto médio, EAM) são úteis para comparar diferentes métodos aplicados ao mesmo conjunto de dados, porém não devem ser utilizadas para comparar dados com diferentes escalas. Para esses casos existem as medidas baseadas em erros percentuais (e.g., o erro percentual absoluto médio, EPAM), que são independentes da escala, mas têm a desvantagem de poderem ser infinitas ou indefinidas caso a série possua observações nulas. Outras alternativas para lidar com os problemas de escala podem ser as medidas de erros relativos ou as medidas de erros escalados.

Pode-se em algumas situações observar a inadequabilidade de várias medidas de precisão, por exemplo em séries temporais que contêm zeros e/ou observações negativas e que, por isso, são representativas de muitas séries reais. Hyndman e Koehler (2006) propõem a utilização de medidas de erros escalados para esses casos, nomeadamente o erro escalado absoluto médio (EEAM), que aparenta ser amplamente

aplicável e sempre definido e finito, além de facilmente interpretável. No entanto, há muitas situações em que as medidas tradicionais continuam a ser preferidas, dada a sua simplicidade. Por exemplo, se todas as séries estão na mesma escala, deve optar-se pelo EAM ou, se os dados são positivos e muito maiores do que zero, a utilização do EPAM é recomendada.

Uma vez que não existe uma medida que seja universalmente aceite como a mais adequada para ser aplicada a todas as situações de previsão, devem ser utilizadas várias para melhor avaliar e decidir sobre os diferentes métodos. Analisando os artigos que comparam os diferentes métodos para o segmento do retalho, a medida mais utilizada é o EPAM, seguida do EAM e do REQM.

1.1 *Packages Séries Temporais*

- **stat** package faz parte da base do R: representar e analisar séries temporais; classe *ts* para séries temporais igualmente espaçadas;
- **AER** package for “Applied Econometrics with R”
<http://CRAN.R-project.org/package=AER>;
- **TSA** package for “Time Series Analysis”
<http://CRAN.R-project.org/package=TSA>;
- **forecast** package for “Forecasting Functions for Time Series and Linear Models”
<http://CRAN.R-project.org/package=forecast>;
- **fma** package for data sets from “Forecasting: Methods and Applications” by Makridakis, Wheelwright Hyndman (1998)
<http://CRAN.R-project.org/package=fma>;
- **fpp** package for data sets from “Forecasting: Principles and Practice” by Hyndman and Athanasopoulos (2012)
<http://CRAN.R-project.org/package=fpp>;
- **expsmooth** package for “Forecasting with Exponential Smoothing”
<http://CRAN.R-project.org/package=expsmooth>;
- **lmtest** package for “Testing Linear Regression Models”
<http://CRAN.R-project.org/package=lmtest>.

CRAN Task View: Time Series Analysis (mantido pelo Hyndman)

<https://cran.r-project.org/web/views/TimeSeries.html>

Ver também em *Econometrics task view* e *Finance task view*.

Capítulo 2

Séries temporais

Neste Capítulo são introduzidas algumas noções sobre séries temporais e sobre processos estocásticos, fundamentais para a compreensão dos conteúdos programáticos desta unidade curricular.

2.1 Conceito de série temporal

Uma *série temporal* é um conjunto de observações medidas sequencialmente no tempo. Essas medições podem ser feitas continuamente no tempo ou apenas em momentos específicos, geralmente igualmente espaçados (e.g., dias, meses, trimestres, anos, etc.). Por convenção, esses dois tipos de séries são designados de séries temporais *contínuas* e *discretas*, respetivamente, independentemente da natureza da variável medida, que tanto pode ser discreta como contínua, em ambos os casos Chatfield (2000, 2004). Além disso, as séries temporais podem também ser classificadas em *univariadas*, se são constituídas por observações de uma só variável, e em *multivariadas*, se se observarem mais variáveis em cada instante.

Usualmente, para analisar séries temporais contínuas, estas são transformadas em discretas, sendo, para isso, amostradas (ou discretizadas) em intervalos de tempo iguais (e.g., temperatura medida em intervalos horários). Desde que o intervalo de amostragem seja suficientemente pequeno, este procedimento não resulta, em regra, numa perda significativa de informação.

Quanto às séries temporais discretas, estas surgem de diversas formas: além de poderem resultar de uma série temporal contínua por discretização (como visto anteriormente), podem também ocorrer quando são agregados valores de uma variável ao longo de um período de tempo (e.g., total de vendas em meses sucessivos) ou até serem inerentemente discretas (e.g., o dividendo pago por uma empresa em anos

sucessivos).

Independentemente do tipo de série, os dados de séries temporais têm, regra geral, uma propriedade muito particular: observações sucessivas são correlacionadas e, portanto, a análise desses dados deve ter em conta a ordem em que as observações são recolhidas. Efetivamente, no caso de séries temporais univariadas, cada observação pode ser vista como bivariada, considerando que a segunda variável corresponde ao tempo em que esta é observada.

Em termos de aplicações, a previsão de séries temporais é fundamental em diversas áreas: na Economia, e.g., preços de ações diários, totais de exportação em meses sucessivos, rendimentos médios mensais, lucros anuais de uma empresa, etc.; nas Ciências Físicas, nomeadamente na Meteorologia, Geofísica e Ciências Marinhas, e.g. precipitação diária, temperatura do ar medida a cada hora, dia ou mês sucessivo, etc.; no *Marketing*, e.g., planejar a produção com base em previsões de vendas mensais, etc.; nas Ciências Sociais, particularmente na Demografia, e.g., prever mudanças na população daí a 10 ou 20 anos, etc.; na Indústria, designadamente no controlo de processos, e.g., detetar mudanças no desempenho de um determinado processo de fabricação e efetuar as correções necessárias, etc..

2.1.1 Objetivos da análise de uma série temporal

O estudo de uma (ou várias) série temporal tem, dependendo de cada caso, objetivos específicos. No entanto, de uma forma geral, podem mencionar-se como principais motivações (Box et al., 2013; Chatfield, 2000):

- descrever os dados usando estatísticas descritivas e/ou métodos gráficos e compreender o mecanismo gerador da série, i.e., procurar encontrar razões que justifiquem o comportamento da série, monitorizar a sua trajetória, analisar periodicidades relevantes nos dados, etc.;
- encontrar um modelo estatístico adequado para descrever a evolução da série temporal. São vários os modelos propostos, entre os quais se destacam os modelos ARIMA, os modelos de alisamento exponencial, os modelos de regressão múltipla, etc.;
- prever o comportamento futuro da série, o que pode revelar-se extremamente útil na construção e execução de planos a curto, médio ou longo prazos e/ou no controlo de um determinado processo. Estas previsões podem ser a *1-passo*, se realizadas apenas para a observação seguinte, ou *multi-passos*, se englobam várias observações futuras.

Em Economia, a análise de séries temporais é utilizada entre outros propósitos para, (a) estudar a estrutura dinâmica de um processo, (b) investigar as relações dinâmicas entre variáveis, (c) proceder ao ajustamento sazonal de dados económicos, (d) melhorar a análise de regressão quando os erros são correlacionados e (e) produzir previsões pontuais e/ou intervalos de previsão.

2.1.2 Abordagens de análise de uma série temporal

- Técnicas descritivas. Representações gráficas, identificação de padrões, etc.
- Modelos Probabilísticos. Seleção, comparação e adequação de modelos, estimação e previsão. Ferramenta básica é a função de autocorrelação.
- Análise espectral.
- Métodos não paramétricos (alisamento ou suavização).
- Outras abordagens. Modelos de espaço de estados, modelos não lineares, séries multivariadas, estudos longitudinais, processos de longa dependência, modelos para volatilidade, etc.

2.1.3 Componentes de uma série temporal

Segundo Persons (1919), a variação de uma série temporal pode ser decomposta em 4 componentes: a tendência (T), a componente sazonal (S), a componente cíclica (C) e a componente irregular/residual (E). Estas podem ser descritas como Alpuim (1998):

Tendência (T) : é a inclinação que a série temporal apresenta ao longo do tempo, podendo esta ser linear ou não, crescente ou decrescente. A tendência pode ser consequência do facto dos valores observados dependerem de uma componente determinística que é função monótona do tempo, embora para muitos autores esta possa ser de natureza estocástica;

Sazonalidade (S) : corresponde a um padrão de aumento e diminuição que ocorre regularmente na série em períodos específicos, originando oscilações que se repetem (e.g., as vendas a retalho apresentam frequentemente um padrão sazonal mensal, onde cada mês, considerando períodos anuais, mostra constantemente a mesma posição relativamente aos restantes: picos de vendas ocorrem em meses de férias e, nos meses seguintes, verifica-se uma diminuição acentuada;

Hyndman e Athanasopoulos, 2014). O movimento dentro de um período tem, então, duração fixa e é atribuído a fatores “sazonais”, i.e., relacionados com aspetos do calendário (e.g., os meses ou trimestres de um ano ou os dias de uma semana). A sazonalidade pode ser classificada como *aditiva*, se não depende do nível da série, ou *multiplicativa*, quando é proporcional ao mesmo. Esta fonte de variação é muitas vezes removida através de um procedimento designado de ajustamento sazonal (*seasonal adjustment*), uma vez que esta pode omitir outras componentes relevantes, tais como a tendência;

Componente cíclica (C) : é um padrão de flutuação que não apresenta qualquer periodicidade definida (i.e., a sua duração não é fixa e, portanto, o seu comprimento varia frequentemente de ciclo para ciclo) nem causa atribuída a fatores “sazonais”. Por outras palavras, os ciclos são qualquer componente não-sazonal que apresenta um padrão reconhecível (e.g., ciclos económicos). Como estas flutuações são irregulares e se podem prolongar por vários anos, é frequente ignorar esta componente para séries “curtas”. Além disso, como se torna difícil dissociar a componente cíclica da tendência, estas são muitas vezes aglomeradas e referidas como componente de tendência-cíclica (*trend-cycle*);

Aleatoriedade (E) : é a componente que contém qualquer variação não explicada pelas componentes anteriores e representa o ruído aleatório. Quando esta componente é modelada por um processo estocástico de variáveis aleatórias não correlacionadas e identicamente distribuídas, é referida como um ruído branco.

Decomposição de uma série temporal

A maioria dos métodos tradicionais para a análise de séries temporais baseia-se na decomposição da variação da série nas componentes anteriores.

Sejam Y_t o valor da série temporal no tempo t , T_t a tendência no tempo t , S_t a componente sazonal no tempo t e E_t a componente irregular no tempo t . Considere-se que a componente cíclica (C), quando presente, se encontra incorporada na tendência (note-se que esta situação foi referida anteriormente como tendência-cíclica).

O *modelo de decomposição aditivo* descreve cada valor da série temporal como sendo a soma das suas componentes, i.e.,

$$Y_t = T_t + S_t + E_t,$$

enquanto que no *modelo de decomposição multiplicativo* cada observação é o produto

dessas mesmas componentes, i.e.,

$$Y_t = T_t \times S_t \times E_t.$$

Um modelo aditivo é apropriado quando a magnitude das oscilações sazonais não varia com o nível da série. No entanto, se estas aumentam ou diminuem proporcionalmente com a tendência da série, então um modelo multiplicativo é o mais adequado.

Na Economia, a decomposição multiplicativa é a dominante, uma vez que a maioria das séries económicas sazonais apresenta uma variação sazonal que depende do nível da série. Muitas vezes, nesses casos, opta-se por aplicar uma transformação aos dados, nomeadamente a logarítmica, de forma a converter um modelo multiplicativo num modelo aditivo, i.e., aplicando a transformação logarítmica ao modelo multiplicativo e obtendo

$$\log Y_t = \log T_t + \log S_t + \log E_t.$$

Desta forma, pode, então, construir-se um modelo multiplicativo por ajustamento de um modelo aditivo ao logaritmo dos dados. Contudo, deve notar-se que um modelo de decomposição multiplicativo não deve ser implementado para séries temporais de valores negativos ou nulos (Caiado, 2016).

Evidentemente, os modelos de decomposição aditivo e multiplicativo não são as únicas formas de decompor uma série temporal e podem até ser misturados, dando origem a outros modelos que incluem relações tanto aditivas como multiplicativas, e.g., um *modelo multiplicativo com erros aditivos*, i.e.,

$$Y_t = T_t \times S_t + E_t.$$

Uma ferramenta vantajosa no que toca à decomposição de séries temporais são os gráficos de decomposição (introduzidos por Cleveland e Terpenning, 1992), que permitem visualizar graficamente as várias componentes (Makridakis et al., 1998). No ambiente R, a função *decompose* divide a série temporal nas três componentes – sazonalidade (S), tendência (T) e aleatoriedade (E), usando médias móveis. Esta função mostra-se bastante útil, uma vez que permite avaliar as diversas componentes separadamente e, portanto, ajudar a identificar o comportamento individual das mesmas.

Modelos de regressão para a tendência

No contexto de séries temporais, a componente tendência pode ser classificada como estocástica ou determinística.

Quando uma tendência determinística é totalmente explicada por um modelo de regressão, a série dos resíduos deve comportar-se como uma série temporal estacionária na média, não apresentando qualquer tendência remanescente.

(i) Modelo de regressão linear

Ao examinar visualmente uma série temporal, é possível descrever as mudanças que a tendência apresenta ao longo do tempo. No entanto, o comportamento de uma tendência pode ser formalmente explicado por uma regressão sobre uma variável que representa o tempo, t (considera-se que a variável t toma valores de 1 a n para uma série com n observações igualmente espaçadas).

No caso mais simples, a tendência pode ser modelada como uma função linear do tempo, o que corresponde ao caso de um modelo de regressão linear simples em que a variável explicativa é a variável tempo, i.e.,

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t, \quad (2.1)$$

onde β_0 representa o nível da série no tempo $t = 0$, β_1 representa a variação da série temporal associada ao aumento de uma unidade na variável tempo, t é a variável tempo, e ϵ_t é a componente aleatória.

Quando o modelo linear não é suficiente para explicar a tendência da série temporal, pode resultar em previsões imprecisas e/ou apresentar tendência na série dos resíduos.

(ii) Modelo de regressão polinomial

Em algumas séries temporais, a tendência apresenta um comportamento não linear, que não é totalmente explicado pelo modelo mais simples. Nesses casos, é possível adicionar um (ou mais) termo polinomial de ordem superior a 1 ao modelo linear, o que corresponde a modelar a tendência como quadrática, cúbica, etc., isto é,

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \epsilon_t, \quad (2.2)$$

onde t representa a variável tempo definida anteriormente, e p representa o grau do polinómio que descreve a tendência. Nos casos particulares em que a tendência é

quadrática e cúbica tem-se $p = 2$ e $p = 3$, respetivamente.

Modelos de regressão para a sazonalidade

(i) Modelo com indicadores sazonais

Para modelar séries com sazonalidade pode optar-se por uma descrição qualitativa do padrão sazonal, através da especificação de um modelo de regressão que inclua uma variável indicatriz para representar cada um dos s períodos sazonais. Este modelo com indicadores sazonais pode ser escrito como

$$Y_t = T_t + \beta_1 D_1 + \cdots + \beta_s D_s + \epsilon_t, \quad (2.3)$$

onde T_t representa a tendência (em função de t sem o termo constante, β_0), β_1, \dots, β_s são os coeficientes que refletem os s efeitos sazonais e D_i ($i = 1, \dots, s$) são as s variáveis indicatrizes que representam os diferentes períodos sazonais: tomam o valor 1 quando o tempo t pertence ao período i e 0 nos restantes casos. Por exemplo, para dados mensais, se D_1 corresponder às ocorrências no mês de janeiro (i.e., 1, se t ocorre em janeiro e 0, caso contrário), então β_1 só é tido em consideração para observações registadas nesse mês.

Intuitivamente, este modelo pode ser visto como um modelo linear onde existe um nível distinto para cada período, que representa o seu efeito.

(ii) Modelo sazonal harmónico

Nos modelos com indicadores sazonais, é atribuído um parâmetro (e uma variável) a cada período sazonal. No entanto, os efeitos sazonais geralmente variam de forma suave e contínua, pelo que pode ser mais eficiente e realista usar uma função de suavização em vez de indicadores sazonais. Assim, uma outra abordagem para modelar efeitos sazonais passa por ajustar um modelo sazonal harmónico, que recorre a funções seno e cosseno para descrever as oscilações observadas ao longo do tempo (Cowpertwait e Metcalfe, 2009).

Uma onda sinusoidal de frequência f (ciclos por intervalo amostral), amplitude A e constante de fase ϕ , pode ser escrita como

$$A \sin(2\pi ft + \phi) = \alpha_c \cos(2\pi ft) + \alpha_s \sin(2\pi ft),$$

onde $\alpha_s = A \cos(\phi)$ e $\alpha_c = A \sin(\phi)$.

Para uma série temporal de período s existem $s/2$ ciclos possíveis (no caso em

que s não é divisível por 2, deve considerar-se a parte inteira da divisão) e, portanto, pode definir-se o modelo sazonal harmónico

$$Y_t = T_t + \sum_{i=1}^{s/2} (\alpha_i \cos(2\pi it/s) + \beta_i \sin(2\pi it/s)) + \epsilon_t, \quad (2.4)$$

onde T_t representa a tendência (em função de t), α_i e β_i ($i = 1, \dots, s/2$) são os parâmetros desconhecidos de interesse, s é o período sazonal (e.g., $s = 12$ para dados mensais), i é um índice que varia entre 1 e $s/2$, e t é uma variável codificada que representa o tempo (e.g., $t = 1, \dots, 90$ para 90 observações igualmente espaçadas).

Apesar de parecer mais complexo, este modelo pode envolver um menor número de parâmetros relativamente ao modelo de indicadores sazonais, uma vez que, como os coeficientes harmónicos são independentes, todos os coeficientes que não são estatisticamente significativos para o modelo completo podem ser eliminados (Cowpertwait e Metcalfe, 2009).

2.2 Processos estocásticos

Uma série temporal pode ser encarada como um conjunto de observações de um processo estocástico, isto é, de um conjunto de variáveis aleatórias indiciadas no tempo.

Dado um processo estocástico $\{Y(t), t \in \mathcal{T}\}$, uma série temporal é um conjunto de observações do processo estocástico em instantes t_1, t_2, \dots, t_n . Geralmente, considera-se t inteiro (i.e., $t = 0, \pm 1, \pm 2, \dots$) e as observações são feitas em intervalos de tempo regulares, isto é, com a mesma amplitude (Alpuim, 1998).

Definição 1. *Um processo estocástico é qualquer família ou coleção de variáveis aleatórias $Y(t), t \in \mathcal{T}$, em que \mathcal{T} é um conjunto de índices representando o tempo.*

Ao conjunto de índices \mathcal{T} dá-se o nome de *espaço de parâmetros* e ao contradomínio das variáveis aleatórias $Y(t)$ dá-se o nome de *espaço de estados*, que se representa por S . Quanto à natureza de \mathcal{T} , se $\mathcal{T} = \mathbb{Z}$ ou $\mathcal{T} = \mathbb{N}$ diz-se que o processo é de tempo discreto e se $\mathcal{T} = \mathbb{R}$ ou, mais comumente, $\mathcal{T} = \mathbb{R}^+$ diz-se que o processo é de tempo contínuo.

Uma maneira de caracterizar um processo estocástico consiste em especificar a distribuição de probabilidade conjunta de n variáveis aleatórias $(Y(t_1), \dots, Y(t_n))$ para todos os inteiros n e quaisquer pontos t_1, \dots, t_n . No entanto, esta forma

de definir um processo estocástico é complicada e, na prática, não é viável. Assim, uma alternativa mais simples de descrever um processo estocástico é através dos momentos do processo, em particular os primeiro e segundo momentos, designados por valor médio $\mu(t) = E[Y(t)]$ e função de autocovariância $\gamma(t_1, t_2) = E[(Y(t_1) - \mu(t_1))(Y(t_2) - \mu(t_2))]$, respetivamente. A variância $\sigma^2(t) = Var[Y(t)]$ é um caso particular da função de autocovariância (quando $t_1 = t_2$) mas, por si só não é suficiente para definir os segundos momentos de uma sequência de variáveis aleatórias (Chatfield, 2004).

A série de valores observados, que compõem a série temporal, é considerada apenas uma única realização (ou trajetória) de um processo estocástico, de entre todas as possíveis. Ou seja, o processo estocástico é um só, mas as séries que ele pode gerar são infinitas. Assim, a relação entre uma série temporal e o processo estocástico que a gerou é semelhante à relação entre uma amostra e a população de onde foi extraída. Em termos gerais, a análise de séries temporais consiste em fazer inferências sobre um processo estocástico desconhecido tendo como informação disponível uma única realização observada.

Os processos estocásticos dividem-se em estacionários e não estacionários. Nesta Secção introduzem-se os dois tipos de estacionariedade (forte e fraca), alguns procedimentos que permitem transformar processos não estacionários em estacionários e outras ferramentas essenciais para a posterior modelação das séries temporais (e.g., as funções de autocorrelação, FAC, e autocorrelação parcial, FACP, e o processo de ruído branco).

2.2.1 Processos estocásticos estacionários

Em termos gerais, os processos estacionários traduzem a situação em que o sistema se apresenta num estado de equilíbrio estatístico em torno de um nível médio fixo, i.e., tem propriedades probabilísticas que são estáveis ou invariantes ao longo do tempo (Murteira et al., 1993).

Definição 2. *Um processo estocástico $\{Y(t), t \in \mathcal{T}\}$ diz-se estritamente estacionário (ou fortemente estacionário) se e só se a distribuição conjunta de $(Y(t_1), \dots, Y(t_n))$ é igual à distribuição conjunta de $(Y(t_1 + \delta), \dots, Y(t_n + \delta))$ qualquer que seja o n -úplo (t_1, \dots, t_n) e para qualquer δ , i.e.,*

$$F_{(Y(t_1), \dots, Y(t_n))}(y_1, \dots, y_n) = F_{(Y(t_1 + \delta), \dots, Y(t_n + \delta))}(y_1, \dots, y_n)$$

em todos os pontos (y_1, \dots, y_n) .

Desta forma, pode dizer-se que um processo fortemente estacionário usufrui da propriedade de que a distribuição de um qualquer conjunto de margens se mantém a mesma, quando estas são sujeitas a uma translação no tempo (Alpuim, 1998).

A estacionariedade no sentido estrito é uma propriedade demasiado exigente e, na maioria dos casos, de difícil verificação. Os processos estacionários de 2.^a ordem (ou fracamente estacionários) obedecem a uma propriedade mais fraca mas que, grosso modo, descreve o mesmo tipo de comportamento (Murteira et al., 1993).

Definição 3. *Um processo $\{Y(t), t \in \mathcal{T}\}$ diz-se estacionário de 2.^a ordem (ou fracamente estacionário) se e só se todos os momentos até à 2.^a ordem de $(Y(t_1), \dots, Y(t_n))$ existem e são iguais aos momentos correspondentes até à 2.^a ordem de $(Y(t_1 + \delta), \dots, Y(t_n + \delta))$. Logo, num processo fracamente estacionário:*

1. o valor médio não depende de t , i.e., $\mu(t) = \mu$;
2. a variância não depende de t , i.e., $\sigma^2(t) = \sigma^2$;
3. a covariância de $Y(t_1)$ e $Y(t_2)$ depende apenas do desfaseamento $t_2 - t_1$, i.e., $Cov[Y(t_1), Y(t_2)] = \gamma(|t_2 - t_1|)$.

Está implícito, pelas condições de estacionariedade fraca, que se assume, nesse caso, que os momentos até à 2.^a ordem existem e são finitos. Assim, pelas Definições 2 e 3, se $Y(t)$ é estritamente estacionário e os seus momentos até à 2.^a ordem são finitos, então $Y(t)$ também é estacionário de 2.^a ordem. A recíproca pode não se verificar. No entanto, se a série temporal $Y(t)$ tem distribuição Normal, então a estacionariedade fraca é equivalente à estacionariedade forte.

De agora em diante, considerem-se apenas processos estacionários de 2.^a ordem, designados simplesmente por processos estacionários, e Y_t a representar um processo estocástico, independentemente do tempo t ser discreto ou contínuo.

Funções de autocovariância, autocorrelação e autocorrelação parcial

Definição 4. *Para um processo estacionário, define-se a função de autocovariância*

$$\gamma_k = Cov[Y_t, Y_{t+k}] = E[(Y_t - \mu)(Y_{t+k} - \mu)],$$

que mede a intensidade com que covariam (se acompanham) pares de valores do processo separados por um intervalo (lag) de amplitude k .

A função de autocovariância γ_k é definida para $k \in \mathbb{R}$ se o processo é de tempo contínuo e para $k \in \mathbb{Z}$ se for de tempo discreto, i.e., $k = 0, \pm 1, \pm 2, \dots$. Além disso, esta função satisfaz as seguintes propriedades:

1. $\gamma_0 = Cov[Y_t, Y_t] = Var[Y_t] = \sigma^2$;
2. $\gamma_k = \gamma_{-k}$, i.e., a função é par e dispensa a representação gráfica para $k < 0$;
3. $|\gamma_k| \leq \gamma_0$, como consequência da desigualdade de Cauchy-Schwarz¹;
4. A função γ_k é semidefinida positiva, i.e., para qualquer conjunto de números reais $\alpha_1, \dots, \alpha_n$ e instantes de tempo t_1, \dots, t_n ,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma(|t_i - t_j|) \geq 0.$$

Definição 5. Para um processo estacionário, define-se a função de autocorrelação (FAC)

$$\rho_k = Corr[Y_t, Y_{t+k}] = \frac{Cov[Y_t, Y_{t+k}]}{\sqrt{Var[Y_t]Var[Y_{t+k}]}} = \frac{Cov[Y_t, Y_{t+k}]}{Var[Y_t]} = \frac{\gamma_k}{\gamma_0},$$

que mede a correlação entre pares de valores do processo separados por um intervalo (lag) de amplitude k .

A representação gráfica de ρ_k em função de k designa-se por correlograma. O comportamento do correlograma dá indicações cruciais sobre as características da série e constitui um auxiliar importante na identificação do modelo que lhe é mais adequado (como será visto no Capítulo 3). Na maioria dos casos, o aumento de k traduz-se no decrescimento de ρ_k (e, obviamente, de γ_k). De facto, à medida que a amplitude do intervalo (k) aumenta é de esperar que a capacidade de memória do processo seja limitada, e, portanto, que no momento $t + k$ se encontre pouco refletido o que se passou no momento t (Murteira et al., 1993). Desta forma, espera-se que a correlação temporal diminua ($\rho_k \rightarrow 0$), quando o desfasamento entre duas observações aumenta ($k \rightarrow +\infty$).

Intuitivamente, pode interpretar-se ρ_k como uma medida da semelhança entre cada realização e a mesma realização deslocada k unidades de tempo (Murteira et al., 1993).

¹Desigualdade de Cauchy-Schwarz: $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$.

Tal como no caso da função de autocovariância, a função de autocorrelação ρ_k pode estar definida para $k \in \mathbb{R}$ ou para $k \in \mathbb{Z}$, consoante o processo for de tempo contínuo ou discreto, respetivamente. Além disso, esta função satisfaz as seguintes propriedades:

1. $\rho_0 = \text{Corr}[Y_t, Y_t] = 1$;
2. $\rho_k = \rho_{-k}$, i.e., a função é par e dispensa a representação gráfica para $k < 0$;
3. $|\rho_k| \leq 1$, como consequência da desigualdade de Cauchy-Schwarz;
4. A função ρ_k é semidefinida positiva, i.e., para qualquer conjunto de números reais $\alpha_1, \dots, \alpha_n$ e instantes de tempo t_1, \dots, t_n ,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho(|t_i - t_j|) \geq 0.$$

Como os processos estacionários se caracterizam pelos parâmetros já referidos, a estimação dos mesmos tem muita importância. Se se considerar um conjunto de n observações de um processo estacionário Y_t durante um certo período de tempo, ou seja, Y_1, Y_2, \dots, Y_n , podem utilizar-se os estimadores clássicos dos parâmetros (Alpuim, 1998), i.e.,

- para a média μ usar $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$;
- para a autocovariância γ_k usar $\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$;
- para a autocorrelação ρ_k usar $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$.

Por vezes, além de se estudar a correlação de uma forma global, interessa também investigar a correlação parcial que existe entre Y_t e Y_{t+k} quando se fixam as variáveis intermédias $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$, i.e., a correlação simples entre Y_t e Y_{t+k} depois de eliminar o efeito que as variáveis intermédias exercem sobre elas (Caiado, 2016).

Sejam $E[Y_t] = 0$ e $\text{Var}[Y_t] = 1$ e considere-se a regressão linear múltipla de Y_{t+k} sobre Y_{t+k-1}, \dots, Y_t , i.e.,

$$Y_{t+k} = \phi_{k1} Y_{t+k-1} + \dots + \phi_{kk} Y_t + \epsilon_{t+k}, \quad (2.5)$$

onde ϕ_{kj} , $j = 1, \dots, k$, são os coeficientes do modelo de regressão linear que se considera ter erros gaussianos. O valor ϕ_{kk} é o coeficiente de correlação do modelo de regressão linear (2.5) onde $\{\epsilon_t, t \in \mathbb{Z}\}$ tem distribuição Normal de parâmetros $(0, \sigma^2)$ e ϵ_{t+k} é independente de $\{Y_{t+k-j}, j \geq 1\}$. O coeficiente ϕ_{kk} exprime a variação em Y_{t+k} que acompanha em média uma variação unitária em Y_t quando $Y_{t+1}, \dots, Y_{t+k-1}$ são constantes; tal variação pode interpretar-se como a correlação parcial entre Y_t e Y_{t+k} .

Multiplicando, então, ambos os membros de (2.5) por Y_{t+k-j} , $j = 1, \dots, k$, calculando os valores esperados e dividindo por ρ_0 obtém-se um sistema constituído pelas equações

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{k-j}, \quad j = 1, 2, \dots, k; \quad (2.6)$$

resolve-se o sistema em ordem a ϕ_{kj} , $j = 1, 2, \dots, k$, utilizando a regra de Cramer, e obtém-se, assim, a função de autocorrelação parcial, ϕ_{kk} .

Uma definição alternativa para esta função é a função de autocorrelação parcial (FACP):

Definição 6. O conjunto de autocorrelações parciais de desfasamento (lag) k é dado por $\{\phi_{kk} : k = 1, 2, \dots\}$ onde

$$\phi_{kk} = \text{Corr}[X_t, X_{t+k} | X_{t+1}, X_{t+2}, \dots, X_{t+k-1}] = \frac{|P_k^*|}{|P_k|}$$

e P_k^* é a matriz $k \times k$ de autocorrelações onde a última coluna é substituída por $[\rho_1 \ \rho_2 \ \dots \ \rho_k]^T$. A matriz P_k é dada por

$$P_k = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{bmatrix}.$$

Resolvendo o sistema constituído pelas equações em (2.6) ou seguindo a Definição 6, obtêm-se as seguintes propriedades:

$$\phi_{11} = \rho_1; \quad \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}; \quad \phi_{33} = \frac{\rho_3(1 - \rho_1^2) + \rho_1(\rho_1^2 + \rho_2^2 - 2\rho_2)}{(1 - \rho_2)(1 + \rho_2 - 2\rho_1^2)}.$$

Ruído branco

Um processo estocástico $\{\epsilon_t, t \in \mathbb{Z}\}$ diz-se um processo puramente aleatório ou processo de ruído branco quando é formado por uma sucessão de variáveis aleatórias não correlacionadas e identicamente distribuídas, de média e variância constantes, i.e., um processo estocástico diz-se um ruído branco se e só se satisfaz as seguintes condições:

1. $E[\epsilon_t] = \mu_\epsilon$ (usualmente $\mu_\epsilon = 0$);
2. $Var[\epsilon_t] = \sigma_\epsilon^2$;
3. $Cov(\epsilon_t, \epsilon_{t+k}) = \gamma_k = 0, \quad k = \pm 1, \pm 2, \dots$

Nota: Se $\sigma = 1$, então trata-se de um ruído branco reduzido.

Se as variáveis aleatórias seguem uma distribuição Normal (i.e., $\epsilon_t \sim N(\mu_\epsilon, \sigma_\epsilon^2)$), então o processo é designado de ruído branco gaussiano.

Um ruído branco é, então, um processo estacionário cujas funções de autocorrelação (FAC) e autocorrelação parcial (FACP) são nulas para todo o $k \neq 0$. Na Figura 2.1 encontra-se representada uma trajetória de um processo de ruído branco e as respetivas FAC e FACP empíricas.

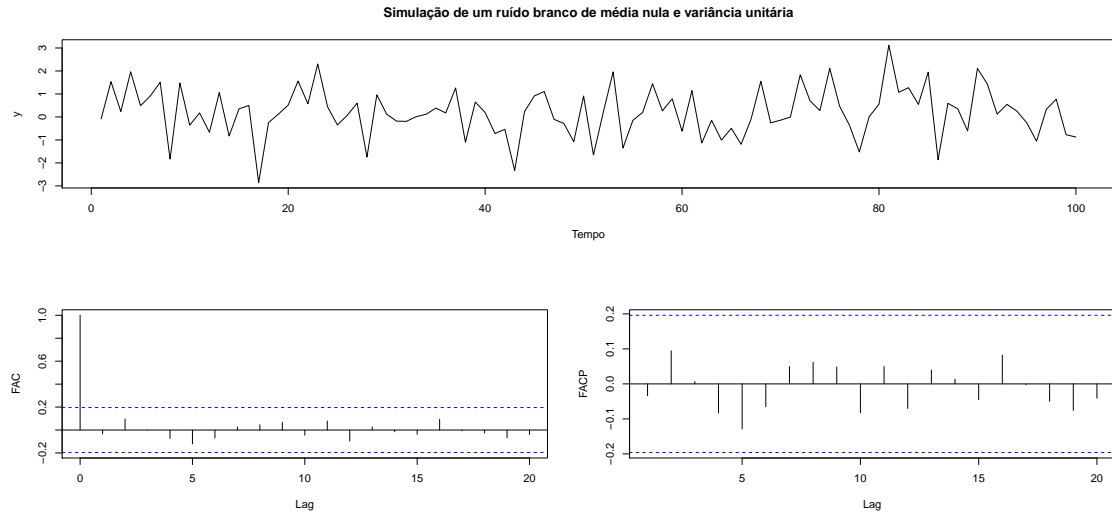


Figura 2.1: Simulação de um ruído branco e respetivas FAC e FACP empíricas.

Embora o ruído branco dificilmente se observe em séries reais, este desempenha um papel fundamental na construção de modelos probabilísticos ou estocásticos. Além disso, um bom modelo de previsão deve ser aquele que produz erros de previsão com comportamento análogo a um ruído branco, isto porque um ruído branco é imprevisível (Caiado, 2016).

Série de médias móveis

Os métodos de médias móveis são adequados para avaliar a tendência de uma série temporal e são muito sensíveis às alterações recentes do seu comportamento. Utilizam a informação relativa aos termos da série numa vizinhança de alguns períodos para estimar o seu nível corrente, o que permite a filtragem das flutuações de carácter aleatório associados ao operador da média (Caiado, 2016).

2.2.2 Processos estocásticos não estacionários

Numa série estacionária, os valores futuros serão similares aos do passado e, por isso, a estacionariedade é um importante pressuposto quando se pretende obter previsões com base em observações passadas. Alguns modelos de previsão de séries temporais assumem que a série já é ou pode ser transformada numa série estacionária.

Muitas séries temporais, nomeadamente as associadas a fenómenos ambientais ou económicos, são não estacionárias. Um processo pode ser não estacionário por a média e/ou a variância serem funções do tempo e não constantes. Uma série estacionária em média não é necessariamente estacionária em variância. Para ultrapassar este problema pode recorrer-se a transformações que estabilizam a média e/ou a variância convertendo, assim, uma série temporal não estacionária numa série estacionária. No caso de se estar perante uma série não estacionária em média nem em variância deve proceder-se, em primeiro lugar, à estabilização da variância e só depois da média (Murteira et al., 1993; Caiado, 2016). Na Figura 2.2 podem ver-se alguns tipos de não estacionariedade: do lado esquerdo, uma série não estacionária em variância, ainda que estacionária em média; e do lado direito, uma série com tendência linear (não estacionária em média) e variância crescente com a tendência (não estacionária em variância).

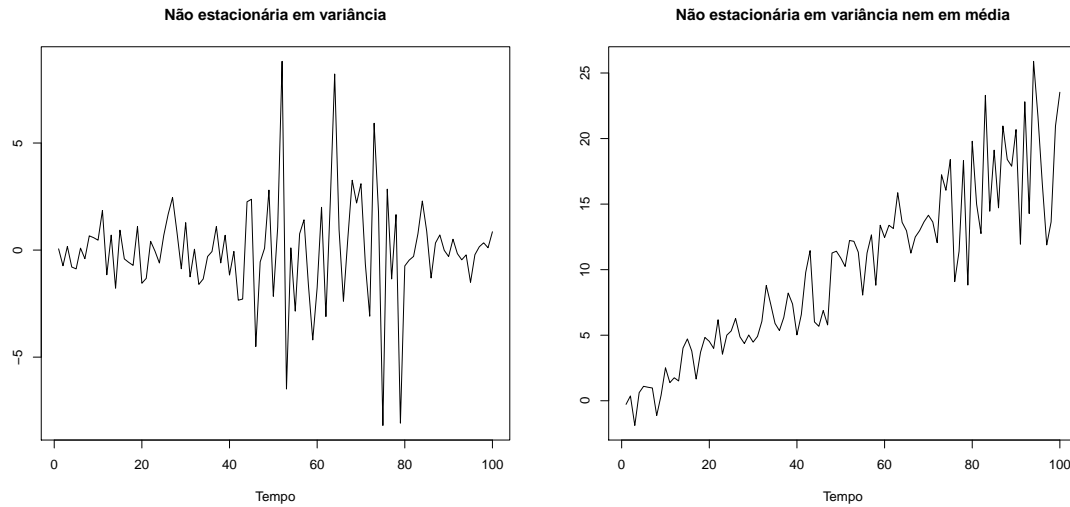


Figura 2.2: Duas realizações de séries temporais não estacionárias.

VER EXEMPLOS

Obtenha o gráfico do seguinte modelo (simulado) estacionário em média, mas não em variância, em que

$$Y_t = \begin{cases} N(0, 1^2) & , \quad t \leq t_0 \\ N(0, 5^2) & , \quad t > t_0 \end{cases} .$$

Em muitas situações, há procedimentos que procuram “retirar” a tendência e a sazonalidade a uma série temporal, permitindo, assim, que se atinja a estacionariedade. Estes métodos seguem os modelos de decomposição anteriormente descritos (Secção 2.1.3) e consistem, naturalmente, na estimação das componentes tendência (T_t) e sazonalidade (S_t) através de funções determinísticas (ou outras abordagens), de modo que a série após “remoção” dessas componentes passe a ser bem modelada por um processo estacionário (Alpuim, 1998).

No entanto, em muitas séries temporais é possível realizar outro tipo de transformações, que permitem transformar séries não estacionárias em estacionárias.

Transformações para a estacionariedade

Uma forma de estabilizar a média consiste em usar processos de diferenciação (regular), que resultam da aplicação do operador diferença ∇ , definido como $\nabla Y_t = Y_t - Y_{t-1}$, à série temporal não estacionária. Assim, se uma série, Y_t , for não estacionária, pode pensar-se em transformá-la numa série estacionária, aplicando

uma diferenciação de primeira ordem, i.e.,

$$\nabla Y_t = Y_t - Y_{t-1} , \quad t = 2, 3, \dots, n.$$

Se a diferenciação de 1.^a ordem não for suficiente para obter uma série estacionária, podem obter-se as diferenças de 2.^a ordem, que correspondem às diferenças das primeiras diferenças da série original, i.e.,

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2} , \quad t = 3, 4, \dots, n.$$

O operador de diferenciação de ordem d , para qualquer inteiro $d \geq 1$, consiste em diferenciar a série d vezes, ou seja,

$$\nabla^d Y_t = \nabla(\nabla^{d-1} Y_t) , \quad t = d + 1, \dots, n.$$

Contudo, a diferenciação inapropriada de uma série já estacionária é indesejável e deve evitar-se. De facto, o objetivo é determinar a série estacionária obtida pela menor diferenciação, uma vez que a variância aumenta com a diferenciação. De uma forma geral, se a série transformada $\nabla^{d_0} Y_t$ é estacionária, então, para qualquer $d > d_0$, a série $\nabla^d Y_t$ é também estacionária, mas tem maior variância. Conclui-se, portanto, que se deve evitar a sobrediferenciação para não introduzir variação indesejada na série transformada. Efetivamente, caso este procedimento seja adequado, as diferenciações de 1.^a e 2.^a ordens devem ser suficientes para obter uma série estacionária.

Se uma série temporal é diferenciada uma vez e a série diferenciada é estacionária, então diz-se que a série original é integrável de ordem 1 e representa-se por $I(1)$. Em geral, se a série for diferenciada d vezes, é integrável de ordem d ou $I(d)$. Por convenção, se $d = 0$, o processo $I(0)$ é um processo estacionário.

Para estabilizar a variância de uma série não estacionária, pode utilizar-se um método de transformação paramétrica, conhecido como transformação de Box-Cox, baseado na seguinte expressão

$$Z_t = T(Y_t) = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} , & \lambda \neq 0 \\ \log Y_t , & \lambda = 0 \end{cases} ,$$

onde os valores de λ estão no intervalo $[-1; 1]$. As transformações mais recorrentes encontram-se sumariadas na Tabela 2.1.

Tabela 2.1: Transformações usuais de Box-Cox.

λ	Transformação
-1	$Z_t = \frac{1}{Y_t}$
-0,5	$Z_t = \frac{1}{\sqrt{Y_t}}$
0	$Z_t = \log Y_t$
0,5	$Z_t = \sqrt{Y_t}$
1	$Z_t = Y_t$

No entanto, note-se que algumas destas transformações (nomeadamente a do logaritmo) apenas estão definidas para séries de valores positivos. Para contornar esse problema, pode encontrar-se uma constante c tal que $Y_t + c > 0$ e só depois aplicar as transformações.

Como já foi referido, a transformação logarítmica permite converter o efeito sazonal multiplicativo em aditivo, sendo, também por isso, uma das transformações mais usadas dentro da família Box-Cox. Além disso, uma vez que esta transformação permite estabilizar as diferenças entre os valores da série temporal (i.e., a sua variância), ela também pode ser utilizada para atenuar o efeito de possíveis *outliers*.

É importante lembrar que, sempre que se aplica alguma transformação aos dados, todas as previsões geradas pelo modelo selecionado estarão nas unidades transformadas. Uma vez ajustado o modelo e estimados os parâmetros, devem, então, ser revertidas as transformações de modo a obter previsões nas unidades originais.

Passeio aleatório

Considere-se o seguinte modelo de tendência estocástica,

$$Y_t = Y_{t-1} + \epsilon_t, \quad (2.7)$$

onde ϵ_t é um ruído branco. Este modelo é vulgarmente conhecido por modelo de passeio aleatório (*random walk*), uma vez que descreve a série no momento t com base no seu valor passado (no momento $t - 1$) mais um choque aleatório.

Em geral, os “passeios aleatórios” caracterizam-se por movimentos de tendência crescente ou decrescente em períodos longos, seguidos de mudanças bruscas de sentido imprevisível (Caiado, 2016). Contudo, uma vez que as aparentes tendências são apenas o resultado da existência de um elevado número de valores positivos em relação ao número de valores negativos (no caso de tendências crescentes) ou vice-versa (no caso de tendências decrescentes), o valor médio de um processo aleatório

é constante. De facto, se Y_0 for conhecido, o modelo de passeio aleatório pode ser reescrito como

$$Y_t = Y_0 + \sum_{i=1}^t \epsilon_i, \quad (2.8)$$

e, calculando o valor esperado de Y_t , conclui-se que este é igual a Y_0 e, portanto, constante. Apesar disso, o modelo de passeio aleatório é um processo não estacionário, uma vez que a sua variância depende de t . No entanto, aplicando uma transformação, nomeadamente uma diferenciação de 1.^a ordem, $\nabla Y_t = \epsilon_t$, é possível obter um ruído branco, que é um processo estacionário.

Este modelo é muito utilizado na modelação e previsão de séries financeiras não estacionárias e na análise de importantes relações económicas. Por exemplo, nos mercados financeiros, as previsões dos preços das ações têm um comportamento do tipo “passeio aleatório”, no sentido em que as melhores previsões para o dia de amanhã correspondem aos preços observados hoje (Caiado, 2016).

Na Figura 2.3 pode observar-se uma série simulada com 100 observações de um modelo de passeio aleatório, com ϵ_t um ruído branco gaussiano de média nula e variância unitária, bem como o resultado da diferenciação de 1.^a ordem da mesma. O correlograma de um passeio aleatório apresenta correlações positivas que decaem muito lentamente para zero.

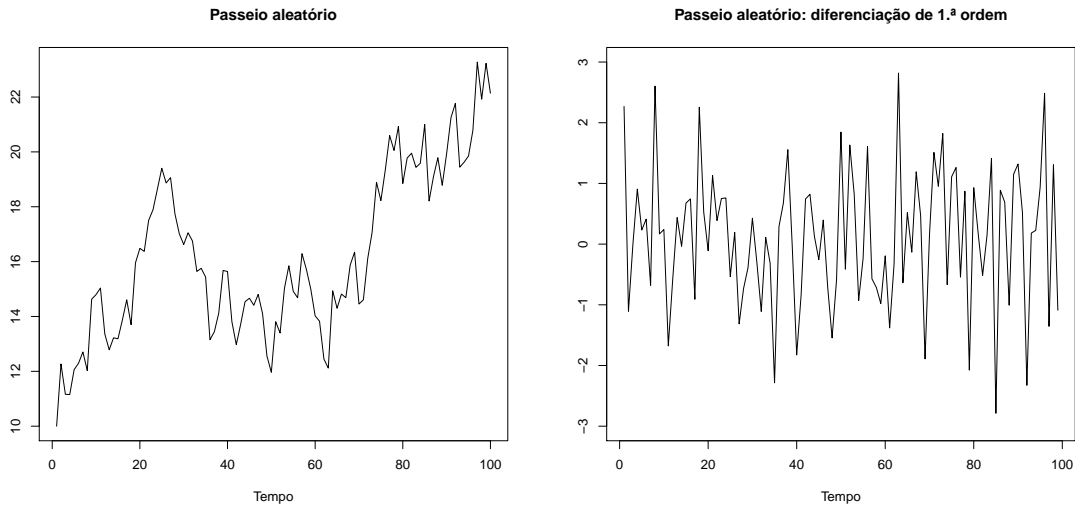


Figura 2.3: Simulação de um modelo de passeio aleatório e as suas diferenças de 1.^a ordem.

Um exemplo de um processo estocástico não estacionário é designado por passeio aleatório (*random walk*): uma partícula em movimento, ocupando inicialmente a posição Y_0 , é observada numa sucessão discreta de pontos, $t = 1, 2, 3, \dots$; em cada momento $t = k$, a partícula apresenta um “salto” de amplitude ϵ_k . Assim, no

momento t a posição da partícula é dada por

$$Y_t = Y_0 + \epsilon_1 + \epsilon_2 + \cdots + \epsilon_t \quad \text{ou} \quad Y_t = Y_{t-1} + \epsilon_t,$$

onde ϵ_t , $t = 1, 2, 3, \dots$ é uma sucessão de variáveis aleatórias não correlacionadas. Diz-se que a partícula realiza um passeio aleatório quando $E[\epsilon_t] = 0$, $t = 1, 2, 3, \dots$. Neste caso, a média, a variância e a covariância dependem do tempo histórico, i.e., da data em que se inicia o passeio (Murteira et al., 1993).

Passeio aleatório com *drift*

O modelo de passeio aleatório com *drift* é uma extensão do modelo anterior, quando se adiciona um termo constante a_0 , i.e.,

$$Y_t = a_0 + Y_{t-1} + \epsilon_t, \tag{2.9}$$

onde ϵ_t é um ruído branco. Neste caso, considera-se que a tendência é parcialmente determinística e parcialmente estocástica. Por esse motivo, o valor médio deste processo depende de t , não sendo constante. De facto, se Y_0 for conhecido, o modelo de passeio aleatório com *drift* pode ser reescrito como

$$Y_t = Y_0 + a_0 t + \sum_{i=1}^t \epsilon_i, \tag{2.10}$$

e, calculando o valor esperado de Y_t , conclui-se que este é igual a $Y_0 + a_0 t$ e, portanto, dependente de t . Desta forma, também o modelo de passeio aleatório com *drift* é não estacionário, podendo ser transformado em estacionário se se aplicar uma diferenciação de 1.^a ordem.

Na Figura 2.4 pode observar-se uma série simulada com 100 observações de um modelo de passeio aleatório com *drift*, onde ϵ_t é um ruído branco gaussiano de média nula e variância unitária e $a_0 = 0,5$, bem como o resultado da diferenciação de 1.^a ordem da mesma.

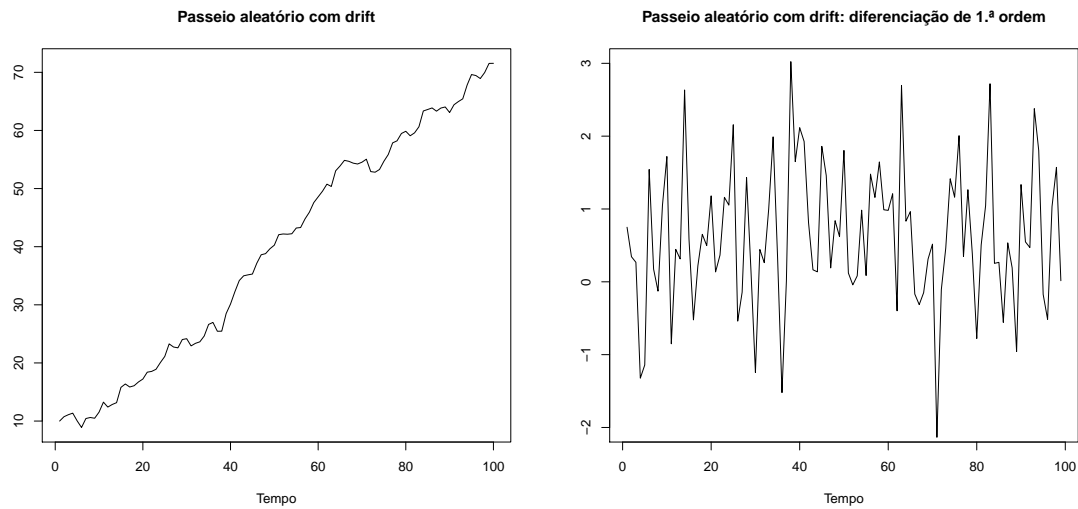


Figura 2.4: Simulação de um modelo de passeio aleatório com *drift* e as suas diferenças de 1.^a ordem.

Usualmente, neste tipo de processos, a tendência determinística é dominante. No entanto, o aumento da variância de ϵ_t ou a diminuição do valor absoluto de a_0 podem mascarar a existência de uma tendência determinística.

Análise de estacionariedade

Uma forma elementar de analisar a estacionariedade de uma série é representar graficamente os dados ao longo do tempo. No entanto, esta análise é subjetiva e, apesar de útil, deve ser confirmada através de testes estatísticos formais. Existem vários testes para realizar este estudo, baseados, na sua maioria, em encontrar uma raiz unitária. Segundo a literatura, deve utilizar-se mais do que um teste de raiz unitária de modo a avaliar a estacionariedade da série. Alguns dos testes mais utilizados são os testes de Dickey-Fuller (DF), Dickey-Fuller Aumentado (*Augmented Dickey Fuller*, ADF), Phillips-Perron (PP) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Dentro destes testes salientam-se dois tipos: aqueles cuja hipótese a testar é a presença de uma raiz unitária (ou seja, a não estacionariedade) e, no caso de esta não ser rejeitada, são fornecidas informações sobre o número de diferenciações necessárias para atingir a estacionariedade (DF, ADF e PP); e aqueles cuja hipótese nula é a estacionariedade (KPSS)

Teste de Dickey-Fuller (1979)

Considere-se um processo definido por

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad -1 \leq \phi \leq 1, \quad (2.11)$$

onde ϵ_t é um ruído branco. Este processo é estacionário se $|\phi| < 1$ (trata-se de um processo autorregressivo de ordem 1, como se verá na Secção 3.1.1). No entanto, quando $\phi = 1$ este processo é um passeio aleatório, o que, como foi visto anteriormente, equivale a fazer uma diferenciação de 1.^a ordem. Assim, se $\phi = 1$, pode dizer-se que a série é não estacionária.

O processo descrito pela equação (2.11) pode ser escrito na forma das diferenças, ou seja,

$$Y_t = \phi Y_{t-1} + \epsilon_t \Leftrightarrow Y_t - Y_{t-1} = \phi Y_{t-1} - Y_{t-1} + \epsilon_t \quad (2.12)$$

$$\Leftrightarrow \nabla Y_t = (\phi - 1)Y_{t-1} + \epsilon_t \quad (2.13)$$

$$\Leftrightarrow \nabla Y_t = \delta Y_{t-1} + \epsilon_t, \quad (2.14)$$

onde $\delta = \phi - 1$ e ϵ_t é um processo estacionário. Assim, as hipóteses para o teste DF em termos de δ são dadas por

$$H_0 : \delta = 0 \quad vs \quad H_1 : -2 < \delta < 0.$$

De notar que, se $\delta = 0$ então $Y_t \sim I(1)$. Isto é, a série das primeiras diferenças é estacionária e a série original é um passeio aleatório. Se $-2 < \delta < 0$, então o processo Y_t é estacionário. Com base nesta modificação, Dickey e Fuller (1979) propõem o teste

$$H_0 : \delta = 0 \quad vs \quad H_1 : \delta < 0.$$

Se a hipótese H_0 não é rejeitada então o processo Y_t é não estacionário, i.e., $Y_t \sim I(1)$. No caso de $\delta < 0$ então $\phi - 1 < 0 \Leftrightarrow \phi < 1$ e, tendo em conta as condições de estacionariedade para este processo, a série Y_t é estacionária.

Além da equação (2.14), Dickey e Fuller consideram também duas equações alternativas:

$$\nabla Y_t = a_0 + \delta Y_{t-1} + \epsilon_t \quad (2.15)$$

$$\nabla Y_t = a_0 + a_1 t + \delta Y_{t-1} + \epsilon_t \quad (2.16)$$

A diferença entre as três equações reside, essencialmente, na presença/ausência de uma constante a_0 e/ou de um termo determinístico $a_1 t$. Assim, a rejeição da hipó-

tese nula, $H_0 : \delta = 0$, considerando a equação (2.15), implica que o processo Y_t é estacionário de média não nula. No caso da equação (2.16), a indicação é de que $Y_t \sim I(0)$ com uma tendência determinística.

Ao aplicar o teste DF deve verificar-se se a sequência ϵ_t é não correlacionada, uma vez que, quando este pressuposto não se verifica, estes testes tendem a exibir um fraco comportamento. Caso isso aconteça, o teste DF não pode ser aplicado e deve optar-se por uma versão do teste para ordens superiores, ou seja, o teste ADF.

Teste de Dickey-Fuller Aumentado

Muitas séries apresentam uma estrutura mais complexa do que a captada pelo modelo (2.11). Para lidar com esses casos, surge uma variante do teste DF capaz de integrar modelos mais complexos. Esta inovação deve-se a Said e Dickey (1984) e, desde então, este tem sido um dos testes mais usados no estudo da estacionariedade de séries temporais.

Considere-se, neste caso, um processo mais complexo que o apresentado em (2.11) definido por

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t. \quad (2.17)$$

Assim, repetindo o processo da (2.12) a (2.14), obtém-se

$$\nabla Y_t = \delta Y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \nabla Y_{t-j} + \epsilon_t, \quad (2.18)$$

onde $\delta = \sum_{i=1}^p \phi_i - 1$, $\gamma_j = -\sum_{i=j+1}^p \phi_i$, $\nabla Y_{t-j} = Y_{t-j} - Y_{t-j-1}$ e ϵ_t é um ruído branco. Esta decomposição separa o modelo (2.17) em dois termos: Y_{t-1} e as $p-1$ primeiras diferenças. No caso em que Y_t é $I(1)$ (passeio aleatório), esta separação envolve uma componente $I(1)$ e $p-1$ componentes $I(0)$ (estacionárias). Na terminologia comum, diz-se que o modelo original foi aumentado por $p-1$ componentes de primeiras diferenças, dando origem à designação $ADF(p-1)$.

Para o modelo (2.18), a existência de uma raiz unitária é garantida se $\sum_{i=1}^p \phi_i = 1$, i.e., se $\delta = 0$. Assim, as hipóteses a testar são, também neste caso,

$$H_0 : \delta = 0 \quad vs \quad H_1 : \delta < 0$$

e a não rejeição da hipótese nula implica a não estacionariedade de Y_t . Como este teste é unilateral à esquerda, a hipótese nula é rejeitada a um nível de significância α se a estatística de teste for inferior ou igual ao quantil $(1-\alpha)100\%$ da distribuição correspondente (valor crítico).

Além da equação (2.18) e à semelhança do que acontece no teste DF, são propostas duas outras equações

$$\nabla Y_t = a_0 + \delta Y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \nabla Y_{t-j} + \epsilon_t, \quad (2.19)$$

$$\nabla Y_t = a_0 + a_1 t + \delta Y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \nabla Y_{t-j} + \epsilon_t, \quad (2.20)$$

cujas diferenças reside, novamente, na presença/ausência de uma constante a_0 e/ou de um termo determinístico $a_1 t$.

Um dos principais problemas do teste ADF é decidir qual o número de termos a incluir na equação a ser testada, ou seja, o valor de p . Para essa escolha existem várias estratégias, entre as quais:

1. Escolher o valor de p empiricamente. A ideia é incluir um número suficiente de termos de forma que a estrutura dos resíduos não apresente correlação;
2. Minimizar algum critério de informação como, por exemplo, o critério de informação de Akaike (*Akaike Information Criterion*, AIC);
3. Usar a regra proposta por Ny e Perron (1995): primeiro, define-se um limite máximo para p , p_{max} ; de seguida, calcula-se o teste ADF considerando $p = p_{max}$; se o valor absoluto da estatística t para testar a significância da diferença de ordem p for maior do que 1,6, define-se $p = p_{max}$ e prossegue-se com o teste; caso contrário, reduz-se p em uma unidade e repete-se o processo. Para determinar p_{max} pode utilizar-se

$$p_{max} = \left\lceil 12 \left(\frac{T}{100} \right)^{1/4} \right\rceil,$$

onde $[x]$ representa a parte inteira de x e T o número de observações. Esta será a estratégia adotada para definir o valor de p .

Independentemente da estratégia escolhida, o importante é averiguar se os resíduos do modelo atuam como um ruído branco antes da aplicação do teste.

Teste de Phillips-Perron

O teste de Phillips-Perron (PP) (Phillips (1987) e Perron (1988)) difere do teste ADF essencialmente na estrutura permitida para os erros, ϵ_t . De facto, enquanto

que o teste ADF exige que estes sejam independentes, o teste PP permite que os erros sejam correlacionados e, possivelmente, heterocedásticos. Assim, as equações a testar são iguais às apresentadas para o teste ADF, com a única diferença que ϵ_t deixa de ser um ruído branco. As hipóteses a testar são também idênticas e, portanto, a rejeição da hipótese nula implica a estacionariedade do processo.

Teste de KPSS

O teste de KPSS, proposto por Kwiatkowski, Phillips, Schmidt e Shin (1992), serve, tal como os anteriores, para avaliar a estacionariedade de um processo. No entanto, neste caso, as hipóteses a testar estão trocadas, ou seja,

H_0 : O processo é estacionário *vs* H_1 : O processo é não estacionário.

Se se considerar o processo Y_t , a equação deste teste decompõe Y_t numa soma de três componentes: uma tendência determinística (T_t), um passeio aleatório (μ_t) e um erro estacionário (u_t), i.e.,

$$Y_t = T_t + \mu_t + u_t, \quad (2.21)$$

$$\mu_t = \mu_{t-1} + \epsilon_t, \quad (2.22)$$

onde ϵ_t é um ruído branco. Este teste é unilateral à direita e, portanto, a hipótese nula é rejeitada a um nível de significância α se a estatística de teste for superior ou igual ao quantil $(1 - \alpha)100\%$ da distribuição correspondente (valor crítico).

Capítulo 3

Modelos e métodos de previsão

A distinção entre modelo, representação matemática da estrutura estocástica de uma série temporal através de uma equação ou sistema de equações, e método, procedimento para calcular previsões, nem sempre foi clara.

Um modelo estatístico determina um processo gerador dos dados, que pode ser usado para obter toda a distribuição de probabilidade para um instante futuro. Os modelos utilizados para as séries temporais são processos estocásticos e, portanto, controlados por leis de probabilidade. Além das previsões pontuais para o futuro, um modelo também permite calcular os intervalos de previsão, com um nível de confiança associado.

Um método de previsão é um procedimento para calcular previsões a partir de valores presentes e passados. Como tal, pode ser simplesmente um algoritmo e não depender de um modelo de probabilidade subjacente ou, alternativamente, surgir da identificação de um modelo específico para os dados fornecidos e da localização de previsões condicionadas a esse modelo (Chatfield, 2000). De acordo com (Chatfield, 2004), os métodos de previsão podem ser classificados em três grupos:

Métodos subjetivos, que, tal como o nome indica, envolvem alguma subjetividade, baseada em julgamento, intuição, conhecimento comercial ou qualquer outra informação relevante;

Métodos univariados, onde as previsões dependem apenas dos valores presentes e passados de uma série temporal única;

Métodos multivariados, em que as previsões para uma dada variável dependem, pelo menos em parte, dos valores de uma ou mais variáveis de séries temporais adicionais, designadas de preditores ou variáveis explicativas.

Na prática, um método de previsão pode resultar da combinação de várias abordagens como, por exemplo, quando métodos univariados ou multivariados envolvem alguma subjetividade, de forma a ter em conta informações externas que dificilmente se expressam através de um modelo matemático.

A escolha do método depende de uma variedade de considerações, incluindo:

- o objetivo do cálculo das previsões;
- o tipo de série temporal e a presença/ausência de componentes como a tendência e/ou a sazonalidade;
- a dimensão da série temporal (n);
- o horizonte de previsão (h);
- o conhecimento e a experiência do analista;
- e a disponibilidade dos programas informáticos.

De facto, existe uma enorme variedade de métodos de previsão que se podem aplicar a uma série temporal, cada um com as suas capacidades e limitações. Em cada caso, deve escolher-se o método que pareça mais adequado, nunca excluindo a hipótese de experimentar outros métodos. Nesta unidade curricular são estudadas duas abordagens distintas: a clássica de Box-Jenkins e a não paramétrica de Holt-Winters. Estas metodologias permitem modelar tendências e flutuações sazonais presentes em séries temporais e, dentro dos métodos de previsão mais tradicionais, são frequentemente utilizadas em estudos comparativos na literatura.

3.1 Modelos SARIMA e a metodologia Box-Jenkins

Os modelos SARIMA (*seasonal autoregressive integrated moving average*), introduzidos por Box e Jenkins (1970), permitem modelar e prever séries temporais estacionárias e não estacionárias, descrevendo a série Y_t como função dos seus valores passados e como combinação linear de uma sucessão de choques aleatórios.

Dentro dos modelos SARIMA, podem considerar-se vários modelos mais simples: o modelo autorregressivo (*autoregressive*, AR), que considera que o comportamento da série pode ser explicado através do seu passado; o modelo de médias móveis (*moving average*, MA), que explica a série temporal através de uma sucessão de choques aleatórios; e ainda o modelo autorregressivo e de médias móveis (*autoregressive*

moving average, ARMA), que, tal como o nome indica, se trata de uma combinação dos dois modelos anteriores. Estes modelos são úteis para séries estacionárias, revelando-se, no entanto, insuficientes para modelar casos de não estacionariedade. Nesses casos, deve optar-se pelos modelos integrados (ARIMA) ou, para séries que apresentam sazonalidade, pelos equivalentes sazonais (SARIMA).

A metodologia Box-Jenkins trata-se de um processo iterativo de três fases (identificação, estimação e diagnóstico), que facilita a escolha do modelo SARIMA que melhor descreve a série temporal em questão. Na primeira fase, tal como o nome indica, pretende-se identificar, através da análise da representação gráfica dos dados e das respetivas FAC e FACP empíricas, o modelo SARIMA mais apropriado. Os parâmetros do modelo escolhido são estimados na segunda fase (estimação) e avaliados quanto à sua significância na fase de diagnóstico. Também na última fase se avalia o comportamento dos resíduos, que se deve assemelhar a um ruído branco. Estas três fases são aplicadas iterativamente até que o modelo final não possa ser melhorado.

No entanto, a fase de identificação do modelo é bastante subjetiva e complexa, resultando, muitas vezes, em modelos distintos para a mesma análise. Para contornar esta situação, muitos *softwares* já possuem algoritmos automatizados de seleção, que permitem não só que a metodologia possa ser utilizada por leigos, mas também que os resultados obtidos por diferentes analistas sejam os mesmos.

A maior vantagem destes modelos é que, ao contrário do que acontece no alisamento exponencial e na regressão linear, estes têm em consideração uma das principais características dos dados de séries temporais: a dependência temporal, ou seja, a autocorrelação. Desta forma, os modelos SARIMA são apropriados quando se pode assumir que existe algum tipo de relação entre o passado e o futuro, sendo, no entanto, isso que os torna pouco recomendados para previsões a longo prazo e/ou previsões de séries com mudanças bruscas de comportamento.

Uma série estacionária é completamente especificada por sua função de média, variância e autocorrelação, uma das principais (e mais subjetivas) tarefas da análise de Box-Jenkins é identificar um modelo apropriado a partir da função de autocorrelação de amostra.

Dentro dos modelos ARIMA, podem considerar-se vários modelos mais simples: o modelo autorregressivo (AR), que considera que o comportamento da série pode ser explicado através do seu passado; o modelo de médias móveis (MA) que explica a série temporal através de uma sucessão de choques aleatórios; e ainda o modelo autorregressivo e de médias móveis (ARMA), que, tal como o nome indica, se trata de

uma combinação dos dois modelos anteriores. Para lidar com séries com sazonalidade existe o modelo SARIMA, que é uma generalização do modelo ARIMA para esses casos.

Mais detalhes sobre estes modelos e a metodologia Box-Jenkins podem ser consultados em Box et al. (2016).

3.1.1 Modelos de processos estacionários

Uma série estacionária fica completamente definida pelas suas funções média, variância e de autocorrelação. Fazendo uso desta característica, pretende-se, com a metodologia Box-Jenkins, identificar um modelo com base no comportamento da função de autocorrelação empírica.

Os processos ARMA são considerados como um grupo bastante diversificado e de grande fiabilidade na modelação de inúmeras séries temporais estacionárias. No entanto, para a modelação de séries que apresentem oscilações bruscas ao longo do tempo, este tipo de processos revela-se insuficiente. Ainda assim, os processos ARMA têm especial importância na modelação de séries não estacionárias, uma vez que estas são facilmente convertidas em estacionárias através de transformações adequadas.

Processo autorregressivo de ordem p , $AR(p)$

Os processos autorregressivos que pertencem à classe dos modelos mais usados no estudo de séries temporais estacionárias, baseiam-se no pressuposto de que a observação da variável no instante t se relaciona, de forma linear, com as observações nos instantes anteriores. Assim, o processo Y_t diz-se um processo autorregressivo de ordem p , $AR(p)$, quando satisfaz a equação

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t, \quad (3.1)$$

onde ϵ_t é um ruído branco de média nula, independente de Y_{t-k} para todo o $k \geq 1$. De facto, Y_t pode ser vista como uma variável dependente que é explicada através de uma regressão linear múltipla, em que as observações em p instantes anteriores funcionam como variáveis explicativas e ϕ_i são os coeficientes de cada Y_{t-i} .

Alternativamente, a representação de um processo $AR(p)$ pode ser feita através do operador atraso B^k , que se define como sendo $B^k Y_t = Y_{t-k}$. Com efeito, a equação (3.1) pode ser reescrita como

$$\Phi_p(B)Y_t = \epsilon_t, \quad (3.2)$$

onde $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinómio autorregressivo de ordem p . Tendo em consideração as p raízes (reais ou complexas), $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$, da equação característica $\Phi_p(B) = 0$, torna-se possível fatorizar o polinómio autorregressivo do seguinte modo

$$\Phi_p(B) = \prod_{i=1}^p (1 - G_i B). \quad (3.3)$$

Para que o processo seja estacionário é condição necessária e suficiente que as raízes da equação característica sejam todas de módulo maior do que a unidade, ou, de forma equivalente, que $|G_i| < 1$, para $i = 1, 2, \dots, p$. Qualquer processo autorregressivo que seja estacionário é também invertível, o que, em termos práticos, significa que a dependência do passado se vai atenuando à medida que o passado se torna mais remoto.

Se o processo Y_t é um processo $AR(p)$, então a sua função de autocorrelação parcial, ϕ_{kk} , é igual a zero para todo o $k > p$. Assim, a FACP de um processo $AR(p)$ apresenta, graficamente, uma queda brusca para zero a partir do $lag\ p + 1$, enquanto que a respetiva FAC tem um decaimento exponencial ou sinusoidal amortecido para zero. Na Figura 3.1 encontra-se representado um processo autorregressivo de ordem 2, $AR(2)$, e as respetivas FAC e FACP empíricas. Este processo é estacionário, uma vez que as raízes da equação característica $1 - 0,6B + 0,3B^2 = 0$ são as duas, em módulo, aproximadamente iguais a 1,83.

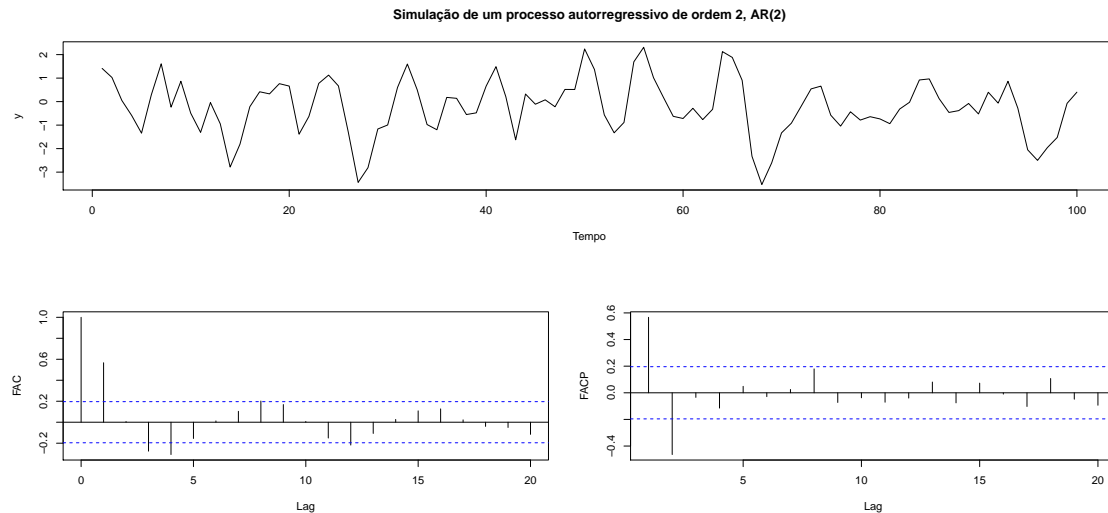


Figura 3.1: Simulação de um processo autorregressivo de equação $(1 - 0,6B + 0,3B^2)Y_t = \epsilon_t$, e respetivas FAC e FACP empíricas.

$$(\phi_1 = 0,6; \phi_2 = -0,3)$$

Processo de médias móveis de ordem q , $MA(q)$

O processo Y_t diz-se um processo de médias móveis de ordem q , $MA(q)$, quando assume a expressão

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad (3.4)$$

ou

$$Y_t = \Theta_q(B) \epsilon_t, \quad (3.5)$$

onde ϵ_t é um ruído branco de média nula e $\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ é o polinómio de médias móveis de ordem q . Pretende-se, através destes processos, exprimir Y_t em termos de um processo mais simples, como é o ruído branco. Assim, um processo de médias móveis de ordem q define-se, em cada instante t , como a média ponderada das $q + 1$ observações de um processo de ruído branco. Desta forma, graças à estacionariedade inerente ao ruído branco, os processos de médias móveis são sempre estacionários. Adicionalmente, um processo de médias móveis é invertível se puder ser escrito como um processo autorregressivo estacionário de ordem infinita. Para garantir a invertibilidade do processo, basta que, à semelhança do que acontece no caso da estacionariedade de processos autorregressivos, as raízes da equação característica $\Theta_q(B) = 0$ se encontrem todas fora do círculo unitário, isto é, sejam, em módulo, todas superiores a 1.

Se o processo Y_t é um processo $MA(q)$, então a sua função de autocorrelação, ρ_k , é igual a zero para todo o $k > q$, e, por isso, a FAC de um processo $MA(q)$ apresenta, graficamente, uma queda brusca para zero a partir do *lag* $q + 1$. No que respeita à FACP, esta exhibe um decaimento exponencial ou sinusoidal amortecido para zero tendo, portanto, a mesma estrutura que a FAC de um processo $AR(q)$. Na Figura 3.2 encontra-se representado um processo de médias móveis de ordem 2, $MA(2)$, e as respetivas FAC e FACP empíricas. Este processo é invertível, uma vez que as raízes da equação característica $1 - 0,7B + 0,1B^2 = 0$ são iguais a 2 e a 5.

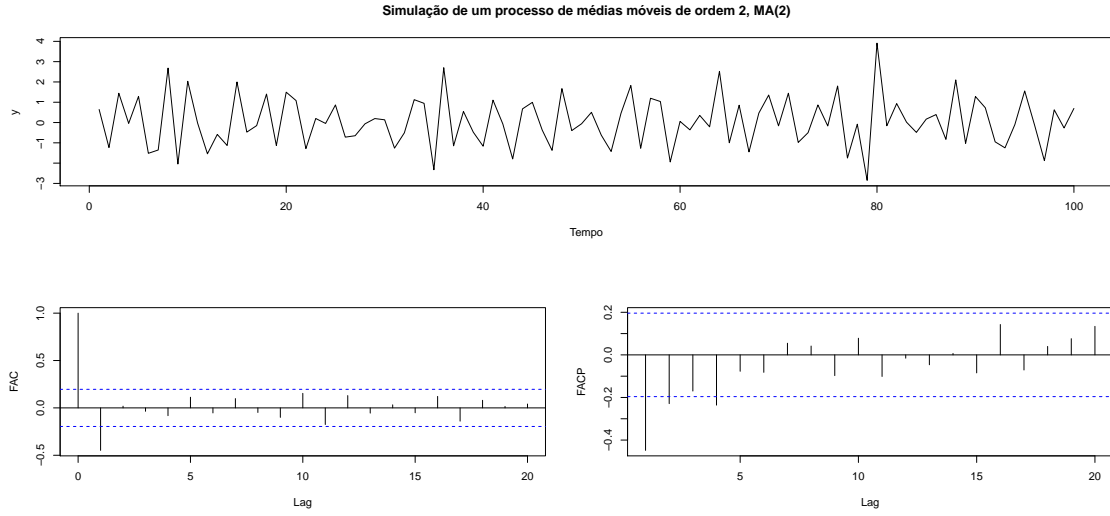


Figura 3.2: Simulação de um processo de médias móveis de equação $Y_t = (1 - 0,7B + 0,1B^2)\epsilon_t$, e respectivas FAC e FACP empíricas.

$$(\theta_1 = -0,7; \theta_2 = 0,1)$$

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

Os processos estacionários e invertíveis podem ser representados quer na forma autorregressiva quer na forma de médias móveis. Contudo, pode acontecer que qualquer um destes processos tenha uma representação com um número excessivo de parâmetros, o que pode conduzir a uma perda de eficiência na sua estimação (Caiado, 2016). Se assim for, pode construir-se um modelo mais parcimonioso que inclua tanto termos autorregressivos como de médias móveis. Este modelo designa-se de processo misto autorregressivo e de médias móveis de ordens p e q e representa-se por ARMA(p, q).

Assim, o processo Y_t diz-se um processo autorregressivo e de médias móveis de ordens p e q , ARMA(p, q), se satisfaz a equação

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \quad (3.6)$$

ou a equação

$$\Phi_p(B)Y_t = \Theta_q(B)\epsilon_t, \quad (3.7)$$

onde ϵ_t é um ruído branco de média nula, independente de Y_{t-k} para todo o $k \geq 1$, $\Phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ e $\Theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ são os polinómios autorregressivo e de médias móveis de ordens p e q , respetivamente.

A FAC e a FACP de um processo ARMA(p, q) resultam da combinação das

respetivas funções dos processos $AR(p)$ e $MA(q)$. Recorde-se que a FAC de um processo $MA(q)$ é insignificante a partir do lag $q + 1$, o mesmo acontecendo para a FACP de um processo $AR(p)$ depois do lag p . **Desta forma, um processo $ARMA(p, q)$ diz-se de ordem k quando a FAC e a FACP decrescem exponencialmente ou sinusoidalmente para zero a partir de $k > \max(p, q)$.** Na Figura 3.3 encontra-se representado um processo autorregressivo e de médias móveis, $ARMA(2, 2)$, e as respetivas FAC e FACP empíricas. Este processo é estacionário e invertível.

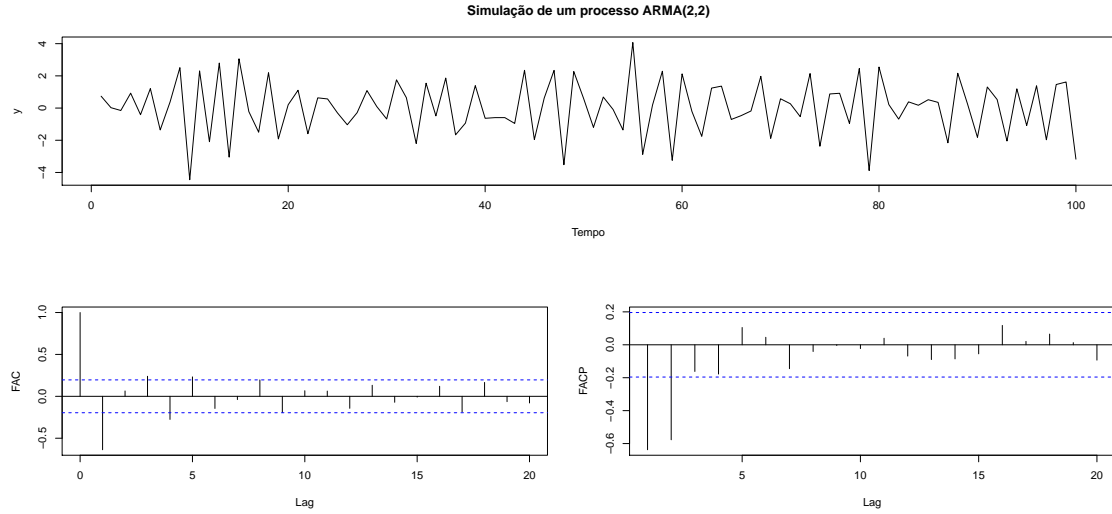


Figura 3.3: Simulação de um processo autorregressivo e de médias móveis, $ARMA(2, 2)$, de equação $(1 + 0,5B + 0,3B^2)Y_t = (1 - 0,7B + 0,1B^2)\epsilon_t$, e respetivas FAC e FACP empíricas.

$$(\theta_1 = -0,7; \theta_2 = 0,1; \phi_1 = -0,5; \phi_2 = -0,3)$$

Este processo é estacionário e invertível, uma vez que as raízes das equações características $1 + 0,5B + 0,3B^2 = 0$ e $1 - 0,7B + 0,1B^2 = 0$ são, em módulo, superiores a 1.

De facto, uma vez que o processo $\text{ARMA}(p, q)$ é uma combinação dos processos $\text{AR}(p)$ e $\text{MA}(q)$, a estacionariedade e a invertibilidade do processo ficam garantidas se as raízes das equações características $\Phi_p(B) = 0$ e $\Theta_q(B) = 0$ são, em módulo, maiores do que a unidade. Como é evidente, estes processos generalizam os processos anteriormente mencionados e, por exemplo, um processo $\text{ARMA}(p, 0)$ é equivalente a um processo $\text{AR}(p)$, o mesmo acontecendo com um $\text{ARMA}(0, q)$ relativamente a um $\text{MA}(q)$.

Processo autorregressivo e de médias móveis sazonal, $\text{SARMA}(P, Q)_s$

Algumas séries temporais exibem, por vezes, uma componente sazonal ou uma componente periódica de período s , caracterizada pela existência de correlação significativa entre observações desfasadas em s unidades temporais, $Y_t, Y_{t-s}, Y_{t-2s}, \dots$. De forma a contemplar essa correlação, existem os processos sazonais, que permitem modelar séries que apresentem um comportamento periódico.

O processo autorregressivo e de médias móveis sazonal (*seasonal autoregressive moving average*, SARMA) de ordens P e Q , $\text{SARMA}(P, Q)_s$, define-se através da equação

$$Y_t = \nu_1 Y_{t-s} + \dots + \nu_P Y_{t-Ps} + \epsilon_t + \eta_1 \epsilon_{t-s} + \dots + \eta_Q \epsilon_{t-Qs} \quad (3.8)$$

ou

$$N_P(B^s)Y_t = H_Q(B^s)\epsilon_t, \quad (3.9)$$

onde ϵ_t é um ruído branco de média nula, independente de Y_{t-ks} para todo o $k \geq 1$, $N_P(B^s) = 1 - \nu_1 B^s - \dots - \nu_P B^{Ps}$ e $H_Q(B^s) = 1 + \eta_1 B^s + \dots + \eta_Q B^{Qs}$ são os polinómios autorregressivo e de médias móveis sazonais de ordens P e Q , respetivamente. Tanto a FAC como a FACP do modelo $\text{SARMA}(P, Q)_s$ apresentam um comportamento do tipo exponencial ou sinusoidal amortecido sobre os *lags* múltiplos de s , mantendo-se nulas nos restantes *lags*. Na Figura 3.4 encontra-se representado um processo autorregressivo e de médias móveis sazonal de período 12, $\text{SARMA}(1, 1)_{12}$, e as respetivas FAC e FACP empíricas.

Os processos autorregressivos sazonais (*seasonal autoregressive*, SAR) de ordem P e os processos de médias móveis sazonais (*seasonal moving average*, SMA) de ordem Q são casos particulares dos processos $\text{SARMA}(P, Q)_s$, que se designam por $\text{SAR}(P)_s$ e $\text{SMA}(Q)_s$, respetivamente. A FAC do processo $\text{SAR}(P)_s$ tem um decaimento exponencial ou sinusoidal amortecido sobre os *lags* múltiplos de s , anulando-se nos restantes, e a FACP tem uma queda brusca para zero a partir do P -ésimo *lag* múltiplo de s . Já a FAC do $\text{SMA}(Q)_s$ apresenta uma queda brusca para zero a

partir do Q -ésimo *lag* múltiplo de s , enquanto que a FACP tem uma queda lenta para zero sobre os *lags* múltiplos de s (Caiado, 2016).

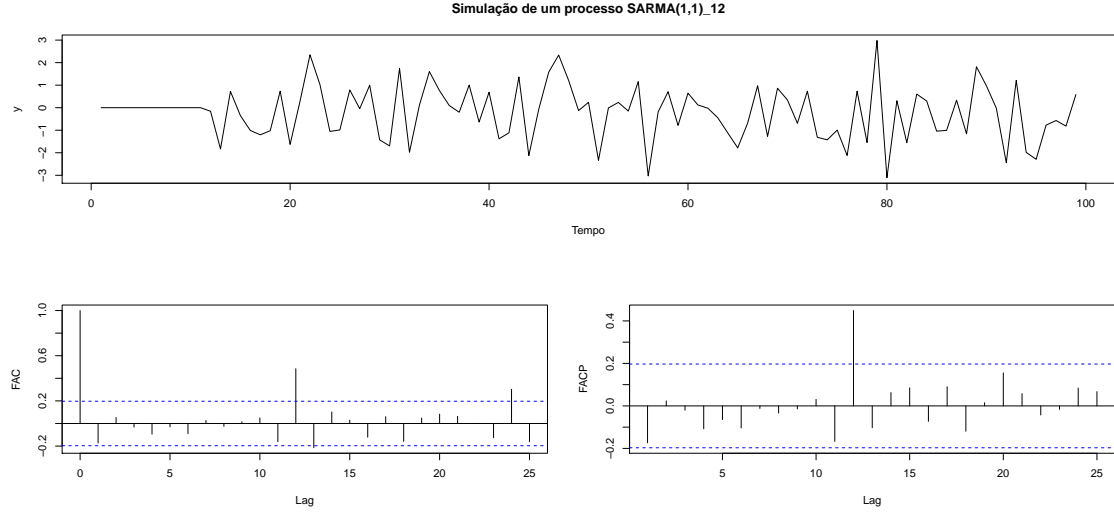


Figura 3.4: Simulação de um processo autorregressivo e de médias móveis sazonal, $\text{SARMA}(1, 1)_{12}$, de equação $(1 - 0,65B^{12})Y_t = (1 - 0,25B^{12})\epsilon_t$, e respetivas FAC e FACP empíricas.

$$(\nu_1 = 0,65; \eta_1 = -0,25)$$

Além disso, a combinação dos processos $\text{SARMA}(P, Q)_s$ com os processos $\text{ARMA}(p, q)$ permite obter o processo multiplicativo, com componente sazonal e não sazonal, $\text{SARMA}(p, q)(P, Q)_s$, com a equação

$$\Phi_p(B)N_P(B^s)Y_t = \Theta_q(B)H_Q(B^s)\epsilon_t, \quad (3.10)$$

onde $\Phi_p(B)$ e $N_P(B^s)$ correspondem aos polinómios anteriormente descritos e representam os polinómios autorregressivos regular (ou não sazonal) e sazonal, $\Theta_q(B)$ e $H_Q(B^s)$ representam os polinómios de médias móveis regular e sazonal, todos iguais aos anteriormente descritos. Quando as raízes destes quatro polinómios se encontram fora do círculo unitário ficam garantidas a estacionariedade e a invertibilidade.

Os comportamentos das FAC e FACP destes processos resultam também da combinação das características das respetivas funções para os processos $\text{ARMA}(p, q)$ e $\text{SARMA}(P, Q)_s$. Por exemplo, num processo $\text{SARMA}(1, 0)(1, 0)_{12}$, a FAC pode apresentar um decaimento sinusoidal amortecido para zero sobre os *lags* não sazonais e sazonais e a FACP apresentar autocorrelações significativas nos *lags* 1 e 12 e quedas bruscas para zero nos *lags* não sazonais e sazonais seguintes, como é possível observar na Figura 3.5 (Caiado, 2016).

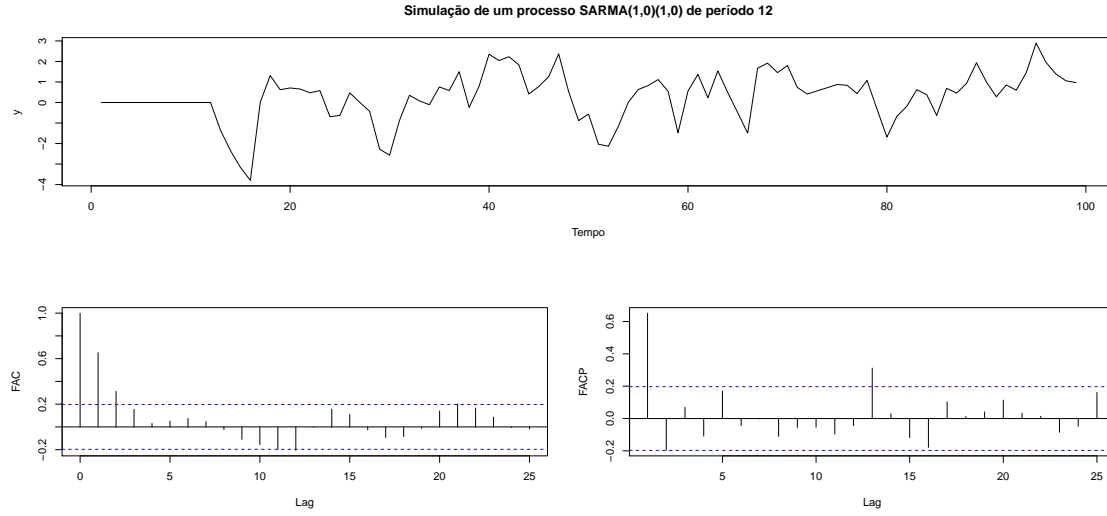


Figura 3.5: Simulação de um processo multiplicativo, $\text{SARMA}(1,0)(1,0)_{12}$, de equação $(1 - 0,7B)(1 + 0,25B^{12})Y_t = \epsilon_t$, e respectivas FAC e FACP empíricas.

$$(\phi_1 = 0,7; \nu_1 = -0,25)$$

Os comportamentos das FAC e FACP destes processos resultam também da combinação das características das respetivas funções para os processos $\text{ARMA}(p, q)$ e $\text{SARMA}(P, Q)_s$.

3.1.2 Modelos de processos não estacionários

Na prática, a maioria das séries temporais é não estacionária. Quando isso acontece, é necessário remover dos dados as fontes de variação não estacionárias (e.g., tendência, sazonalidade), de forma a possibilitar o ajustamento de um modelo estacionário. Como visto na Secção 2.2.2, se a série temporal observada for não estacionária na média, pode aplicar-se uma (ou várias) diferenciação (regular) à mesma. Assim, se se substituir Y_t por $\nabla^d Y_t$ na equação (3.7), obtém-se um modelo capaz de descrever séries não estacionárias (modelo ARIMA). Este tipo de modelo é designado de modelo “integrado”, uma vez que o modelo estacionário que é ajustado aos dados diferenciados deve ser somado ou “integrado” para fornecer um modelo para os dados não estacionários. Além disso, estes modelos podem, à semelhança dos modelos ARMA, ser generalizados para incluir termos sazonais, dando origem aos modelos SARIMA.

Os processos ARIMA podem ser generalizados para incluir termos sazonais, dando origem aos processos SARIMA.

Muitas das séries temporais não apresentam um comportamento correspondente

a trajetórias de um processo estacionário. Recorde-se que uma série é estacionária se apresenta um comportamento que se desenvolve aleatoriamente ao longo do tempo em redor de uma média constante, apresentando um equilíbrio estável. Mas se estas apresentam uma média μ_t que não é constante, uma variância σ_t^2 que não é constante, ou ainda ambas as características então diz-se que a série é não estacionária. É o caso de séries económicas e financeiras (e.g., índice da bolsa de valores de Lisboa, taxas de juro e câmbio, inflação, etc) que em geral apresentam tendência crescente ou decrescente ao longo do tempo e ainda podem apresentar variabilidade.

Processo autorregressivo e de médias móveis integrado, ARIMA(p, d, q)

O processo Y_t diz-se um processo autorregressivo e de médias móveis integrado, ARIMA(p, d, q), quando assume a expressão

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d Y_t = (1 + \theta_1 B + \cdots + \theta_q B^q) \epsilon_t \quad (3.11)$$

ou

$$\Phi_p(B) \nabla^d Y_t = \Theta_q(B) \epsilon_t, \quad (3.12)$$

onde $\nabla^d Y_t = (1 - B)^d Y_t$, com $d \geq 1$, é a série estacionária depois de diferenciada d vezes, $\phi_1, \phi_2, \dots, \phi_p$ são os parâmetros autorregressivos, $\theta_1, \theta_2, \dots, \theta_q$ são os parâmetros de médias móveis e $\Phi_p(B)$ e $\Theta_q(B)$ são os polinómios autorregressivo e de médias móveis regulares.

Por se tratar de um processo não estacionário, um processo deste tipo apresenta uma FAC com coeficientes positivos e decaimento muito lento para zero, pelo que a necessidade de uma diferenciação (regular) é facilmente identificável. Na Figura 3.6 encontra-se representado um processo autorregressivo e de médias móveis integrado, ARIMA(2, 1, 1), e as respetivas FAC e FACP empíricas.

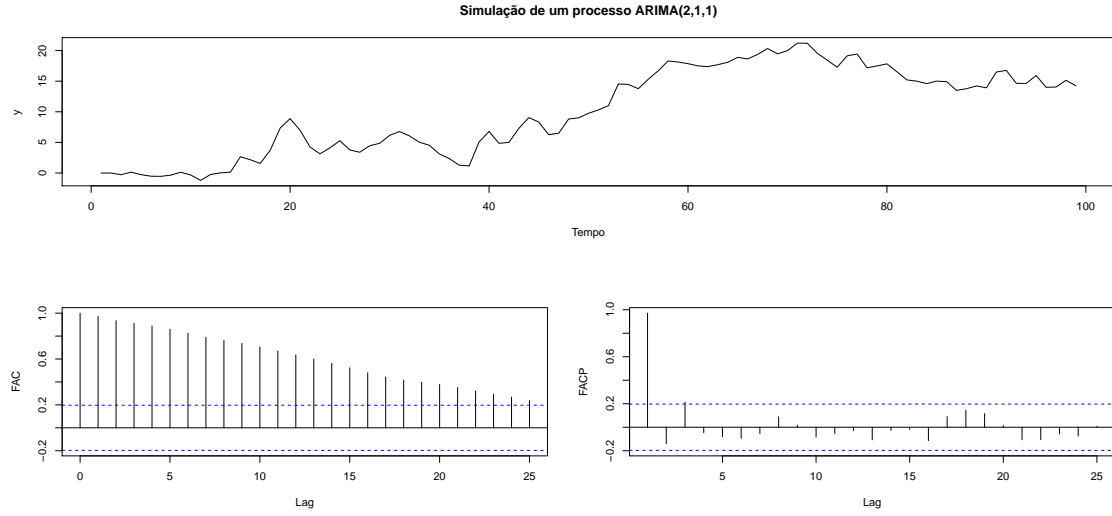


Figura 3.6: Simulação de um processo autorregressivo e de médias móveis integrado, ARIMA(2, 1, 1), de equação $(1 + 0,2B + 0,3B^2)(1 - B)Y_t = (1 + 0,7B)\epsilon_t$, e respetivas FAC e FACP empíricas.

$$(\phi_1 = -0,2; \phi_2 = -0,3; \theta_1 = 0,7)$$

As formas mais simples e mais frequentes do processo ARIMA são ARIMA(0, 1, 0), ARIMA(1, 1, 0), ARIMA(0, 1, 1) e ARIMA(1, 1, 1) (Caiado, 2016).

Processo autorregressivo e de médias móveis integrado sazonal, SARIMA(p, d, q)(P, D, Q)_s

Em muitas séries temporais sazonais é de esperar que a componente sazonal esteja de alguma forma relacionada com as componentes não sazonais. Isto é, se as observações vizinhas de uma série, $Y_t, Y_{t-1}, Y_{t-2}, \dots$, estão relacionadas, é muito provável que as observações vizinhas espaçadas em s unidades temporais, $Y_t, Y_{t-s}, Y_{t-2s}, \dots$, também estejam relacionadas. Assim sendo, pode estender-se o processo ARIMA(p, d, q) a um processo multiplicativo integrado sazonal, que generaliza todos os processos apresentados anteriormente e se representa por SARIMA(p, d, q)(P, D, Q)_s.

Antes de se introduzir a expressão do modelo completo, é conveniente apresentar-se a noção de diferenciação sazonal. Uma diferenciação sazonal não é mais do que a diferença entre a observação no instante t e a observação que ocorre s momentos antes, ou seja, a observação no instante $t - s$. Desta forma, quando uma série apresenta um comportamento periódico repetitivo, pode-se aplicar uma diferenciação sazonal, isto é,

$$\nabla_s Y_t = Y_t - Y_{t-s} = (1 - B^s)Y_t.$$

A série resultante desta diferenciação corresponde, então, à mudança entre observações separadas por períodos de tempo s . Por exemplo, para uma série mensal, com $s = 12$, a série resultante de uma diferenciação sazonal representa a mudança que ocorre de ano para ano.

À semelhança do que acontece para a diferenciação (regular) introduzida na Secção 2.2.2, a diferenciação sazonal pode ser aplicada a uma série D vezes, dando origem ao operador de diferenciação sazonal de ordem D , para qualquer inteiro $D \geq 1$, que se define por

$$\nabla_s^D Y_t = (1 - B^s)^D Y_t.$$

Assim, um processo Y_t diz-se um processo autorregressivo e de médias móveis integrado sazonal, $\text{SARIMA}(p, d, q)(P, D, Q)_s$, quando satisfaz a equação

$$\Phi_p(B)N_P(B^s)\nabla^d\nabla_s^D Y_t = \Theta_q(B)H_Q(B^s)\epsilon_t, \quad (3.13)$$

em que $\Phi_p(B)$, $N_P(B^s)$, $\Theta_q(B)$ e $H_Q(B^s)$ são os polinómios já referidos, d e D são as ordens de diferenciação das partes regular e sazonal, respetivamente.

Em geral, a necessidade de uma diferenciação sazonal pode ser indicada quando a FAC de um processo decai lentamente nos *lags* múltiplos de s e é insignificante nos restantes. Na Figura 3.7 encontra-se representado um processo $\text{SARIMA}(2, 1, 1)(1, 1, 1)_{12}$ e as respetivas FAC e FACP empíricas.

Na maioria das aplicações práticas, os valores de p, q, P e Q são quase sempre inferiores ou iguais a 2, enquanto que os valores de d e D usualmente assumem os valores inteiros 0 ou 1 (Caiado, 2016).

A FAC de um processo SARIMA é uma mistura das suas componentes regular e sazonal. Nos *lags* baixos o comportamento da FAC deve assemelhar-se ao da sua parte regular e nos *lags* sazonais deve observar-se apenas o efeito da parte sazonal. Em torno destes últimos pode observar-se a interação entre as partes regular e sazonal, que se manifesta na repetição em ambos os lados de cada *lag* sazonal da função FAC da parte regular.

Também a FACP revela influência tanto da parte regular como da parte sazonal. Nos primeiros *lags*, o comportamento é idêntico ao da parte regular e nos *lags* sazonais a FACP reflete a parte sazonal. À direita de cada *lag* sazonal deve observar-se a FACP da parte regular, enquanto que à esquerda se deve notar o efeito da FAC da parte regular.

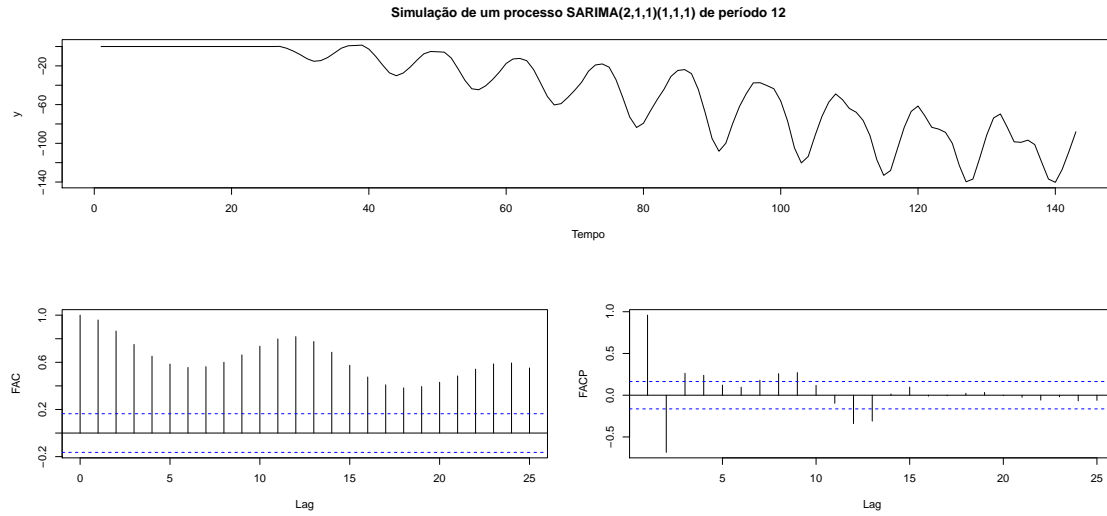


Figura 3.7: Simulação de um processo autorregressivo e de médias móveis integrado sazonal, $\text{SARIMA}(2, 1, 1)(1, 1, 1)_{12}$, de equação $(1 - 1,2B + 0,8B^2)(1 - 0,3B^{12})(1 - B)(1 - B^{12})Y_t = (1 + 0,4B)(1 + 0,7B^{12})\epsilon_t$, e respetivas FAC e FACP empíricas.

3.1.3 Metodologia Box-Jenkins

Baseados nos trabalhos de Yule (1926) e Wold (1938), Box e Jenkins (1970) introduziram uma abordagem prática e sistemática para a construção de modelos SARIMA, a conhecida metodologia Box-Jenkins. Esta metodologia não é mais do que um processo de modelação iterativo em três fases: identificação do modelo, estimação dos parâmetros e análise de diagnóstico (ou validação do modelo). Este processo é tipicamente repetido várias vezes até que um modelo satisfatório seja selecionado. O modelo final pode, então, ser usado para fins de previsão.

A ideia base por detrás da identificação do modelo é que, se uma série temporal é gerada a partir de um processo SARIMA, então deve ter algumas propriedades teóricas de autocorrelação. Assim, ao comparar os padrões empíricos de autocorrelação com os teóricos, é frequentemente possível identificar um ou vários potenciais modelos para a série temporal a estudar. Box e Jenkins (1970) propuseram, então, usar a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) como ferramentas básicas para identificar as ordens do modelo SARIMA.

Descrevem-se, de seguida, as três fases da metodologia Box-Jenkins, esquematizadas na Figura 3.8: identificação, estimação e diagnóstico.

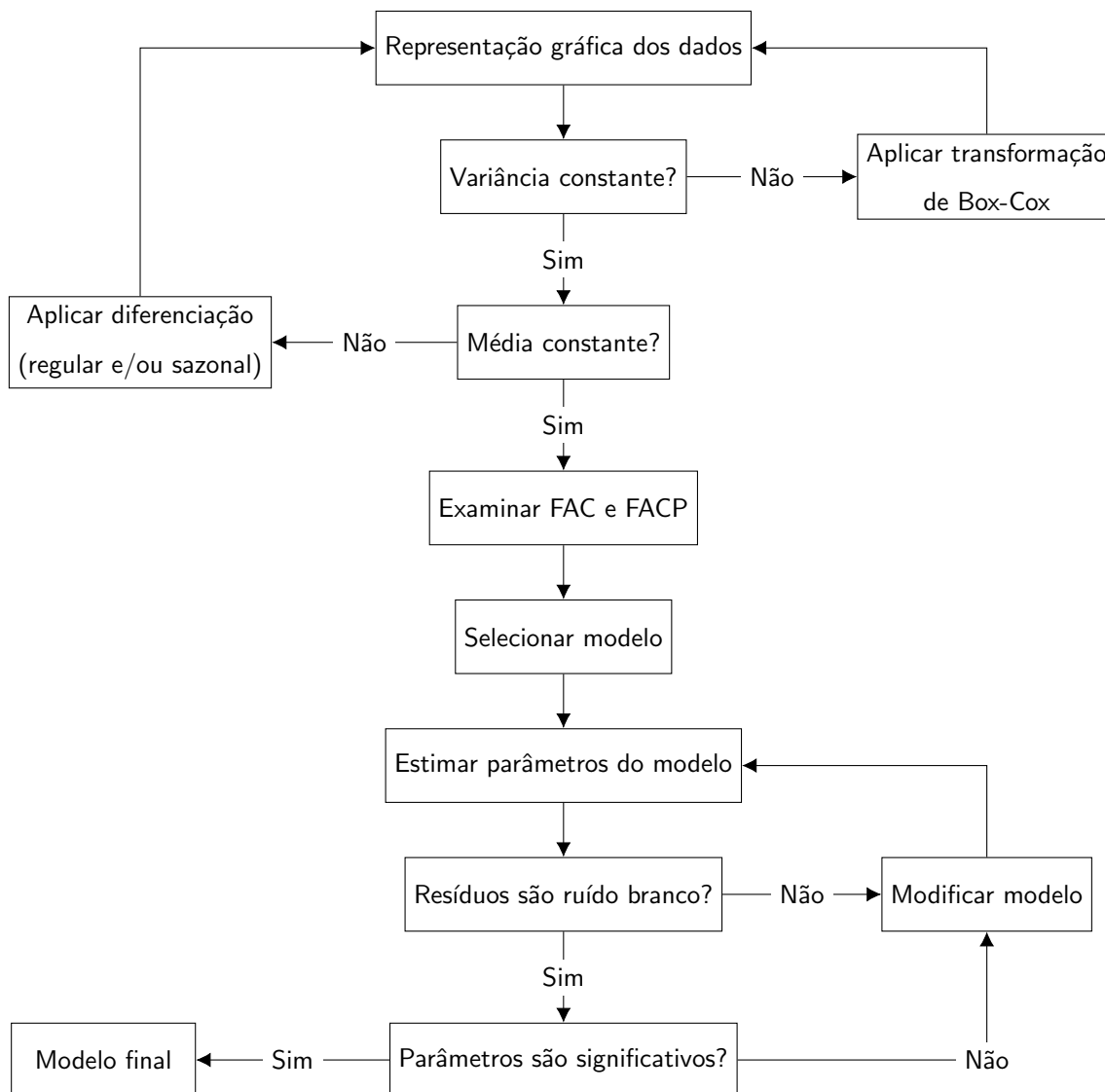


Figura 3.8: Esquema da metodologia Box-Jenkins.

Identificação

A primeira etapa de modelação de uma série temporal consiste na identificação de um modelo $SARIMA(p, d, q)(P, D, Q)_s$ que descreva a relação existente entre as suas observações. Esta etapa compreende três importantes passos na tentativa de identificação dos valores dos inteiros d, s, D, p, q, P e Q :

Passo 1: Representação gráfica da série e sua estacionarização

O estudo de uma qualquer série temporal deve começar pela análise detalhada da sua representação gráfica, a fim de se poder identificar a existência ou não de tendência, variância não constante, sazonalidade, *outliers* e outras

fontes de não estacionariedade. No caso da série ser não estacionária, deve, então, proceder-se à sua estacionarização através das transformações adequadas: para a estabilização da variância recorrer a transformações Box-Cox; para a eliminação da tendência usar a diferenciação regular; e para a eliminação de movimentos periódicos, a diferenciação sazonal. Recorde-se que, se a estabilização da variância for necessária, esta deve ser realizada antes de qualquer outra transformação.

Passo 2: Estimação das FAC e FACP da série original

Analisar o comportamento das FAC e FACP da série original pode ser útil para complementar a informação obtida através da representação gráfica (passo 1), nomeadamente no que diz respeito à utilização de diferenciações. Por exemplo, um decaimento lento para zero na FAC de uma série temporal pode indicar a necessidade da aplicação de uma diferenciação não sazonal (regular).

Passo 3: Estimação das FAC e FACP da série estacionária e identificação dos inteiros p, q, P e Q

Depois de se identificarem as ordens de diferenciação, d e D , e o período, s , a escolha do modelo que descreve a série temporal só fica concluída quando são determinados os inteiros p, q, P e Q . Uma das formas de identificar esses inteiros é através da comparação do comportamento das FAC e FACP empíricas com o das FAC e FACP teóricas dos modelos $\text{ARMA}(p, q)$ e $\text{SARMA}(p, q)(P, Q)_s$. A Tabela 3.1 resume os padrões teóricos apresentados pelas FAC e FACP dos modelos AR, MA, ARMA, SAR, SMA, SARMA e SARMA multiplicativo.

Estimação

Uma vez identificados os modelos candidatos a descrever a série em estudo, segue-se a etapa de estimação dos seus parâmetros. Nesta fase, é imprescindível o auxílio de um *software* estatístico adequado, dado que a estimação dos parâmetros requer a aplicação de um conjunto de métodos numéricos e de cálculos computacionais com alguma complexidade. Os dois principais métodos de estimação dos parâmetros do modelo SARIMA são o método da máxima verosimilhança e o método dos mínimos quadrados.

A ideia base do método da máxima verosimilhança consiste em determinar os valores dos parâmetros que tornam mais verosímil a ocorrência de um conjunto de observações idênticas aquelas de que efetivamente se dispõe. Este método obtém

estimativas dos parâmetros através de um processo iterativo em que se maximiza a função de verosimilhança dos estimadores (mais detalhes em Box et al., 2016). Em geral, os *softwares* estatísticos começam por obter estimativas iniciais dos parâmetros, seguindo-se um processo iterativo até se obterem os valores ótimos dos parâmetros.

O método dos mínimos quadrados é, provavelmente, o método estatístico mais utilizado na estimação de modelos. Contudo, em modelos $\text{ARMA}(p, q)$ este método não permite obter estimadores consistentes com os verdadeiros parâmetros. A única exceção ocorre nos casos em que os modelos têm apenas componente autorregressiva (modelos $\text{AR}(p)$), em que os estimadores se podem obter por minimização da soma de quadrados dos resíduos (Caiado, 2016).

Tabela 3.1: Padrões teóricos das FAC e FACP dos modelos $\text{SARMA}(p, q)(P, Q)_s$.

Modelo	FAC	FACP
$\text{AR}(p)$	Decaimento exponencial ou sinusoidal amortecido para zero	Queda brusca para zero a partir do $\text{lag } p + 1$
$\text{MA}(q)$	Queda brusca para zero a partir do $\text{lag } q + 1$	Decaimento exponencial ou sinusoidal amortecido para zero
$\text{ARMA}(p, q)$	Decaimento exponencial ou sinusoidal amortecido para zero	Decaimento exponencial ou sinusoidal amortecido para zero
$\text{SAR}(P)_s$	Decaimento exponencial ou sinusoidal amortecido para zero sobre os $\text{lags } s, 2s, \dots$	Queda brusca para zero a partir do $\text{lag } (P + 1)s$
$\text{SMA}(Q)_s$	Queda brusca para zero a partir do $\text{lag } (Q + 1)s$	Decaimento exponencial ou sinusoidal amortecido para zero sobre os $\text{lags } s, 2s, \dots$
$\text{SARMA}(P, Q)_s$	Decaimento exponencial ou sinusoidal amortecido para zero sobre os $\text{lags } s, 2s, \dots$	Decaimento exponencial ou sinusoidal amortecido para zero sobre os $\text{lags } s, 2s, \dots$
$\text{SARMA}(p, q)(P, Q)_s$	Decaimento exponencial ou sinusoidal amortecido para zero	Decaimento exponencial ou sinusoidal amortecido para zero

Diagnóstico

Após se identificar o modelo SARIMA e estimar os respetivos parâmetros, é necessário verificar se o modelo é realmente adequado. A fase de diagnóstico engloba duas importantes etapas: a avaliação da qualidade das estimativas obtidas e a avaliação da qualidade do ajustamento do modelo às observações da série em estudo.

Na avaliação da qualidade das estimativas obtidas interessa, sobretudo, analisar a significância estatística dos parâmetros estimados. Para tal, a cada parâmetro, diga-se β_i (onde $i = 1, \dots, m$ e m representa o número de parâmetros estimados), deve aplicar-se um teste de hipóteses que avalie a necessidade (ou não) de incluir esse parâmetro no modelo. Interessa, assim, testar a hipótese de que β_i é estatisticamente nulo, isto é, $H_0 : \beta_i = 0$. A rejeição desta hipótese acontece, a um nível de significância α , quando a estatística T associada ao coeficiente estimado for, em valor absoluto, superior ou igual ao quantil $1 - \alpha/2$ de uma distribuição t -Student com $n - m$ graus de liberdade, (onde n representa o número de observações), isto é,

$$|T| = \left| \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \right| \geq t_{1-\alpha/2; n-m}.$$

De forma equivalente, a hipótese nula é rejeitada, a um nível de significância α , quando o valor de prova obtido é inferior ou igual a α .

Tendo sempre presente o princípio da parcimónia, devem incluir-se no modelo apenas os parâmetros que possam considerar-se significativamente diferentes de zero (Caiado, 2016).

No que diz respeito à avaliação da qualidade do ajustamento do modelo SARIMA, deve proceder-se à análise do comportamento dos respetivos resíduos (mais detalhes na Secção 3.1.4). Se os resíduos tiverem comportamento semelhante a um ruído branco, pode dizer-se que o modelo estimado descreve bem a série em estudo.

Um modelo que não satisfaça os critérios aplicados em alguma das duas etapas, deve ser rejeitado. Nesses casos, as informações recolhidas durante a avaliação podem sugerir indicações que ajudem na formulação de um novo modelo.

Se o objetivo final da análise é o cálculo de previsões, deve proceder-se ao estudo comparativo, através do cálculo dos respetivos erros de previsão para um período pós-amstral ou não utilizado na estimação.

3.1.4 Análise dos resíduos

Após a modelação de uma série temporal, os resíduos, que correspondem à informação não captada pelo modelo, podem ser calculados através da diferença entre os valores observados e os valores estimados correspondentes, ou seja,

$$e_t = Y_t - \hat{Y}_t.$$

Como referido anteriormente, um bom modelo deve gerar resíduos com o comportamento idêntico ao de um ruído branco, e, portanto, estes devem apresentar média nula e satisfazer o pressuposto da não correlação. Adicionalmente, para a construção de intervalos de previsão, torna-se pertinente verificar se os resíduos têm variância constante e apresentam uma distribuição aproximadamente Normal (Hyndman e Athanasopoulos, 2014).

Distribuição Normal

A condição de normalidade pode ser avaliada quer por análise gráfica, quer por testes estatísticos, ou, idealmente, por ambos.

No caso das representações gráficas, as mais usuais são o histograma e o *QQ-plot*. Para indicar a normalidade, o histograma deve aproximar-se do comportamento da função densidade de uma distribuição Normal. No que respeita ao *QQ-plot*, uma vez que se trata de uma representação gráfica dos quantis reais e dos teóricos, este deve apresentar um conjunto de pontos que se posicione mais ou menos sobre uma reta (aproximem da bissetriz dos quadrantes ímpares ($y = x$)).

Para uma verificação rigorosa, os testes estatísticos mais comuns são o teste de Shapiro-Wilk (para amostras de pequenas dimensões, com menos de 50 observações) e o teste de Kolmogorov-Smirnov. Em ambos é testada a hipótese nula “os erros seguem uma distribuição Normal”.

Não correlação

Relativamente ao pressuposto de não correlação, este deve ser verificado tanto individualmente como de forma conjunta.

Individualmente, as autocorrelações dos resíduos podem ser avaliadas através da observação da FAC que, se o modelo for apropriado, deve apresentar um comportamento semelhante ao da FAC de um ruído branco, ou seja, com autocorrelações não significativamente diferentes de zero.

Para testar várias autocorrelações como um grupo pode recorrer-se a um teste de Portmanteau. Um dos mais utilizados dentro desta classe de testes é o de Ljung-Box, cuja estatística de teste Q é definida por

$$Q = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{n-j}$$

e segue aproximadamente uma distribuição do Qui-Quadrado com $k - m$ graus de liberdade (com k a corresponder ao número de autocorrelações a serem testadas e m ao número de parâmetros estimados). Em caso de rejeição da hipótese nula, $H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$, conclui-se que o modelo escolhido não é apropriado. Quanto ao valor de k , não existe um critério específico para a sua escolha, pelo que se compreende que a melhor abordagem passa por realizar o teste para vários valores diferentes. Hyndman e Athanasopoulos (2014) sugerem que, para séries não sazonais, k seja igual a 10 e, para séries sazonais, o dobro do período sazonal. No entanto, a maioria das referências sugere que devem ser explorados vários valores de k .

Média nula e variância constante

A condição exigida à média dos resíduos pode ser facilmente verificada através de um teste estatístico muito comum: o teste t para o valor médio. Neste teste, rejeita-se a hipótese da média ser nula se

$$\left| \frac{\bar{e}\sqrt{n}}{s_e} \right| \geq t_{1-\alpha/2;n-1},$$

onde \bar{e} corresponde à média dos resíduos, n à dimensão da amostra, s_e ao desvio padrão dos resíduos e $t_{1-\alpha/2;n-1}$ ao valor do quantil $1 - \alpha/2$ de uma distribuição t -Student com $n - 1$ graus de liberdade. De forma equivalente, a hipótese de média nula é rejeitada, a um nível de significância α , quando o valor de prova obtido é inferior ou igual a α . No entanto, este teste só deve ser aplicado quando os pressupostos de normalidade e não correlação (independência, no caso de não se rejeitar a normalidade) se verificam.

A estabilidade da variância ou homocedasticidade pode ser avaliada, visualmente, através da análise do gráfico dos resíduos ao longo do tempo.

3.1.5 Seleção de modelos

Em qualquer análise estatística de dados, podem encontrar-se vários modelos adequados para descrever o fenómeno em estudo. Mais concretamente, na modelação de uma série temporal pode existir mais do que um modelo que verifique os diferentes critérios de avaliação do diagnóstico, o que torna difícil a tarefa de escolher o melhor modelo. Assim sendo, devem procurar-se critérios de seleção de modelos que tomem em consideração as estatísticas baseadas nos resíduos do modelo ajustado.

Um critério plausível para escolher o melhor modelo SARIMA poderia ser escolher o modelo que fornece a menor soma dos quadrados dos erros (ou erro quadrático médio) ou o maior valor para a função de verosimilhança. No entanto, esta abordagem nem sempre funciona porque, muitas vezes, o erro quadrático médio pode ser reduzido e a função de verosimilhança aumentada simplesmente pelo aumento do número de parâmetros no modelo.

Para solucionar esta questão, a função de verosimilhança deve ser penalizada por cada parâmetro adicional no modelo, ou seja, se o parâmetro extra não melhorar o valor da função de verosimilhança mais do que o valor da penalização, esse parâmetro não deve ser acrescentado ao modelo.

São vários os critérios, baseados na função de verosimilhança, existentes na literatura, sendo os mais utilizados o critério de informação de Akaike (AIC) e o critério de informação Bayesiano (*Bayesian information criterion*, BIC).

Alguns dos critérios de seleção mais comuns são o critério de informação de Akaike (AIC), o critério de informação de Akaike corrigido (AICc) e o critério de informação Bayesiano (BIC). Estes critérios integram duas componentes, uma que consiste no logaritmo da função de verosimilhança, que decresce quando o número de parâmetros estimados aumenta, e outra mais “penalizadora”, que aumenta à medida que o número de parâmetros também aumenta. O que estes critérios consideram é, então, uma situação de equilíbrio entre as duas componentes.

Critério de informação de Akaike

Considere-se que um modelo com $m = p + q + P + Q$ parâmetros foi ajustado a uma série com n observações. Para avaliar a qualidade do ajustamento, Akaike (1974) introduziu um critério baseado na quantidade de informação, definido por

$$\text{AIC} = -2\log L + 2m, \quad (3.14)$$

onde L é a função de verosimilhança.

Nem todos os *softwares* estatísticos possuem a capacidade de determinar o AIC ou a função de verosimilhança L e, por isso, nem sempre é possível encontrar o AIC exato para um determinado modelo. No entanto, uma aproximação útil para o AIC é obtida através da aproximação

$$-2 \log L \approx n(1 + \log 2\pi) + n \log \sigma^2,$$

onde σ^2 representa a variância dos resíduos. Esta variância é facilmente estimada por qualquer *software* estatístico, permitindo, assim, que o AIC possa ser encontrado aproximadamente através da fórmula

$$\text{AIC} \approx n(1 + \log 2\pi) + n \log \hat{\sigma}^2 + 2m. \quad (3.15)$$

Às vezes, o primeiro termo em (3.15) é omitido por ser igual para todos os modelos.

Note-se que o AIC não tem muito significado por si só e, por isso, só é útil em comparação com o AIC de outro modelo ajustado ao mesmo conjunto de dados. Assim, deve escolher-se o modelo que tenha o menor AIC, considerando que uma diferença de valores de duas unidades (2) ou menos não é substancial. Nesses casos, deve optar-se pelo modelo mais simples, seja pela parcimónia, ou para obter um melhor ajustamento do modelo.

Critério de informação Bayesiano

O critério de informação Bayesiano, proposto por Schwarz (1978), define-se como

$$\text{BIC} = -2 \log L + m \log(n), \quad (3.16)$$

onde L é a função de verosimilhança, m é o número de parâmetros do modelo e n é a dimensão da amostra.

Ao contrário do AIC, o BIC depende da dimensão da amostra (n) pelo que, para $\log(n) > 2$, isto é, para uma amostra de dimensão superior a 7, a penalização do BIC é superior à penalização do AIC. Consequentemente, a minimização do BIC leva, em geral, à seleção de modelos com um menor número de parâmetros do que os obtidos pela minimização do critério AIC, evitando, de certa forma, a sobrestimação do número de componentes.

3.1.6 Previsão pontual e intervalar

Depois de se escolher o modelo que melhor descreve a série temporal, pode prosseguir-se para o cálculo de previsões, sejam estas pontuais ou intervalares.

As previsões pontuais podem ser facilmente determinadas fazendo uso da própria expressão do modelo escolhido. De facto, para obter previsões a h -passos, isto é, para um instante $t + h$, basta calcular a esperança condicionada aos valores observados, ou seja, $E[Y_{t+h}|Y_1, Y_2, \dots, Y_t]$. Hyndman e Athanasopoulos (2014) descrevem este processo em apenas três passos:

1. Expandir a equação do modelo de forma a obter Y_t separado dos restantes termos (ou seja, Y_t num dos membros da equação e os restantes termos no outro);
2. Reescrever a equação substituindo t por $t + h$ (considera-se que o modelo é estimado usando as t observações);
3. Substituir as observações futuras pelas respetivas previsões, os erros futuros por zero e os erros passados pelos resíduos correspondentes.

Iniciando o processo para uma previsão a 1-passo, isto é, para $h = 1$, e repetindo-o para $h = 2, 3, \dots$, é, então, possível obter todas as previsões pretendidas.

Os intervalos de previsão usuais são construídos com base em estimativas do desvio padrão das próprias previsões. Partindo, assim, do pressuposto que os erros são independentes e seguem uma distribuição Normal, a previsão intervalar para o instante $t + h$ é dada por

$$\left(\hat{y}_{t+h|t} - z_{1-\alpha/2} \hat{\sigma}_h, \hat{y}_{t+h|t} + z_{1-\alpha/2} \hat{\sigma}_h \right),$$

onde z é o quantil da distribuição Normal padrão, $1 - \alpha$ corresponde ao nível de confiança do intervalo e $\hat{\sigma}_h$ é a estimativa do desvio padrão da previsão para o passo h . Os níveis de confiança mais utilizados em intervalos de previsão são 90%, 95% e 99%, aos quais correspondem valores de z iguais a 1,645, 1,960 e 2,576, respetivamente.

Em geral, os intervalos de previsão aumentam conforme o horizonte de previsão, h , aumenta. No entanto, em modelos estacionários (isto é, com $d = 0$) as sucessões dos limites inferiores e superiores são convergentes e, portanto, para horizontes distantes, os intervalos de previsão terão amplitudes idênticas (Hyndman e Athanasopoulos, 2014).

Mais detalhes sobre estes e outros processos para a obtenção de previsões, tanto pontuais como intervalares, podem ser consultados em Box (2016).

3.2 Métodos de alisamento exponencial

Os métodos de alisamento exponencial integram um conjunto de métodos de previsão adaptativos, i.e., capazes de acompanhar mudanças no nível, declive e/ou padrão sazonal e, por isso, são considerados dos mais eficientes quando aplicados a séries cujas componentes, tendência e sazonalidade, variam com o tempo. Nestes métodos, as previsões não são mais do que uma combinação ponderada das observações anteriores, onde às observações mais recentes é atribuído um “peso” maior do que às mais antigas. Por outras palavras, quanto mais recente for a observação, maior é a sua contribuição na previsão e, portanto, pode dizer-se que a ponderação decresce exponencialmente com a antiguidade das observações (daí a designação de *métodos de alisamento exponencial*).

Os primeiros desenvolvimentos destes métodos surgem no final da década de 50 do século XX. A ideia parece ser introduzida por Robert Brown, em 1944, enquanto trabalhava para a Marinha dos Estados Unidos da América como analista de Investição Operacional. Brown é contratado para desenvolver um modelo matemático de controlo de incêndios em submarinos, que se mostra, essencialmente, um alisamento exponencial simples de dados contínuos. Durante essa década, ele estende o método para séries discretas e desenvolve métodos que permitem lidar com tendência e sazonalidade (descritos em Brown (1959, 1963)).

Também durante a mesma década, Charles Holt desenvolve um método de alisamento exponencial, diferente do de Brown na forma de lidar com as componentes de tendência e de sazonalidade. Os métodos sazonais propostos por Holt (1957) são, posteriormente, testados por Winters (1960) para dados empíricos, daí a atual designação de métodos de Holt-Winters.

De acordo com as diferentes formas que a tendência e a sazonalidade podem assumir (nula, aditiva ou multiplicativa), Pegels (1969) constrói, então, uma primeira classificação dos métodos de alisamento exponencial (Tabela 3.2).

Tabela 3.2: Classificação dos métodos de alisamento exponencial proposta por Pegels (1969).

Tendência	Sazonalidade		
	N (Nula)	A (Aditiva)	M (Multiplicativa)
N (Nula)	N,N	N,A	N,M
A (Aditiva)	A,N	A,A	A,M
M (Multiplicativa)	M,N	M,A	M,M

Mais tarde, e como consequência dos trabalhos de Gardner(1985), Hyndman (2002) e Taylor (2003), surgem novos métodos de alisamento exponencial, que levam ao alargamento da classificação inicial proposta por Pegels. O aparecimento de uma componente “amortecida” resulta na introdução de dois novos tipos de tendência, que culminam numa classificação com 15 métodos (ignorando a componente erro, Tabela 3.3).

Tabela 3.3: Classificação atual dos métodos de alisamento exponencial.

Tendência	Sazonalidade		
	N (Nula)	A (Aditiva)	M (Multiplicativa)
N (Nula)	N,N	N,A	N,M
A (Aditiva)	A,N	A,A	A,M
Ad (Aditiva amortecida)	Ad,N	Ad,A	Ad,M
M (Multiplicativa)	M,N	M,A	M,M
Md (Multiplicativa amortecida)	Md,N	Md,A	Md,M

Alguns dos métodos referidos na Tabela 3.3 apresentam recorrentemente outras designações. Por exemplo, o método que corresponde à célula (N,N) é usualmente denominado de método de alisamento exponencial simples, enquanto que o método linear de Holt corresponde à entrada (A,N). Já as células (A,A) e (A,M) representam o método de Holt-Winters nas suas duas versões, aditiva e multiplicativa, respetivamente. As restantes entradas da Tabela 3.3 correspondem a outros métodos menos utilizados em previsão.

Para cada um dos 15 métodos apresentados, existem dois modelos em espaço de estados possíveis, um que considera erros aditivos e outro erros multiplicativos, perfazendo um total de 30 modelos. Não se deve, no entanto, confundir métodos de alisamento exponencial com modelos em espaço de estados. Um método de alisamento exponencial trata-se de um algoritmo que produz apenas previsões pontuais. O respetivo modelo em espaço de estados fornece as mesmas previsões pontuais, mas também disponibiliza uma estrutura para determinar intervalos de confiança, entre outras propriedades. Uma vez que a distinção entre erros aditivos e multiplicativos não altera as estimativas pontuais, antes do aparecimento dos modelos em espaço de estados, a componente erro foi muitas vezes ignorada Hyndman (2008).

Os métodos de alisamento exponencial têm vindo a destacar-se devido à sua versatilidade, rapidez, fiabilidade e facilidade de automatização. Ainda assim, deve-se referir que a propriedade mais importante destes métodos é a robustez.

As vantagens dos métodos de alisamento exponencial em relação a outros métodos de previsão são a simplicidade de utilização, o baixo custo de computação e a facilidade de implementação em situações reais em que é necessário estabelecer previsões para muitas séries simultaneamente. No entanto, apesar de simples e versátil, a implementação destes métodos exige alguns cuidados, nomeadamente com os processos de inicialização das suas componentes e constantes de alisamento. A principal desvantagem dos métodos de alisamento exponencial prende-se com o facto de estes serem métodos não paramétricos e, por isso, não permitirem realizar inferências estatísticas de uma forma direta. Assim, a construção e interpretação de intervalos de previsão deve ser feita com algum cuidado sempre que os erros de previsão não tiverem um comportamento próximo de uma distribuição Normal de média zero. Em particular, os intervalos de previsão deixam de ser válidos sempre que os erros de previsão forem correlacionados entre si Caiado (2016).

Na Tabela 3.4 encontra-se descrita a notação utilizada no alisamento exponencial.

Tabela 3.4: Notação para o alisamento exponencial.

Símb.	Definição	Símb.	Definição
\hat{Y}_t	Previsão a 1-passo para o instante t	h	Horizonte de previsão
α	Constante de alisamento associada ao nível	l_t	Nível no instante t
β	Constante de alisamento associada ao declive	b_t	Declive no instante t
γ	Constante de alisamento associada à sazonalidade	s_t	Fator sazonal no instante t
ϕ	Parâmetro de amortecimento da tendência	s	Período sazonal

Apresentam-se, de seguida, os métodos de alisamento exponencial mais utilizados (método de alisamento exponencial simples, método linear de Holt e métodos de Holt-Winters, aditivo e multiplicativo).

3.2.1 Método de alisamento exponencial simples

O método de alisamento exponencial simples trata-se de um método de previsão a curto prazo, adequado para séries simples, i.e., que não apresentem ciclicidade, tendência e/ou sazonalidade.

O método de alisamento exponencial simples, desenvolvido por Brown (1959), considera que a previsão para o período seguinte, \hat{Y}_{t+1} , não é mais do que a previsão para o período corrente, \hat{Y}_t , ajustada pelo respetivo erro de previsão, $e_t = Y_t - \hat{Y}_t$,

i.e.,

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t), \quad (3.17)$$

onde α é uma constante de alisamento que varia entre 0 e 1. Quando α toma um valor próximo de 1, a previsão para o período seguinte sofre um ajustamento substancial através do erro de previsão, enquanto que, quando α se encontra próximo de 0, esse ajustamento é muito reduzido Hyndman (2008).

De forma equivalente, a equação (3.17) escreve-se como

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t. \quad (3.18)$$

A previsão \hat{Y}_{t+1} pode, então, ser interpretada como sendo uma média pesada da mais recente previsão e da mais recente observação, uma vez que se trata de uma ponderação entre ambas, com pesos $1 - \alpha$ e α , respetivamente. Além disso, desenvolvendo esta última equação, é fácil mostrar que \hat{Y}_{t+1} representa uma média móvel pesada de todas as observações anteriores, com pesos exponencialmente decrescentes Hyndman (2008). Efetivamente, se se substituir na equação (3.18) o valor de \hat{Y}_t por $\alpha Y_{t-1} + (1 - \alpha)\hat{Y}_{t-1}$ obtém-se

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + (1 - \alpha)^2\hat{Y}_{t-1},$$

e, repetindo esse processo sucessivamente, tem-se

$$\hat{Y}_{t+1} = (1 - \alpha)^t \hat{Y}_1 + \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i Y_{t-i}.$$

Assim, pode dizer-se que, para cada observação, o peso decresce exponencialmente dependendo do valor do parâmetro α : quando α tem um valor grande, é atribuído um peso mais elevado às observações mais recentes; e quando α tem um valor pequeno, são os dados históricos que têm mais peso. Portanto, o valor do parâmetro α pode ser utilizado para ajustar a sensibilidade da previsão a mudanças no nível da série: quanto maior o valor de α , maior é o peso atribuído a Y_t e mais sensível é a previsão às mudanças no nível.

Uma das principais limitações de métodos recursivos prende-se com a inicialização do processo. De facto, se uma nova previsão depende sempre de uma previsão anterior, será necessário iniciar o processo com um valor pré-definido para a primeira previsão. Desta forma, Makridakis (1998) recomendam considerar como previsão inicial a primeira observação, ou seja, $\hat{Y}_1 = Y_1$. Alternativamente, propõem usar a

média dos primeiros quatro ou cinco valores da série temporal. No ambiente R, a inicialização do processo é feita usando a primeira opção, ou seja, considerando que o valor da previsão inicial corresponde à primeira observação.

Quanto à previsão para períodos futuros, a mais de um passo, pode dizer-se que, uma vez que este método é adequado apenas para séries sem tendência, sazonalidade ou qualquer outro tipo de padrão subjacente, a função de previsão é “constante”, i.e., $\hat{Y}_{t+h} = \hat{Y}_{t+1}$, com $h = 2, 3, \dots$.

Existe, ainda, outra forma de escrever a equação (3.18) que, embora mais “pesada”, se revela útil para a extensão deste método (Secções 3.2.2 e 3.2.3). Assim, para se obter a equação do nível, basta considerar que o nível da série é igual à previsão, i.e., $\hat{Y}_{t+1} = l_t$ e, portanto, $l_t = \alpha Y_t + (1 - \alpha)l_{t-1}$ e $\hat{Y}_{t+h} = l_t$.

Note-se que, neste caso, o nível da série é igual à previsão, e, por isso, a equação proposta por Brown (1959) também estima o nível da série no instante t . Para tal, basta substituir na expressão (5.2) \hat{y}_{t+1} por l_t , i.e., $l_t = \alpha y_t + (1 - \alpha)l_{t-1}$.

3.2.2 Método linear de Holt

Holt (1957) estende o método de alisamento exponencial simples, desenvolvendo um método que permite lidar com séries temporais que apresentem tendência: o método linear de Holt (também conhecido apenas por método de Holt). Este método funciona de forma idêntica ao anterior, sendo, no entanto, necessária a atualização de duas componentes em cada período t , o nível e o declive, dando, por isso, origem a duas equações de atualização:

$$\textbf{Nível: } l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (3.19)$$

$$\textbf{Declive: } b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}, \quad (3.20)$$

para α e β com valores compreendidos entre 0 e 1.

Analisando a expressão (3.21), pode dizer-se que a existência de tendência impossibilita que se façam previsões só com base na estimativa do nível, sendo necessário um segundo termo, hb_t , para projetar essa tendência para o futuro:

$$\textbf{Previsão: } \hat{Y}_{t+h} = l_t + hb_t, \quad h = 1, 2, \dots \quad (3.21)$$

No que respeita à inicialização do processo, o método de Holt exige duas estimativas, uma para l_1 e outra para b_1 . Assim, Makridakis et al. (1998) recomendam que se defina $\hat{l}_1 = Y_1$ e, para a estimativa de b_1 , sugerem duas alternativas: $\hat{b}_1 = Y_2 - Y_1$

ou $\hat{b}_1 = (Y_4 - Y_1)/3$. Outra opção, quando os dados históricos estão disponíveis, passa por utilizar a regressão e, pelo método dos mínimos quadrados, determinar \hat{l}_1 e \hat{b}_1 a partir de alguns valores iniciais da série temporal. Já no ambiente R, os valores iniciais para l_1 e b_1 são Y_2 e $Y_2 - Y_1$, respetivamente.

Brown (1963) também propõe um método que permite lidar com tendências: o método de alisamento exponencial duplo. Este método consiste na aplicação do método de alisamento exponencial simples duas vezes, utilizando a mesma constante de alisamento. Para proceder ao alisamento duplo de Brown utilizam-se as equações de atualização

$$m_t = \alpha Y_t + (1 - \alpha)m_{t-1}, \quad (3.22)$$

$$d_t = \alpha m_t + (1 - \alpha)d_{t-1}, \quad (3.23)$$

onde m_t é a série de alisamento exponencial simples, d_t a série de alisamento exponencial duplo e α a constante de alisamento.

As previsões a h -passos são obtidas através da expressão (3.21), considerando que as estimativas do nível e do declive são dadas por:

$$\text{Nível: } l_t = 2m_t - d_t \quad (3.24)$$

$$\text{Declive: } b_t = \frac{\alpha}{1 - \alpha}(m_t - d_t). \quad (3.25)$$

Comparando as equações (3.19) e (3.20) com (3.24) e (3.25), pode ver-se que quando $\alpha = \beta$, o método de Holt e o método de alisamento exponencial duplo são equivalentes (Caiado, 2016). Desta forma, pode concluir-se que o método de alisamento exponencial duplo é um caso particular do método de Holt.

3.2.3 Método de Holt-Winters

O método de Holt-Winters trata-se de uma extensão do método de Holt proposta para situações em que, além da tendência, a série também apresenta sazonalidade. De forma a incorporar a componente sazonal, surge a necessidade de se considerar um terceiro parâmetro ($0 < \gamma < 1$) e, conseqüentemente, é introduzida uma terceira equação de atualização.

Na verdade, existem dois métodos de Holt-Winters distintos, dependendo se a sazonalidade se encontra na sua forma aditiva ou multiplicativa. Na sua forma aditiva, o método adapta-se a séries com amplitude sazonal independente ao nível da série, enquanto que, na sua forma multiplicativa, o método é utilizado em séries

cuja amplitude sazonal é proporcional ao nível da série. Os modelos de decomposição para os métodos aditivo e multiplicativo são, neste caso, o modelo de decomposição aditivo ($Y_t = T_t + S_t + \epsilon_t$) e o modelo de decomposição multiplicativo com erros aditivos ($Y_t = T_t \times S_t + \epsilon_t$), respetivamente. Podem ver-se na Tabela 3.5 as equações que constituem cada um dos métodos, com $h_s^+ = [(h - 1) \bmod s] + 1$.

Tabela 3.5: Equações do método de Holt-Winters.

Método aditivo	Método multiplicativo
$l_t = \alpha(Y_t - s_{t-s}) + (1 - \alpha)(l_{t-1} + b_{t-1})$	$l_t = \alpha \frac{Y_t}{s_{t-s}} + (1 - \alpha)(l_{t-1} + b_{t-1})$
$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$	$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$
$s_t = \gamma(Y_t - l_t) + (1 - \gamma)s_{t-s}$	$s_t = \gamma \frac{Y_t}{l_t} + (1 - \gamma)s_{t-s}$
$\hat{Y}_{t+h} = l_t + hb_t + s_{t-s+h_s^+} \quad h = 1, 2, \dots$	$\hat{Y}_{t+h} = (l_t + hb_t)s_{t-s+h_s^+}, \quad h = 1, 2, \dots$

À semelhança do que acontece para os métodos de alisamento exponencial já apresentados, para inicializar o método de Holt-Winters são precisos valores iniciais para o nível, para o declive e para os índices de sazonalidade. Note-se que, segundo Makridakis et al. (1998), existem agora exigências sobre o número mínimo de observações: para determinar as estimativas iniciais para o nível e os índices de sazonalidade são necessárias, no mínimo, s observações (s corresponde ao período sazonal); e para as estimativas iniciais do declive são precisas, pelo menos, $2s$ observações. As expressões para a inicialização dos dois métodos de Holt-Winters (aditivo e multiplicativo) encontram-se na Tabela 3.6.

Tabela 3.6: Inicialização do método de Holt-Winters.

Modelo aditivo	Modelo multiplicativo
$\hat{l}_s = \frac{1}{s} \sum_{i=1}^s Y_i$	$\hat{l}_s = \frac{1}{s} \sum_{i=1}^s Y_i$
$\hat{b}_s = \frac{1}{s^2} (\sum_{i=s+1}^{2s} Y_i - \sum_{i=1}^s Y_i)$	$\hat{b}_s = \frac{1}{s^2} (\sum_{i=s+1}^{2s} Y_i - \sum_{i=1}^s Y_i)$
$\hat{s}_i = Y_i - \hat{l}_s$, para $i = 1, \dots, s$	$\hat{s}_i = \frac{Y_i}{\hat{l}_s}$, para $i = 1, \dots, s$

Para inicializar estes métodos, o ambiente R utiliza a função *decompose* para realizar uma decomposição da série nas suas componentes – tendência e sazonalidade –, através de um processo de médias móveis, e assim determinar os valores iniciais para l_s , b_s e s_i , $i = 1, \dots, s$ (\hat{l}_s e \hat{b}_s são determinadas através de uma regressão linear simples da tendência das primeiras observações).

Os métodos de Holt-Winters, além de permitirem a modelação de séries que apresentem tendência e sazonalidade, generalizam os métodos de alisamento exponencial

já referidos. De facto, tanto o método de alisamento exponencial simples como o método de Holt podem ser derivados através das equações apresentadas, considerando os parâmetros β e γ (e os valores iniciais para o declive e para os fatores sazonais, respetivamente) como sendo nulos, sempre que se revele oportuno, i.e., sempre que a série não apresente tendência ($\beta = 0$) nem sazonalidade ($\gamma = 0$).

3.2.4 Outros métodos de alisamento exponencial

Os métodos de alisamento exponencial já estudados (simples, Holt, duplo e Holt-Winters) podem ser modificados de forma a captarem melhor o comportamento de algumas séries temporais.

Como referido anteriormente, com os trabalhos de Gardner e McKenzie (1985), Hyndman et al. (2002) e Taylor (2003), são introduzidos os métodos de alisamento exponencial com tendência amortecida, particularmente úteis na previsão de séries temporais com taxas de crescimento não constantes ao longo do tempo e de efeitos amortecidos (Caiado, 2016).

Para séries não sazonais, pode usar-se o *método de Holt-Winters com tendência aditiva amortecida* (Ad,N), cujas equações de atualização são:

$$\textbf{Nível: } l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \quad (3.26)$$

$$\textbf{Declive: } b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1}, \quad (3.27)$$

onde α , β e ϕ variam entre 0 e 1. A previsão a h -passos é, neste caso, dada por

$$\textbf{Previsão: } \hat{Y}_{t+h} = l_t + \sum_{i=1}^h \phi^i b_t, \quad h = 1, 2, \dots \quad (3.28)$$

Assim, o crescimento para a previsão a 1-passo é ϕb_t e o crescimento é amortecido por um fator de ϕ para cada período de tempo futuro adicional (Hyndman et al., 2008). Quando $\phi = 1$ este método é equivalente ao método linear de Holt.

Para séries sazonais, podem destacar-se dois métodos de tendência amortecida: o *método de Holt-Winters multiplicativo com tendência aditiva amortecida* (Ad,M) e o *método de Holt-Winters aditivo com tendência aditiva amortecida* (Ad,A). Note-se que, quando $\phi = 1$, estes métodos são equivalentes aos métodos de Holt-Winters multiplicativo e aditivo, respetivamente. Podem ver-se na Tabela 3.7 as equações que constituem cada um dos métodos (aditivo e multiplicativo), com $h_s^+ = [(h - 1) \bmod s] + 1$ e $h = 1, 2, \dots$.

Tabela 3.7: Equações do método de Holt-Winters com tendência aditiva amortecida.

Método aditivo (Ad,A)	Método multiplicativo (Ad,M)
$l_t = \alpha(Y_t - s_{t-s}) + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$	$l_t = \alpha \frac{Y_t}{s_{t-s}} + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$
$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1}$	$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1}$
$s_t = \gamma(Y_t - l_t) + (1 - \gamma)s_{t-s}$	$s_t = \gamma \frac{Y_t}{l_t} + (1 - \gamma)s_{t-s}$
$\hat{Y}_{t+h} = l_t + \sum_{i=1}^h \phi^i b_t + s_{t-s+h_s^+}$	$\hat{Y}_{t+h} = (l_t + \sum_{i=1}^h \phi^i b_t) s_{t-s+h_s^+}$

Um método de tendência amortecida é, então, apropriado quando é detetada uma tendência que aparenta não se prolongar no futuro. As equações deste tipo de método fazem o que o nome indica: atenuam a tendência à medida que o horizonte de previsão aumenta, o que, geralmente, melhora a precisão das previsões, principalmente para horizontes de previsão longínquos (Hyndman et al., 2008).

3.2.5 Intervalos de previsão

Depois de se obter previsões pontuais, pode ser útil construir um intervalo de previsão para cada instante temporal. Efetivamente, como o futuro é desconhecido e as previsões não passam disso mesmo, a indicação de uma estimativa intervalar permite quantificar a incerteza associada às previsões pontuais.

Os intervalos de previsão são, usualmente, baseados na medida do erro quadrático médio (EQM),

$$\text{EQM} = \frac{1}{n} \sum_{t=1}^n e_t^2 = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2,$$

que constitui uma estimativa da variância do erro de previsão a 1-passo (i.e., com horizonte temporal de uma unidade temporal). Se se admitir a hipótese de que os erros de previsão têm distribuição aproximadamente Normal de média zero, pode, então, construir-se um intervalo de previsão aproximado para cada instante do tempo.

Assim, supondo a normalidade dos erros de previsão, o intervalo de previsão a 1-passo para o instante t pode ser obtido através da expressão (Makridakis et al., 1998)

$$\left(\hat{Y}_t - z_{1-\alpha/2} \sqrt{\text{EQM}}, \hat{Y}_t + z_{1-\alpha/2} \sqrt{\text{EQM}} \right),$$

onde z é o quantil da distribuição Normal padrão e $1 - \alpha$ corresponde ao nível de confiança do intervalo.

Se o objetivo da previsão, no que se refere ao horizonte temporal, for de médio ou longo prazo, devem calcular-se previsões a h -passos utilizando exclusivamente a

informação disponível da série até ao momento t . Neste caso, a medida do EQM passa a ser determinada com base nos erros de previsão a h -passos, dando lugar à expressão

$$\text{EQM}_{(h)} = \frac{1}{n-h} \sum_{t=h+1}^n (Y_t - \hat{Y}_{t|t-h})^2, \quad (3.29)$$

onde h é o horizonte temporal. Nos casos em que a série temporal apresenta sazonalidade, a expressão (3.29) deve ser substituída por

$$\text{EQM}_{(h)} = \frac{1}{n-h-s+1} \sum_{t=h+s}^n (Y_t - \hat{Y}_{t|t-h})^2,$$

em que s é o período sazonal. Desta forma, para construir intervalos de previsão empíricos (ou aproximados) multi-passos (h unidades temporais) deve utilizar-se a expressão

$$\left(\hat{Y}_{t+h} - z_{1-\alpha/2} \sqrt{\text{EQM}_{(h)}}, \hat{Y}_{t+h} + z_{1-\alpha/2} \sqrt{\text{EQM}_{(h)}} \right). \quad (3.30)$$

Como referido anteriormente, estes intervalos constituem uma medida da incerteza das previsões obtidas através dos diferentes métodos de previsão. De facto, a previsão intervalar é uma das partes mais importantes da fundamentação estatística, uma vez que as previsões não são exatas e os intervalos destacam isso mesmo.

3.3 Medidas de avaliação

Como referido neste Capítulo, são vários os métodos existentes para se fazer previsão de séries temporais, desde os mais complexos, que envolvem diferentes parâmetros, aos métodos mais simples e de fácil compreensão. No entanto, o facto de se usar procedimentos mais complexos nem sempre significa uma melhoria na qualidade das previsões. Então, como escolher o melhor método de previsão?

Vários fatores devem ser considerados aquando na escolha do método de previsão, incluindo a precisão das previsões, a facilidade de utilização, de interpretação e de implementação do método, a poupança de custos, a flexibilidade, etc.. Chatfield (2000) acrescenta, ainda, a experiência do analista, as características da série em estudo, entre outros.

As medidas de avaliação são, em muitos dos casos de análise de previsões, o principal critério de seleção de um método de previsão. Através destas medidas pretende-se avaliar a eficácia de um modelo ou método de previsão, mostrando o quanto ele é capaz de reproduzir a série temporal em estudo e de obter previsões

que venham a revelar-se o mais “corretas” possível. No caso das previsões, é a precisão da previsão futura que mais importa.

Hyndman e Koehler (2006) classificam as medidas de avaliação em quatro tipos distintos:

Medidas dependentes da escala (*scale-dependent*): tal como o nome indica, estas medidas de precisão possuem uma escala que depende da escala dos dados. São bastante utilizadas e muito úteis para comparar diferentes métodos aplicados ao mesmo conjunto de dados, não devendo, no entanto, ser escolhidas, por exemplo, para comparar dados com diferentes escalas. As medidas dependentes da escala mais usadas baseiam-se em erros absolutos (e.g., erro absoluto médio, EAM) ou erros quadráticos (e.g., erro quadrático médio, EQM), o que evita que erros negativos e positivos se “compensem” uns aos outros;

Medidas de erros percentuais: o erro percentual é dado por $p_t = 100e_t/Y_t$. Os erros percentuais têm a vantagem de ser independentes da escala e, portanto, são usados com frequência para comparar o desempenho de previsão entre diferentes séries. Contudo, as medidas baseadas neste tipo de erros têm a desvantagem de ser infinitas ou indefinidas se $Y_t = 0$, para qualquer t no período de interesse, e de ter uma distribuição extremamente assimétrica quando Y_t está próximo de zero. Outro problema com estas medidas, frequentemente ignorado, é o facto de ser assumido um zero significativo (e.g., não faz sentido aplicar este tipo de erros para medir a precisão de previsões de temperatura nas escalas Fahrenheit ou Celsius);

Medidas de erros relativos: uma alternativa aos erros percentuais no cálculo de medidas independentes da escala envolve a divisão de cada erro pelo erro correspondente obtido através de um método de previsão de referência (usualmente, o método de referência é o método *naïve* em que \hat{Y}_t é igual à última observação). Por não dependerem da escala dos dados, as medidas de erros relativos também são recomendadas para avaliar a precisão das previsões entre várias séries. No entanto, quando os erros são pequenos, o uso do método *naïve* como referência não é adequado, uma vez que envolve divisão por zero;

Medidas de erros escalados (*scale-free* ou *scaled*): estas medidas, propostas por Hyndman e Koehler (2006), são também independentes da escala e, por isso, podem ser utilizadas tanto para comparar diferentes métodos de previsão aplicados a uma única série como para comparar a precisão das previsões entre

diversas séries. As medidas de erros escalados envolvem a divisão de cada erro pelo EAM *in-sample* (EAM calculado para a amostra de treino) do método *naïve* (passeio aleatório). A única circunstância em que estas medidas são infinitas ou indefinidas é quando todas as observações históricas são iguais.

Um dos aspetos fundamentais do exercício de previsão diz respeito à escolha da amostra de treino (*training set*) para ajustar o modelo e da amostra de teste (*test set*) para testar a sua qualidade preditiva. Apesar de um bom modelo de previsão não ser necessariamente o modelo que melhor se ajusta aos dados da amostra de treino, a avaliação da sua qualidade preditiva (isto é, a avaliação da precisão das suas previsões) só pode ser feita com uma amostra de observações que não tenha sido utilizada na sua estimação, ou seja, com uma amostra de teste (Caiado, 2016).

Desta forma, uma série temporal é geralmente dividida em série de treino, usada para estimar o modelo, e série de teste, para avaliar as previsões. Assim, se se considerar uma série temporal de dimensão n , $\{Y_1, Y_2, \dots, Y_n\}$, e um horizonte temporal de previsão, h , tem-se a seguinte partição da série:

$$\underbrace{Y_1, Y_2, \dots, Y_{n-h}}_{\text{série de treino}}, \underbrace{Y_{n-h+1}, \dots, Y_n}_{\text{série de teste}}.$$

Usualmente, a série de teste corresponde a 20% da série original, apesar deste valor depender da dimensão da amostra e do horizonte temporal pretendido para a previsão futura. É aconselhado escolher uma série de teste de dimensão igual ou superior ao horizonte temporal pretendido.

Em termos práticos, consideram-se duas formas de avaliar a qualidade preditiva: previsão a 1-passo (ou previsão estática) e previsão multi-passos à frente. A previsão a 1-passo, tal como o nome indica, prevê uma unidade temporal à frente ($h = 1$) da última observação. Isto é, se Y_t é a observação no instante t e \hat{Y}_t a sua estimativa obtida usando as observações Y_1, Y_2, \dots, Y_{t-1} , então \hat{Y}_t é a previsão a 1-passo de Y_t . Desta forma, o erro de previsão a um-passo, $e_t = y_t - \hat{y}_t$, não é mais do que a diferença entre a observação y_t e a previsão feita usando todas as observações até y_{t-1} . Da mesma forma, utilizando todas as observações até ao tempo t (inclusive), é possível obter a previsão h -passos à frente, $\hat{Y}_{t+h|t}$. Generalizando, a previsão a h passos à frente (previsão multi-passos), \hat{y}_{t+h} , corresponde à previsão para y_{t+h} obtida usando as mesmas observações, y_1, y_2, \dots, y_{t-1} .

Os erros de previsão indicam se a metodologia de previsão é apropriada, sendo, por isso, importante medir a sua magnitude. Um erro de previsão é a diferença entre o valor observado e a sua previsão. De forma geral, o erro de previsão (a h -passos)

pode, então, ser escrito como $e_{t+h} = Y_{t+h} - \hat{Y}_{t+h|t}$. Se se considerar $h = 1$, ou seja, a previsão a 1-passo, o erro de previsão a 1-passo é dado simplesmente por $e_t = Y_t - \hat{Y}_t$. No entanto, os erros de previsão não devem ser confundidos com os resíduos do modelo. Isto porque os resíduos são calculados na série de treino e baseiam-se apenas em previsões a 1-passo, enquanto que os erros de previsão utilizam a série de teste e podem envolver previsões multi-passos (Hyndman e Athanasopoulos, 2014).

3.3.1 Medidas de avaliação utilizadas

Não existe nenhuma medida que seja consensual para ser aplicada a todas as situações de previsão e, por isso, é aconselhada a utilização de vários critérios para avaliar a qualidade preditiva de um modelo ou método de previsão. Desta forma, foram escolhidas cinco medidas para proceder à avaliação da qualidade preditiva dos métodos adotados: uma medida dependente da escala, EQM, e a sua correspondente na mesma escala dos dados, REQM; uma medida percentual, EPAM; uma medida escalada, EEAM; e uma medida alternativa, estatística U de Theil.

Erro quadrático médio

O erro quadrático médio (EQM), já definido na Secção 3.2.5, traduz o valor médio dos desvios ao quadrado entre os valores observados e as previsões para os instantes $1, 2, \dots, n$, ou seja,

$$\text{EQM} = \frac{1}{n} \sum_{t=1}^n e_t^2 = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2.$$

Trata-se de uma medida dependente da escala dos dados, uma vez que é calculada através do erro de previsão a 1-passo, $e_t = Y_t - \hat{Y}_t$. Muitas vezes, a raiz do erro quadrático médio, $\text{REQM} = \sqrt{\text{EQM}}$, é preferida ao EQM porque permite reduzir a grandeza dos valores para a mesma escala dos dados.

Na comparação entre métodos de previsão, e uma vez que se $Y_t = \hat{Y}_t$ se obtém $\text{EQM} = 0$, considera-se que o método mais preciso é o que apresenta menor EQM (e, conseqüentemente, menor REQM).

Historicamente, a REQM e o EQM são medidas bastante utilizadas. No entanto, estas medidas são mais sensíveis a *outliers* do que outras do mesmo tipo, como, por exemplo, o erro absoluto médio

$$\text{EAM} = \frac{1}{n} \sum_{t=1}^n |e_t| = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|.$$

Erro percentual absoluto médio

O erro percentual absoluto médio, EPAM, representa a percentagem média do erro de previsão em relação à grandeza das observações, sendo, portanto, uma medida de erros percentuais, que se define como

$$\text{EPAM} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \quad (\%).$$

Tal como acontece no EQM, se $Y_t = \hat{Y}_t$ tem-se $\text{EPAM} = 0\%$ e, portanto, quanto menor for o EPAM mais preciso é o método de previsão.

Apesar de possibilitar a comparação da precisão entre diversas séries, esta medida não pode ser calculada quando existem zeros na série. Além disso, quando as observações se aproximam de zero, o EPAM apresenta valores extremos. De forma a contornar estes problemas foi introduzida uma outra medida, o erro percentual absoluto médio simétrico (EPAMS), definido por Caiado (2016)

$$\text{EPAMS} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t + \hat{Y}_t} \right| \times 200 \quad (\%).$$

No entanto, mesmo solucionando os inconvenientes da utilização do EPAM, o EPAMS tem sido criticado por alguns autores, nomeadamente Hyndman e Koehler (2006), por produzir erros extremos na presença de zeros na amostra.

Erro escalado absoluto médio

O erro escalado absoluto médio, EEAM, proposto por Hyndman e Koehler (2006), permite a comparação relativa de qualquer método de previsão com o método *naïve* (referência), seguindo, de certa forma, a ideia por detrás das medidas de erros relativos. O EEAM define-se como

$$\text{EEAM} = \frac{1}{n} \sum_{t=1}^n |q_t| = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{\frac{1}{n-1} \sum_{t=2}^n |Y_t - Y_{t-1}|} \right|.$$

Nesta medida, o erro escalado, q_t , é o quociente entre o erro de previsão, $e_t = Y_t - \hat{Y}_t$, e o erro absoluto médio da previsão *naïve*, $\text{EAMN} = \sum_{t=2}^n |Y_t - Y_{t-1}| / (n-1)$. Na presença de séries de dados com sazonalidade, o denominador do erro escalado deverá ser substituído pelo erro absoluto médio da previsão *naïve* sazonal, $\text{EAMNs} = \sum_{t=s+1}^n |Y_t - Y_{t-s}| / (n-s)$, onde $s = 12$ para dados mensais (Caiado, 2016).

O EEAM pode ser utilizado para comparar a qualidade preditiva de séries com

diferentes escalas, uma vez que a grandeza q_t é independente da escala dos dados. Além disso, esta medida é sempre definida e finita, exceto no caso em que todas as observações são iguais.

Quanto à interpretação do EEAM, valores superiores a 1 indicam que as previsões para o método adotado são menos precisas, em média, do que as *naïve* e, portanto, quanto mais próximo de zero maior é a precisão do método. Desta forma, na comparação entre diferentes métodos de previsão, considera-se que o mais preciso é o que apresenta o menor EEAM.

Estatística U de Theil

A estatística U de Theil combina duas características presentes em algumas medidas: por um lado, possibilita a comparação relativa de qualquer método de previsão com o método *naïve* (referência), o que remete para o que é feito nas medidas de erros relativos; e por outro, à semelhança do que acontece no EQM, os erros são elevados ao quadrado e, portanto, é dado maior peso aos grandes erros em detrimento dos mais pequenos. Matematicamente, esta estatística é definida como (Makridakis et al., 1998)

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} (f_{t+1} - a_{t+1})^2}{\sum_{t=1}^{n-1} (a_{t+1})^2}}, \quad (3.31)$$

onde $f_{t+1} = (\hat{Y}_{t+1} - Y_t)/Y_t$ e $a_{t+1} = (Y_{t+1} - Y_t)/Y_t$, que correspondem às mudanças relativas prevista e real, respetivamente. Substituindo na expressão (3.31) os valores de f_{t+1} e a_{t+1} obtém-se, numa versão simplificada,

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}.$$

Analisando a expressão da estatística U, podem retirar-se as seguintes ilações:

- $U = 0$ se $f_{t+1} = a_{t+1}$, para $t = 1, 2, \dots, n - 1$, i.e., quando as previsões são exatas;
- $U = 1$ se $f_{t+1} = 0$, i.e., quando os erros de previsão do método adotado são iguais aos erros obtidos pelo método *naïve*. Neste caso, pode dizer-se que o método adotado e o método *naïve* são igualmente precisos;
- $U > 1$ se f_{t+1} se encontra na direção oposta (sinais diferentes) de a_{t+1} , uma vez que o numerador será maior do que o denominador. Neste caso, o método

de previsão adotado é menos preciso do que o método *naïve* e, portanto, não deve ser o escolhido;

- $U < 1$ se f_{t+1} se encontra na mesma direção (sinais iguais) de a_{t+1} . Assim, o método de previsão adotado é mais preciso do que o método *naïve* e, portanto, deve ser o escolhido. Quanto menor o valor de U , melhor é o método de previsão relativamente ao método *naïve*.

Tal como acontece no EEAM, também no caso da estatística U se considera que, na comparação entre diferentes métodos de previsão, o mais preciso é o que apresenta o valor de U mais baixo, desde que o método de referência utilizado seja sempre o mesmo (neste caso, o método *naïve*, cuja previsão corresponde à última observação disponível).

Lista das principais funções no ambiente R

Tabela 8: Principais funções no ambiente R.

Função	Package	Descrição
acf	stats	Calcula a função de autocorrelação (FAC)
arima	stats	Ajusta um modelo ARIMA a uma série temporal univariada
arima.sim	stats	Simula processos ARIMA
Box.test	stats	Calcula a estatística de teste Box-Pierce ou Ljung-Box para examinar a hipótese nula de independência para uma determinada série temporal
decompose	stats	Decompõe uma série temporal nas componentes tendência, sazonalidade e aleatoriedade, através de um processo de médias móveis
eqm_h	(nenhum)	Devolve o erro quadrático médio a h passos para um dado modelo de Holt-Winters (função implementada)
fitted	stats	Extraí os valores estimados por um dado modelo
forecast	forecast	Calcula previsões e intervalos de previsão para modelos de séries temporais
HoltWinters	stats	Aplica o método de Holt-Winters a uma determinada série temporal
hw_est_h	(nenhum)	Devolve as estimativas a h passos para um dado modelo de Holt-Winters (função implementada)
int_prev_h	(nenhum)	Devolve as previsões pontuais e intervalares para cada horizonte temporal h para um dado modelo de Holt-Winters (função implementada)
ks.test	stats	Aplica o teste de Kolmogorov-Smirnov a uma ou duas amostras
lm	stats	Ajusta um modelo de regressão linear aos dados
mape	Metrics	Calcula o erro percentual absoluto médio (EPAM)
mase	Metrics	Calcula o erro escalado absoluto médio (EEAM)
mse	Metrics	Calcula o erro quadrático médio (EQM)
ndiffs	forecast	Estima o número de diferenças regulares necessárias para tornar uma determinada série estacionária
nsdiffs	forecast	Estima o número de diferenças sazonais necessárias para tornar uma determinada série estacionária
pacf	stats	Calcula a função de autocorrelação parcial (FACP)
polyroot	base	Encontra zeros de um polinómio real ou complexo
predict	stats	Calcula previsões e intervalos de previsão para um dado modelo ajustado
rmse	Metrics	Calcula a raiz do erro quadrático médio (REQM)
sarima	astsa	Realiza uma avaliação de diagnóstico a modelos ARIMA
shapiro.test	stats	Aplica o teste de normalidade de Shapiro-Wilk
sim_sarima	sarima	Simula processos SARIMA
t.test	stats	Aplica o teste t-Student
ts	stats	Cria objetos em formato time-series
ur.df	urca	Aplica o teste de Dickey-Fuller Aumentado (ADF)
ur.kpss	urca	Aplica o teste KPSS
Ustat	(nenhum)	Calcula o valor da estatística U de Theil (função implementada)