



**ASSOCIAÇÃO DE POLITÉCNICOS DO NORTE (APNOR)
INSTITUTO POLITÉCNICO DO PORTO**

**Gestão da Cadeia de Abastecimento de Bens de Consumo
Baseada em Modelos de Previsão Lineares**

Carlos Manuel Sousa Barbosa

*Dissertação apresentada ao Instituto Politécnico do Porto para obtenção do Grau de Mestre em
Logística*

Orientada por: Professora Doutora Patrícia Alexandra Gregório Ramos

Porto, Novembro de 2014



Gestão da Cadeia de Abastecimento de Bens de Consumo Baseada em Modelos de Previsão Lineares

Carlos Manuel Sousa Barbosa

Orientada por: Professora Doutora Patrícia Alexandra Gregório Ramos

Porto, Novembro de 2014

RESUMO

A previsão de vendas é fundamental para o sucesso das operações da cadeia de abastecimento de qualquer distribuidor do comércio a retalho. Previsões erradas poderão conduzir a aprovisionamentos escassos ou excessivos afetando diretamente o lucro da empresa e a sua posição competitiva no mercado.

O objetivo deste trabalho foi comparar o desempenho de previsão dos modelos de espaço de estados e dos modelos ARIMA quando aplicados a um vasto conjunto de séries de vendas de bens de consumo do comércio a retalho.

Para este trabalho a empresa Jerónimo Martins disponibilizou as vendas diárias, compreendidas entre 2 de janeiro de 2007 e 31 de julho de 2012, de todos os produtos de quatro categorias distintas de uma loja Pingo Doce com uma dimensão de 1500 m².

Em ambas as metodologias de previsão foram utilizados procedimentos automáticos de seleção do melhor modelo, baseados no critério de informação de Akaike, essenciais quando está em análise um vasto conjunto de séries.

Os resultados mostram que o desempenho de previsão dos modelos ARIMA é indiscutivelmente superior ao dos modelos de espaço de estados quando julgados pelo EPAM e que as previsões multi-passo são de um modo geral mais corretas do que as previsões 1-passo à frente situação que não é surpreendente visto que as previsões multi-passo incorporam dados históricos mais recentes.

Também foi avaliado o desempenho de ambas as metodologias na produção de intervalos de previsão. Os resultados mostram que ambas as metodologias ETS e ARIMA produzem probabilidades de cobertura que estão muito próximas das taxas nominais. A metodologia ETS produz melhores probabilidades de cobertura em ambos os intervalos de previsão de 80% e 95% para ambas as previsões 1-passo à frente e multi-passo.

Palavras - chave: Previsão, Modelos ARIMA, Modelos de espaço de estados, Avaliação do desempenho de previsão, Gestão das operações da cadeia de abastecimento

ABSTRACT

Sales forecasting is crucial to the success of the supply chain operations of any retail distributor. Wrong predictions can lead to excessive or scarce supplies directly affecting the profit of the company and its competitive position in the market.

The purpose of this study was to compare the forecasting performance of state space models and ARIMA models when applied to a wide range of sales series of consumer goods of retail trade.

For this work the company Jerónimo Martins provided daily sales, between January 2, 2007 and July 31, 2012, of all products from four distinct categories of a Pingo Doce store with size of 1500 m².

In both forecasting methodologies automatic procedures for selecting the best model (based on the Akaike information criterion) were used which are essential when analyzing a wide range of series.

The results show that the forecasting performance of ARIMA models is arguably superior to the state space models when judged by EPAM and that the multi-step forecasts are generally more accurate than the 1-step forecasts which is not surprising since the multi-step forecasts incorporate more recent historical data.

We also assessed the performance of both approaches in the production of prediction intervals. The results show that both ETS and ARIMA methods produce coverage probabilities that are very close to the nominal rates of 80% and 95%. The ETS methodology produces better coverage probabilities than ARIMA on both 1-step and multi-step forecasts.

Keywords: Forecasting, ARIMA Models, State space models, Forecast accuracy evaluation, Supply chain operations management

ABREVIATURAS E SÍMBOLOS

AE	Alisamento Exponencial
AES	Alisamento Exponencial Simples
AIC	Critério de Informação de Akaike (em inglês)
AIC _c	Critério de Informação de Akaike corrigido (em inglês)
AR	Autoregressivo
ARMA	Autoregressivo e de Médias Móveis
ARIMA	Autoregressivo e de Médias Móveis Integrado
BIC	Critério de Informação Bayesiano (em inglês)
EAM	Erro Absoluto Médio
EEAM	Erro Escalado Absoluto Médio
EM	Erro Médio
EPAM	Erro Percentual Absoluto Médio
EPM	Erro Percentual Médio
EQM	Erro Quadrático Médio
ETS	“Exponential Smoothing”
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
IP	Intervalos de Previsão
MA	Médias Móveis
NID	Normal e Independentemente Distribuído
REQM	Raiz Quadrada do Erro Quadrático Médio
SARIMA	Autoregressivo e de Médias Móveis Integrado Estritamente Sazonal
SQE	Soma de Quadrados dos Erros

ÍNDICE

RESUMO	I
ABSTRACT	II
ABREVIATURAS E SÍMBOLOS.....	III
ÍNDICE DE FIGURAS.....	VI
ÍNDICE DE TABELAS	VII
1. INTRODUÇÃO	1
1.1 Enquadramento e motivação.....	1
1.2 Objetivos do trabalho	2
1.3 Estrutura do documento	3
2- MÉTODOS DE PREVISÃO.....	4
2.1. Introdução	4
2.2. Conceitos básicos de previsão	4
2.3 Alisamento exponencial.....	12
2.3.1 Alisamento exponencial simples	12
2.3.2 Método de tendência linear de Holt.....	14
2.3.3 Métodos de tendência amortecida	16
2.3.4 Método sazonal de Holt-Winters.....	17
2.3.5 Taxonomia dos métodos de alisamento exponencial	19
2.3.6 Modelos de espaço de estados	22
2.3.7 Seleção de modelos ETS.....	27
2.3.8 Previsão com modelos ETS	28
2.4 Modelos ARIMA.....	29
2.4.1 Estacionaridade e diferenciação	29
2.4.2 Modelos autoregressivos e de médias móveis.....	34
2.4.3 Modelos ARIMA não sazonais	36
2.4.4 Modelos ARIMA estritamente sazonais	38
2.4.5 Modelo ARIMA multiplicativo sazonal.....	39
2.4.6 Seleção de modelos ARIMA	40
2.4.7 Previsão com modelos ARIMA.....	41
3. CASO DE ESTUDO	43
3.1 Empresa Jerónimo Martins.....	43
3.2 Análise exploratória dos dados.....	43

4. MODELAÇÃO E PREVISÃO	48
4.1 O “ <i>software</i> ” estatístico R	48
4.2 Modelação e diagnóstico	49
4.2.1 Modelo ETS.....	49
4.2.2 Modelo ARIMA	50
4.3 Resultados da avaliação do desempenho de previsão	51
5. CONCLUSÕES	56
REFERÊNCIAS BIBLIOGRÁFICAS	58

ÍNDICE DE FIGURAS

Figura 1: Séries temporais simuladas (Caiado, 2011).	30
Figura 2: Cronograma do número de produtos distintos vendidos diariamente entre 2 de janeiro de 2007 e julho de 2012.	44
Figura 3: Gráfico sazonal do número de produtos distintos vendidos mensalmente entre janeiro de 2007 e julho de 2012.	44
Figura 4: Gráfico sazonal do número de produtos distintos vendidos mensalmente entre janeiro de 2007 e julho de 2012 com identificação do valor médio de cada mês.....	45
Figura 5: Cronograma das vendas mensais de bens de consumo de uma loja Pingo Doce entre janeiro de 2007 e julho de 2012.	47
Figura 6: Ambiente de modelação e previsão.....	48

ÍNDICE DE TABELAS

Tabela 1: Taxonomia dos métodos de alisamento exponencial (adaptado de Hyndman (2012)).	19
Tabela 2: Fórmulas recursivas dos métodos de alisamento exponencial (Hyndman, 2012).....	20
Tabela 3: Estratégia para escolha dos valores iniciais de alguns métodos de alisamento exponencial (adaptado de Hyndman, 2012).....	21
Tabela 4: Modelos de espaços de estados com erros aditivos (Hyndman, 2012).....	25
Tabela 5: Modelos de espaço de estados com erros multiplicativos (Hyndman, 2012).	26
Tabela 6: EPAM (%) para as previsões fora da amostra (agosto 2011 a julho 2012)	53
Tabela 7: Cobertura dos intervalos de previsão de 80% (agosto 2011 a julho 2012).	54
Tabela 8: Previsão de cobertura dos intervalos de 95% para previsões fora da amostra (agosto 2011 a julho 2012).....	55

1. INTRODUÇÃO

1.1 Enquadramento e motivação

A previsão de vendas é uma das questões mais importantes que está para além de todas as decisões estratégicas e de planeamento de qualquer organização empresarial do comércio a retalho (Hogarth, 1981). A existência de previsões de vendas corretas é fundamental para o sucesso das operações da cadeia de abastecimento de qualquer distribuidor do comércio a retalho (Johnson, 1994). Previsões erradas poderão conduzir a aprovisionamentos escassos ou excessivos afetando diretamente o lucro da empresa e a sua posição competitiva no mercado (Gugarati, 2003; Hill, 2008).

O responsável pela previsão deve executar previsões de vendas de acordo com o tipo de produtos/serviços, fornecendo-as posteriormente aos gestores que elaboram os planos operacionais das áreas funcionais da organização, marketing e vendas, contabilidade e finanças, produção e compras, e logística e distribuição (Brown, 1956; Brown, 1963).

Assim, e por esta ordem, o gestor de marketing e vendas deve ter conhecimento dos planos de marketing dos atuais e dos novos produtos a comercializar. A previsão deve ter em conta o preço, as alterações no produto, as campanhas de marketing e publicidade, a sazonalidade e as variações conjunturais. As previsões estabelecem objetivos a serem cumpridos pela equipa de vendas. O gestor de marketing e vendas deve também motivar a sua equipa de forma a esta atingir ou mesmo superar os objetivos.

O gestor da área financeira deve ter conhecimento das previsões para proceder à correta elaboração das contas de exploração, orçamentos previsionais de custos e proveitos. Os orçamentos, apesar de terem carácter anual, habitualmente contêm metas temporais até 5 anos.

O gestor de compras e produção tem a responsabilidade de planear e organizar a aquisição de mercadorias, matérias-primas e tudo o que for necessário à produção e comercialização. Este processo vai ser inerente às vendas de cada artigo/serviço. Logo deve ter em consideração as previsões de vendas para evitar gastos desnecessários.

A logística garante o transporte, o armazenamento e o registo de encomendas, de forma a garantir o melhor serviço ao cliente. Assim, o gestor da logística deve ter conhecimento da previsão de vendas para, no seu caso, aprimorar a escolha e quantidades de produtos a dispensar, decidindo quando e quais os destinatários dos produtos (Caiado, 2002).

Na sua maioria, os gestores efetuam previsões demasiado otimistas, subestimando a incerteza do comportamento futuro da organização, originando assim custos que deveriam ser evitados (Caiado, 2008).

A incerteza das previsões é um fator importante a ter em consideração na gestão de *stock* de produtos acabados e de compras (Brown, 1959). Um aspeto que aumenta a incerteza das previsões é um elevado número de consumidores/produtos. Uma fragilidade da previsão é o facto de os consumidores serem facilmente influenciáveis pelo ambiente que o rodeia, sendo isso notório nos dias que correm, com a crise instalada no nosso país.

Os empresários têm receio de lançar novos produtos e chegam mesmo a vender produtos atuais a preços muito baixos, tornando a previsão das vendas incerta, visto que os produtos ficam sujeitos às alterações de fatores, nomeadamente o preço, que não foram tidos em conta aquando da realização das previsões. Contudo, qualquer que seja a circunstância ou o horizonte temporal, a previsão é um importante auxílio ao planeamento eficaz e eficiente (Winklhofer, 1996).

1.2 Objetivos do trabalho

As vendas no comércio de bens e serviços pertencem a um tipo especial de séries temporais que normalmente contêm ambos os padrões de tendência e sazonalidade, para além de outros aspetos, apresentando desafios para o desenvolvimento eficaz de modelos de previsão.

Os modelos de espaço de estados e os modelos ARIMA são as duas metodologias mais utilizadas para previsão de séries temporais, proporcionando abordagens complementares do problema. Enquanto os modelos de espaço de estados são baseados na descrição da tendência e sazonalidade dos dados, os modelos ARIMA são baseados na descrição das autocorrelações dos dados.

O objetivo deste trabalho é comparar o desempenho de previsão dos modelos de espaço de estados e dos modelos ARIMA quando aplicados a um vasto conjunto de séries de vendas de bens de consumo do comércio a retalho. Ao melhor do nosso conhecimento, este trabalho apresenta o primeiro teste aos modelos de espaço de estados na previsão de um vasto conjunto de séries deste tipo.

Para este trabalho a empresa Jerónimo Martins disponibilizou as vendas diárias, compreendidas entre 2 de janeiro de 2007 e 31 de julho de 2012, de todos os produtos de quatro categorias distintas de uma loja Pingo Doce com uma dimensão de 1500 m².

Um dos interesses manifestados pela empresa Jerónimo Martins foi o da determinação do erro de uma previsão mensal das vendas de cada um dos produtos durante o horizonte temporal de um ano, com o objetivo de planear as compras a médio prazo de acordo com as condições dos fornecedores, gerir o espaço disponível em armazém ao longo do tempo, gerir a logística de abastecimento às lojas de acordo com os respetivos espaços, etc, ou seja fazer a gestão da cadeia de abastecimento.

Este trabalho pretendeu dar resposta a essa pretensão da empresa confrontando as duas principais metodologias de previsão – os modelos de espaço de estados e os modelos ARIMA.

1.3 Estrutura do documento

O presente trabalho está estruturado em cinco capítulos.

O primeiro capítulo enquadra a previsão no contexto da gestão da cadeia de abastecimento referindo o seu impacto nas várias áreas funcionais, especifica os objetivos do trabalho e refere a estrutura desta dissertação.

O segundo capítulo introduz de uma forma breve os conceitos básicos de previsão e os dois métodos de previsão utilizados no caso em estudo: os modelos de espaço de estados e os modelos ARIMA.

No terceiro capítulo é apresentado o caso de estudo desenvolvido neste trabalho. É feita uma breve apresentação da empresa Jerónimo Martins e é apresentada uma análise exploratória dos dados em estudo.

O quarto capítulo descreve como foram utilizadas as duas metodologias no caso em estudo, apresentando os resultados de modelação e previsão para o conjunto de séries estudado.

Finalmente o quinto capítulo apresenta as principais conclusões do trabalho e tece alguns desenvolvimentos futuros.

2- MÉTODOS DE PREVISÃO

2.1. Introdução

Neste capítulo apresentam-se os dois principais métodos de previsão que são utilizados no trabalho desenvolvido: Alisamento Exponencial e Modelos ARIMA (Hamilton, 1994; Chatfield, 2001; Caiado, 2011). Antes de se apresentarem os aspectos fundamentais destas metodologias de previsão, serão referidos os conceitos básicos de previsão nomeadamente: as etapas do processo de previsão, as principais estatísticas descritivas utilizadas, as medidas de avaliação dos erros de previsão e os intervalos de previsão.

2.2. Conceitos básicos de previsão

Uma série temporal consiste num conjunto de observações de uma variável feitas em períodos sucessivos de tempo, durante um determinado intervalo. São exemplos a cotação diária de ações, a venda semanal de um produto, o número mensal de dormidas em hotelaria numa região, o lucro anual de uma empresa, a temperatura mínima, máxima e média diária de uma cidade, etc (Enders, 1995, Mentzer, 1997; Hyndman, 2012; Ramos, 2012).

A análise de séries temporais considera os seguintes padrões de comportamento. A tendência consiste no andamento mais notório da série durante um longo período de tempo. Os movimentos oscilatórios ou cíclicos estão associados às fases de expansão e recessão dos sistemas económicos (Pindyck, 1998; Peña, 2001). Em ciclos longos, as componentes de tendência e cíclica são difíceis de separar, pelo que se podem tomar como uma única componente (tendência-cíclica). A sazonalidade consiste nas oscilações periódicas que ocorrem semanalmente, mensalmente, trimestralmente ou anualmente. Podem estar associadas com as estações do ano (temperatura do ar, consumo de água/eletricidade, turismo), medidas administrativas (início e fim do ano escolar), tradições e costumes sociais ou culturais (aumento das vendas no período natalício) ou com as variações do calendário (número de dias úteis do mês/semana, número de sábados no mês).

Etapas da previsão

O processo de previsão de uma série temporal é constituído pelas seguintes etapas (Hyndman, 2012; Ramos, 2012):

- Definição do problema – a primeira etapa no processo de previsão consiste em especificar o problema a ser analisado. Questionando profissionais, quadros técnicos e responsáveis por recolha de dados, desenvolvimento, manutenção e gestão de bases de dados, é fundamental perceber como e por quem é que as previsões serão utilizadas, e como é que estas se encaixam na organização.
- Recolha de informação – o processo de recolha de dados compreende a obtenção dos dados numéricos históricos e de informação adicional acerca do comportamento desses dados. Ocasionalmente, dados muito antigos poderão não ter utilidade devido às modificações sofridas entretanto pela organização.
- Análise exploratória – esta etapa inicia-se com uma representação gráfica dos dados tendo em vista a identificação de padrões consistentes, movimentos de tendência explícita, sazonalidade evidente, movimentos cíclicos, pontos de viragem e eventualmente observações anómalas. A representação gráfica dos dados de uma série ao longo do tempo designa-se por cronograma.
- Modelação – esta etapa consiste na seleção e especificação do modelo que irá traduzir o comportamento dos dados históricos em estudo. O modelo de previsão a utilizar depende dos dados históricos disponíveis, da magnitude das relações entre a variável de previsão e as variáveis explicativas e a forma como irão ser utilizadas as previsões. Habitualmente, neste processo compara-se o desempenho de vários modelos candidatos.
- Previsão – depois de escolhido o modelo e estimados os seus parâmetros, este é usado para obter as previsões. O desempenho do modelo pode ser avaliado logo que os dados do período de previsão fiquem disponíveis.

A utilização de um modelo de previsão nunca é definitiva, sendo necessária uma constante reavaliação da sua qualidade nos últimos instantes observados.

Estatísticas descritivas

A informação contida nos dados históricos de séries temporais é habitualmente caracterizada por um conjunto de indicadores estatísticos.

As medidas estatísticas univariadas e bivariadas mais utilizadas em previsão são as que se apresentam de seguida (Brockwell, 1991; Hamilton, 2006, Makridakis, 1989, Makridakis, 1998; Murteira, 1994; Murteira, 2010).

Sejam x_t , $t = 1, 2, \dots, N$ os valores observados de um conjunto de dados univariados.

- A média ou valor médio do conjunto de dados é dada por (Hyndman, 2012; Ramos, 2012)

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = (x_1 + x_2 + \dots + x_N) / N \quad (1)$$

- Considerando o conjunto de dados ordenado (por ordem crescente ou decrescente), a mediana é o valor que divide o conjunto em duas partes iguais dado por (Hyndman, 2012; Ramos, 2012)

$$\tilde{x} = \begin{cases} \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par} \\ x_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar} \end{cases} \quad (2)$$

- Os percentis são uteis para descrever a distribuição do conjunto de dados.
- A variância amostral é uma medida relativa à variabilidade (dispersão) do conjunto de dados definida por (Hyndman, 2012; Ramos, 2012)

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3)$$

- O desvio padrão amostral é a raiz quadrada positiva da variância amostral, ou seja

$$S = +\sqrt{S^2}$$

O desvio padrão tem a vantagem, em relação à variância, de ser expresso nas unidades das observações.

- A covariância amostral, ou simplesmente covariância, é uma estatística bivariada que mede a intensidade com que covariam pares de valores de duas variáveis.

Sendo $y_t, t=1,2,\dots,N$ os valores observados de um outro conjunto de dados univariados, a covariância entre as variáveis x e y é definida por (Hyndman, 2012; Ramos, 2012)

$$\text{cov}_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (4)$$

As unidades em que a covariância é expressa tornam habitualmente a sua interpretação difícil. A sua magnitude também depende das unidades envolvidas.

- O coeficiente de correlação linear, ou simplesmente correlação, resolve o problema de escala da covariância. A correlação entre duas variáveis x e y é definida por (Hyndman, 2012; Ramos, 2012)

$$r_{xy} = \frac{\text{cov}_{xy}}{s_x s_y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5)$$

A correlação mede o grau de relação ou associação linear entre duas variáveis. Ao contrário da covariância, a correlação é adimensional. O seu valor varia entre -1 e 1. Um valor negativo indica uma relação negativa - à medida que uma variável aumenta a outra diminui; um valor positivo indica uma relação positiva - à medida que uma variável aumenta a outra também aumenta.

- A Autocorrelação é uma estatística univariada que mede a correlação entre pares de valores de uma série temporal desfasados em 1, 2, ou mais períodos. A autocorrelação para o desfasamento (*lag*) $k \in \mathbb{Z}$ é definida por (Hyndman, 2012; Ramos, 2012)

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (6)$$

onde T é o nº de observações da série temporal e $k \leq T-1$.

Ao conjunto das autocorrelações r_1, r_2, \dots, r_p ($p \leq T-1$) de uma série temporal chama-se Função de AutoCorrelação (FAC). Ao gráfico da FAC chama-se correlograma.

Séries temporais que não evidenciam autocorrelação denominam-se “ruído branco”. Espera-se que 95% dos picos da FAC de um ruído branco se encontrem entre $\pm 2/\sqrt{T}$.

- É também conveniente medir a correlação entre y_t e y_{t-k} , para diferentes valores de k , depois de eliminado o efeito que sobre eles exercem os valores intermédios $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$.

Esta correlação condicional designa-se por Função de Autocorrelação Parcial (FACP). A FACP pode ser estimada através do método recursivo seguinte. Inicializa-se com $p_{11} = r_1$ e depois recursivamente calcula-se

$$p_{k+1,k+1} = \frac{r_{k+1} - \sum_{j=1}^k p_{kj} r_{k+1-j}}{1 - \sum_{j=1}^k p_{kj} r_j} \quad (7)$$

onde

$$p_{k+1,j} = p_{kj} - p_{k+1,k+1} p_{k,k+1-j}, \quad j = 1, 2, \dots, k \quad (8)$$

Avaliação dos erros de previsão

É incorreto avaliar o desempenho de previsão de um modelo exclusivamente pelo ajuste do modelo aos dados históricos, ou seja, a avaliação do desempenho deve ser efetuada usando dados históricos que não foram utilizados no ajuste do modelo (Hyndman, 2012).

Para o efeito, tipicamente, divide-se o conjunto de dados em dois conjuntos:

- O Conjunto de treino, que é utilizado para fazer o ajuste do modelo de previsão;
- O Conjunto de teste, que é utilizado para avaliar o desempenho do modelo de previsão ajustado.

Habitualmente, o conjunto de teste consiste nos últimos 20% do conjunto de dados, podendo este valor ser ajustado em função do número de observações disponíveis e do horizonte

temporal das previsões (Dalrymple, 1975; Lawrence, 1992; Mentzer, 1984; Mentzer, 1995; Ramos, 2012).

É importante ter-se em atenção os seguintes aspetos (Hyndman, 2012):

- Um modelo que se ajusta bem aos dados pode ter um mau desempenho na previsão;
- Um ajuste perfeito pode sempre ser obtido desde que o modelo tenha o número suficiente de parâmetros. No entanto, um sobre-ajuste do modelo (utilização de um número excessivo de parâmetros) é tão prejudicial como uma identificação incorreta dos padrões dos dados.

Seja y_i o valor observado no instante i e \hat{y}_i a previsão de y_i .

O erro de previsão é definido por (Hyndman, 2012; Ramos, 2012)

$$e_i = y_i - \hat{y}_i \quad (9)$$

As medidas de avaliação dos erros de previsão baseadas em e_i estão dependentes da escala dos dados, não podendo ser utilizadas para comparação com outras séries expressas em escalas diferentes.

As medidas dependentes da escala mais utilizadas são o erro médio (EM), a raiz quadrada do erro quadrático médio (REQM) e o erro absoluto médio (EAM) definindo-se da forma seguinte (Hyndman, 2012; Ramos, 2012):

$$\text{EM} = \text{média}(e_i) \quad \text{REQM} = \sqrt{\text{média}(e_i^2)} \quad \text{EAM} = \text{média}(|e_i|) \quad (10)$$

Os erros percentuais têm a vantagem de serem adimensionais e portanto são habitualmente utilizados para comparar desempenhos de previsão relativos a diferentes conjuntos de dados.

As medidas mais habituais são o erro percentual médio (EPM) e erro percentual absoluto médio (EPAM) definidas da forma seguinte (Hyndman, 2012; Ramos, 2012):

$$\text{EPM} = \text{média}(p_i) \quad \text{EPAM} = \text{média}(|p_i|) \quad \text{onde } p_i = \frac{e_i}{y_i} \times 100 \quad (11)$$

O erro escalado absoluto médio (EEAM) pode ser usado, em alternativa ao EPAM, para comparar desempenhos de previsão relativos a conjuntos de dados expressos em escalas diferentes. A ideia é escalar os erros com o EAM de um método de previsão básico do conjunto de treino.

Para séries temporais não-sazonais e sazonais o erro escalado q_j é definido respetivamente por (Hyndman, 2012; Ramos, 2012)

$$q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^P |y_t - y_{t-1}|} \text{ e } q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^P |y_t - y_{t-m}|}, \text{ com EEAM} = \text{média}(|q_i|) \quad (12)$$

onde P é o número de observações do conjunto de treino e m é o período de sazonalidade.

O EEAM é inferior/superior a 1 se a previsão é melhor/pior do que a previsão média do método básico de previsão para o conjunto de treino.

Intervalos de previsão

Depois de obtidas as previsões da série em estudo podem-se determinar os intervalos de previsão (IP) para cada instante de tempo estimando o limite inferior e superior entre os quais se espera que o respetivo valor desconhecido se encontre com uma determinada probabilidade, normalmente elevada.

Os IP baseiam-se no erro quadrático médio, EQM = média(e_t^2), fornecendo uma estimativa da variância do erro de previsão.

Assumindo a hipótese de que os erros de previsão têm distribuição aproximadamente Normal de média zero, o IP aproximado para cada instante de tempo é (Hyndman, 2012; Ramos, 2012)

$$\left[\hat{y}_t - z\sqrt{\text{EQM}}, \hat{y}_t + z\sqrt{\text{EQM}} \right] \quad (13)$$

onde z é um multiplicador que limita o IP e que corresponde a uma determinada probabilidade ou nível de confiança.

Os níveis de confiança mais utilizados em intervalos de previsão são 80%, 90%, 95% e 99%, a que correspondem valores de z respetivamente iguais a 1.282, 1.645, 1.960 e 2.576.

Diagnóstico dos resíduos

Em previsão um resíduo é a diferença entre o valor observado e o respetivo valor obtido pelo modelo de previsão ajustado ao conjunto de dados.

Num modelo de previsão é essencial que (Hyndman, 2012; Ramos, 2012):

- Os resíduos sejam não correlacionados.
Se isso não acontecer, os resíduos contêm ainda informação que deve ser utilizada no cálculo das previsões.
- Os resíduos tenham média nula.
Se isso não acontecer, as previsões são enviesadas.

Qualquer método de previsão que não satisfaça estas propriedades pode ser melhorado.

A verificação destas propriedades é importante para analisar se o método de previsão está a utilizar corretamente toda a informação contida nos dados, mas não para selecionar um método de previsão de entre vários candidatos.

Para além daquelas propriedades é vantajoso (mas não necessário) que (Hyndman, 2012; Ramos, 2012):

- Os resíduos tenham variância constante.
- Os resíduos tenham distribuição aproximadamente normal.

O cálculo de intervalos de previsão fica facilitado se estas duas propriedades forem satisfeitas.

Para além da análise visual à FAC dos resíduos para deteção de autocorrelações significativas, podem ser efetuados testes de hipóteses ao conjunto de valores r_k como um grupo, designados por testes de *Portmanteau*.

Estes testes não são individuais aos valores r_k visto que se o fossem, pelo seu elevado número, era provável que pelo menos um deles desse um falso positivo, levando-nos a concluir que os resíduos ainda continham alguma autocorrelação, quando de facto não continham.

Assim, testa-se se as primeiras h autocorrelações são significativamente diferentes das que se esperaria obter de um ruído branco.

O teste de *Portmanteau* mais utilizado é teste de *Ljung-Box* (Ljung, 1978). Este teste é baseado na estatística (Hyndman, 2012; Ramos, 2012)

$$Q = T(T + 2) \sum_{k=1}^h \frac{r_k^2}{T - k} \quad (14)$$

onde h é o *lag* máximo considerado e T é o número de observações.

Sugere-se $h = 10$ para dados não sazonais e $h = 2m$ para dados sazonais (m é o período de sazonalidade). Caso h seja superior a $T/5$ sugere-se $h = T/5$, onde T é o nº de observações da série temporal (Hyndman, 2012).

Se as autocorrelações fossem de um ruído branco então Q teria uma distribuição Qui-quadrado com $h - K$ graus de liberdade, onde K é o número de parâmetros do modelo de previsão. E rejeitaríamos a hipótese nula ($H_0 : r_1 = r_2 = \dots = r_h = 0$) com um nível de significância α se o valor de Q excedesse o quantil de ordem $(1 - \alpha)$ da distribuição Qui-quadrado com $h - K$ graus de liberdade (isto é, se o valor de prova ≤ 0.05 , para $\alpha = 0.05$).

2.3 Alisamento exponencial

Os métodos de alisamento exponencial (AE) surgiram na década de 50 por Brown (1956,1959,1963), Holt (1957) e Winters (1960) e são métodos de previsão que utilizam combinações ponderadas das observações passadas para prever valores futuros.

O termo alisamento exponencial significa que os pesos são exponencialmente decrescentes com a antiguidade das observações, ou seja, quanto mais recente é a observação maior é o respetivo peso (Gardner, 1995; Gardner, 2006; Caiado, 2011; Ramos, 2012; Hyndman, 2012).

Por exemplo:

- As vendas de produtos não sazonais num determinado momento, em geral, estão mais correlacionadas com as vendas observadas nos instantes imediatamente anteriores do que com as observações mais remotas;
- As vendas de produtos sazonais correlacionam-se mais com as vendas nos períodos homólogos mais recentes do que com as observações sazonais mais antigas ou mesmo com as observações não sazonais mais recentes.

2.3.1 Alisamento exponencial simples

O mais simples dos métodos de alisamento exponencial é o método de alisamento exponencial simples (AES). Este método é adequado para previsão de séries temporais que não possuam nem um padrão de tendência nem um padrão de sazonalidade.

Neste método, as previsões são calculadas usando médias ponderadas onde os pesos decrescem exponencialmente com a antiguidade das observações:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \alpha(1-\alpha)^3 y_{T-3} + \dots, \quad (15)$$

onde $0 \leq \alpha \leq 1$ é o parâmetro de alisamento, ou amortecimento.

A previsão 1-passo à frente para o instante $T+1$ é a média ponderada de todas as observações y_1, y_2, \dots, y_T da série. A taxa de crescimento dos pesos é controlada pelo parâmetro α .

Uma representação alternativa na forma recursiva para o AES é a forma de média ponderada, onde a previsão para o instante $t+1$ é igual à média ponderada entre a observação mais recente y_t e a previsão mais recente $\hat{y}_{t|t-1}$,

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1} \quad t=1,2,\dots,T \quad (16)$$

onde $0 \leq \alpha \leq 1$ é o parâmetro de alisamento.

Para se iniciar o processo admite-se que a previsão de y_1 , $\hat{y}_{1|0}$, é l_0 .

Uma outra representação alternativa para o AES é a forma de componente. No método de AES existe apenas uma componente, que é o nível da série l_t . Outros métodos podem incluir também uma componente de tendência b_t e uma componente sazonal s_t .

As representações em forma de componente dos métodos de AE incluem uma equação da previsão e uma equação do alisamento para cada uma das componentes existentes no método.

A representação na forma de componente do método de AES é

$$\begin{aligned} \text{Equação de previsão:} \quad & \hat{y}_{t+1|t} = l_t \\ \text{Equação do nível:} \quad & l_t = \alpha y_t + (1-\alpha)l_{t-1} \end{aligned} \quad (17)$$

para $t=1,2,\dots,T$, onde l_t é o nível (ou valor alisado) da série no instante t .

A equação da previsão mostra que o valor previsto para o instante $t+1$ é o nível estimado no instante t . A equação de alisamento para o nível dá o nível estimado da série em cada instante t . Aplicando a equação da previsão para o instante T vem

$$\hat{y}_{T+1|T} = l_T \quad (18)$$

a previsão para o instante $T+1$ (previsão a 1-passo á frente, fora da amostra), que é a ultima estimativa do nível.

As previsões para um horizonte temporal de h -passos à frente vêm

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = l_T, \quad h = 2, 3, \dots \quad (19)$$

Atente-se que estas previsões só serão adequadas se a série não possuir nem um padrão de tendência nem um padrão de sazonalidade.

A 3ª forma de AES, designada por forma de correção do erro, obtém-se rearranjando a equação do nível

$$\begin{aligned} l_t &= l_{t-1} + \alpha(y_t - l_{t-1}) \\ &= l_{t-1} + \alpha e_t \end{aligned} \quad (20)$$

para $t = 1, 2, \dots, T$, onde $e_t = y_t - l_{t-1} = y_t - \hat{y}_{t|t-1}$.

Repare-se que os erros de previsão dentro da amostra conduzem ao ajustamento/correção do nível estimado. Quanto mais próximo de 1 for o valor de α maiores (mais rápidos) serão os ajustamentos.

A aplicação de qualquer método de AE requer a inicialização do processo de alisamento.

No caso do AES é necessário especificar um valor inicial para l_0 . Para além disso, é necessário também escolher o valor dos parâmetros de alisamento. No caso do AES existe apenas um parâmetro - α . Os valores iniciais e os parâmetros de alisamento de qualquer método de AE podem ser estimados minimizando a Soma dos Quadrados dos Erros (SQE) de previsão a 1-passo à frente dentro da amostra, ou seja,

$$\min \text{SQE} = \sum_{t=1}^T (y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^T e_t^2 \quad (21)$$

o que envolve a resolução de um problema de minimização não linear utilizando uma ferramenta de otimização (optim/R ou Solver/MS-Excel).

Habitualmente, no caso do AES, especifica-se $l_0 = y_1$ como valor inicial do processo de otimização (ver mais à frente a Tabela 3).

2.3.2 Método de tendência linear de Holt

Holt (1947) estendeu o AES para permitir a previsão de dados com tendência. O seu método envolve a equação de previsão e duas equações de alisamento (uma para o nível e outra para a tendência):

$$\begin{aligned}
 \text{Equação da previsão:} & \quad \hat{y}_{t+h|t} = l_t + hb_t \\
 \text{Equação do nível:} & \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
 \text{Equação da tendência:} & \quad b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}
 \end{aligned} \tag{22}$$

para $t=1,2,\dots,T$, onde

- l_t é a estimativa do nível da série no instante t ,
- b_t é a estimativa da tendência (declive) da série no instante t ,
- $0 \leq \alpha \leq 1$ é o parâmetro de alisamento do nível,
- $0 \leq \beta^* \leq 1$ é o parâmetro de alisamento da tendência,
- $h=1$ é o passo da previsão dentro da amostra.

A equação do nível mostra que l_t é a média ponderada entre a observação y_t e a previsão a 1-passo à frente dentro da amostra para o instante. A equação da tendência mostra que b_t é a média ponderada entre a tendência estimada no instante t , baseada em $l_t - l_{t-1}$, e a estimativa anterior da tendência b_{t-1} .

Ao contrário do AES, a equação da previsão já não é plana. A previsão h -passos à frente (fora da amostra) obtém-se adicionando a última estimativa do nível com a última estimativa da tendência multiplicada por h :

$$\hat{y}_{T+h|T} = l_T + hb_T, \quad h=1,2,3,\dots \tag{23}$$

Logo as previsões são função linear de h .

Uma variante do método de tendência linear de Holt é obtida permitindo que o nível e o declive sejam multiplicados (em vez de adicionados):

$$\begin{aligned}
 \text{Equação da previsão:} & \quad \hat{y}_{t+h|t} = l_t b_t^h \\
 \text{Equação do nível:} & \quad l_t = \alpha y_t + (1-\alpha)l_{t-1} b_{t-1} \\
 \text{Equação da tendência:} & \quad b_t = \beta^* \frac{l_t}{l_{t-1}} + (1-\beta^*)b_{t-1}
 \end{aligned} \tag{24}$$

para $t=1,2,\dots,T$, onde l_t é a estimativa do nível da série no instante t .

2.3.3 Métodos de tendência amortecida

As previsões geradas pelo método de tendência linear de Holt apresentam indefinidamente uma tendência constante (crescente ou decrescente). As previsões geradas pelo método de tendência exponencial são ainda mais extremas ao apresentar um crescimento, ou decrescimento, exponencial. A evidência empírica mostra que estes métodos tendem a sobreprever, especialmente para horizontes de previsão longos.

Gardner e McKenzie (1985) introduziram no método de tendência linear de Holt um parâmetro que “amortece” a tendência convergindo-a para uma linha plana, resultando o método de tendência amortecida aditiva. Em conjugação com os parâmetros de alisamento $0 \leq \alpha \leq 1$ e $0 \leq \beta^* \leq 1$, este método de tendência amortecida aditiva inclui o parâmetro de amortecimento $0 \leq \phi \leq 1$:

$$\begin{aligned}
 \text{Equação da previsão:} \quad & \hat{y}_{t+h|t} = l_t(\phi + \phi^2 + \dots + \phi^h)b_t \\
 \text{Equação do nível:} \quad & l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1}) \\
 \text{Equação da tendência:} \quad & b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1}
 \end{aligned} \tag{25}$$

para $t=1,2,\dots,T$.

As previsões convergem para $l_t + \phi b_t / (1-\phi)$ quando $h \rightarrow \infty$, para qualquer $0 \leq \phi \leq 1$. Consequentemente, as previsões de curto prazo são amortecidas e as previsões de longo prazo são constantes.

Motivado pelo sucesso do desempenho das previsões do método de tendência amortecida aditiva, Taylor (2003) introduziu um parâmetro de amortecimento $0 \leq \phi \leq 1$ no método de tendência exponencial, resultando o método de tendência amortecida multiplicativa:

$$\begin{aligned}
 \text{Equação da previsão:} \quad & \hat{y}_{t+h|t} = l_t b_t^{(\phi + \phi^2 + \dots + \phi^h)} \\
 \text{Equação do nível:} \quad & l_t = \alpha y_t + (1-\alpha)l_{t-1} b_{t-1}^\phi \\
 \text{Equação da tendência:} \quad & b_t = \beta^* \frac{l_t}{l_{t-1}} + (1-\beta^*)b_{t-1}^\phi
 \end{aligned} \tag{26}$$

para $t=1,2,\dots,T$.

Este método irá produzir previsões ainda mais conservativas do que o método de tendência amortecida aditiva, quando comparado com o método linear de Holt.

2.3.4 Método sazonal de Holt-Winters

Holt (1957) e Winters (1960) estenderam o método de tendência linear de Holt para captar sazonalidade.

O método sazonal de Holt-Winters compreende a equação de previsão e três equações de alisamento – uma para o nível l_t , uma para a tendência b_t e uma para a sazonalidade s_t , com parâmetros de alisamento α, β^* e γ , respectivamente.

O período de sazonalidade, isto é, o período de tempo regular (número de observações) em que o fenómeno periódico se repete, é denotado por m . Por exemplo, se a sazonalidade é anual, para dados mensais $m = 12$ e para dados semestrais $m = 2$.

Este método é apropriado para séries que apresentam tendência linear e movimentos sazonais.

Existem duas variantes deste método que diferem na natureza da componente sazonal.

- O método aditivo é utilizado quando as variações sazonais são aproximadamente constantes ao longo da série. Neste método, a componente sazonal é expressa em termos absolutos nas unidades da série, e a equação do nível é sazonalmente ajustada subtraindo a componente sazonal. Em cada período de sazonalidade a soma dos componentes sazonais é aproximadamente 0.
- O método multiplicativo é utilizado quando as variações sazonais variam proporcionalmente com o nível da série. Neste método, a componente sazonal é expressa em termos relativos (percentagem), e a equação do nível é sazonalmente ajustada dividindo a componente sazonal. Em cada período de sazonalidade a soma das componentes sazonais é aproximadamente m .

A forma de componente para o método de Holt-Winters aditivo é:

$$\begin{aligned}
 \text{Equação de previsão:} \quad & \hat{y}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+} \\
 \text{Equação do nível:} \quad & l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\
 \text{Equação da tendência:} \quad & b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \\
 \text{Equação da sazonalidade:} \quad & s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \quad (27)
 \end{aligned}$$

para $t=1,2,\dots,T$, onde

- l_t é a estimativa do nível da série no instante t ,
- b_t é a estimativa da tendência (declive) da série no instante t ,
- s_t é a estimativa da sazonalidade da série no instante t ,
- $0 \leq \alpha \leq 1$ é o parâmetro de alisamento do nível,
- $0 \leq \beta^* \leq 1$ é o parâmetro de alisamento da tendência,
- $0 \leq \gamma \leq 1$ é o parâmetro de alisamento da sazonalidade,
- $h=1$ é o passo da previsão dentro da amostra, com $h_m^+ = [(h-1) \bmod m] + 1$.

A equação do alisamento para o nível mostra uma média ponderada entre a observação ajustada de sazonalidade ($y_t - s_{t-m}$) e a previsão não sazonal ($l_{t-1} + b_{t-1}$) para o instante t .

A equação do alisamento para a tendência é idêntica à equação da tendência do método de tendência linear de Holt.

A equação do alisamento para a componente sazonal mostra uma média ponderada entre o índice sazonal corrente ($y_t - l_{t-1} - b_{t-1}$) e o índice sazonal do instante homólogo do período de sazonalidade anterior (isto é, m instantes antes) s_{t-m} .

A forma de componente para o método de Holt-Winters multiplicativo é:

$$\begin{aligned}
 \text{Equação de previsão:} \quad & \hat{y}_{t+h|t} = (l_t + hb_t)s_{t-m+h_m^+} \\
 \text{Equação do nível:} \quad & l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \\
 \text{Equação da tendência:} \quad & b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \\
 \text{Equação da sazonalidade:} \quad & s_t = \gamma \frac{y_t}{l_{t-1} + b_{t-1}} + (1-\gamma)s_{t-m} \tag{28}
 \end{aligned}$$

para $t=1,2,\dots,T$.

Um método que é frequentemente o melhor método de previsão para dados sazonais é o método de Holt-Winters com uma tendência amortecida aditiva e sazonalidade multiplicativa.

A forma de componente deste método é:

$$\begin{aligned}
 \text{Equação de previsão:} \quad & \hat{y}_{t+h|t} = [l_t + (\phi + \phi^2 + \dots + \phi^h)b_t]s_{t-m+h_m^+} \\
 \text{Equação do nível:} \quad & l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + \phi b_{t-1}) \\
 \text{Equação da tendência:} \quad & b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1} \\
 \text{Equação da Sazonalidade:} \quad & s_t = \gamma \frac{y_t}{l_{t-1} + \phi b_{t-1}} + (1-\gamma)s_{t-m}
 \end{aligned} \tag{29}$$

para $t=1,2,\dots,T$.

2.3.5 Taxonomia dos métodos de alisamento exponencial

Considerando todas as combinações possíveis da componente de tendência e da componente sazonal obtêm-se 15 métodos de alisamento exponencial diferentes, como se pode ver na Tabela 1.

Tabela 1: Taxonomia dos métodos de alisamento exponencial (adaptado de Hyndman (2012)).

Componente de tendência		Componente sazonal		
		N (Nenhuma)	A (Aditiva)	M (Multiplicativa)
N	(Nenhuma)	(N,N)	(N,A)	(N,M)
A	(Aditiva)	(A,N)	(A,A)	(A,M)
A _d	(Amortecida aditiva)	(A _d ,N)	(A _d ,A)	(A _d ,M)
M	(Multiplicativa)	(M,N)	(M,A)	(M,M)
M _d	(Amortecida multiplicativa)	(M _d ,N)	(M _d ,A)	(M _d ,M)

Atente-se para a taxonomia dos métodos apresentados:

(N,N) = Alisamento exponencial simples

(A,N) = Método de tendência linear de Holt

(M,N) = Método de tendência exponencial

(A_d,N) = Método de tendência amortecida aditiva

(M_d,N) = Método de tendência amortecida multiplicativa

(A,A) = Método de Holt-Winters aditivo

(A,M) = Método de Holt-Winters multiplicativo

(A_d,M) = Método amortecido de Holt-Winters

Cada método é identificado por um par de letras (T,S) que especifica, respectivamente, o tipo da componente de tendência e o tipo da componente sazonal.

A Tabela 2 mostra as fórmulas recursivas para aplicação dos 15 métodos de alisamento exponencial possíveis.

Cada célula inclui a equação de previsão, para a geração das previsões h -passos à frente, e as equações de alisamento para aplicação do método.

Tabela 2: Fórmulas recursivas dos métodos de alisamento exponencial (Hyndman, 2012).

T	S		
	N	A	M
N	$\hat{y}_{t+h t} = l_t$ $l_t = \alpha y_t + (1-\alpha)l_{t-1}$	$\hat{y}_{t+h t} = l_t + s_{t-m+h_m^*}$ $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)l_{t-1}$ $s_t = \gamma(y_t - l_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h t} = l_t s_{t-m+h_m^*}$ $l_t = \alpha(y_t / s_{t-m}) + (1-\alpha)l_{t-1}$ $s_t = \gamma(y_t / l_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h t} = l_t + hb_t$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h t} = l_t + hb_t + s_{t-m+h_m^*}$ $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h t} = (l_t + hb_t)s_{t-m+h_m^*}$ $l_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (l_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
A _d	$\hat{y}_{t+h t} = l_t + \phi_h b_t$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h t} = l_t + \phi_h b_t + s_{t-m+h_m^*}$ $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - l_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h t} = (l_t + \phi_h b_t)s_{t-m+h_m^*}$ $l_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(l_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t / (l_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h t} = l_t b_t^h$ $l_t = \alpha y_t + (1-\alpha)l_{t-1} b_{t-1}$ $b_t = \beta^*(l_t / l_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h t} = l_t b_t^h + s_{t-m+h_m^*}$ $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)l_{t-1} b_{t-1}$ $b_t = \beta^*(l_t / l_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - l_{t-1} b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h t} = l_t b_t^h s_{t-m+h_m^*}$ $l_t = \alpha(y_t / s_{t-m}) + (1-\alpha)l_{t-1} b_{t-1}$ $b_t = \beta^*(l_t / l_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (l_{t-1} b_{t-1})) + (1-\gamma)s_{t-m}$
M _d	$\hat{y}_{t+h t} = l_t b_t^{\phi_h}$ $l_t = \alpha y_t + (1-\alpha)l_{t-1} b_{t-1}^{\phi}$ $b_t = \beta^*(l_t / l_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h t} = l_t b_t^{\phi_h} + s_{t-m+h_m^*}$ $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}^{\phi})$ $b_t = \beta^*(l_t / l_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - l_{t-1} - b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h t} = l_t b_t^{\phi_h} s_{t-m+h_m^*}$ $l_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}^{\phi})$ $b_t = \beta^*(l_t / l_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t / (l_{t-1} + b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

Onde:

- l_t é a estimativa do nível da série no instante t ,
- b_t é o declive da série no instante t ,
- s_t é a componente sazonal da série no instante t ,
- m é o número de observações do período de sazonalidade,
- α, β^*, γ e ϕ são parâmetros de alisamento com $\phi_h = \phi + \phi^2 + \dots + \phi^h$ e $h_m^+ = [(h+1) \bmod m] + 1$.

A Tabela 3 apresenta estratégias para a escolha dos valores iniciais de alguns métodos de alisamento exponencial.

Não se recomenda que estas estratégias sejam utilizadas diretamente; elas são uteis apenas para a especificação dos valores iniciais do processo de otimização (Hyndman, 2012).

Tabela 3: Estratégia para escolha dos valores iniciais de alguns métodos de alisamento exponencial (adaptado de Hyndman, 2012).

Método	Valores iniciais
(N,N)	$l_0 = y_1$
(A,N) (A _d ,N)	$l_0 = y_1, b_0 = y_2 - y_1$
(M,N) (M _d ,N)	$l_0 = y_1, b_0 = y_2 / y_1$
(A,A) (A _d ,A)	$l_0 = \frac{1}{m}(y_1 + \dots + y_m)$ $b_0 = \frac{1}{m} \left(\frac{y_{m+1} - y_1}{m} + \dots + \frac{y_{m+m} - y_m}{m} \right)$ $s_0 = y_m - l_0, s_{-1} = y_{m-1} - l_0, \dots, s_{-m+1} = y_1 - l_0$
(A,M) (A _d ,M)	$l_0 = \frac{1}{m}(y_1 + \dots + y_m)$ $b_0 = \frac{1}{m} \left(\frac{y_{m+1} - y_1}{m} + \dots + \frac{y_{m+m} - y_m}{m} \right)$ $s_0 = y_m / l_0, s_{-1} = y_{m-1} / l_0, \dots, s_{-m+1} = y_1 / l_0$

2.3.6 Modelos de espaço de estados

Nesta secção iremos definir modelos estatísticos com base nos métodos de AE apresentados que, para além de gerarem as mesmas previsões pontuais,

- geram também intervalos de previsão e
- permitem a utilização de um critério objetivo de seleção de modelos candidatos.

Cada modelo estatístico, designado por modelo de espaço de estados, consiste numa equação da medida (ou observação) que descreve os dados, e uma ou mais equações dos estados (ou transições) que descrevem como as componentes ou estados (nível, tendência, sazonalidade) não observados variam com o tempo (Hyndman, 2012, Ramos, 2012).

Para cada método existem dois modelos, um com erros aditivos e um com erros multiplicativos. As previsões pontuais produzidas pelos dois modelos são iguais, se estes usarem os mesmos valores para as constantes de alisamento. Contudo irão gerar sempre intervalos de previsão diferentes (Hyndman, 2012).

Cada modelo é identificado por um terno de letras (E, T, S) que especifica o tipo de cada componente:

- **Erro** – cujas possibilidades são {A, M},
- **Tendência** – cujas possibilidades são {N, A, A_d , M, M_d },
- **Sazonal** – cujas possibilidades são {N, A, M}.

Existem 30 modelos de espaço de estados:

- 15 com erros aditivos e
- 15 com erros multiplicativos.

Para designar cada modelo utiliza-se a label ETS(, ,), onde ETS significa “*Exponential Smoothing*”.

Vamos deduzir de seguida as equações do modelo ETS(A,N,N) que tem subjacente o método de AES.

Tal com referido na Equação (20), a forma de correção do erro do AES é

$$l_t = l_{t-1} + \alpha e_t \quad (30)$$

onde $e_t = y_t - l_{t-1}$ e $\hat{y}_{t|t-1} = l_{t-1}$.

Então, $e_t = y_t - \hat{y}_{t|t-1}$ representa o erro de previsão a 1-passo e podemos escrever

$$y_t = l_{t-1} + e_t \quad (31)$$

Para tornar isto num modelo de espaço de estados basta especificar a distribuição de probabilidade de e_t .

Para um modelo com erros aditivos, assume-se que os erros de previsão a 1-passo e_t são ruído branco com distribuição normal de média 0 e variância σ^2 (Hyndman, 2012; Ramos, 2012)

$$e_t = \varepsilon_t \sim \text{NID}(0, \sigma^2) \quad (32)$$

onde NID significa “Normal e Independentemente Distribuído”. As equações do modelo podem assim ser escritas da forma (Hyndman, 2012; Ramos, 2012)

$$\begin{aligned} \text{Equação da medida:} \quad & y_t = l_{t-1} + \varepsilon_t \\ \text{Equação do estado:} \quad & l_t = l_{t-1} + \alpha \varepsilon_t \end{aligned} \quad (33)$$

Estas equações, em conjunto com a distribuição estatísticas dos erros, constituem o modelo de espaço de estados inovativos sob AES com erros aditivos – ETS(A,N,N).

O termo “inovativos” vem do facto de todas as equações neste tipo de especificação usarem o mesmo processo de erro aleatório ε_t .

A equação da medida mostra a relação entre a observação e o estado não observado. A equação da transição mostra a evolução ao longo do tempo. Neste caso, a observação y_t é uma função linear do nível l_{t-1} , a parte previsível de y_t , e do erro aleatório ε_t , a parte não previsível de y_t .

Noutros modelos de espaço de estados esta relação pode ser não linear.

A influência do parâmetro de alisamento α é a mesma que para os métodos de alisamento exponencial. Neste caso α controla o grau de variação dos sucessivos níveis. Quanto maior for o valor de α mais rápidas são as mudanças do nível; quanto menor for o valor de α mais lentas são as mudanças do nível. No limite inferior, com $\alpha = 0$, o nível da série não varia ao longo do tempo. No extremo superior, com $\alpha = 1$, o modelo reduz-se a $y_t = y_{t-1} + \varepsilon_t$.

Analogamente, pode-se definir o modelo ETS(M,N,N) escrevendo os erros de previsão a 1-passo aleatórios como erros relativos (Hyndman, 2012; Ramos, 2012)

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \quad (34)$$

onde $\varepsilon_t \sim \text{NID}(0, \sigma^2)$. Substituindo $\hat{y}_{t|t-1} = l_{t-1}$ em (34) obtém-se $y_t = l_{t-1} + l_{t-1}\varepsilon_t$.

Da equação (20) tem-se que $e_t = y_t - \hat{y}_{t|t-1}$, logo usando (34) vem

$$e_t = l_{t-1}\varepsilon_t \quad (35)$$

Substituindo (35) na equação (20) obtém-se $l_t = l_{t-1}(1 + \alpha\varepsilon_t)$.

Então, o modelo ETS(M,N,N) vem (Hyndman, 2012; Ramos, 2012)

$$\text{Equação da medida: } y_t = l_{t-1}(1 + \varepsilon_t)$$

$$\text{Equação do estado: } l_t = l_{t-1}(1 + \alpha\varepsilon_t) \quad (36)$$

De forma análoga podem obter-se as equações dos restantes modelos de espaço de estados subjacentes a cada um dos métodos de AE.

A Tabela 4 apresenta as equações dos modelos de espaço de estados com erros aditivos e a Tabela 5 apresenta as equações dos modelos de espaço de estados com erros multiplicativos.

Tabela 4: Modelos de espaços de estados com erros aditivos (Hyndman, 2012).

T	S		
	N	A	M
N	$y_t = l_{t-1} + \varepsilon_t$ $l_t = l_{t-1} + \alpha\varepsilon_t$	$y_t = l_{t-1} + s_{t-m} + \varepsilon_t$ $l_t = l_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = l_{t-1}s_{t-m} + \varepsilon_t$ $l_t = l_{t-1} + \alpha\varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t / l_{t-1}$
A	$y_t = l_{t-1} + b_{t-1} + \varepsilon_t$ $l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (l_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t / (l_{t-1} + b_{t-1})$
A_d	$y_t = l_{t-1} + \phi b_{t-1} + \varepsilon_t$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = l_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (l_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t / (l_{t-1} + \phi b_{t-1})$
M	$y_t = l_{t-1}b_{t-1} + \varepsilon_t$ $l_t = l_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t / l_{t-1}$	$y_t = l_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $l_t = l_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = l_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $l_t = l_{t-1}b_{t-1} + \alpha\varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t / (l_{t-1}s_{t-m})$ $s_t = s_{t-m} + \gamma\varepsilon_t / (l_{t-1}b_{t-1})$
M_d	$y_t = l_{t-1}b_{t-1}^\phi + \varepsilon_t$ $l_t = l_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t / l_{t-1}$	$y_t = l_{t-1}b_{t-1}^\phi + s_{t-m} + \varepsilon_t$ $l_t = l_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = l_{t-1}b_{t-1}^\phi s_{t-m} + \varepsilon_t$ $l_t = l_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t / s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t / (l_{t-1}s_{t-m})$ $s_t = s_{t-m} + \gamma\varepsilon_t / (l_{t-1}b_{t-1}^\phi)$

Tabela 5: Modelos de espaço de estados com erros multiplicativos (Hyndman, 2012).

T	S		
	N	A	M
N	$y_t = l_{t-1}(1 + \varepsilon_t)$ $l_t = l_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (l_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $l_t = l_{t-1} + \alpha(l_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(l_{t-1} + s_{t-m})\varepsilon_t$	$y_t = l_{t-1}s_{t-m}(1 + \varepsilon_t)$ $l_t = l_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (l_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $l_t = l_{t-1} + b_{t-1} + \alpha(l_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(l_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (l_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A_d	$y_t = (l_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $l_t = (l_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (l_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha(l_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(l_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (l_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $l_t = (l_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = l_{t-1}b_{t-1}(1 + \varepsilon_t)$ $l_t = l_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (l_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $l_t = l_{t-1}b_{t-1} + \alpha(l_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(l_{t-1}b_{t-1} + s_{t-m})\varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma(l_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = l_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $l_t = l_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M_d	$y_t = l_{t-1}b_{t-1}^\phi(1 + \varepsilon_t)$ $l_t = l_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$y_t = (l_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \varepsilon_t)$ $l_t = l_{t-1}b_{t-1}^\phi + \alpha(l_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(l_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t / l_{t-1}$ $s_t = s_{t-m} + \gamma(l_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$y_t = l_{t-1}b_{t-1}^\phi s_{t-m}(1 + \varepsilon_t)$ $l_t = l_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

onde $\beta = \alpha\beta^*$, $0 < \alpha < 1$, $0 < \beta < \alpha$ and $0 < \gamma < 1 - \alpha$.

Na prática o parâmetro de amortecimento ϕ é restrito a $0.8 < \phi < 0.98$ para evitar dificuldades numéricas na estimação do modelo.

Uma alternativa à estimação dos parâmetros minimizando a SQE é maximizar a verossimilhança.

A verossimilhança é a probabilidade de um conjunto de dados ser originário de um determinado modelo (Hyndman, 2012; Ramos, 2012). Logo, um bom modelo está associado a uma verossimilhança elevada.

Para um modelo com erros aditivos, maximizar a verosimilhança conduz aos mesmos resultados que minimizar SQE. No entanto, para modelos com erros multiplicativos obtêm-se resultados diferentes (Hyndman, 2012; Ramos, 2012).

Os parâmetros de alisamento α , β , γ e ϕ , e os estados iniciais $l_0, b_0, s_0, s_{-1}, \dots, s_{-m+1}$, dos modelos ETS são habitualmente estimados através de um “*software*” de previsão maximizando a verosimilhança.

2.3.7 Seleção de modelos ETS

A grande vantagem de usar modelos ETS é a possibilidade de utilizar um critério objetivo de seleção de modelos candidatos.

Os critérios seguintes podem ser utilizados para determinar qual dos 30 modelos ETS é o mais apropriado para uma dada série temporal (Akaike, 1974).

- O Critério de Informação de Akaike (AIC em inglês) é definido por

$$AIC = -2\log(L) + 2k \quad (37)$$

- O Critério de Informação de Akaike corrigido (AIC_c em inglês), adequado para amostras de dados pequenas, é definido por

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{T-k} \quad (38)$$

- O Critério de Informação Bayesiano (BIC em inglês) é definido por

$$BIC = AIC + k[\log(T) - 2] \quad (39)$$

onde L é a verosimilhança do modelo e k é o nº total de parâmetros e estados iniciais estimados.

O melhor modelo é aquele que apresenta menores valores de AIC, AIC_c e BIC.

Algumas combinações (E,T,S) podem conduzir a dificuldades numéricas.

Os modelos que podem causar essas instabilidades são nomeadamente: ETS(M,M,A), ETS(M,M_d,A), ETS(A,N,M), ETS(A,A,M), ETS(A,A_d,M), ETS(A,M,N), ETS(A,M,A), ETS(A,M,M), ETS(A,M_d,N), ETS(A,M_d,A) e ETS(A,M_d,M).

Habitualmente, estas combinações não são consideradas no processo de seleção de um modelo (Hyndman, 2013).

Os modelos com erros multiplicativos são úteis quando os dados são estritamente positivos, mas não são numericamente estáveis quando os dados contêm zeros ou valores negativos.

Consequentemente, habitualmente os modelos com erros multiplicativos não são considerados se os dados da série temporal não são estritamente positivos. Nesse caso, apenas os 5 modelos completamente aditivos são considerados (Hyndman, 2013).

Relembrando, o Teste de Ljung-Box utiliza a FAC dos resíduos para testar a hipótese nula conjunta $H_0 : r_1 = r_2 = \dots = r_h = 0$ com a estatística de teste Q (ver Equação 14). No caso dos modelos ETS, rejeitamos a hipótese nula com um nível de significância α se o valor de Q exceder o quantil de ordem $(1-\alpha)$ da distribuição Qui-quadrado com $h-K$ graus de liberdade, onde K é o número de parâmetros do modelo, sendo igual à soma

- do número de parâmetros de alisamento $(\alpha, \beta, \gamma, \phi)$ estimados, com
- o número de estados iniciais $(l_0, b_0, s_0, s_{-1}, \dots, s_{-m+1})$ estimados.

2.3.8 Previsão com modelos ETS

Previsões pontuais são obtidas através dos modelos iterando as equações para $t = T+1, T+2, \dots, T+h$ e fazendo $\varepsilon_t = 0$ para $t > T$.

Por exemplo, para o modelo ETS(M,A,N)

$$y_{T+1} = (l_T + b_T)(1 + \varepsilon_{T+1}) \quad (40)$$

Logo,

$$\hat{y}_{T+1|T} = l_T + b_T \quad (41)$$

Analogamente,

$$\begin{aligned} y_{T+2} &= (l_{T+1} + b_{T+1})(1 + \varepsilon_{T+2}) \\ &= [(l_T + b_T)(1 + \alpha\varepsilon_{T+1}) + b_T + \beta(l_T + b_T)\varepsilon_{T+1}](1 + \varepsilon_{T+2}) \end{aligned} \quad (42)$$

Logo,

$$\hat{y}_{T+2|T} = l_T + 2b_T \quad (43)$$

E assim sucessivamente.

Atente-se que estas previsões são idênticas às previsões do método de tendência linear de Holt e também às do modelo ETS(A,A,N) (assumindo o mesmo valor das constantes de alisamento).

A outra grande vantagem de usar modelos ETS é a possibilidade de obter intervalos de previsão. Os IP irão diferir entre o modelo com erros aditivos e o modelo com erros multiplicativos sob o mesmo método. Para alguns modelos existem fórmulas algébricas para a determinação dos respectivos IP. Uma abordagem mais genérica, que pode ser aplicada a todos os modelos, é simular caminhos amostrais futuros, condicionados à última estimativa dos dados dos estados, e obter IP a partir dos percentis desses caminhos futuros simulados. Esta opção está habitualmente disponível num “*software*” de previsão (Hyndman, 2012).

2.4 Modelos ARIMA

Os modelos ARIMA (AutoRegressivos e de Médias Móveis Integrados) consistem numa outra abordagem estatística para previsão de séries temporais (Box, 1970; Box 1976).

Aliás, o Alisamento Exponencial e os modelos ARIMA são as duas metodologias mais utilizadas para previsão de séries temporais, proporcionando abordagens complementares do problema.

Enquanto os modelos de alisamento exponencial são baseados na descrição da tendência e sazonalidade dos dados, os modelos ARIMA são baseados na descrição das autocorrelações dos dados.

Antes de introduzirmos os modelos ARIMA é conveniente abordar o conceito de estacionaridade e a técnica de diferenciação de séries temporais.

2.4.1 Estacionaridade e diferenciação

A estacionaridade de uma série temporal implica média e variância constantes, e covariância independente do tempo, dependendo apenas do desfasamento temporal (Wei, 1990).

Na prática, a estacionaridade de uma série temporal observa-se quando os dados:

- não apresentam tendência crescente ou decrescente nem movimentos periódicos, e
- flutuam em torno de uma média constante, independente do tempo, e a variância das flutuações não se altera ao longo do tempo.

Atente-se que uma série temporal apenas com comportamento cíclico (sem qualquer padrão previsível no longo prazo) é estacionária, visto que os ciclos são aleatórios sem comprimento fixo.

Na Figura 1 pode observar-se uma série não estacionária em média e uma série estacionária em média.

Uma série estacionária em média não é necessariamente estacionária em variância e covariância. Contudo, uma série que não é estacionária em média também não é estacionária em variância e covariância (Wei, 2007).

Na Figura 1 encontram-se ainda uma série não estacionária em variância (ainda que estacionária em média) e uma série com tendência linear (não estacionária em média) e variância crescente com a tendência (não estacionária em variância).

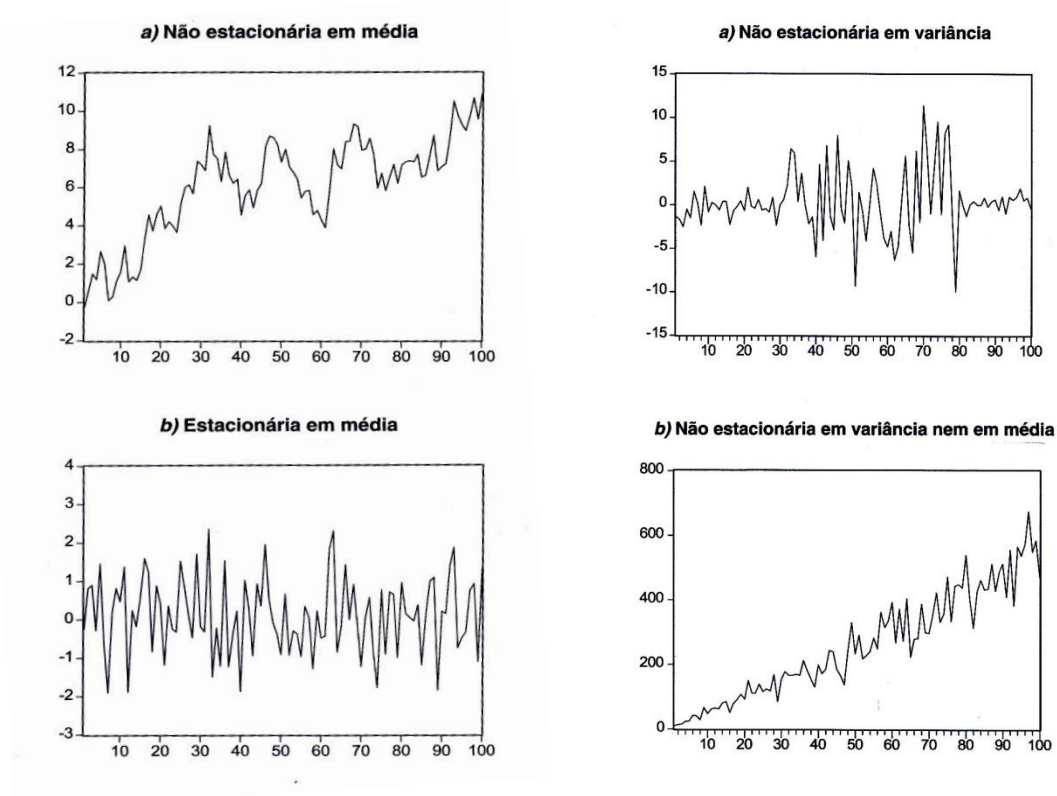


Figura 1: Séries temporais simuladas (Caiado, 2011).

O correlograma também é útil para analisar a estacionaridade de uma série.

O correlograma de uma série estacionária apresenta um decaimento para zero relativamente rápido, enquanto o correlograma de uma série não estacionária apresenta um decaimento

relativamente lento. Habitualmente, r_1 é elevado e positivo para séries não estacionárias (Wei, 2007).

Para estabilizar a variância de uma série não estacionária em variância pode utilizar-se a família de transformações de Box-Cox (Hyndman, 2012)

$$w_t = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y_t), & \lambda = 0 \end{cases} \quad (44)$$

onde λ é um parâmetro real, habitualmente entre -1 e 2.

Com um pequeno ajuste, a transformação logarítmica de base e e as transformações potências estão incluídas na família de Box-Cox.

Para estabilizar a média de uma série não estacionária em média podem utilizar-se transformações de diferenciação (Hyndman, 2012).

A diferenciação simples (ou de 1ª ordem) de uma série temporal consiste em obter a diferença entre observações consecutivas, isto é

$$y'_t = y_t - y_{t-1} \quad (45)$$

A série diferenciada terá somente $T-1$ valores, visto que não é possível calcular a diferença y'_t para a 1ª observação.

Por vezes, a série diferenciada ainda não é estacionária, e pode ser necessário diferenciá-la novamente para obter uma série estacionária.

As diferenças de 2ª ordem de uma série correspondem às diferenças das primeiras diferenças (diferenças de 1ª ordem ou diferenças simples)

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \quad (46)$$

Neste caso y''_t terá $T-2$ valores.

Quando assim acontece, modeliza-se a série das diferenças de 2ª ordem y''_t , que corresponde à “variação das variações” dos dados originais. Na prática, quase nunca é necessário ir para além das diferenças de 2ª ordem (Hyndman, 2012).

Quando uma série apresenta um comportamento periódico repetitivo pode-se aplicar uma transformação de diferenciação sazonal, definida por

$$y'_t = y_t - y_{t-m} \quad (47)$$

onde y_{t-m} é a observação homóloga de y_t relativa ao período sazonal anterior e m é o período de sazonalidade. Neste caso y'_t terá $T - m$ valores.

Por vezes é necessário efetuar uma diferenciação simples e uma diferenciação sazonal para obter dados estacionários.

Uma diferenciação sazonal seguida de uma diferenciação simples é definida por

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-m}) - (y_{t-1} - y_{t-1-m}) = y_t - y_{t-1} - y_{t-m} + y_{t-m-1} \quad (48)$$

Facilmente se mostra que a ordem pela qual se faz a diferenciação sazonal e a diferenciação simples é indiferente (o resultado obtido é o mesmo). No entanto, se os dados apresentam um forte padrão sazonal, recomenda-se que a diferenciação sazonal seja feita em primeiro lugar visto que, por vezes, a série resultante já é estacionária e não é necessário aplicar a diferenciação simples. Se a diferenciação simples é aplicada em primeiro lugar, ainda haverá depois sazonalidade para remover.

Um Teste de raiz unitária é um Teste de Hipóteses estatístico de estacionaridade concebido para determinar se a diferenciação de uma série é necessária.

Os Testes de raízes unitárias mais habituais para determinar o número de diferenciações simples necessárias para tornar uma série estacionária são (Hyndman, 2012):

- Teste de Dickey-Fuller aumentado (habitualmente designado teste “Augmented Dickey-Fuller”)

Hipótese nula: a série não é estacionária.

Valores de prova elevados sugerem não estacionaridade e valores de prova pequenos sugerem estacionaridade.

Para um nível de significância de α a diferenciação é necessária se o valor de prova $< \alpha$.

- Teste de Kwiatkowski-Phillips-Schmidt-Shin (habitualmente designado teste KPSS)

Hipótese nula: a série é estacionária.

Valores de prova elevados sugerem estacionaridade e valores de prova pequenos sugerem não estacionaridade.

Para um nível de significância de α a diferenciação é necessária se o valor de prova $> \alpha$.

O Teste de raiz unitária mais habitual para determinar o número de diferenciações sazonais necessárias para tornar uma série estacionária é (Hyndman, 2012):

- Teste de Osborn-Chui-Smith-Birchenhall (habitualmente designado teste OCSB)

Hipótese nula: a série não é estacionária.

Valores de prova elevados sugerem não estacionaridade e valores de prova pequenos sugerem estacionaridade.

Para um nível de significância de α a diferenciação é necessária se o valor de prova $> \alpha$.

O operador atraso B (*Backshift*) é uma ferramenta útil para a construção de modelos de séries temporais. É definido por

$$By_t = y_{t-1} \quad (49)$$

Ou seja, aplicar B a uma série y_t tem o efeito de “atrasar” os dados 1 instante. Aplicar B duas vezes a y_t tem o efeito de “atrasar” os dados 2 instantes

$$B^2 y_t = B(By_t) = By_{t-1} = y_{t-2} \quad (50)$$

E para dados sazonais, por exemplo de periodicidade mensal, $m=12$, a série referente ao período homólogo do ano anterior pode escrever-se na forma

$$B^{12} y_t = y_{t-12} \quad (51)$$

O operador atraso B é conveniente para descrever o processo de diferenciação. Uma diferença simples pode ser escrita na forma

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t \quad (52)$$

De forma análoga, uma diferença de 2ª ordem pode ser escrita na forma

$$y''_t = y_t - 2y_{t-1} + y_{t-2} = (1-2B+B^2)y_t = (1-B)^2 y_t \quad (53)$$

De um modo geral, uma diferença de ordem d pode ser escrita na forma

$$(1-B)^d y_t \quad (54)$$

Esta notação é muito útil para combinar diferenças, já que o operador pode ser tratado aplicando regras algébricas.

Por exemplo, uma diferenciação sazonal seguida de uma diferenciação simples pode ser escrita na forma

$$(1-B)(1-B^m)y_t = (1-B-B^m+B^{m+1})y_t = y_t - y_{t-1} - y_{t-m} + y_{t-m-1} \quad (55)$$

Atente-se às seguintes considerações relativamente às operações de estabilização:

- A transformação de estabilização da variância, se for necessária, deve ser feita antes da transformação de estabilização da média (i.e. antes da diferenciação);
- A transformação de estabilização da variância só é aplicável a séries positivas (contudo, é sempre possível adicionar uma quantidade positiva suficientemente grande para tornar a série positiva, sem afetar a estrutura das correlações).

2.4.2 Modelos autoregressivos e de médias móveis

Num modelo de autoregressão prevemos a variável em estudo usando uma combinação linear de valores passados dessa variável. O termo autoregressão indica que é uma regressão da variável com ela própria (Schwartz, 1978).

O modelo AutoRegressivo de ordem p , ou modelo $AR(p)$, pode ser escrito numa das formas seguintes (Shumway, 2000)

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

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(y_t - \mu) = \varepsilon_t \quad (56)$$

onde μ é a média de y_t e $c = \mu(1 - \phi_1 - \dots - \phi_p)$, ε_t é um ruído branco, p é um inteiro não negativo e $\phi_1, \phi_2, \dots, \phi_p$ são parâmetros reais. Diferentes parâmetros $\phi_1, \phi_2, \dots, \phi_p$ resultam em diferentes padrões de séries temporais.

Habitualmente restringem-se os modelos autoregressivos a séries estacionárias, e nesse caso são impostas condições aos parâmetros:

- Para um modelo $AR(1)$: $-1 < \phi_1 < 1$.
- Para um modelo $AR(2)$: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.
- Para $p \geq 3$ as condições são bastante mais complexas.

Num modelo de médias móveis prevemos a variável em estudo usando uma combinação linear dos erros de previsão passados.

O modelo de médias móveis de ordem q , ou modelo MA(q), pode ser escrito em uma das formas seguintes

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

$$y_t = \mu + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t \quad (57)$$

onde μ é a média de y_t , ε_t é um ruído branco, q é um inteiro não negativo e $\theta_1, \theta_2, \dots, \theta_q$ são parâmetros reais. Diferentes parâmetros $\theta_1, \theta_2, \dots, \theta_q$ resultam em diferentes padrões de séries temporais.

Pode mostrar-se que um processo AR estacionário y_t de ordem finita é equivalente a um processo MA de ordem infinita (Wei, 2007).

A relação inversa também se verifica se o processo MA for invertível, isto é, se satisfizer as seguintes condições, que são idênticas às condições de estacionaridade:

- Para um modelo MA(1): $-1 < \theta_1 < 1$.
- Para um modelo MA(2): $-1 < \theta_2 < 1$, $\theta_1 + \theta_2 < 1$, $\theta_2 - \theta_1 < 1$.
- Para $q \geq 3$ as condições são bastante mais complexas.

Assim, um processo MA invertível de ordem finita é equivalente a um processo AR de ordem infinita.

Conclui-se então que os processos estacionários e invertíveis podem ser representados quer na forma autoregressiva quer na forma de médias móveis.

Contudo, pode acontecer que um qualquer destes processos tenha uma representação com um nº excessivo de parâmetros, que pode conduzir a uma perda de eficiência da sua estimação. Se assim for, pode construir-se um modelo mais parcimonioso que inclua ambas as partes autoregressiva e de médias móveis.

Esse modelo designa-se por modelo autoregressivo e de médias móveis de ordem p e q , ARMA(p, q), e representa-se em uma das formas seguintes

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

$$(1 - \phi_1 B - \dots - \phi_p B^p)(y_t - \mu) = (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (58)$$

onde μ é a média de y_t e $c = \mu(1 - \phi_1 - \dots - \phi_p)$, ε_t é um ruído branco, p e q são inteiros não negativos, ϕ_1, \dots, ϕ_p são parâmetros autoregressivos e $\theta_1, \dots, \theta_q$ são parâmetros de médias móveis.

O modelo ARMA será

- estacionário se os parâmetros autoregressivos satisfizerem as condições de estacionaridade, e
- invertível se os parâmetros de médias móveis satisfizerem as condições de invertibilidade.

2.4.3 Modelos ARIMA não sazonais

Combinando diferenciação com modelos ARMA obtemos os modelos ARMA integrados ou modelos ARIMA – modelos AutoRegressivos e de Médias Móveis Integrados.

O modelo ARIMA(p, d, q) não sazonal tem representação (Shumway, 2000)

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (59)$$

onde

- p, d e q são inteiros não negativos, sendo d o nº de diferenciações simples necessárias para tornar a série estacionária;
- $c = \mu(1 - \phi_1 - \dots - \phi_p)$ e μ é a média de $(1 - B)^d y_t$;
- ϕ_1, \dots, ϕ_p são os parâmetros autoregressivos;
- $\theta_1, \dots, \theta_q$ são os parâmetros de médias móveis;
- ε_t é um ruído branco.

O modelo será estacionário e/ou invertível se os parâmetros AR e/ou MA satisfizerem as respectivas condições. Habitualmente assume-se $c = 0$ quando $d > 1$.

Habitualmente, não é possível determinar os valores de p e q apropriados exclusivamente através do cronograma. No entanto, por vezes, é possível determinar esses valores através da FAC e da FACP.

Uma série temporal poderá ser descrita por um modelo ARIMA($p, d, 0$) se

- a FAC dos dados diferenciados tem um decaimento exponencial ou sinusoidal amortecido para zero;
- a FACP dos dados diferenciados tem um pico significativamente diferente de zero no *lag* p e nenhum pico significativamente diferente de zero a partir daí.

Uma série temporal poderá ser descrita por um modelo ARIMA($0, d, q$) se

- a FAC dos dados diferenciados tem um pico significativamente diferente de zero no *lag* q e nenhum pico significativamente diferente de zero a partir daí;
- a FACP dos dados diferenciados tem um decaimento exponencial ou sinusoidal amortecido para zero.

Se p e q são ambos positivos então a FAC e a FACP não auxiliam na determinação dos valores de p e q apropriados. Ambas as FAC e FACP têm um decaimento exponencial ou sinusoidal amortecido para zero.

2.4.4 Modelos ARIMA estritamente sazonais

Algumas séries temporais exibem, por vezes, uma significativa correlação entre observações desfasadas em m períodos, $y_t, y_{t-m}, y_{t-2m}, \dots$. Os modelos AutoRegressivos e de Médias Móveis Integrados estritamente sazonais de ordens P e Q – SARIMA $(P, D, Q)_m$, contemplam essa correlação e têm a forma (Shumway, 2000)

$$\Phi_p(B^m)(1-B^m)^D y_t = c + \Theta_q(B^m)\varepsilon_t \quad (60)$$

com

$$\Phi_p(B^m) = 1 - \Phi_1 B^m - \Phi_2 B^{2m} - \dots - \Phi_p B^{pm}$$

$$\Theta_q(B^m) = 1 + \Theta_1 B^m + \Theta_2 B^{2m} + \dots + \Theta_q B^{qm}$$

onde,

- P, D e Q são inteiros não negativos, sendo D o nº de diferenciações sazonais necessárias para tornar a série estacionária;
- $m > 1$ é o período de sazonalidade;

- $c = \mu(1 - \Phi_1 - \dots - \Phi_p)$ e μ é a média de $(1 - B^m)^D y_t$;
- Φ_1, \dots, Φ_p são os parâmetros autoregressivos;
- $\Theta_1, \dots, \Theta_Q$ são os parâmetros de médias móveis;
- ε_t é um ruído branco.

Habitualmente assume-se $c = 0$ quando $D > 1$.

Uma série temporal poderá ser descrita por um modelo AutoRegressivo estritamente sazonal de ordem P , isto é um modelo SARIMA $(P, D, 0)_m$, se:

- a FAC tem um decaimento exponencial ou sinusoidal amortecido sobre os *lags* múltiplos de m , anulando-se nos restantes *lags*;
- a FACP tem uma queda brusca para zero a partir do P -ésimo *lag* múltiplo de m .

Uma série temporal poderá ser descrita por um modelo de Médias Móveis estritamente sazonal de ordem Q , isto é um modelo ARIMA $(0, D, Q)_m$, se:

- a FAC tem uma queda brusca para zero a partir do Q -ésimo *lag* múltiplo de m ;
- a FACP tem um decaimento exponencial ou sinusoidal amortecido sobre os *lags* múltiplos de m , anulando-se nos restantes *lags*.

Se P e Q são ambos positivos então a FAC e a FACP não auxiliam na determinação dos valores de P e Q apropriados. Ambas as FAC e FACP têm um decaimento exponencial ou sinusoidal amortecido para zero sobre os *lags* múltiplos de m , anulando-se nos restantes *lags*.

2.4.5 Modelo ARIMA multiplicativo sazonal

Em muitas séries temporais sazonais será de esperar que a componente sazonal esteja de alguma forma relacionada com as componentes não sazonais. Isto é, se as observações vizinhas da série, $y_t, y_{t-1}, y_{t-2}, \dots$, estão relacionadas, é muito provável que as observações vizinhas espaçadas de m períodos, $y_m, y_{t-m}, y_{t-2m}, \dots$, também estejam relacionadas. Sendo assim, combinando (59) e (60) obtém-se o modelo multiplicativo sazonal integrado ARIMA $(p, d, q) \times (P, D, Q)_m$ (Shumway, 2000)

$$\Phi_p(B^m)\phi_p(B)(1-B)^d(1-B^m)^D y_t = c + \theta_q(B)\Theta_Q(B^m)\varepsilon_t \quad (61)$$

com

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p, \quad \theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q$$

$$\Phi_p(B^m) = 1 - \Phi_1 B^m - \dots - \Phi_p B^{Pm}, \quad \Theta_Q(B^s) = 1 + \Theta_1 B^m + \dots + \Theta_Q B^{Qm}$$

onde, $c = \mu(1 - \phi_1 - \dots - \phi_p)(1 - \Phi_1 - \dots - \Phi_p)$ e μ é a média de $(1 - B)^d (1 - B^m)^D y_t$ e ε_t é um ruído branco de média zero. Habitualmente assume-se $c = 0$ quando $d + D \geq 2$.

Uma vez identificados o(s) modelo(s) candidato(s) a descrever a série em estudo, segue-se a etapa de estimação dos respetivos parâmetros, isto é $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_Q$ e c .

Dado que esta estimação requer a aplicação de um conjunto de métodos numéricos e de cálculos com alguma complexidade, é habitualmente levada a cabo utilizando um “*software*” adequado.

Os dois principais métodos de estimação dos parâmetros de modelos ARIMA são o método da máxima verosimilhança e o método dos mínimos quadrados.

Tal como já foi referido, o método da máxima verosimilhança consiste em determinar os valores dos parâmetros do modelo que tornam verosímil a ocorrência de um conjunto de observações idênticas àquelas de que efetivamente se dispõe.

As estimativas desses parâmetros são obtidas através de um processo iterativo em que se maximiza a função de verosimilhança dos estimadores (ver mais detalhes em Box (1994)).

O método dos mínimos quadrados é talvez o método estatístico mais utilizado na estimação de modelos. Contudo, em modelos ARIMA (p, d, q) este método não permite obter estimadores consistentes com os verdadeiros parâmetros. Exceto no caso dos modelos AR(p) em que os estimadores se podem obter por minimização da soma dos quadrados dos resíduos.

2.4.6 Seleção de modelos ARIMA

Identificados vários modelos candidatos é necessário escolher o melhor modelo.

Os critérios de seleção mais habituais são (Akaike, 1974):

- O Critério de Informação de Akaike (AIC) definido por

$$AIC = -2 \log(L) + 2(p + q + P + Q + k + 1) \quad (62)$$

onde, L é a verossimilhança dos dados e $k=1$ se $c \neq 0$ ou $k=0$ se $c=0$ (entre parêntesis tem-se o n° de parâmetros do modelo, incluindo σ^2 , a variância dos resíduos).

- O Critério de Informação de Akaike corrigido (AIC_c), adequado para amostras de dados pequenas, é definido por

$$AIC_c = AIC + \frac{2(p+q+P+Q+k+1)(p+q+P+Q+k+1+2)}{T-p-q-P-Q-k-4} \quad (63)$$

- O Critério de Informação Bayesiano (BIC em inglês) é definido por

$$BIC = AIC + \log(T)(p+q+P+Q+k-1) \quad (64)$$

O melhor modelo é aquele que apresenta menores valores de AIC, AIC_c e BIC.

Relembrando, o Teste de Ljung-Box utiliza a FAC dos resíduos para testar a hipótese nula conjunta $H_0 : r_1 = r_2 = \dots = r_h = 0$ com a estatística de teste Q (ver Equação 14).

No caso dos modelos ARIMA, rejeitamos a hipótese nula com um nível de significância α se o valor de Q exceder o quantil de ordem $(1-\alpha)$ da distribuição Qui-quadrado com $h-K$ graus de liberdade, onde K é o número de parâmetros do modelo, e se

- o modelo é não-sazonal, $K = p+q+1$ ou $K = p+q$, respectivamente, se o modelo contém, ou não, a constante c .
- o modelo é sazonal, $K = p+q+P+Q+1$ ou $K = p+q+P+Q$, respectivamente, se o modelo contém, ou não, a constante c .

2.4.7 Previsão com modelos ARIMA

As previsões pontuais (1-passo à frente) de modelos ARIMA são calculadas de acordo com as iterações seguintes:

1. Expande-se a equação do modelo de modo a que y_t fique isolado no 1° membro;
2. Reescreve-se a equação substituindo-se t por $T+h$;
3. No 2° membro da equação do modelo, substituem-se as observações futuras pelas suas previsões, os erros futuros por zero e os erros passados pelos resíduos correspondentes.

Começando em $h=1$, estes passos são depois repetidos para $h=2,3,\dots$ até que todas as previsões 1-passo à frente tenham sido calculadas.

Ilustremos o processo de obtenção de previsões 1-passo à frente para o caso do modelo ARIMA(3,1,1) cuja expressão é

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)y_t = (1 + \theta_1 B)\varepsilon_t \quad (65)$$

Começamos por expandir o 1º membro da equação do modelo

$$\left[1 - (1 + \phi_1)B + (\phi_1 - \phi_2)B^2 + (\phi_2 - \phi_3)B^3 + \phi_3 B^4\right]y_t = (1 + \theta_1 B)\varepsilon_t \quad (66)$$

Aplicando o operador atraso vem

$$y_t - (1 + \phi_1)y_{t-1} + (\phi_1 - \phi_2)y_{t-2} + (\phi_2 - \phi_3)y_{t-3} + \phi_3 y_{t-4} = \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (67)$$

e finalmente, isolando y_t no 1º membro obtém-se

$$y_t = (1 + \phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3} - \phi_3 y_{t-4} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (68)$$

completando o 1º passo.

Passando ao 2º passo, substitui-se t por $T+1$ na Equação 68 obtendo-se

$$y_{T+1} = (1 + \phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} - \phi_3 y_{T-3} + \varepsilon_{T+1} + \theta_1 \varepsilon_T \quad (69)$$

Assumindo que temos observações até ao instante T , todos os valores do 2º membro são conhecidos exceto ε_{T+1} , que é substituído por 0, e ε_T que é substituído pelo último resíduo observado $\hat{\varepsilon}_T$

$$\hat{y}_{T+1|T} = (1 + \phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} - \phi_3 y_{T-3} + \theta_1 \hat{\varepsilon}_T \quad (70)$$

A previsão de y_{T+2} é obtida substituindo t por $T+2$ na Equação 68 obtendo-se

$$\hat{y}_{T+2|T} = (1 + \phi_1)\hat{y}_{T+1|T} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi_3 y_{T-2} \quad (71)$$

E o processo continua desta forma para todos os instantes de tempo futuros que se desejarem.

3. CASO DE ESTUDO

3.1 Empresa Jerónimo Martins

A Jerónimo Martins é um grupo internacional com sede em Portugal que opera na distribuição e fabrico de alimentos e serviços.

Envolvendo vendas a grosso e a retalho, o Grupo Jerónimo Martins é o líder na distribuição alimentar em Portugal, com as marcas Pingo Doce (líder em supermercados) e Recheio (líder em cash & carry), tendo também redes de lojas alimentares na Polónia (Biedronka) e na Colômbia (Ara) (Jerónimo Martins, 2013).

3.2 Análise exploratória dos dados

O objetivo deste trabalho é comparar o desempenho de previsão dos modelos ETS e ARIMA quando aplicados a um vasto conjunto de séries de vendas de bens de consumo do retalhista português Jerónimo Martins.

Ao melhor do nosso conhecimento, este trabalho apresenta o primeiro teste aos modelos ETS na previsão de um vasto conjunto de séries de vendas do comércio a retalho.

Para este trabalho a empresa Jerónimo Martins disponibilizou as vendas diárias, compreendidas entre 2 de janeiro de 2007 e 31 de julho de 2012, de todos os produtos de quatro categorias distintas de uma loja Pingo Doce com uma dimensão de 1500 m².

O conjunto em estudo contém um total de 817 produtos pertencentes às categorias de:

- “Água” com 118 produtos,
- “Higiene Oral” com 212 produtos,
- “Produtos Básicos” com 268 produtos e
- “Sobremesas e Gelados” com 219 produtos.

A Figura 2 mostra o cronograma do número de produtos distintos vendidos diariamente entre 2 de janeiro de 2007 e 31 de julho de 2012. Pode observar-se que este número apresenta uma tendência de subida com uma forte sazonalidade anual.

A Figura 3 mostra o gráfico sazonal do número de produtos distintos vendidos mensalmente entre janeiro de 2007 e julho de 2012. Também através deste gráfico se constata que o número mensal de produtos distintos vendidos aumentou gradualmente desde 2007 até 2010. No ano

de 2011 o aumento não foi tão significativo como nos anos anteriores, tendo havido mesmo uma ligeira diminuição em alguns meses do ano. No 1º semestre de 2012 observa-se uma ligeira tendência de descida em relação ao ano anterior com as exceções dos meses de fevereiro e julho.

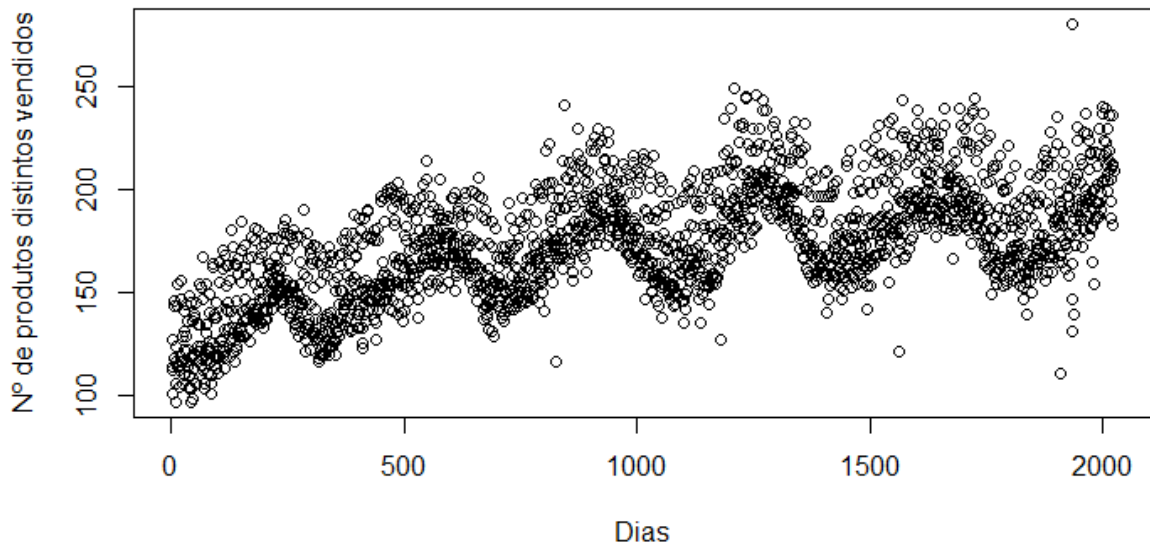


Figura 2: Cronograma do número de produtos distintos vendidos diariamente entre 2 de janeiro de 2007 e julho de 2012.

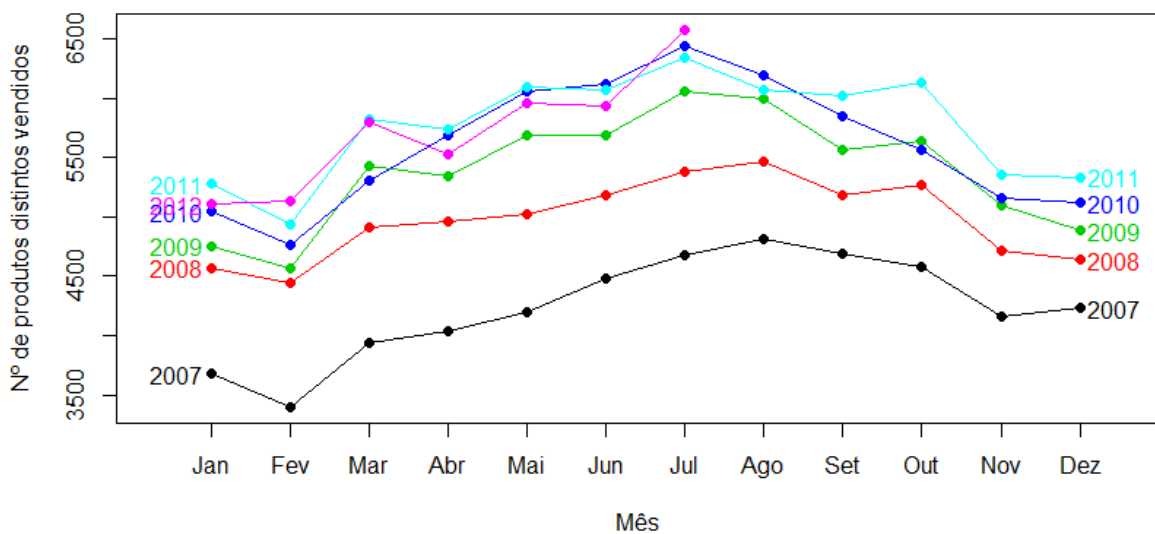


Figura 3: Gráfico sazonal do número de produtos distintos vendidos mensalmente entre janeiro de 2007 e julho de 2012.

A Figura 4 mostra um outro gráfico sazonal do número de produtos distintos vendidos mensalmente entre janeiro de 2007 e julho de 2012, onde se pode observar o comportamento dos valores em cada mês ao longo dos anos e o respetivo valor médio.

Este gráfico mostra que o número de produtos distintos vendidos mensalmente vai aumentando gradualmente desde o início do ano (com a exceção do mês de fevereiro onde se observa uma diminuição) até ao mês de julho, descendo a partir daí até ao final do ano.

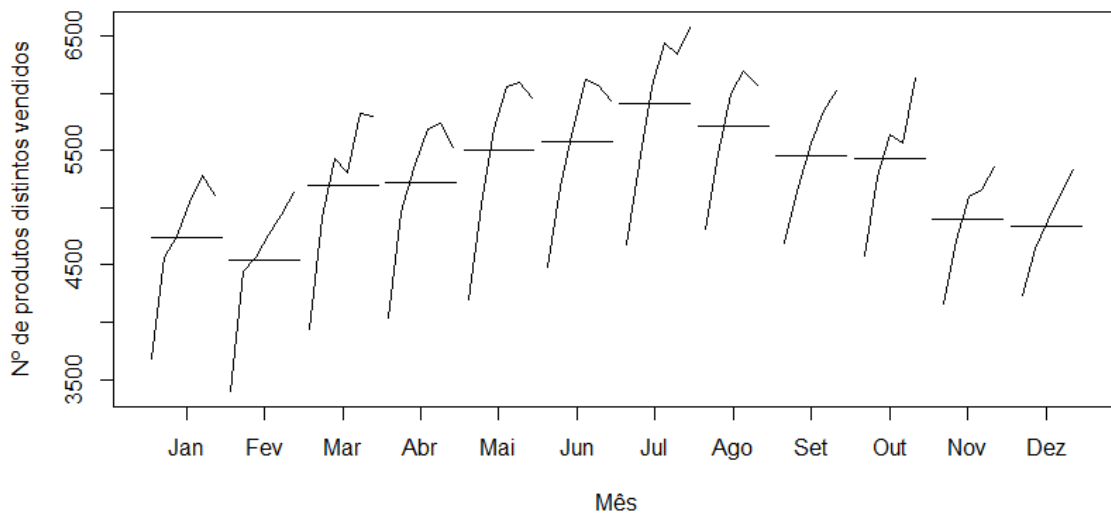


Figura 4: Gráfico sazonal do número de produtos distintos vendidos mensalmente entre janeiro de 2007 e julho de 2012 com identificação do valor médio de cada mês.

Um dos interesses manifestados pela empresa Jerónimo Martins foi o da determinação do erro de uma previsão mensal das vendas de cada um dos produtos durante o horizonte temporal de um ano, com o objetivo de planear as compras a médio prazo de acordo com as condições dos fornecedores, gerir o espaço disponível em armazém ao longo do tempo, gerir a logística de abastecimento às lojas de acordo com os respetivos espaços, etc, ou seja fazer a gestão da cadeia de abastecimento.

Este trabalho pretendeu dar resposta a essa pretensão da empresa confrontando as duas principais metodologias de previsão – os modelos de espaço de estados e os modelos ARIMA.

Ao agregar mensalmente o número de unidades vendidas de cada produto constatou-se que apenas 67 produtos continham vendas nos 67 meses observados (janeiro de 2007 a julho de 2012). Os restantes tinham falhas em diversos meses, havendo nomeadamente 33 produtos que tinham apenas unidades vendidas em um dos meses do período observado.

Para não comprometer o objetivo do trabalho, que era comparar o desempenho de previsão dos modelos ETS e ARIMA quando utilizados para prever um vasto conjunto de séries de vendas do comércio a retalho, foram utilizados apenas os 67 produtos cujo número de unidades vendidas em cada um dos meses do período observado era não nulo.

Por um lado, este número seria suficiente para levar a cabo a referida comparação e por outro, em termos de modelação seria conveniente que as séries utilizadas correspondessem a um cenário real e completo do comportamento das vendas de cada um dos produtos ao longo do tempo.

Refira-se que destes 67 produtos, 5 pertencem à categoria “Água”, 12 pertencem à categoria “Higiene Oral”, 25 pertencem à categoria “Produtos Básicos” e 25 pertencem à categoria “Sobremesas e Gelados”.

Para ilustrar a variedade de séries analisadas, a Figura 5 mostra o cronograma das vendas mensais entre janeiro de 2007 e julho de 2012 (67 observações) de 6 do conjunto das 67 séries estudadas.

Como se pode observar, tal como a maioria das séries de vendas do comércio a retalho, estas séries são não-estacionárias apresentando tendência e/ou padrões sazonais e/ou movimentos cíclicos, correspondendo uma grande variedade de comportamentos complexos que permitem testar o desempenho de previsão dos modelos ETS e ARIMA.

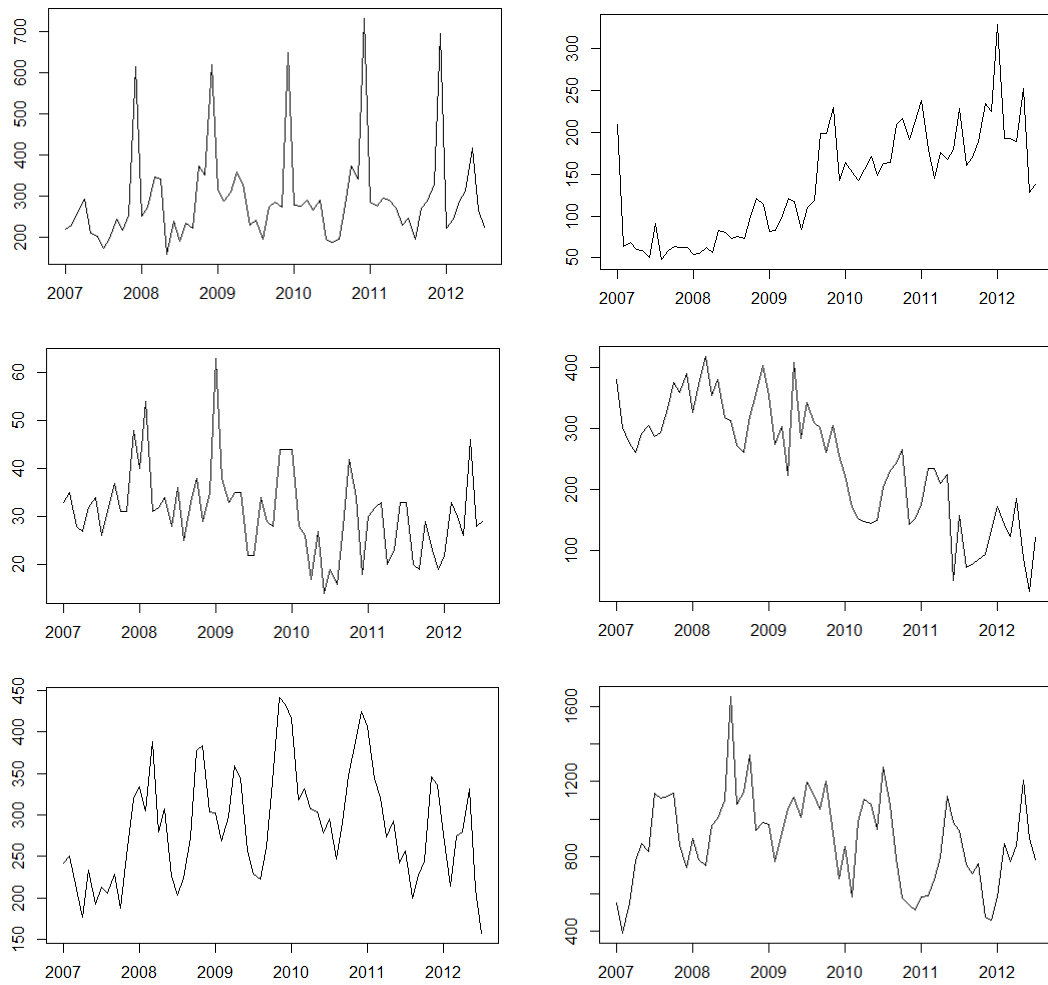


Figura 5: Cronograma das vendas mensais de bens de consumo de uma loja Pingo Doce entre janeiro de 2007 e julho de 2012.

4. MODELAÇÃO E PREVISÃO

4.1 O “*software*” estatístico R

Todo o trabalho de modelação e previsão foi desenvolvido usando o “*software*” livre estatístico R (R Development Core Team, 2013).

Este “*software*” consiste num conjunto de “*packages*” cada um deles orientado para o tratamento e resolução de problemas específicos.

Cada “*package*” consiste num conjunto de funções que quando invocadas executam rotinas de programação.

O “*software*” R pode ser utilizado através do programa “RStudio”, cujo “*layout*” se mostra na Figura 6, que consiste num ambiente de programação mais amigável do que o ambiente de programação disponibilizado pelo R.

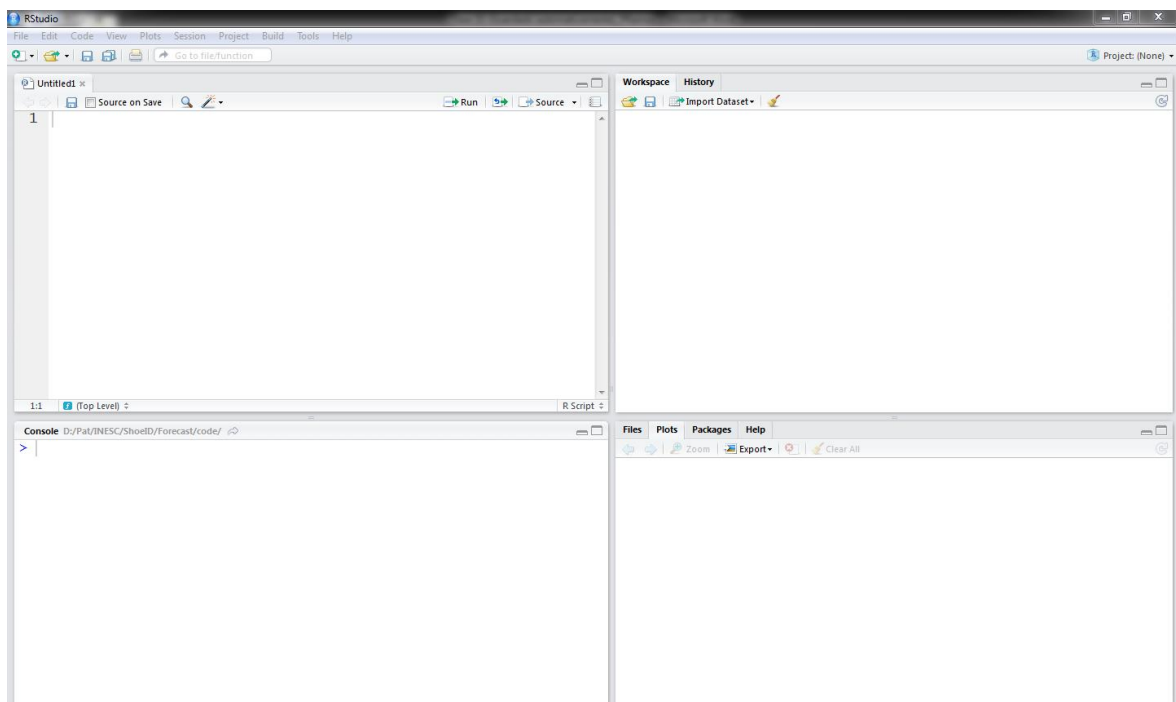


Figura 6: Ambiente de modelação e previsão.

Para a análise das séries temporais em estudo foi utilizado o “*package Forecast*” que tem implementados os métodos referidos no Capítulo 2 (Hyndman, 2012, Cryer, 2008). Os resultados dessa modelação encontram-se na secção que se segue.

4.2 Modelação e diagnóstico

Tal como referido no Capítulo 2, é incorreto avaliar o desempenho da previsão exclusivamente pelo ajuste do modelo aos dados históricos. A avaliação do desempenho deve ser efetuada usando dados históricos que não foram utilizados no ajuste do modelo.

Assim, para o efeito, a série de vendas mensais compreendidas entre janeiro de 2007 e julho de 2012 de cada um dos produtos (67 observações) foi dividida em um conjunto de treino e um conjunto de teste:

- O conjunto de treino consistiu nas vendas mensais compreendidas entre janeiro de 2007 e julho de 2011 (as primeiras 55 observações);
- O conjunto de teste consistiu nas vendas mensais compreendidas entre agosto de 2011 e julho de 2012 (as últimas 12 observações).

Os dados do conjunto de treino foram utilizados para o processo de modelação e diagnóstico. Os dados do conjunto de teste foram utilizados para a avaliação do desempenho de previsão de cada um dos modelos.

4.2.1 Modelo ETS

O conjunto de treino de cada uma das séries de vendas mensais dos 67 produtos (55 observações) foi ajustado a todos os modelos ETS admissíveis nomeadamente: ETS(A,N,N), ETS(A,N,A), ETS(A,A,N), ETS(A,A,A), ETS(A,A_d,N), ETS(A,A_d,A), ETS(M,N,N), ETS(M,N,A), ETS(M,N,M), ETS(M,A,N), ETS(M,A,A), ETS(M,A,M), ETS(M,A_d,N), ETS(M,A_d,A), ETS(M,A_d,M), ETS(M,M,N), ETS(M,M,M), ETS(M,M_d,N), ETS(M,M_d,M).

Os parâmetros de alisamento e os estados iniciais de cada modelo ajustado foram estimados maximizando a verosimilhança do modelo (para mais detalhes consultar Hyndman (2012)).

De entre todos os modelos ETS ajustados foi selecionado aquele que possuía o menor valor do Critério de Informação de Akaike corrigido – AIC_c (ver Equação 38) e que simultaneamente passou o teste estatístico de diagnóstico de resíduos de Ljung-Box com um nível de significância de 5% (ver Equação 14) e $h = 15$.

O modelo ETS escolhido para cada uma das séries de vendas mensais dos 67 produtos foi depois utilizado para efetuar as previsões e os intervalos de previsão para o período do conjunto de teste, i.e. de agosto de 2011 a julho de 2012 (12 instantes).

4.2.2 Modelo ARIMA

Para cada uma das séries de vendas mensais dos 67 produtos foi determinado um modelo ARIMA utilizando o respetivo conjunto de treino (55 observações).

A seleção do modelo ARIMA adequado foi levada a cabo utilizando o algoritmo proposto por Hyndman (2012) constituído pelos passos seguintes:

1. Determinar o número de diferenciações sazonais D ($D = 0,1$) necessárias para tornar a séries estacionária usando o teste OCSB; Diferenciar sazonalmente a série, se for o caso.
2. Determinar o número de diferenciações simples d , necessárias para tornar a série estacionária, fazendo repetidos testes KPSS à série diferenciada sazonalmente (se $D = 1$) ou à série original (se $D = 0$); Efetuar as diferenciações simples determinadas à série diferenciada sazonalmente ou à série original, se for o caso.
3. Escolher os valores de p, q, P e Q da forma seguinte:

- a. Selecionar o modelo com o menor valor do Critério de Informação de Akaike corrigido – AICc (ver Equação 63) de entre os modelos seguintes:

ARIMA(0, d ,0) se $m = 1$ ou ARIMA(0, d ,0)×(0, D ,0) _{m} se $m > 1$;

ARIMA(1, d ,0) se $m = 1$ ou ARIMA(1, d ,0)×(1, D ,0) _{m} se $m > 1$;

ARIMA(0, d ,1) se $m = 1$ ou ARIMA(0, d ,1)×(0, D ,1) _{m} se $m > 1$;

ARIMA(2, d ,2) se $m = 1$ ou ARIMA(2, d ,2)×(1, D ,1) _{m} se $m > 1$;

Se $d + D \leq 1$ estes modelos são ajustados com $c \neq 0$, caso contrário $c = 0$.

O modelo selecionado em (a) é designado por modelo corrente.

- b. Considerar no máximo as seguintes 13 variações do modelo corrente:
 - i. p, q, P e Q variam individualmente em ± 1 ;
 - ii. p e q variam simultaneamente em ± 1 ;
 - iii. P e Q variam individualmente em ± 1 ;
 - iv. c é incluído se o modelo corrente tem $c = 0$, ou excluído se o modelo corrente tem $c \neq 0$.

Se um destes modelos tiver menor valor de AICc do que o modelo corrente então esse passa a ser o modelo corrente.

- c. Repetir 3b. enquanto um novo modelo corrente for encontrado.

Atente-se que neste algoritmo, dados anuais são considerados não sazonais ($m=1$) e todos os outros são considerados dados sazonais. Por exemplo, para dados mensais ou trimestrais $m=12$ e $m=4$, respetivamente.

Para evitar problemas de convergência ou proximidade de raízes unitárias, os modelos ajustados por este algoritmo estão restringidos às condições seguintes:

- Os valores de p e q estão limitados a um valor máximo (o valor por defeito é 5).
- Os valores de P e Q estão limitados a um valor máximo (o valor por defeito é 2).
- Um modelo que esteja próximo de ser não invertível ou não estacionário é rejeitado. Especificamente, se alguma das raízes de $\Phi_p(B^m)\phi_p(B)$ ou de $\theta_q(B)\Theta_Q(B^m)$ for inferior a 1.001 em valor absoluto, o respetivo modelo é rejeitado
- Se surgir algum erro durante a execução da rotina de otimização não linear usada na estimação do modelo, este é rejeitado. A ideia subjacente é que um modelo difícil de ajustar não é provavelmente um bom modelo para traduzir o comportamento dos dados.

Este algoritmo retornará sempre um modelo válido visto que o conjunto dos modelos admissíveis é finito e pelo menos o modelo do passo 3a. com p, q, P e Q iguais 0 é aceite.

Todos os modelos seleccionados para cada um dos conjuntos de treino das séries de vendas mensais dos 67 produtos passaram o teste estatístico de diagnóstico de resíduos de Ljung-Box com um nível de significância de 5% (ver Equação 14) e $h=15$.

O modelo ARIMA escolhido para cada uma das séries de vendas mensais dos 67 produtos foi depois utilizado para efetuar previsões e intervalos de previsão para o período do conjunto de teste, i.e. de agosto de 2011 a julho de 2012 (12 instantes).

4.3 Resultados da avaliação do desempenho de previsão

Os dois modelos ETS e ARIMA escolhidos para cada uma das séries de vendas mensais dos 67 produtos foram utilizados para efetuar previsões 1-passo e multi-passo à frente para o período do conjunto de teste, i.e. de agosto de 2011 a julho de 2012.

As previsões 1-passo à frente (para os 12 instantes do conjunto de teste) foram obtidas, utilizando cada um dos modelos ajustados ao conjunto de treino, através dos procedimentos descritos nas Secções 2.3.8 (para o modelo ETS) e 2.4.7 (para o modelo ARIMA) com $h=12$.

Obtidas as 12 previsões pontuais, relativas a cada um dos modelos ETS e ARIMA, para cada uma das séries de vendas mensais dos 67 produtos, foi calculado o respetivo EPAM – erro percentual absoluto médio, utilizando a Equação 11 mencionada na Secção 2.2.

O valor médio dos 67 erros percentuais absolutos médios relativo a cada um dos modelos ETS e ARIMA encontra-se na Tabela 6. Esta medida de avaliação de erros de previsão é uma das mais adequadas para comparar desempenhos de previsão relativos a conjuntos de dados expressos em escalas distintas.

Seja T o número total de observações da série, N o número de observações do conjunto de treino e h o passo à frente das previsões multi-passo.

No caso de estudo presente $T=67$ e $N=55$.

As previsões multi-passo à frente para cada uma das séries de vendas mensais dos 67 produtos foram obtidas usando o algoritmo seguinte:

Para $h=1$ até $T-N$

Para $i=1$ até $T-N-h+1$

Selecionar a observação do instante $N+h+i-1$ para conjunto de teste

Utilizar as observações até ao instante $N+i-1$ para estimar o modelo

Calcular o erro h -passo(s) à frente relativo à previsão do instante $N+h+i-1$

Calcular o EPAM baseado nos erros de previsão h -passo(s) à frente obtidos

O valor médio dos 67 erros percentuais absolutos médios obtidos para cada um dos h -passos à frente das previsões multi-passo (com $h=1$ até 12) relativo a cada um dos modelos ETS e ARIMA encontra-se na Tabelas 6.

Deve salientar-se que nas previsões multi-passo o modelo é estimado recursivamente em cada passo i usando as observações até ao instante $N+i-1$. Atente-se que neste caso para cada série, o número de previsões multi-passo 1- passo à frente são 12, o número de previsões

multi-passo 2- passos à frente são 11, o número de previsões multi-passo 3- passos à frente são 10, e assim sucessivamente.

Tabela 6: EPAM (%) para as previsões fora da amostra (agosto 2011 a julho 2012)

Modelo	Previsão 1-passo à frente	Passo à frente das previsões multi-passo (<i>h</i>)											
		1	2	3	4	5	6	7	8	9	10	11	12
ARIMA	64.83	44.31	53.20	60.96	67.74	67.42	68.26	59.38	50.65	49.29	41.78	44.35	45.76
ETS	132.93	44.53	60.69	83.59	102.31	107.51	113.90	92.81	73.32	69.21	52.40	53.67	52.56

Ambas as previsões 1-passo à frente e multi-passo estão em linha com o interesse manifestado pela empresa Jerónimo Martins de utilizar esta informação com o objetivo de planear as compras a médio prazo de acordo com as condições dos fornecedores, gerir o espaço disponível em armazém ao longo do tempo, gerir a logística de abastecimento às lojas de acordo com os respetivos espaços, etc, ou seja fazer a gestão da cadeia de abastecimento.

Os resultados mostram que o desempenho de previsão dos modelos ARIMA é indiscutivelmente superior ao do dos modelos ETS quando julgados pelo EPAM.

Quer as previsões 1-passo à frente quer as previsões multi-passo dos modelos ARIMA são consistentemente melhores do que as dos modelos ETS. A diferença de desempenho nas previsões 1-passo à frente é 51%. Nas previsões multi-passo a diferença de desempenho aumenta com o aumento do passo à frente até ao passo à frente 6 e depois diminui até ao passo à frente 12. As diferenças de desempenho são 0%, 12%, 27%, 34%, 37%, 40%, 36%, 31%, 29%, 20%, 17% e 13%, respetivamente. Em ambos os modelos o valor do EPAM tende a aumentar com o aumento do passo à frente até ao passo à frente 6 e depois tende a diminuir até ao passo à frente 12.

Estes resultados também mostram que as previsões multi-passo são de um modo geral mais corretas do que as previsões 1-passo à frente (com a exceção dos passos à frente 4, 5 e 6 nos modelos ARIMA). Esta situação não é surpreendente visto que as previsões multi-passo incorporam dados históricos mais recentes.

Observa-se que os valores do EPAM obtidos são relativamente elevados. Esta situação foi analisada e concluiu-se que se devia ao facto do EPAM de determinadas séries (cujo comportamento em certos instantes era inesperado) ser bastante elevado. Por exemplo, no caso das previsões 1-passo à frente, determinando o valor médio do EPAM apenas das séries cujo EPAM é inferior a 50% (um total de 40 séries), este valor diminui (de 64.83%) para 27.57% no caso dos modelos ARIMA e diminui (de 132.93%) para 36.17% no caso dos modelos ETS, erros bastante mais razoáveis. Conclui-se assim que, como trabalho futuro, dever-se-ão utilizar modelos avançados de previsão que considerem informações adicionais tais como o preço do produto e a existência ou não de promoções que “explicarão” o comportamento inesperado em certos instantes e farão diminuir os erros de previsão (Godfrey, 1978; Bowerman, 2005).

A determinação de estimativas da incerteza das previsões é um aspeto essencial em previsão que é frequentemente ignorado.

Também foi avaliado o desempenho de ambas as metodologias na produção de intervalos de previsão. As Tabelas 7 e 8 mostram a percentagem média de vezes que os intervalos de previsão nominais de 80% e 95%, respetivamente, contêm as observações reais para as previsões 1-passo à frente e multi-passo.

Os resultados mostram que ambas as metodologias ETS e ARIMA produzem probabilidades de cobertura que estão muito próximas das taxas nominais. A metodologia ETS produz melhores probabilidades de cobertura em ambos os intervalos de previsão de 80% e 95% para ambas as previsões 1-passo à frente e multi-passo.

Tabela 7: Cobertura dos intervalos de previsão de 80% (agosto 2011 a julho 2012).

Modelo	Previsão 1-passo à frente	Passo à frente das previsões multi-passo (h)											
		1	2	3	4	5	6	7	8	9	10	11	12
ARIMA	75	74	75	75	74	73	75	75	77	76	77	76	66
ETS	80	75	78	78	78	80	80	82	84	82	81	86	78

Tabela 8: Previsão de cobertura dos intervalos de 95% para previsões fora da amostra (agosto 2011 a julho 2012)

Modelo	Previsão 1-passo à frente	Passo à frente das previsões multi-passo (h)											
		1	2	3	4	5	6	7	8	9	10	11	12
ARIMA	92	89	91	91	90	89	92	91	90	90	89	89	90
ETS	94	90	91	90	90	91	93	94	93	93	92	95	97

5. CONCLUSÕES

A previsão de vendas é fundamental para o sucesso das operações da cadeia de abastecimento de qualquer distribuidor do comércio a retalho. Previsões erradas poderão conduzir a aprovisionamentos escassos ou excessivos afetando diretamente o lucro da empresa e a sua posição competitiva no mercado.

As vendas no comércio de bens e serviços pertencem a um tipo especial de séries temporais que normalmente contêm ambos os padrões de tendência e sazonalidade, para além de outros aspetos, apresentando desafios para o desenvolvimento eficaz de modelos de previsão.

O objetivo deste trabalho foi comparar o desempenho de previsão dos modelos de espaço de estados e dos modelos ARIMA quando aplicados a um vasto conjunto de séries de vendas de bens de consumo do comércio a retalho.

Para este trabalho a empresa Jerónimo Martins disponibilizou as vendas diárias, compreendidas entre 2 de janeiro de 2007 e 31 de julho de 2012, de todos os produtos de quatro categorias distintas de uma loja Pingo Doce com uma dimensão de 1500 m².

Um dos interesses manifestados pela empresa Jerónimo Martins foi o da determinação do erro de uma previsão mensal das vendas de cada um dos produtos durante o horizonte temporal de um ano, com o objetivo de planear as compras a médio prazo de acordo com as condições dos fornecedores, gerir o espaço disponível em armazém ao longo do tempo, gerir a logística de abastecimento às lojas de acordo com os respetivos espaços, etc, ou seja fazer a gestão da cadeia de abastecimento. Este trabalho pretendeu dar resposta a essa pretensão da empresa confrontando as duas principais metodologias de previsão – os modelos de espaço de estados e os modelos ARIMA.

A série de vendas mensais compreendidas entre janeiro de 2007 e julho de 2012 de cada um dos produtos foi dividida em um conjunto de treino e um conjunto de teste. Os dados do conjunto de treino foram utilizados para o processo de modelação e diagnóstico. Os dados do conjunto de teste foram utilizados para a avaliação do desempenho de previsão de cada um dos modelos.

O conjunto de treino de cada uma das séries de vendas mensais dos produtos foi ajustado a todos os modelos ETS admissíveis. De entre todos os modelos ETS ajustados foi selecionado aquele que possuía o menor valor do AIC_c e que simultaneamente passou o teste estatístico de diagnóstico de resíduos de Ljung-Box com um nível de significância de 5%.

A seleção do modelo ARIMA adequado a cada uma das séries de vendas mensais dos produtos estudados utilizando o respectivo conjunto de treino foi levada a cabo utilizando o algoritmo proposto por Hyndman (2012).

Atente-se que em ambos os casos foram utilizados procedimentos automáticos de seleção do melhor modelo, necessários quando está em análise um vasto conjunto de séries.

Os dois modelos ETS e ARIMA escolhidos para cada uma das séries de vendas mensais dos produtos em análise foram utilizados para efetuar previsões 1-passo e multi-passo à frente para o período do conjunto de teste.

Os resultados mostram que o desempenho de previsão dos modelos ARIMA é indiscutivelmente superior ao do dos modelos ETS quando julgados pelo EPAM e que as previsões multi-passo são de um modo geral mais corretas do que as previsões 1-passo à frente situação que não é surpreendente visto que as previsões multi-passo incorporam dados históricos mais recentes.

Observa-se que os valores do EPAM obtidos são relativamente elevados concluindo-se que, como trabalho futuro, dever-se-ão utilizar modelos avançados de previsão que considerem informações adicionais tais como o preço do produto e a existência ou não de promoções que eventualmente façam diminuir os erros de previsão.

Também foi avaliado o desempenho de ambas as metodologias na produção de intervalos de previsão. Os resultados mostram que ambas as metodologias ETS e ARIMA produzem probabilidades de cobertura que estão muito próximas das taxas nominais. A metodologia ETS produz melhores probabilidades de cobertura em ambos os intervalos de previsão de 80% e 95% para ambas as previsões 1-passo à frente e multi-passo.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. (1974). “A new look at the statistical model identification”, IEEE Trans. On Automatic Control 19, 716-722.
- BOX, G. P. E. e JENKINS, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden-day, San Francisco.
- BOX, G. P. E. e JENKINS, G. M. (1976). Time series Analysis: Forecasting and Control, 2nd ed., Holden-day, San Francisco.
- BOX, G. P. E., JENKINS, G. M. e REINSEL, G. (1994). Time Series Analysis: Forecasting and Control, 3rd ed., Prentice-Hall.
- BOWERMAN, B., O'CONNELL, R. e KOEHLER, A. (2005). Forecasting, Time Series and Regression, 4^a edição, Duxbury, Thomson Brooks/Cole, Belmont.
- BROCKWELL, P. J. e DAVIS, R. A. (1991). Time Series: Theory and Methods. 2nd ed., Springer, New York.
- BROWN, R (1956). “Exponential smoothing for predicting demand”, 10th National Meeting of the Operations Research Society of America, San Francisco.
- BROWN, R. (1959). Statistical Forecasting for Inventory Control, McGraw-Hill, New York.
- BROWN, R. (1963). Smoothing, Forecasting, and prediction of Discrete Time series, Prentice-Hall, New Jersey.
- CAIADO, J. (2002). “Métodos de previsão em Economia e Gestão: Desenvolvimentos recentes”, Inforbanca, XIV, 53, 18-19 e 44-47.
- CAIADO, A. C. e CAIADO, J. (2008). Gestão de Instituições Financeiras, 2^a edição, Edições Sílabo, Lisboa.
- CAIADO, J. (2011). Métodos de Previsão em Gestão Com Aplicações em Excel, Edições Sílabo, Lisboa.
- CHATFIELD, C. (2001). “A survey of recent developments in forecasting methods”, 5th Time Series Workshop, Arrábida, Portugal.
- CRYER, J. D., CHAN, K-S (2008). Time series analysis with applications in R, 2nd ed., Springer.

- DALRYMPLE, D. J. (1975). "Sales forecasting methods and accuracy", Business Horizons, pp. 69-73.
- ENDERS, W. (1995). Applied Econometric Time Series. Wiley, New York.
- GARDNER, E. S., Jr. e MCKENZIE, Ed. (1985). "Forecasting trends in time series", Management Science, 31, 1237-1246.
- GARDNER, E. S. (1995). "Exponential smoothing: The state of the art", Journal of Forecasting, 4, 1-28.
- GARDNER, E. S. (2006). "Exponential smoothing: The state of the art - Part II", Journal of Forecasting, 22, 637-666.
- GODFREY, L. G. (1978). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables", Econometrica, 46, 1293-1301.
- GUGARATI, D. (2003). Basic Econometrics, 4ª edição, McGraw-Hill, Boston.
- HAMILTON, J. D. (1994). Time Series Analysis, Princeton University Press.
- HAMILTON, L. C. (2006). Statistics with STATA, Duxbury, Thomson Brooks/Cole.
- HILL R.C., GRIFFITHS, W. E. e LIM, G. C. (2008). Principles of Econometrics, 3ª ed., John Wiley & Sons, New York.
- HOGARTH, R. e MAKRIDAKIS, S. (1981). "Forecasting and planning: an evolution", Management Science, 27, 115-138.
- HOLT, C. C. (1957). "Forecasting seasonal and trends by exponentially weighted moving averages", Office of Naval Research, Research Memorandum No.52.
- HYNDMAN R., ATHANASOPOULOS G. (2012). Forecasting: Principles and practice. <http://otexts.com/fpp/>.
- Jerónimo Martins. Endereço URL: <http://www.jeronimomartins.pt/> (acedido em julho 2013).
- JOHNSON. R, E. M., KURTZ, D. e SCHEVING, E. (1994). Sales Management - Concepts, Practices and Cases, McGraw-Hill, Singapore International Editions.
- LAWERENCE, M. J. e O'CONNOR, M. J. (1992). "Exploring judgmental forecasting", International Journal of Forecasting, 8, 15-26.

- LJUNG, G. e BOX, G. (1978). “On a measure of lack of fit in time series models”, *Biometrika*, 66, 297-303.
- MAKRIDAKIS, S. e WHEELRIGHT, S. C. (1989). *Forecasting Methods for Management*, 5ª ed., John Wiley & Sons, New York.
- MAKRIDAKIS, S., WHEELRIGHT, S. C. e HYNDMAN, R. (1998). *Forecasting: Methods and Applications*, 3ª ed., John Wiley & Sons, New York.
- MENTZER, J. T. e Cox, J. E. (1984). “Familiarity, application and performance of sales forecasting techniques”, *Journal of Forecasting*, 3, 27 -36.
- MENTZER, J. T. e KAHN, K. B. (1995). “Forecasting technique familiarity, satisfaction, usage, and application”, *Journal of Forecasting*, 14, 465-476.
- MENTZER, J. T. e KAHN, K. B. (1997). “State of sales forecasting systems in corporate America”, *Journal of Business Forecasting*, 16, 6-13.
- MURTEIRA, B. J., MULLER, D. e TURKMAN, F. (1994). *Análise de Sucessões Cronológicas*, McGraw-Hill, Lisboa.
- MURTEIRA, B. J., RIBEIRO, C. S., ANDRADE E SILVA, J. e PIMENTA, C. (2010). *Introdução à Estatística*, Escolar Editora, Lisboa.
- PEÑA, D., TIAO, G. e TSAY, R. (2001). *A Course in Time Series Analysis*, Wiley, New York.
- PINDYCK, R. e RUBINFELD, D. (1998). *Econometric Models and Economic Forecasts*, 4th ed., McGraw-Hill, Boston.
- R DEVELOPMENT CORE TEAM. Free *software* environment for statistical computing and graphics. <http://www.r-project.org/> (acedido em junho 2013).
- RAMOS, P. (2012). “Apontamentos da Unidade Curricular de Métodos Quantitativos”. ISCAP-IPP.
- SCHWARTZ, G.(1978). “Estimating the dimension of a model”, *Annals of Statistics*, 6, 461-464.
- SHUMWAY, R. H. and STOFFER, D. S. (2000). *Time Series Analysis and Its Applications*. Springer-Verlag, New York.
- TAYLOR, J. (2003). Exponential Smoothing with a Damped Multiplicative Trend. *International Journal of Forecasting*, Vol. 19, 715-725.

WEI, W. W. S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley Publishing Company, Redwood City, California.

WEI, W. W. S. (2007). *Time Series Analysis: Univariate and Multivariate Methods*. 2nd ed., Addison-Wesley Publishing Company, Redwood City, California.

WINKLHOFER, H., DIAMANTOPOULOS, A. e WITT, S. F. (1996). “Forecasting practice: a review of the empirical literature and an agenda for future research”, *International Journal of Forecasting*, 12, 193-221.

WINTERS, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science* 6, 324–342.