



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

**Modelação e Previsão de Vendas no Setor do Retalho de Calçado**

**Ana Isabel da Silva Soares de Sousa**

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

**Orientada por:** Professora Doutora Patrícia Alexandra Gregório Ramos

**Porto, Fevereiro de 2013**





# **Modelação e Previsão de Vendas no Setor do Retalho de Calçado**

**Ana Isabel da Silva Soares de Sousa**

**Orientada por:** Professora Doutora Patrícia Alexandra Gregório Ramos

**Porto, Fevereiro de 2013**

## RESUMO

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 principal deste trabalho consiste na comparação do desempenho de duas metodologias de previsão na análise de séries de vendas do setor do retalho de calçado. Os dados analisados consistem em cinco séries temporais relativas às cinco principais categorias de calçado comercializadas pela empresa Foreva: a Bota, o Botim, a Sabrina, a Sandália, e o Sapato. Para este estudo foram disponibilizados pela empresa dados diários das vendas destas categorias até dezembro de 2011. Desde logo foi solicitada pela Foreva uma previsão anual das vendas de cada uma das categorias de calçado para o ano seguinte, tendo a empresa como objetivo a utilização dessas previsões para a especificação do número pares de calçado de cada uma das categorias a adquirir para comercialização. Este trabalho pretendeu dar resposta a esta pretensão da empresa confrontando as duas principais metodologias de previsão – os modelos de espaço de estado e os modelos ARIMA. Os resultados mostram que, de um modo geral, os modelos ARIMA têm melhor desempenho que os modelos de espaço de estado na previsão de séries de vendas do setor do retalho de calçado.

**Palavras - chave:** Previsão, Modelos ARIMA, Modelos de espaço de estado inovativos, Alisamento Exponencial, Vendas, Comércio a Retalho, Calçado

## **ABSTRACT**

Retail sales of goods and services belong to a special type of time series that typically contain both trend and seasonality patterns, in addition to other aspects, presenting challenges for the effective development of prediction models. The main objective of this work consists in comparing the performance of two methods of prediction on analysis of industry sales in the retail of footwear. The data analyzed consist of five series related to the five main categories of footwear marketed by the company Foreva: the Boot, the Bootie, the Flat, the Sandal and the Shoe. For this study were made available by the company daily data from sales of these categories until December 2011. Since then was requested by Foreva a forecast annual sales for each category of footwear for the following year, and the company aimed to use these predictions to the specification of the number of pairs of shoes each category to acquire for marketing. This work aimed to address this claim of the company contrasting the two main methods of forecasting - the state-space models and ARIMA models. The results show that, in general, ARIMA models have better performance than the state-space models in time series forecasting industry sales of retail footwear.

**Keywords:** Forecasting, ARIMA Models, Innovations state space models, Exponential Smoothing, Retail Sales, Shoes

## ABREVIATURAS E SÍMBOLOS

AE	Alisamento Exponencial
AES	Alisamento Exponencial Simples
AIC	CrITÉrio de InformaÇo de Akaike (em inglÊs)
AIC <sub>c</sub>	CrITÉrio de InformaÇo de Akaike corrigido (em inglÊs)
AR	Autoregressivo
ARMA	Autoregressivo e de MÉdias MÓveis
ARIMA	Autoregressivo e de MÉdias MÓveis Integrado
BIC	CrITÉrio de InformaÇo Bayesiano (em inglÊs)
EM	Erro Médio
EAM	Erro Absoluto Médio
EEAM	Erro Escalado Absoluto Médio
EPAM	Erro Percentual Absoluto Médio
EPM	Erro Percentual Médio
EQM	Erro Quadrático Médio
ETS	ExponenTial Smoothing
FAC	FunÇo de AutocorrelaÇo
FACP	FunÇo de AutocorrelaÇo Parcial
IP	Intervalos de Previso
MA	MÉdias MÓveis
NID	Normal e Independentemente Distribuído
REQM	Raiz Quadrada do Erro Quadrático Médio
SAR	Autoregressivo Estritamente Sazonal
SARMA	Autoregressivo e de MÉdias MÓveis Estritamente Sazonal
SARIMA	Autoregressivo e de MÉdias MÓveis Integrado Estritamente Sazonal
SMA	MÉdias MÓveis Estritamente Sazonal
SQE	Soma de Quadrados dos Erros

# ÍNDICE

RESUMO.....	I
ABSTRACT .....	II
ABREVIATURAS E SÍMBOLOS.....	III
ÍNDICE DE FIGURAS.....	VI
ÍNDICE DE TABELAS .....	VIII
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.....	10
2.3.1 Alisamento exponencial simples.....	11
2.3.2 Método de tendência linear de Holt.....	12
2.3.3 Métodos de tendência amortecida .....	13
2.3.4 Método sazonal de Holt-Winters.....	14
2.3.5 Taxonomia dos métodos de alisamento exponencial .....	16
2.3.6 Modelos de espaço de estado .....	18
2.4 Modelos ARIMA .....	22
2.4.1 Estacionaridade e diferenciação.....	23
2.4.2 Modelos autoregressivos e de médias móveis .....	26
2.4.3 Modelos ARIMA não sazonais.....	28
2.4.4 Modelos ARIMA estritamente sazonais.....	29
2.4.5 Modelo ARIMA multiplicativo sazonal .....	30
2.4.6 Metodologia de Box-Jenkins .....	30
3. CASO DE ESTUDO .....	34
3.1 Empresa Foreva.....	34
3.2 Análise exploratória dos dados.....	34
4. MODELAÇÃO E PREVISÃO.....	42
4.1 Modelação e previsão por alisamento exponencial .....	44
4.1.1 Categoria Bota .....	44

4.1.2 Categoria Botim .....	47
4.1.3 Categoria Sabrina .....	50
4.1.4 Categoria Sandália .....	52
4.1.5 Categoria Sapato .....	55
4.2 Modelação e previsão ARIMA .....	58
4.2.1 Categoria Bota .....	58
4.2.2 Categoria Botim .....	62
4.2.3 Categoria Sabrina .....	65
4.2.4 Categoria Sandália .....	68
4.2.5 Categoria Sapato .....	71
4.3-Comparação de resultados .....	75
5. CONCLUSÕES .....	78
REFERÊNCIAS BIBLIOGRÁFICAS .....	79

## ÍNDICE DE FIGURAS

Figura 1: Séries temporais simuladas (Caiado, 2011).....	23
Figura 2: Gráficos da série de vendas da categoria Bota.....	36
Figura 3: Gráficos da série de vendas da categoria Botim.....	37
Figura 4: Gráficos da série de vendas da categoria Sabrina.....	38
Figura 5: Gráficos da série de vendas da categoria Sandália.....	39
Figura 6: Gráficos da série de vendas da categoria Sapato.....	40
Figura 7: Ambiente de modelação e previsão.....	42
Figura 8: Cronograma, FAC e FACP dos resíduos do modelo $ETS(M,A_d,A)$ para a Bota.....	45
Figura 9: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Bota.....	46
Figura 10: Cronograma, FAC e FACP dos resíduos do modelo $ETS(M,A_d,M)$ para o Botim.....	48
Figura 11: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Botim.....	49
Figura 12: Cronograma, FAC e FACP dos resíduos do modelo $ETS(M,N,A)$ para a Sabrina.....	51
Figura 13: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Sabrina.....	52
Figura 14: Cronograma, FAC e FACP dos resíduos do modelo $ETS(A,A,A)$ para a Sandália.....	54
Figura 15: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Sandália.....	55
Figura 16: Cronograma, FAC e FACP dos resíduos do modelo $ETS(A,A,A)$ para o Sapato.....	57
Figura 17: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Sapato.....	58
Figura 18: Cronograma, FAC e FACP dos resíduos do modelo $ARIMA(2,1,5) \times (1,0,1)_{12}$ para a Bota.....	60
Figura 19: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Bota.....	62
Figura 20: Cronograma, FAC e FACP dos resíduos do modelo $ARIMA(5,0,2) \times (0,1,0)_{12}$ para o Botim.....	63
Figura 21: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Botim.....	65
Figura 22: Cronograma, FAC e FACP dos resíduos do modelo $ARIMA(5,0,5) \times (2,1,0)_{12}$ para a Sabrina.....	67
Figura 23: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Sabrina.....	68
Figura 24: Cronograma, FAC e FACP dos resíduos do modelo $ARIMA(3,0,1) \times (1,1,0)_{12}$ para a Sandália.....	70
Figura 25: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Sandália.....	71

Figura 26: Cronograma, FAC e FACP dos resíduos do modelo  $ARIMA(0,1,3) \times (1,0,0)_{12}$  para o Sapato. ....74

Figura 27: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Sapato..75

## ÍNDICE DE TABELAS

Tabela 1: Taxonomia dos métodos de alisamento exponencial (adaptado de Hyndman (2012)). .....	16
Tabela 2: Fórmulas recursivas dos métodos de alisamento exponencial (Hyndman, 2012).....	17
Tabela 3: Estratégia para escolha dos valores iniciais de alguns métodos de alisamento exponencial (adaptado de Hyndman, 2012).....	18
Tabela 4: Modelos de espaço de estado inovativos com erros aditivos (Hyndman, 2012).....	20
Tabela 5: Modelos de espaço de estado com erros multiplicativos (Hyndman, 2012).....	21
Tabela 6: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Bota. ....	44
Tabela 7: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Bota. ....	45
Tabela 8: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Botim.....	47
Tabela 9: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Botim.....	48
Tabela 10: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Sabrina.....	50
Tabela 11: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sabrina.....	51
Tabela 12: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Sandália. ....	53
Tabela 13: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sandália. ....	54
Tabela 14: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Sapato.....	56
Tabela 15: Medidas de avaliação dos erros de previsão dos meses de 2011 da categoria Sapato. .....	57
Tabela 16: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Bota. ....	60
Tabela 17: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Bota. ....	61
Tabela 18: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Botim.....	63

Tabela 19: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Botim.....	64
Tabela 20: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Sabrina.....	66
Tabela 21: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sabrina.....	67
Tabela 22: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Sandália. ....	69
Tabela 23: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sandália. ....	71
Tabela 24: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Sapato.....	73
Tabela 25: Medidas de avaliação dos erros de previsão para os meses do ano de 2011 da categoria Sapato.....	74
Tabela 26: Medidas de avaliação dos erros de previsão para o conjunto de teste dos modelos ETS e ARIMA selecionados para todas as séries de vendas. ....	77

## 1. INTRODUÇÃO

### 1.1 Enquadramento e motivação

A previsão tem atualmente um papel muito importante no planeamento, tomada de decisões e controlo em todas as organizações (Caiado, 2011). A dúvida relacionada com os fenómenos temporais não fornece exatidão do seu comportamento futuro, por isso usa-se a previsão para contornar esta fragilidade. Os planos de negócios estão normalmente sujeitos à previsão de vendas, levando os gestores de marketing a realizar periodicamente planos previsionais de vendas para os produtos/serviços da organização, tendo sempre em consideração as variações da procura e da oferta (Caiado, 2002; Caiado 2008). A utilização de métodos estatísticos, opiniões qualificadas e a experiência garante a minimização do desperdício de recursos. Todas as organizações utilizam a previsão para fazer o seu planeamento, e a sua interpretação é fundamental na elaboração de orçamentos, contas de exploração, gestão de compras, de pessoal e outros. Pois sem previsões é mais difícil atingir objetivos precisos e reais (Johnson, 1994; Hogarth, 1981; Gugarati, 2003; Bowerman, 2005; 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.

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

O objetivo principal deste trabalho consiste na comparação do desempenho de duas metodologias de previsão na análise de séries de vendas do setor do retalho de calçado.

Os dados analisados neste trabalho consistem em cinco séries temporais relativas às cinco principais categorias de calçado comercializadas pela empresa Foreva que são: a Bota, o Botim, a Sabrina, a Sandália, e o Sapato.

Dado que só a partir do ano de 2007 é que a Foreva estabilizou a sua rede de lojas optou-se por considerar as séries de vendas destas categorias de calçado a partir de 1 de janeiro de 2007. Para este estudo foram disponibilizados pela empresa dados diários das vendas destas categorias até dezembro de 2011.

Desde logo foi solicitada pela Foreva uma previsão anual das vendas de cada uma das categorias de calçado para o ano seguinte, tendo a empresa como objetivo a utilização dessas previsões para a especificação do número pares de calçado de cada uma das categorias a adquirir para comercialização.

Este trabalho pretendeu dar resposta a esta pretensão da empresa confrontando as duas principais metodologias de previsão – os modelos de espaço de estado 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 das organizações 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 dois principais métodos de previsão utilizados no caso em estudo: Alisamento Exponencial e Modelos ARIMA.

No terceiro capítulo é apresentado o caso de estudo desenvolvido neste trabalho. É feita uma breve apresentação da empresa FOREVA 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 cada uma das categorias de calçado estudadas.

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

## **2- MÉTODOS DE PREVISÃO**

### **2.1. Introdução**

No presente capítulo abordam-se os dois principais métodos de previsão utilizados neste trabalho: Alisamento Exponencial e Modelos ARIMA (Hamilton, 1994; Chatfield, 2001). Antes de se apresentarem os diversos aspetos destas metodologias de previsão, são referidos os principais 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).

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/electricidade, 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:

- 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).

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

- A média ou valor médio do conjunto de dados é dada por

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

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

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

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

$$S = +\sqrt{S^2} \quad (3)$$

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

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

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

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

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*)  $k \in \mathbb{N}$  é definida por

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

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

Em análise de séries temporais, ou de uma forma mais geral em qualquer análise de dados, frequentemente vários modelos podem representar de forma adequada o conjunto de dados (Dalrymple, 1975). O critério de seleção pode ser baseado nos erros de previsão (Lawerence, 1992, Mentzer, 1984, Mentzer, 1995; Ramos, 2012). Habitualmente, 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.

Tipicamente, o conjunto de teste consiste nas 20% últimas observações do conjunto de dados, podendo este valor ser ajustado em função do número de observações disponíveis e do horizonte temporal das previsões.

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

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

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:

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

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:

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

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

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

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 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 intervalos de previsão baseiam-se no erro quadrático médio,  $EQM = \text{média}(e_i^2)$ , 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 intervalo de previsão aproximado para cada instante de tempo é

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

onde  $z$  é um multiplicador que limita o intervalo de previsão 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:

- 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:

- 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, pela 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

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

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).

Se cada  $r_k$  for aproximadamente zero então  $Q$  será baixo. Se algum dos  $r_k$  for elevado (positivo ou negativo) então  $Q$  será elevado, sugerindo que as autocorrelações não provêm de um ruído branco. Mas qual deverá ser o limite para  $Q$ ? 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  $\alpha$  se o valor de  $Q$  excedesse o  $(1 - \alpha)$ -quantil da distribuição Qui-quadrado com  $h - K$  graus de liberdade (isto é, se o valor- $p \leq 0.05$ , 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; Ramos, 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  $\alpha$ .

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 na 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 á 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  $\alpha$  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 -  $\alpha$ . 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 á 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.

### 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):

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

$$\text{Equação da tendência: } b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \quad (22)$$

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  $h$ :

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

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):

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

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 ao 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 o método de tendência amortecida aditiva inclui o parâmetro de amortecimento  $0 \leq \phi \leq 1$ :

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

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:

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

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  $\alpha, \beta^*$  e  $\gamma$ , respetivamente. 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:

$$\text{Equação de previsão:} \quad \hat{y}_{t+h|t} = (l_t + hb_t)s_{t-m+h_m^+}$$

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

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:

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

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.

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 <sub>d</sub> (Amortecida aditiva)	(A <sub>d</sub> ,N)	(A <sub>d</sub> ,A)	(A <sub>d</sub> ,M)
M (Multiplicativa)	(M,N)	(M,A)	(M,M)
M <sub>d</sub> (Amortecida multiplicativa)	(M <sub>d</sub> ,N)	(M <sub>d</sub> ,A)	(M <sub>d</sub> ,M)

Taxonomia dos métodos estudados:

- (N,N) = Alisamento exponencial simples
- (A,N) = Método de tendência linear de Holt
- (M,N) = Método de tendência exponencial
- (A<sub>d</sub>,N) = Método de tendência amortecida aditiva
- (M<sub>d</sub>,N) = Método de tendência amortecida multiplicativa
- (A,A) = Método de Holt-Winters aditivo
- (A,M) = Método de Holt-Winters multiplicativo
- (A<sub>d</sub>,M) = Método amortecido de Holt-Winters

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

Cada método é identificado por um par de letras (T,S) que especifica, respetivamente, 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:

- $l_t$  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,
- $\alpha, \beta^*, \gamma$  e  $\phi$  são parâmetros de alisamento com  $\phi_h = \phi + \phi^2 + \dots + \phi^h$  e  $h_m^+ = [(h+1) \text{ mod } m] + 1$ .

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

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

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.

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

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

### 2.3.6 Modelos de espaço de estado

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 estado, consiste em uma equação da medida (ou observação) que descreve os dados, e uma ou mais equações do estado (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.

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

- **E**rro – cujas possibilidades são {A, M},
- **T**endência – cujas possibilidades são {N, A, A<sub>d</sub>, M, M<sub>d</sub>},
- **S**azonal – cujas possibilidades são {N, A, M}.

Existem 30 modelos de espaço de estado: 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 estado 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  $\sigma^2$ :

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

onde NID significa “Normal e Independentemente Distribuído”. As equações do modelo podem assim ser escritas da forma

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

Estas equações, em conjunto com a distribuição estatísticas dos erros, constituem o modelo de espaço de estado 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  $\varepsilon_t$ .

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  $\varepsilon_t$ , a parte não previsível de  $y_t$ . Em outros modelos de espaço de estado inovativos esta relação pode ser não linear. A influência do parâmetro de alisamento  $\alpha$  é a mesma que para os métodos de alisamento exponencial. Neste caso  $\alpha$  controla o grau de variação dos sucessivos níveis. Quanto maior for o valor de  $\alpha$  mais rápidas são as mudanças do nível; quanto menor for o valor de  $\alpha$  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:

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

onde  $\boldsymbol{\varepsilon}_t \sim \text{NID}(0, \sigma^2)$ . Substituindo  $\hat{y}_{t|t-1} = l_{t-1}$  em (34) obtém-se  $y_t = l_{t-1} + l_{t-1}\boldsymbol{\varepsilon}_t$ . Da equação (20) tem-se que  $e_t = y_t - \hat{y}_{t|t-1}$ , logo usando (34) vem

$$e_t = l_{t-1}\boldsymbol{\varepsilon}_t \quad (35)$$

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

Então, o modelo ETS(M,N,N) vem

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

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

De forma análoga podem obter-se as equações dos restantes modelos de espaço de estado inovativos subjacentes a cada um dos métodos de AE. A Tabela 4 apresenta as equações dos modelos de espaço de estado inovativos com erros aditivos e a Tabela 5 apresenta as equações dos modelos de espaço de estado inovativos com erros multiplicativos.

T	S		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A <sub>d</sub>	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M <sub>d</sub>	$y_t = \ell_{t-1}b_{t-1}^\phi + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$

Tabela 4: Modelos de espaço de estado inovativos com erros aditivos (Hyndman, 2012).

Uma alternativa à estimação dos parâmetros minimizando a SQE é maximizar a “verosimilhança”. A verosimilhança é a probabilidade de um conjunto de dados ser originário de um determinado modelo. Logo, um bom modelo está associado a uma verosimilhança 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. Os parâmetros de alisamento  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\phi$ , 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.

T	S		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A <sub>d</sub>	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M <sub>d</sub>	$y_t = \ell_{t-1}b_{t-1}^\phi(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t / \ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

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

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 serie 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<sub>c</sub> 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<sub>c</sub> e BIC.

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,

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

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 ETS é a possibilidade de obter intervalos de previsão (IP). 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 respetivos 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.

## 2.4 Modelos ARIMA

Os modelos ARIMA (AutoRegressivos e de Médias Móveis Integrados) consistem em uma outra abordagem estatística para previsão de séries temporais. 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 que 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. 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-se observar uma série não estacionária em média e uma série estacionária em média.

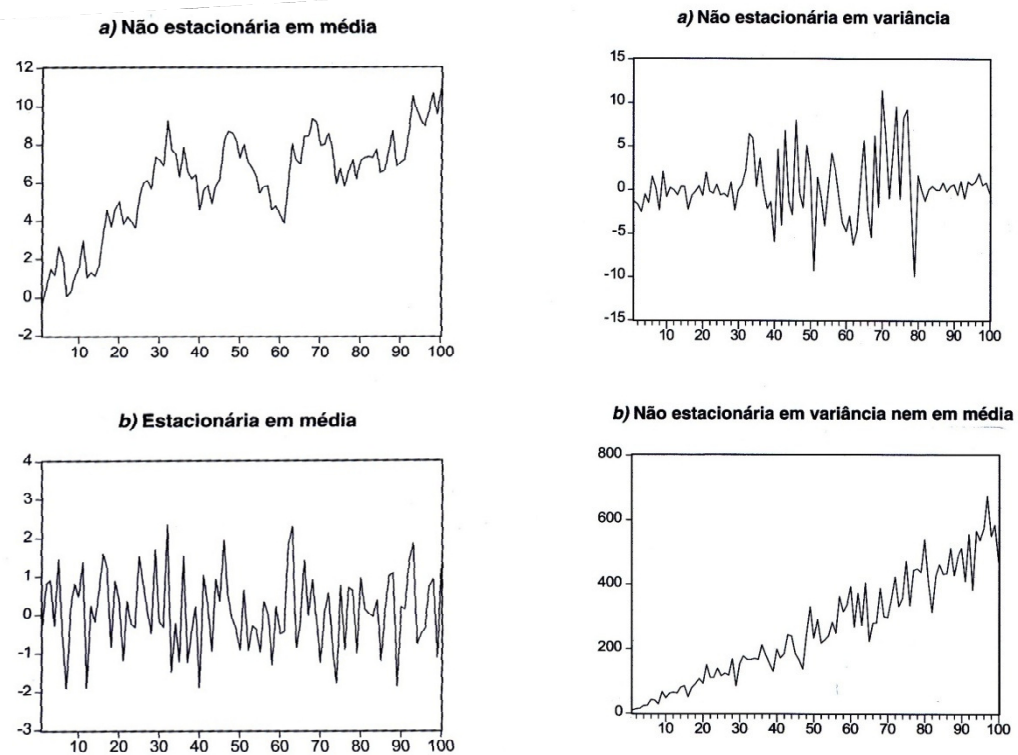


Figura 1: Séries temporais simuladas (Caiado, 2011).

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).

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 que o correlograma de uma série não estacionária apresenta um decaimento relativamente lento. Habitualmente,  $r_1$  é elevado e positivo para séries não estacionárias.

Para estabilizar a variância de uma série não estacionária em variância pode utilizar-se a família de transformações de Box-Cox:

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

onde  $\lambda$  é um parâmetro real, habitualmente entre -1 e 2. Com um pequeno ajuste, a transformação logarítmica de base  $e$  e as transformações potência estão incluídas na família de Box-Cox.

Para estabilizar a média de uma série não estacionária em média podem utilizar-se transformações de diferenciação. 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.

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:

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

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 são:

- Teste de Dickey-Fuller aumentado

Hipótese nula: a série não é estacionária.

Valores  $-p$  elevados sugerem não estacionaridade e valores  $-p$  pequenos sugerem estacionaridade.

Para um nível de significância de 0.05 a diferenciação é necessária se o valor  $-p > 0.05$ .

- Teste de Kwiatkowski-Phillips-Schmidt-Shin

Hipótese nula: a série é estacionária.

Valores  $-p$  elevados sugerem estacionaridade e valores  $-p$  pequenos sugerem não estacionaridade.

Para um nível de significância de 0.05 a diferenciação é necessária se o valor  $-p < 0.05$ .

Também existem Testes de raízes unitárias sazonais cujos detalhes podem ser consultados em Wei (2007).

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. O modelo AutoRegressivo de ordem  $p$ , ou modelo  $AR(p)$ , pode ser escrito em uma 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  $\mu$  é a média de  $y_t$  e  $c = \mu(1 - \phi_1 - \dots - \phi_p)$ ,  $\varepsilon_t$  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

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \\ y_t &= \mu + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t \end{aligned} \quad (57)$$

onde  $\mu$  é a média de  $y_t$ ,  $\varepsilon_t$  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

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \\ (1 - \phi_1 B - \dots - \phi_p B^p)(y_t - \mu) &= (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \end{aligned} \quad (58)$$

onde  $\mu$  é a média de  $y_t$  e  $c = \mu(1 - \phi_1 - \dots - \phi_p)$ ,  $\varepsilon_t$  é um ruído branco,  $p$  e  $q$  são inteiros não negativos,  $\phi_1, \dots, \phi_p$  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 estacionarizar a série;
- $c = \mu(1 - \phi_1 - \dots - \phi_p)$  e  $\mu$  é a média de  $(1 - B^d)y_t$ ;
- $\phi_1, \dots, \phi_p$  são os parâmetros autoregressivos;
- $\theta_1, \dots, \theta_q$  são os parâmetros de médias móveis;
- $\varepsilon_t$  é um ruído branco.

O modelo será estacionário e/ou invertível se os parâmetros AR e/ou MA satisfizerem as respetivas 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  $Q$  – ARIMA( $P, D, Q$ ) $_m$ , 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

$$\begin{aligned} \Phi_p(B^m) &= 1 - \Phi_1 B^m - \Phi_2 B^{2m} - \dots - \Phi_p B^{pm} \\ \Theta_Q(B^m) &= 1 + \Theta_1 B^m + \Theta_2 B^{2m} + \dots + \Theta_Q B^{Qm} \end{aligned}$$

onde,

- $P, D$  e  $Q$  são inteiros não negativos, sendo  $D$  o nº de diferenciações sazonais necessárias para estacionarizar a série;
- $m > 1$  é o período de sazonalidade;
- $c = \mu(1 - \Phi_1 - \dots - \Phi_p)$  e  $\mu$  é a média de  $(1 - B^m)^D y_t$ ;
- $\Phi_1, \dots, \Phi_p$  são os parâmetros autoregressivos;
- $\Theta_1, \dots, \Theta_Q$  são os parâmetros de médias móveis;
- $\varepsilon_t$  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 ARIMA( $P, D, 0$ ) $_m$ , 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

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, & \theta_q(B) &= 1 + \theta_1 B + \dots + \theta_q B^q \\ \Phi_p(B^m) &= 1 - \Phi_1 B^m - \dots - \Phi_p B^{Pm}, & \Theta_Q(B^m) &= 1 + \Theta_1 B^m + \dots + \Theta_Q B^{Qm} \end{aligned}$$

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

#### 2.4.6 Metodologia de Box-Jenkins

De seguida apresenta-se a metodologia de modelação ARIMA de séries temporais, introduzida por Box (1970, 1976) e desenvolvida mais tarde por Box (1994) e Wei (1990, 2007) entre outros, que consiste nas etapas seguintes:

1. Identificação,

2. Estimação,
3. Seleção de modelos,
4. Avaliação do diagnóstico, e
5. Previsão.

### 1. Identificação

A etapa de identificação consiste em identificar um modelo  $ARIMA(p, d, q) \times (P, D, Q)_m$  que descreva a relação existente entre as observações da série temporal (Schwartz, 1978). Esta etapa consiste em três passos na tentativa de identificação dos valores de  $p, d, q, P, D, Q$  e  $m$ .

- a. Construção do cronograma da série e sua estacionarização

O estudo de uma série temporal deve iniciar-se por uma análise do seu cronograma para se verificar a existência ou não de tendência, variância não constante, sazonalidade, *outliers*, e outros fenómenos não estacionários. Segue-se a sua estacionarização através das transformações de estabilização da variância (transformações de Box-Cox e ajustes de calendário), que deverão ser realizadas antes de qualquer transformação, eliminação da tendência (diferenciação não sazonal) e eliminação de movimentos de caráter periódico (diferenciação sazonal). Para a determinação de  $d$  e  $D$  poderão eventualmente utilizar-se Testes de raízes unitárias.

- b. Estimação das FAC e FACP da série original

O comportamento das FAC e FACP da série original pode sugerir algumas transformações de diferenciação. Como já foi referido, um decaimento lento para zero da FAC e uma queda brusca para zero a partir do *lag* 1 da FACP sugerem a aplicação de uma diferenciação simples.

- c. Estimação das FAC e FACP da série estacionária e identificação de  $p, q, P, e Q$

Neste último passo deve proceder-se à escolha de um modelo candidato a descrever o comportamento da série em estudo, isto é, à identificação dos valores de  $p, q, P, e Q$ , comparando o comportamento das FAC e FACP da série estacionária com o comportamento das FAC e FACP dos modelos teóricos.

### 2. Estimação

Uma vez identificados os modelos candidatos a descrever a série em estudo, segue-se a etapa de estimação dos respetivos parâmetros, isto é  $\phi_1, \dots, \phi_p, \Phi_1, \dots, \Phi_P, \theta_1, \dots, \theta_q, \Theta_1, \dots, \Theta_Q$  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. 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 aquelas 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.

### 3. Seleção de modelos

Identificados vários modelos candidatos é necessário escolher o melhor modelo. O critério de seleção pode ser baseado: (1) nos erros de previsão do conjunto de teste – tal como foi descrito na secção 2.2, ou (2) em critérios de informação.

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

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

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

onde,  $L$  é a verosimilhança 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  $\sigma^2$ , a variância dos resíduos).

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

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

- 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.

### 4. Avaliação do diagnóstico

Escolhido o melhor modelo  $ARIMA$  é necessário avaliar:

- A qualidade estatística das estimativas dos parâmetros.

Para o efeito dever-se-á fazer um teste parcial à hipótese nula de que cada parâmetro  $\beta_i$  é estritamente nulo, baseado na distribuição t- Student:

$$|t| = \left| \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \right| > t_{(T-K)} \Rightarrow \text{rejeitar } H_0 : \beta_i = 0 \quad (65)$$

onde  $K$  é o número de parâmetros do modelo de previsão. Devem incluir-se no modelo apenas os parâmetros considerados significativamente diferentes de zero.

- A qualidade do ajuste do modelo às observações da série em estudo.

Para o efeito dever-se-á:

1. Determinar a FAC dos resíduos, e
2. Fazer testes de *Portmanteau* aos resíduos (tal como foi descrito na secção 2.2).  
De acordo com os pressupostos assumidos, os resíduos deverão ter um comportamento idêntico a um ruído branco.

Se o modelo selecionado for adequado deve passar-se à etapa da previsão. Caso contrário tem de se iniciar novamente o processo.

## 5. Previsão

As previsões pontuais de modelos ARIMA são calculadas seguindo os passos seguintes:

1. Expande-se a equação do modelo de modo a que  $y_t$  fique isolado no 1º membro;
2. Reescreve-se a equação substituindo-se  $t$  por  $T + h$  ;
3. No 2º 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 tenham sido calculadas.

### **3. CASO DE ESTUDO**

#### **3.1 Empresa Foreva**

Fundada por Manuel Cajada, a marca Foreva nasceu em Setembro de 1984, abrindo a primeira loja na Rua Guerra Junqueira em Lisboa. Caracteriza-se desde sempre pela oferta de uma gama alargada de produtos para todas as estações do ano e estilos, no início com marca própria e marca do fornecedor e atualmente com exclusividade. A Foreva é essencialmente uma marca especializada em calçado, oferecendo no entanto também uma vasta gama de acessórios, garantindo qualidade aliada a um serviço de excelência ao melhor preço. Tem também à disposição dos clientes uma gama de conforto, da marca Softinos, e uma variedade de produtos para cuidar do calçado da marca Colónil, uma marca especialista em couros.

Desde o princípio a Foreva manteve uma estratégia de crescimento baseada em investimentos seletivos que permitiram assegurar hoje a sua presença nos principais Centros Comerciais do País, dirigida ao tipo de cliente que dá preferência à relação qualidade/preço, situando-se num segmento intermédio do mercado, ocupando, assim, uma posição vantajosa para acompanhar a crescente abertura de Centros Comerciais. Em 1998 a Foreva celebra o seu 1º contrato de Franchising, tendo atualmente já 25 lojas franchisadas. No território português conta com cerca de 70 lojas abertas ao público.

Em Maio de 2005, a Foreva foi adquirida pela empresa Fortunato Frederico & Cª Lda., empresa Líder do Grupo Kyaia, abrindo um novo ramo, o retalho. Em 2011 a Foreva altera a sua imagem de marca e inicia o processo de fusão de duas marcas do grupo transformando 25 das suas lojas Sapatália em Foreva. A marca efetua um esforço permanente de renovação da imagem dos seus espaços de venda, renovando constantemente as suas coleções, criando harmonia entre o espaço e o produto.

Tem agora uma política de preços estáveis, aliada a uma crescente qualidade dos produtos. Preocupa-se com o serviço ao cliente, fornecendo às equipas formação adequada, mantendo o contacto com o cliente constante e oferecendo meios e possibilidades de contacto com a empresa, quer pelo apoio ao cliente quer pela ajuda dos seus colaboradores.

#### **3.2 Análise exploratória dos dados**

Os dados analisados neste trabalho consistem em cinco séries temporais relativas às cinco principais categorias de calçado comercializadas pela Foreva que são:

- Bota,
- Botim,
- Sabrina,

- Sandália, e
- Sapato.

O calçado das categorias Sandália e Sabrina é comercializado essencialmente durante a estação de verão enquanto que o calçado das categorias Bota e Botim é comercializado essencialmente durante a estação de inverno. O calçado da categoria Sapato é comercializado durante todo o ano. A estação de verão tem início a 28 de fevereiro e termina a 29 de setembro de cada ano. A estação de inverno tem início a 30 de setembro de um ano e termina a 27 de fevereiro do ano seguinte. Dado que só a partir do ano de 2007 é que a Foreva estabilizou a sua rede de lojas optou-se por considerar as séries de vendas destas categorias de calçado a partir de 1 de janeiro de 2007. Para este estudo foram disponibilizados dados diários das vendas destas categorias até dezembro de 2011.

Desde logo foi solicitada pela Foreva uma previsão anual das vendas de cada uma das categorias de calçado para o ano seguinte, tendo a empresa como objetivo a utilização dessas previsões para a especificação do número de pares de calçado de cada uma das categorias a adquirir para comercialização. Assim, os dados das vendas diárias de cada uma das categorias de calçado foram agregados mensalmente para ser determinada uma previsão mensal com um horizonte temporal de um ano. Cada uma das séries de vendas de categorias de calçado foi assim constituída por dados mensais entre janeiro de 2007 e dezembro de 2011, perfazendo um total de 60 observações.

Como já foi referido, o primeiro passo na análise de uma série temporal consiste na sua representação gráfica. As Figuras 2 a 6 mostram o cronograma, a caixa de bigodes e um gráfico sazonal das séries em estudo.

À exceção da categoria Sabrina, todas as restantes séries apresentam uma forte sazonalidade anual e conseqüentemente não são estacionárias. A série da categoria Sabrina apresenta uma tendência variável e logo também não é estacionária.

Os gráficos da série da categoria Bota permitem observar um padrão sazonal anual. Fora da estação de inverno, isto é entre abril e agosto de cada ano as vendas são pouco significativas (apesar da estação de inverno só se iniciar a 30 de setembro a Foreva inicia a comercialização de alguns produtos um pouco antes do início da estação para motivar o arranque das vendas; e apesar da mesma estação terminar a 28 de fevereiro, as vendas dos últimos artigos ainda acontecem nos primeiros dias do mês de março). Durante a estação de inverno, as vendas apresentam uma forte tendência crescente entre setembro e dezembro do ano anterior, invertendo essa tendência a partir de janeiro do ano seguinte, que se mantém até março.

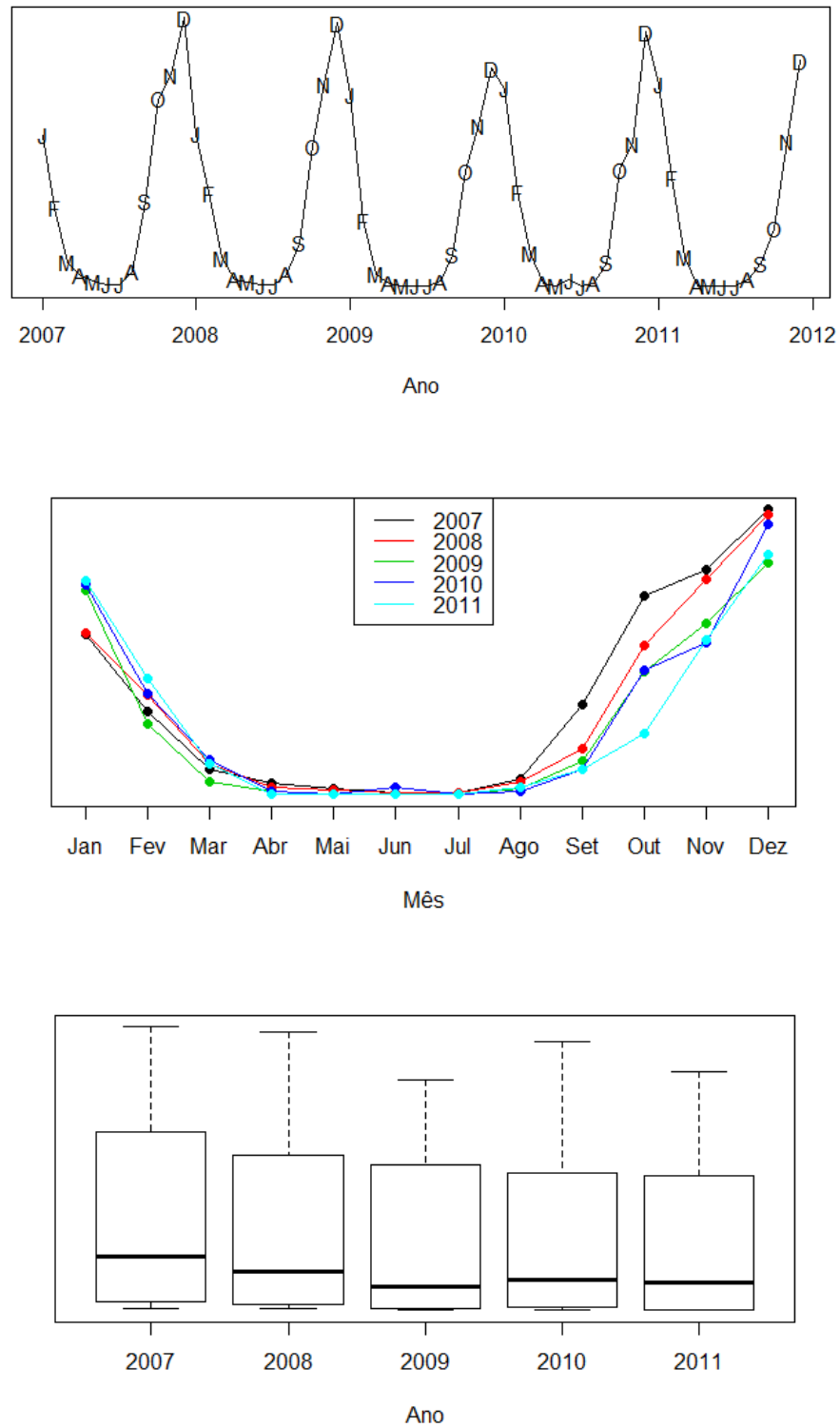


Figura 2: Gráficos da série de vendas da categoria Bota.

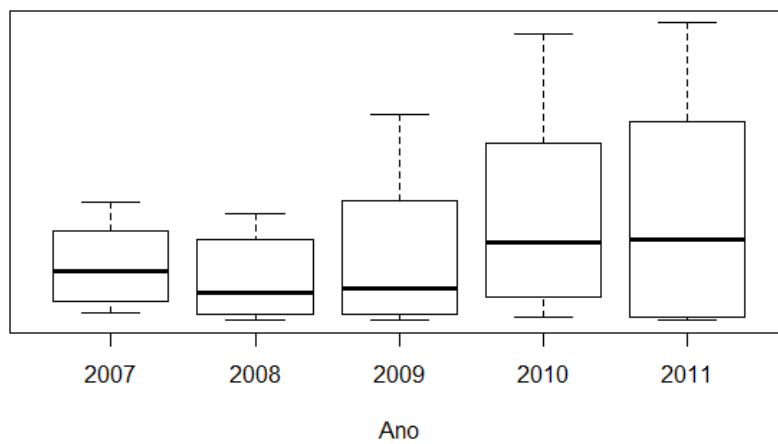
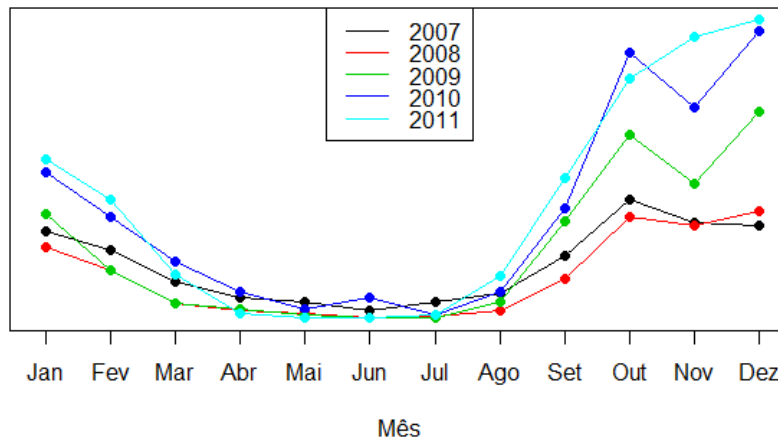
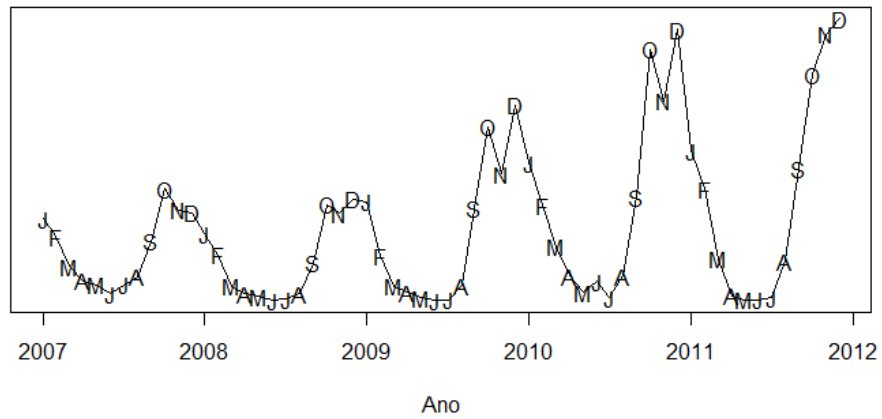


Figura 3: Gráficos da série de vendas da categoria Botim.

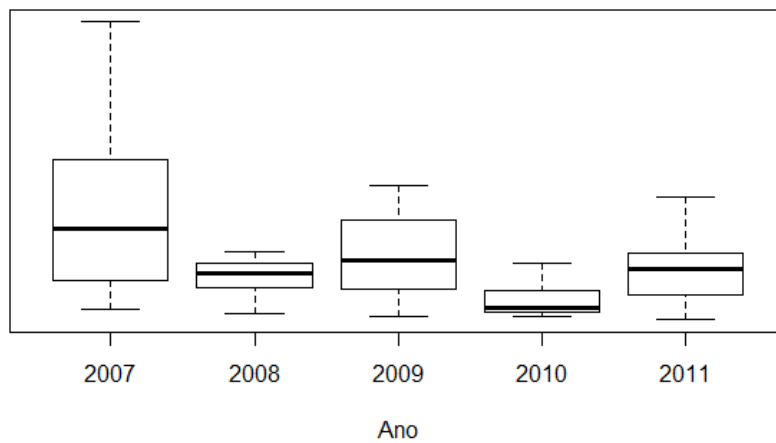
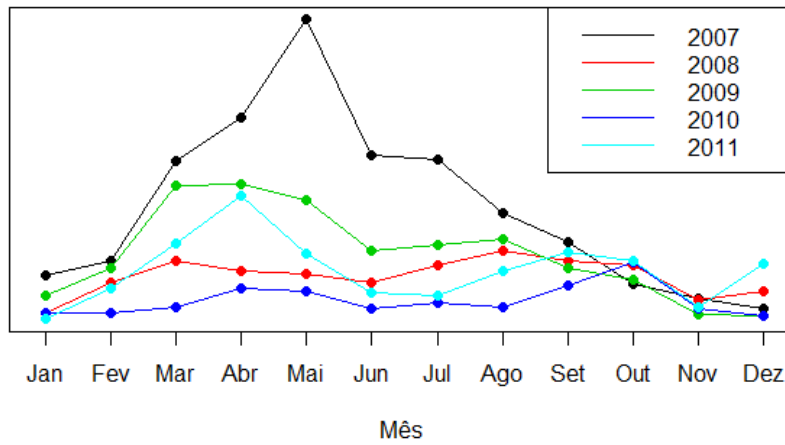
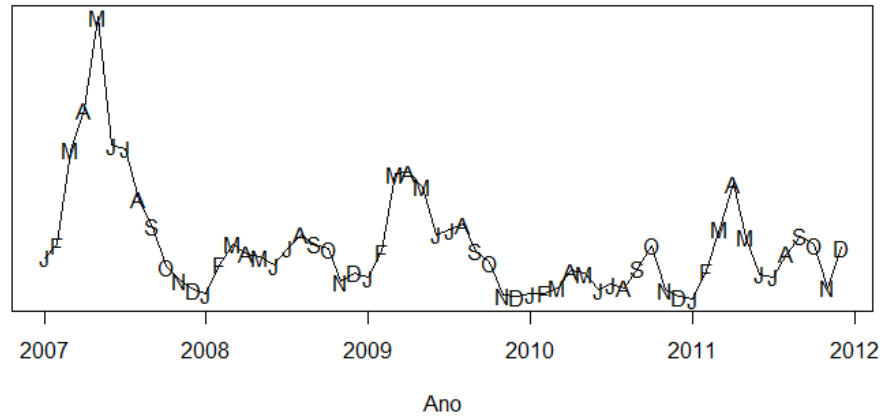


Figura 4: Gráficos da série de vendas da categoria Sabrina.

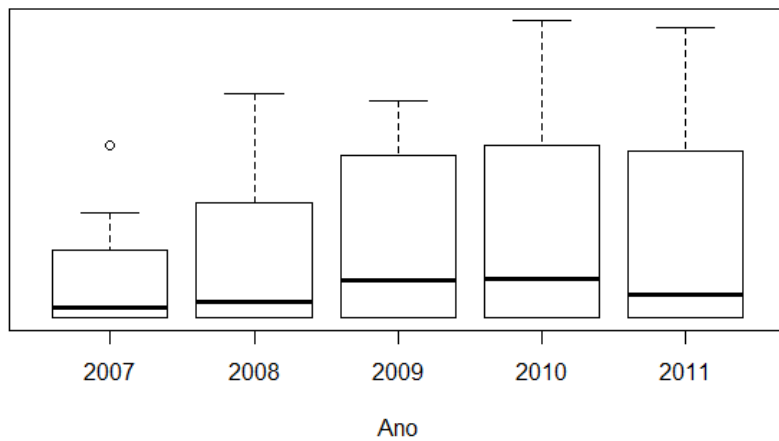
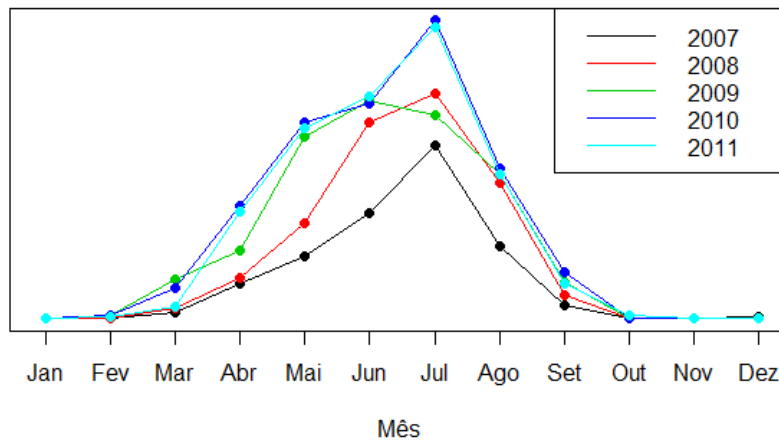
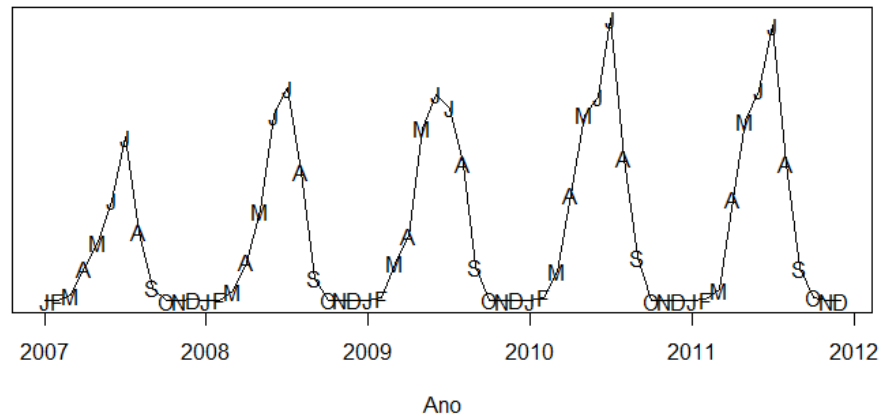


Figura 5: Gráficos da série de vendas da categoria Sandália.

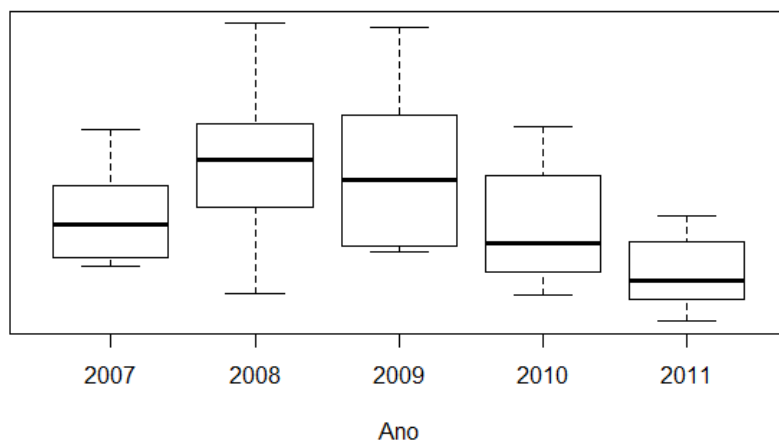
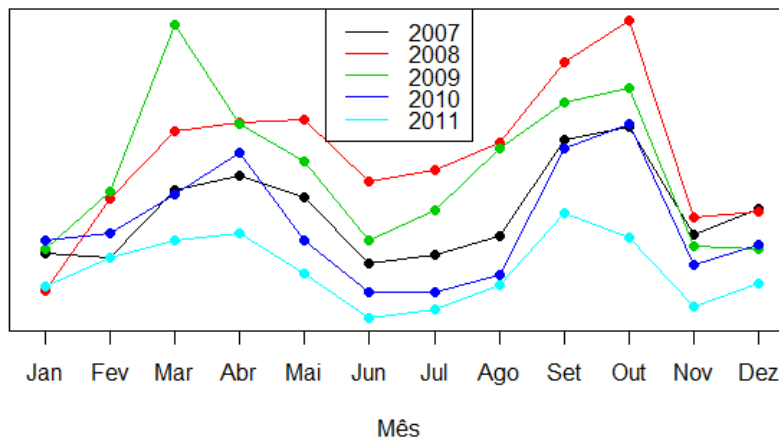
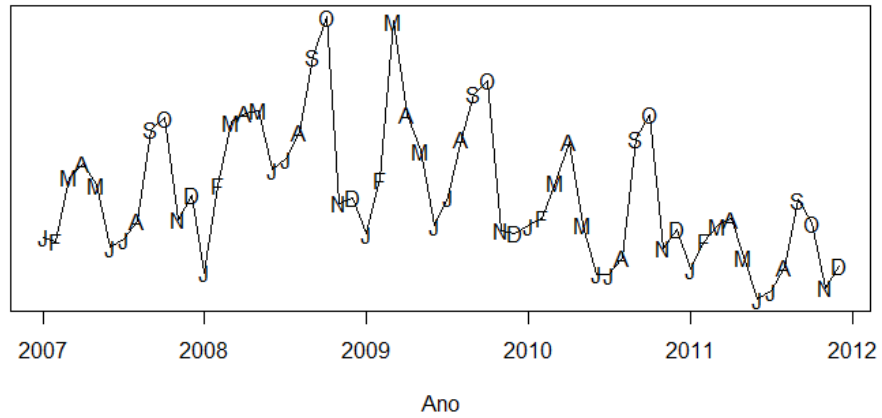


Figura 6: Gráficos da série de vendas da categoria Sapato.

O volume de vendas apresenta uma tendência de descida nos primeiros três anos, subindo ligeiramente no ano de 2010 e mantendo-se praticamente constante em 2011. Ao longo dos cinco anos, observa-se também uma tendência de descida nas vendas entre setembro e dezembro e uma ligeira tendência de subida entre janeiro e março.

Os gráficos da série da categoria Botim também apresentam um forte padrão sazonal anual. Tal como na série da categoria Bota, fora da estação de inverno, isto é entre abril e agosto as vendas são pouco significativas. Durante a estação de inverno, as vendas apresentam uma forte tendência crescente entre setembro e dezembro do ano anterior (quebrando ligeiramente em novembro), invertendo essa tendência a partir de janeiro do ano seguinte, tendência essa que se mantém até março. O volume de vendas apresenta uma ligeira tendência de descida em 2008 invertendo essa tendência a partir de 2009. Este crescimento das vendas a partir do ano de 2009 também é observado de um modo geral entre meses homólogos.

A série da categoria Sabrina apresenta uma grande volatilidade. O ano de 2007 é aquele que regista o maior volume de vendas e o ano de 2010 é aquele que regista o menor volume de vendas. Observa-se em cada ano ao longo do tempo uma alternância na tendência e globalmente o volume de vendas tem vindo a diminuir. Esta série não apresenta um padrão sazonal claro, ao contrário das restantes. É visível uma tendência de subida no início da estação de verão, sendo essa tendência invertida, a grosso modo, a partir de maio. Em alguns anos nomeadamente em 2008, 2010 e 2011 observa-se uma inversão da tendência de descida em julho, consequência provavelmente de um período de promoções mais agressivo para contrariar a quebra de vendas registada durante a estação.

Os gráficos da série da categoria Sandália permitem observar um forte padrão sazonal anual. Fora da estação de verão, isto é entre outubro do ano anterior e fevereiro do ano seguinte as vendas são pouco significativas. Durante a estação de verão, as vendas apresentam uma forte tendência crescente entre março e julho invertendo essa tendência apenas em agosto. O volume de vendas apresenta uma tendência de subida nos primeiros quatro anos e mantendo-se praticamente constante em 2011.

Os gráficos da série da categoria Sapato também permitem observar um forte padrão sazonal anual. Em cada ano, a série apresenta um movimento oscilatório com dois picos máximos em abril e outubro e dois picos mínimos em junho e novembro. O volume de vendas apresenta uma tendência de subida no ano de 2008 invertendo essa tendência em 2009, tendência essa que se tem mantido.

À exceção da série da categoria Sandália, nenhuma das outras apresenta outliers (ver as caixas de bigodes). Uma vez que as séries não são estacionárias em variância porventura será necessária uma transformação logarítmica para estabilizar a mesma.

## 4. MODELAÇÃO E PREVISÃO

Todo o trabalho de modelação e previsão foi desenvolvido usando o *software* livre estatístico R (<http://www.r-project.org/>). 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 7, que consiste num ambiente de programação mais amigável do que o ambiente de programação disponibilizado pelo R.

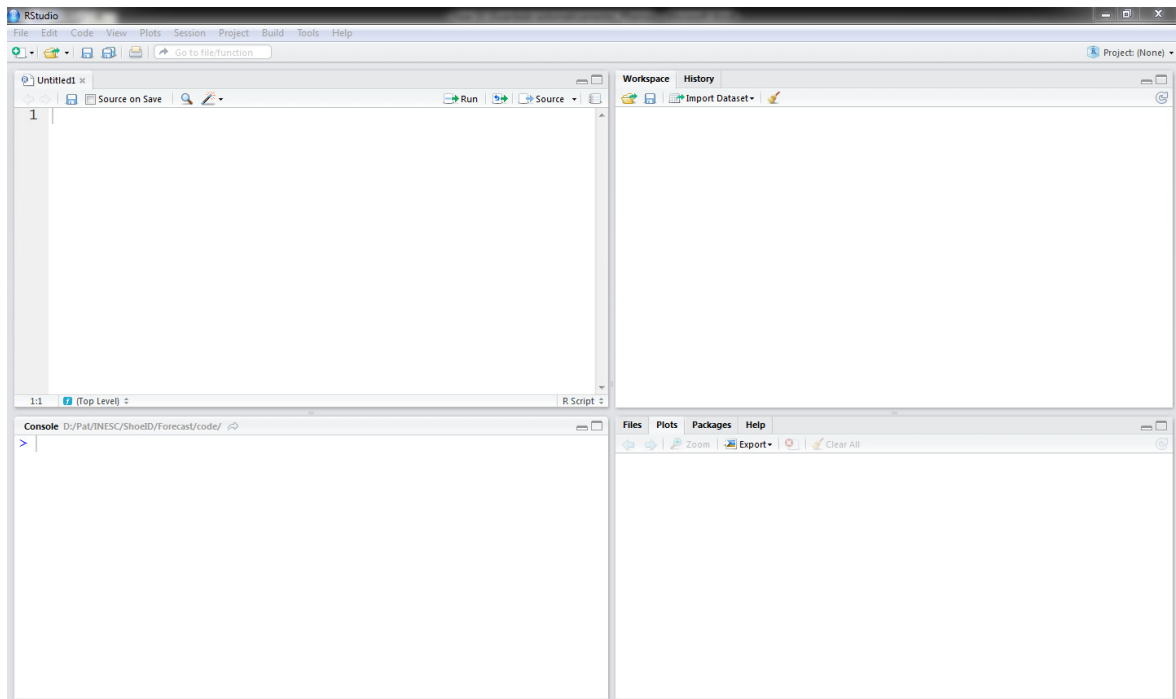


Figura 7: Ambiente de modelação e previsão.

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). Todas as séries foram analisadas usando a metodologia de alisamento exponencial e a metodologia ARIMA. Os resultados dessa modelação encontram-se nas secções que se seguem.

Tal como referido, 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, o conjunto de dados foi dividido em um conjunto de treino e um conjunto de teste. O conjunto de treino consistiu nas vendas mensais dos anos de 2007, 2008, 2009 e 2010 (48 observações); o conjunto de teste consistiu nas vendas mensais do ano de 2011 (12 observações).

Os dados do conjunto de treino foram utilizados para o processo de estimação e diagnóstico. Os dados do conjunto de teste foram utilizados para a seleção do modelo e avaliação do desempenho de previsão.

Para melhor comparar a performance de previsão das duas abordagens, alisamento exponencial e ARIMA, para cada uma das séries temporais foi decidido ajustar todos os modelos de espaço de estado possíveis e todos os modelos  $ARIMA(p, d, q) \times (P, D, Q)_m$  onde  $p$  e  $q$  variaram entre 0 e 5 e  $P$  e  $Q$  variaram entre 0 e 2. Estes valores máximos são genericamente aceites como admissíveis para a variação destes parâmetros.

No caso dos modelos de espaço de estado foram considerados dois tipos de dados:

- Dados *raw*, e
- Dados transformados logaritmicamente.

No caso dos modelos ARIMA foram considerados oito tipos de dados:

- Dados *raw*,
- Dados com diferenciação simples ( $d = 1$ ),
- Dados com diferenciação sazonal ( $D = 1$ ),
- Dados com diferenciação simples e sazonal ( $d = D = 1$ ),
- Dados transformados logaritmicamente,
- Dados transformados logaritmicamente com diferenciação simples ( $d = 1$ ),
- Dados transformados logaritmicamente com diferenciação sazonal ( $D = 1$ ), e
- Dados transformados logaritmicamente com diferenciação simples e sazonal ( $d = D = 1$ ).

Em cada abordagem, para cada tipo de dados, foi selecionado o modelo que:

- Apresentou o menor valor do parâmetro de avaliação de desempenho de previsão REQM para o conjunto de teste e
- Passou o teste estatístico de diagnóstico de resíduos de *Ljung-Box* com um nível de significância de 5%.

Todos os resultados de previsão relativos a dados pré-processados foram convertidos nas suas escalas originais antes de se calcularem as medidas de avaliação dos erros de previsão para que fosse possível a comparação, em relação ao desempenho do ajuste e desempenho de previsão, entre os diferentes modelos e estratégias de transformação dos dados.

## 4.1 Modelação e previsão por alisamento exponencial

### 4.1.1 Categoria Bota

Na Tabela 6 encontram-se as medidas de avaliação dos erros de previsão dos modelos ETS selecionados dentro (conjunto de treino) e fora da amostra (conjunto de teste) para a série de vendas da categoria Bota. Para cada tipo de dados (dados *raw* e dados transformados logaritmicamente) foi selecionado o modelo ETS que: (1) apresentou o menor valor da REQM para o conjunto de teste e que (2) passou o teste de *Ljung-Box* com um nível de significância de 5%. As medidas de avaliação apresentadas na Tabela 6 encontram-se definidas na Secção 2.2.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
ETS(M,A <sub>d</sub> ,A)	-74.11	1024.02	678.32	-8.39%	98.34%	0.29	0.12
Log ETS(A,A,A)	-216.27	1878.67	1081.70	-12.12%	36.51%	0.46	0.16

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ETS (M,A <sub>d</sub> ,A)	-26.15	1682.70	1132.33	35.99%	69.22%
Log ETS (A,A,A)	699.82	1829.88	1140.54	-69.37%	98.20%

Tabela 6: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Bota.

Como se pode observar na Tabela 6, para os dados *raw* foi selecionado o modelo de espaço de estado com erros multiplicativos, tendência amortecida aditiva e sazonalidade aditiva – ETS(M,A<sub>d</sub>,A). Para os dados transformados logaritmicamente foi selecionado o modelo de espaço de estado com erros aditivos, tendência aditiva e sazonalidade aditiva – ETS(A,A,A). Atente-se que ambos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. De entre os dois modelos o que apresenta melhor desempenho de previsão é o ETS(M,A<sub>d</sub>,A), consequentemente este é o modelo ETS selecionado para a previsão da série de vendas da categoria Bota. Acrescente-se que, para além da REQM, todas as restantes medidas de avaliação dos erros de previsão fora da amostra apresentam o menor valor no modelo selecionado, isto é, no modelo ETS(M,A<sub>d</sub>,A). Excetuando o EPAM, todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam o menor valor no modelo selecionado. Deve referir-se que o valor do EPAM fora da amostra, 69.22% é bastante elevado. Repare-se que o modelo ETS(M,A<sub>d</sub>,A) está de acordo com a análise exploratória da série apresentada na Secção 3.2. A Figura 8 mostra o cronograma, a FAC e FACP dos

resíduos do modelo ETS(M,A<sub>d</sub>,A). O cronograma mostra que os resíduos apresentam um comportamento semelhante a um ruído branco, como é esperado. A FAC e a FACP mostram que não existe correlação entre os resíduos, como é desejável.

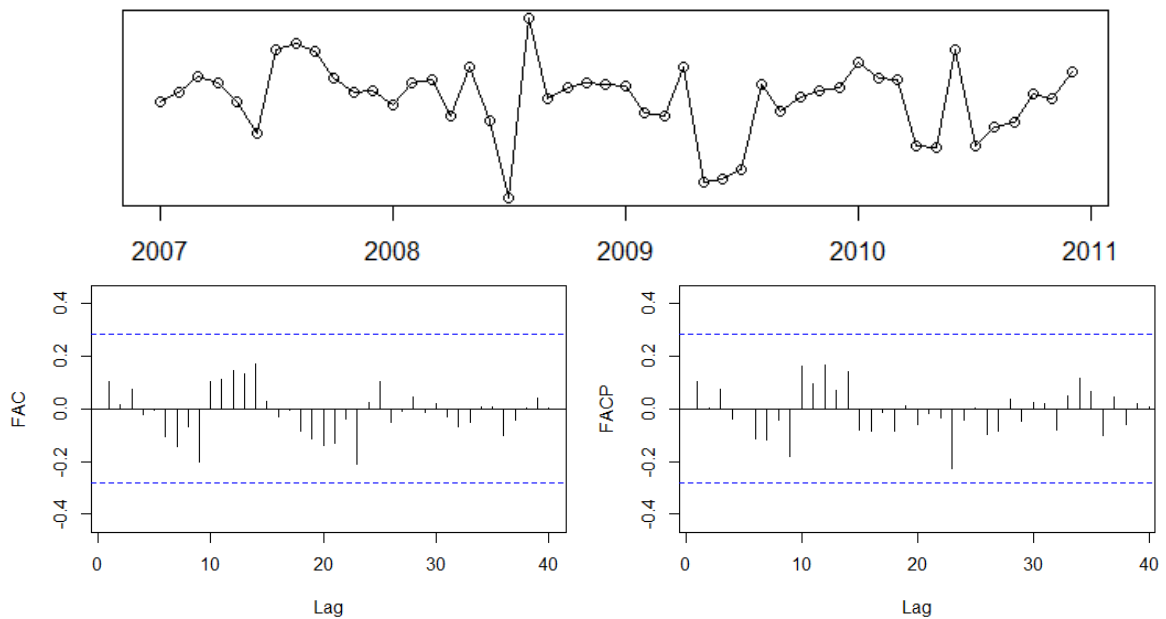


Figura 8: Cronograma, FAC e FACP dos resíduos do modelo ETS(M,A<sub>d</sub>,A) para a Bota.

A Tabela 7 mostra medidas de avaliação dos erros de previsão do modelo ETS(M,A<sub>d</sub>,A) apenas para os meses da estação de inverno do ano de 2011. O erro de previsão foi calculado de acordo com a Equação 9 e o erro percentual foi calculado de acordo com a Equação 11, ambas apresentadas na secção 2.2. O erro percentual mensal varia entre -2.8% (mês de dezembro) e -122.5% (mês de outubro). Pode afirmar-se que os erros percentuais mensais são relativamente elevados.

Mês de 2011	Erro de previsão	Erro percentual
Janeiro	2564	22.0%
Fevereiro	2263	36.0%
Março	1237	73.2%
Setembro	-728	-53.1%
Outubro	-4096	-122.5%
Novembro	-1768	-21.1%
Dezembro	-360	-2.8%
	EAM	EPAM
Estações (Jan-Mar;Set-Dez)	1859.28	47.23%

Tabela 7: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Bota.

Selecioneado o modelo ETS(M,A<sub>d</sub>,A) para a caracterização da série de vendas da categoria Bota, este foi novamente estimado utilizando agora todos os dados históricos (60 observações). No quadro seguinte encontram-se as estimativas dos parâmetros de alisamento desse modelo ETS(M,A<sub>d</sub>,A) e as estimativas dos valores iniciais utilizados. Saliente-se o valor bastante elevado do parâmetro de amortecimento, 0.98.

	$\alpha$	$\beta$	$\gamma$	$\phi$			
Parâmetros de alisamento	0.66	0.10	0.04	0.98			
Estados iniciais	$l_0$	$b_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
	4703.09	522.73	9895.47	5770.48	3487.84	-1932.43	-3907.18
	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
	-4468.68	-4012.74	-4191.05	-4182.77	-3032.23	670.70	5902.60

Na Figura 9 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Bota, (2) dos valores ajustados pelo modelo ETS(M,A<sub>d</sub>,A), e (3) das previsões de vendas da categoria Bota para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo ETS aos dados históricos e o modo como este captou a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se no aumento rápido da amplitude dos intervalos de previsão à medida que o horizonte temporal da previsão aumenta.

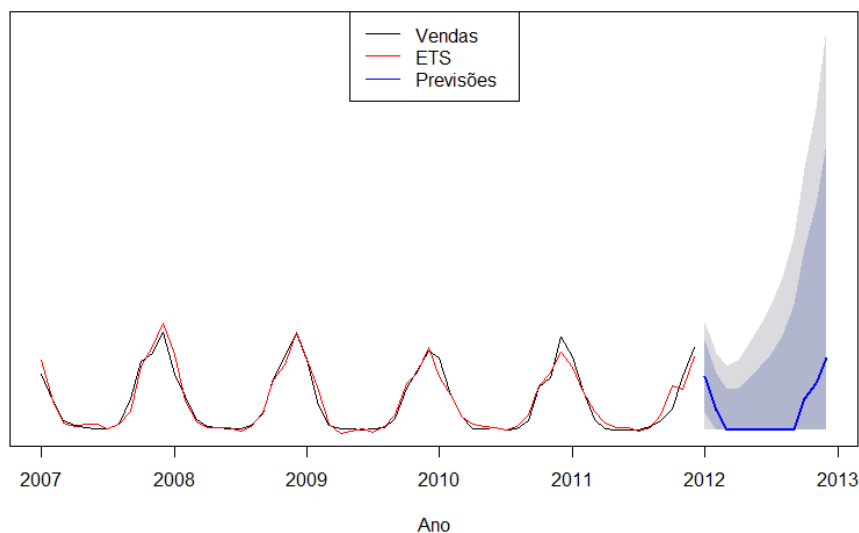


Figura 9: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Bota.

### 4.1.2 Categoria Botim

Na Tabela 8 encontram-se as medidas de avaliação dos erros de previsão dos modelos ETS selecionados dentro e fora da amostra para a série de vendas da categoria Botim. Como se pode observar na Tabela 8, para os dados *raw* foi selecionado o modelo de espaço de estado com erros multiplicativos, tendência amortecida aditiva e sazonalidade multiplicativa – ETS(M,A<sub>d</sub>,M). Para os dados transformados logaritmicamente foi selecionado o modelo de espaço de estado com erros aditivos, sem tendência e sazonalidade aditiva – ETS(A,N,A). Atente-se que o modelo ETS(M,A<sub>d</sub>,M) passou o teste de *Ljung-Box* mas o modelo ETS(A,N,A) não passou este teste. Aliás nenhum dos modelos ETS estimados para os dados transformados logaritmicamente passou esse teste, tendo sido selecionado, de entre todos os modelos, aquele que apresentou o menor valor da REQM para o conjunto de teste. Consequentemente foi selecionado o modelo o ETS(M,A<sub>d</sub>,M) para a previsão da série de vendas da categoria Botim, não só porque apresenta melhor desempenho de previsão mas também porque passou o teste de *Ljung-Box*. Acrescente-se que, para além da REQM, todas as restantes medidas de avaliação dos erros de previsão fora da amostra, à exceção do EM, apresentam o menor valor no modelo selecionado. Excetuando o EM, todas as restantes medidas de avaliação dos erros de previsão dentro da amostra apresentam o maior valor no modelo selecionado. Esta situação é frequente, isto é, nem sempre todas as medidas de avaliação dos erros de previsão dentro ou fora da amostra apresentam o menor valor no modelo selecionado. Observe-se o EPAM fora da amostra da série do Botim, 119.12% é superior ao EPAM fora da amostra da série da Bota, 69.22%. Repare-se que o modelo ETS(M,A<sub>d</sub>,M) está de acordo com a análise exploratória da série apresentada na Secção 3.2.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
ETS(M,A <sub>d</sub> ,M)	4.93	427.68	279.69	-25.13%	49.01%	0.55	0.06
Log ETS(A,N,A)	12.52	426.23	258.40	-8.46%	33.95%	0.51	0.04*

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ETS(M,A <sub>d</sub> ,M)	-52.59	309.32	226.79	-110.87%	119.12%
Log ETS(A,N,A)	31.47	390.23	293.29	-114.62%	127.04%

Tabela 8: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Botim.

A Figura 10 mostra o cronograma, a FAC e FACP dos resíduos do modelo ETS(M,A<sub>d</sub>,M). O cronograma mostra que os resíduos apresentam um comportamento semelhante a um ruído branco, como é esperado. A FAC e a FACP mostram que não existe correlação significativa entre os resíduos, como é desejável. Observa-se que o pico no lag 24 da FACP está praticamente no limite do intervalo não comprometendo o diagnóstico.

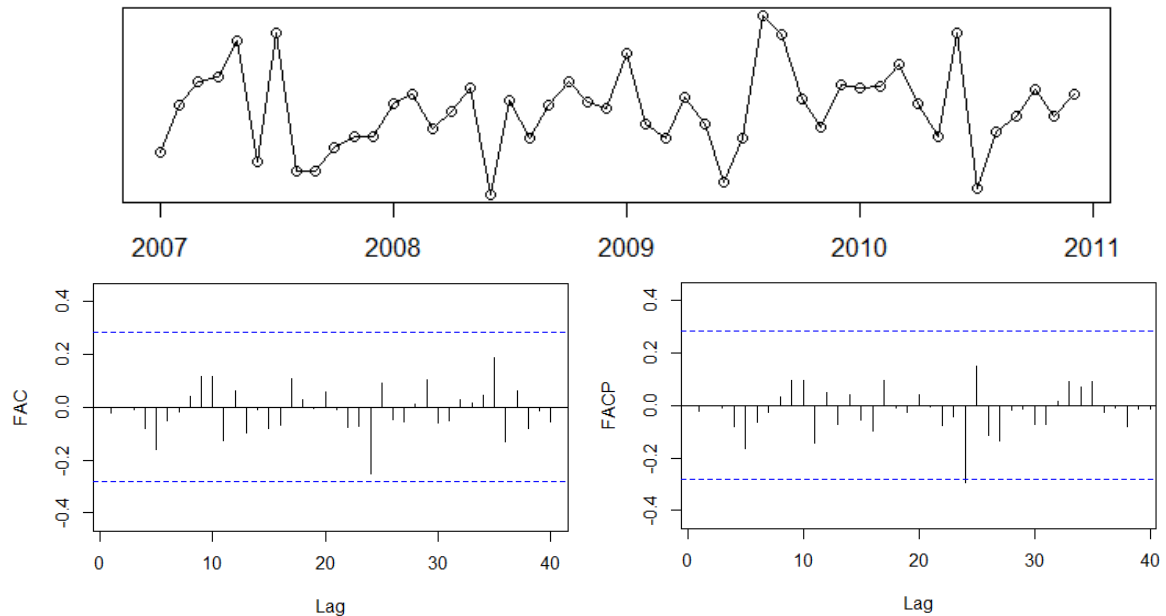


Figura 10: Cronograma, FAC e FACP dos resíduos do modelo ETS(M,A<sub>d</sub>,M) para o Botim. A Tabela 9 mostra medidas de avaliação dos erros de previsão do modelo ETS(M,A<sub>d</sub>,M) apenas para os meses da estação de inverno do ano de 2011. O erro percentual mensal varia entre 2.8% (mês de dezembro) e -21.3% (mês de outubro). Observe-se que os erros percentuais mensais da categoria Botim são bastante mais baixos do que os erros percentuais mensais da categoria Bota.

Mês de 2011	Erro de previsão	Erro percentual
Janeiro	-103	-4.0%
Fevereiro	345	17.8%
Março	72	9.8%
Setembro	-160	-7.0%
Outubro	-833	-21.3%
Novembro	422	9.2%
Dezembro	137	2.8%
	EAM	EPAM
Estações (Jan-Mar;Set-Dez)	295.97	10.27%

Tabela 9: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Botim.

Atente-se que o EPAM relativo aos meses da estação de inverno da série do Botim, 10.27% é bastante inferior ao EPAM relativo aos meses da estação de inverno da série da Bota, 47.23%. No quadro seguinte encontram-se as estimativas dos parâmetros de alisamento do modelo ETS ( $M, A_d, M$ ), estimado utilizando agora todos os dados históricos, e as estimativas dos valores iniciais utilizados. Saliente-se o valor relativamente elevado do parâmetro de amortecimento, 0.88.

	$\alpha$	$\beta$	$\gamma$	$\phi$				
Parâmetros de alisamento	0.84	1.1e-3	3e-04	0.88				
Estados iniciais	$l_0$	$b_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	
	1802.12	-59.34	2.14	1.98	2.61	1.50	0.56	
	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$	
	0.10	0.07	0.10	0.19	0.37	0.89	1.50	

Na Figura 11 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Botim, (2) dos valores ajustados pelo modelo ETS( $M, A_d, M$ ), e (3) das previsões de vendas da categoria Botim para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo ETS aos dados históricos e o modo como este captou a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se no aumento rápido da amplitude dos intervalos de previsão à medida que o horizonte temporal da previsão aumenta.

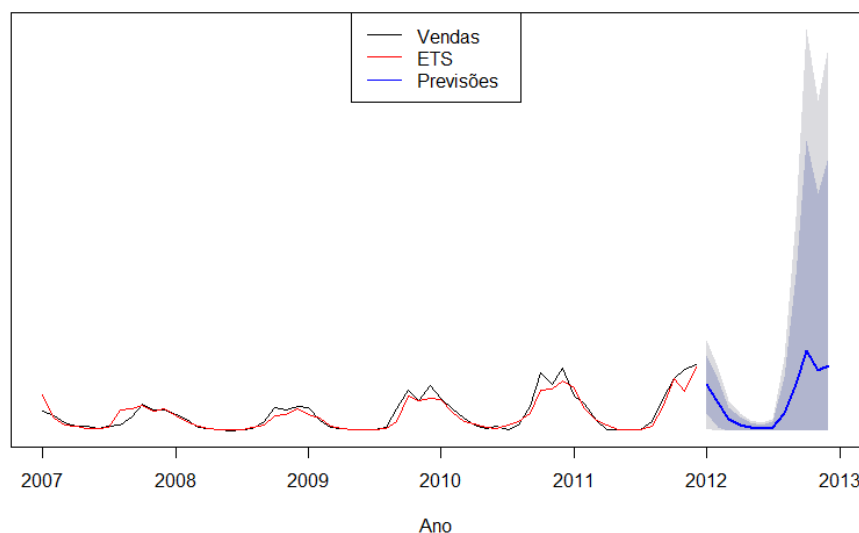


Figura 11: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Botim.

### 4.1.3 Categoria Sabrina

Na Tabela 10 encontram-se as medidas de avaliação dos erros de previsão dos modelos ETS selecionados dentro e fora da amostra para a série de vendas da categoria Sabrina. Como se pode observar na Tabela 10, para os dados *raw* foi selecionado o modelo de espaço de estado com erros multiplicativos, sem tendência e sazonalidade aditiva – ETS(M,N,A). Para os dados transformados logaritmicamente foi selecionado o modelo de espaço de estado com erros aditivos, sem tendência e sazonalidade aditiva – ETS(A,N,A). Atente-se que ambos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. De entre os dois modelos o que apresenta melhor desempenho de previsão é o ETS(M,N,A), conseqüentemente este é o modelo ETS selecionado para a previsão da série de vendas da categoria Sabrina. Acrescente-se que, para além da REQM, todas as restantes medidas de avaliação dos erros de previsão fora da amostra, à exceção do EPAM, apresentam o menor valor no modelo selecionado. Todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam o maior valor no modelo selecionado. Observe-se o EPAM fora da amostra da série da Sabrina, 38.25% é inferior ao EPAM fora da amostra da série do Botim, 119.12% e inferior ao EPAM fora da amostra da série da Bota, 69.22%. Repare-se que o modelo ETS(M,N,A) não está de acordo com a análise exploratória da série apresentada na Secção 3.2.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
ETS(M,N,A)	-29.86	380.08	263.18	-11.98%	34.90%	0.94	0.54
Log ETS(A,N,A)	-25.65	296.25	196.67	-7.04%	23.40%	0.70	0.08

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ETS(M,N,A)	51.71	304.31	224.55	-5.64%	38.25%
Log ETS(A,N,A)	242.76	388.32	285.52	22.40%	34.88%

Tabela 10: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Sabrina.

A Figura 12 mostra o cronograma, a FAC e FACP dos resíduos do modelo ETS(M,N,A). O cronograma mostra que os resíduos apresentam um comportamento relativamente semelhante a um ruído branco. Apesar dos valores dos resíduos relativos aos meses setembro e outubro

de 2010 serem um pouco mais elevados, este facto não compromete o diagnóstico já que a FAC e a FACP não contêm correlações significativas.

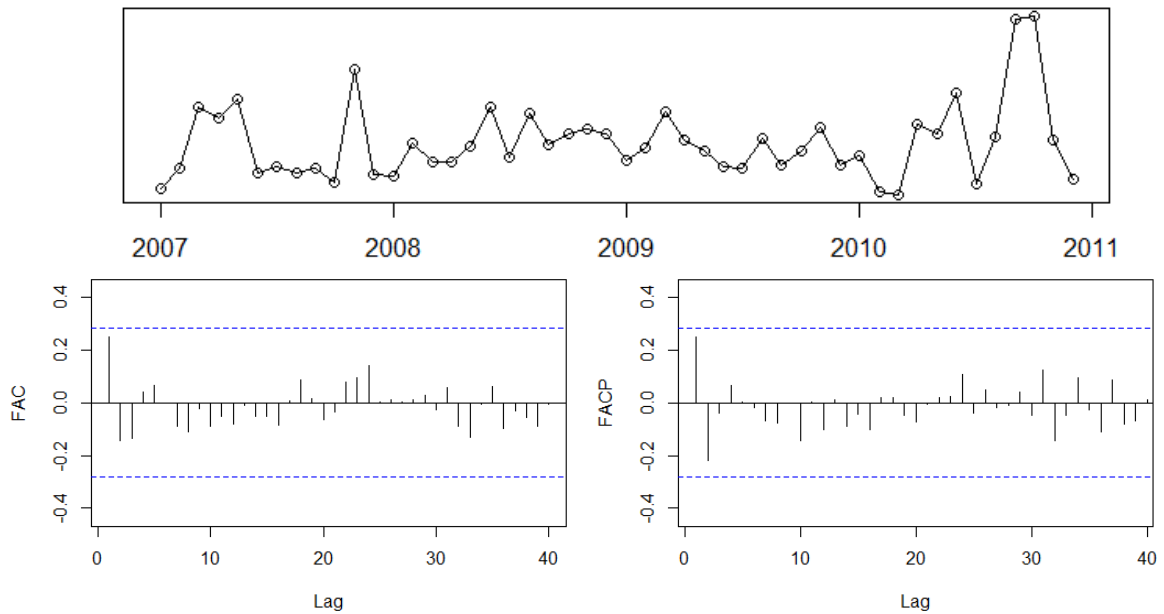


Figura 12: Cronograma, FAC e FACP dos resíduos do modelo ETS(M,N,A) para a Sabrina.

A Tabela 11 mostra medidas de avaliação dos erros de previsão do modelo ETS(M,N,A) apenas para os meses da estação de verão do ano de 2011. O erro percentual mensal varia entre -0.7% (mês de maio) e -126.0% (mês de julho).

Mês de 2011	Erro de previsão	Erro percentual
Março	14	1.4%
Abril	574	36.4%
Maio	-6	-0.7%
Junho	-167	-35.9%
Julho	-553	-126.0%
Agosto	-144	-20.1%
Setembro	161	17.2%
	EAM	EPAM
Estação (Mar-Set)	231.3	33.95%

Tabela 11: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sabrina.

No quadro seguinte encontram-se as estimativas dos parâmetros de alisamento do modelo ETS (M,N,A), estimado utilizando agora todos os dados históricos, e as estimativas dos valores iniciais utilizados.

Na Figura 13 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Sabrina, (2) dos valores ajustados pelo modelo ETS(M,N,A), e (3) das previsões de vendas da categoria Sabrina para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo ETS aos dados históricos e o modo como este captou a tendência. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se no aumento rápido da amplitude dos intervalos de previsão à medida que o horizonte temporal da previsão aumenta.

	$\alpha$	$\gamma$						
Parâmetros de alisamento	0.86	1e-04						
Estados iniciais	$l_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	
	1751.07	-453.50	-475.29	-76.06	-6.69	75.95	154.24	
	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$		
	25.11	314.56	546.41	372.05	-27.42	-449.36		

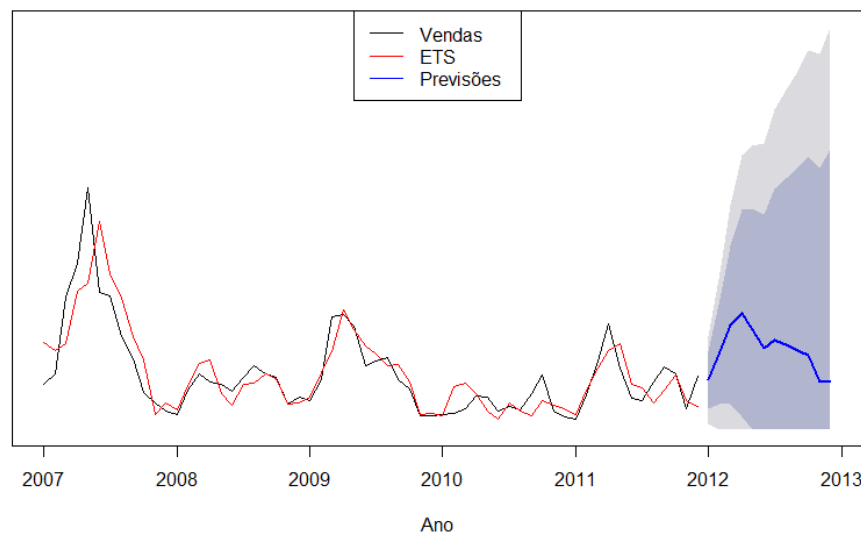


Figura 13: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Sabrina.

#### 4.1.4 Categoria Sandália

Na Tabela 12 encontram-se as medidas de avaliação dos erros de previsão dos modelos ETS selecionados dentro e fora da amostra para a série de vendas da categoria Sandália. Como se pode observar na Tabela 12, para os dados *raw* foi selecionado o modelo de espaço de estado com erros aditivos, tendência aditiva e sazonalidade aditiva – ETS(A,A,A). Para os dados transformados logaritmicamente foi selecionado o modelo de espaço de estado com erros aditivos, sem tendência e sazonalidade aditiva – ETS(A,N,A). Atente-se que o modelo

ETS(A,N,A) passou o teste de *Ljung-Box* mas o modelo ETS(A,A,A) não passou este teste. Aliás nenhum dos modelos ETS estimados para os dados *raw* passou esse teste, tendo sido selecionado, de entre todos os modelos, aquele que apresentou o menor valor da REQM para o conjunto de teste. Apesar de não ter passado o teste de *Ljung-Box* decidiu-se selecionar o modelo o ETS(A,A,A) para a previsão da série de vendas da categoria Sandália porque este apresenta muito melhor desempenho de previsão, com valores da REQM e do EAM consideravelmente inferiores aos do seu rival. Acrescente-se que a REQM, o EAM e o EM fora da amostra apresentam o menor valor no modelo selecionado não acontecendo o mesmo em relação ao EPM e ao EPAM. Excetuando o EM e a REQM, todas as restantes medidas de avaliação dos erros de previsão dentro da amostra apresentam o maior valor no modelo selecionado. Como já foi referido esta situação é frequente. Observe-se o EPAM fora da amostra da série da Sandália, 12889.20% é muito superior ao EPAM fora da amostra da série da Sabrina, 38.25%, ao EPAM fora da amostra da série do Botim, 119.12% e ao EPAM fora da amostra da série da Bota, 69.22%. Repare-se que o modelo ETS(A,A,A) está de acordo com a análise exploratória da série apresentada na Secção 3.2.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
ETS(A,A,A)	-103.63	1802.00	1346.83	466.94%	2638.46%	0.46	0.02*
Log ETS(A,N,A)	-206.10	1938.61	972.40	-22.41%	47.81%	0.33	0.18

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ETS(A,A,A)	505.78	2015.27	1486.89	-12841.70%	12889.20%
Log ETS(A,N,A)	5272.15	8202.03	5272.47	52.36%	84.72%

Tabela 12: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Sandália.

A Figura 14 mostra o cronograma, a FAC e FACP dos resíduos do modelo ETS(A,A,A). Atente-se que apesar deste modelo não ter passado o teste de *Ljung-Box*, o cronograma mostra que os resíduos apresentam um comportamento semelhante a um ruído branco, como é desejável. A FAC e a FACP mostram que não existe correlação significativa entre os resíduos. Observa-se que o pico no *lag* 11 da FAC está apenas um pouco fora do limite do intervalo não comprometendo de modo algum o diagnóstico. A Tabela 13 mostra medidas de avaliação dos erros de previsão do modelo ETS(A,A,A) apenas para os meses da estação de verão do ano de

2011. O erro percentual mensal varia entre 5.0% (mês de agosto) e -107.3% (mês de março). Atente-se que o EPAM relativo aos meses da estação de verão da série da Sabrina, 33.95% é equivalente ao EPAM relativo aos meses da estação de verão da série da Sandália, 32.03%.

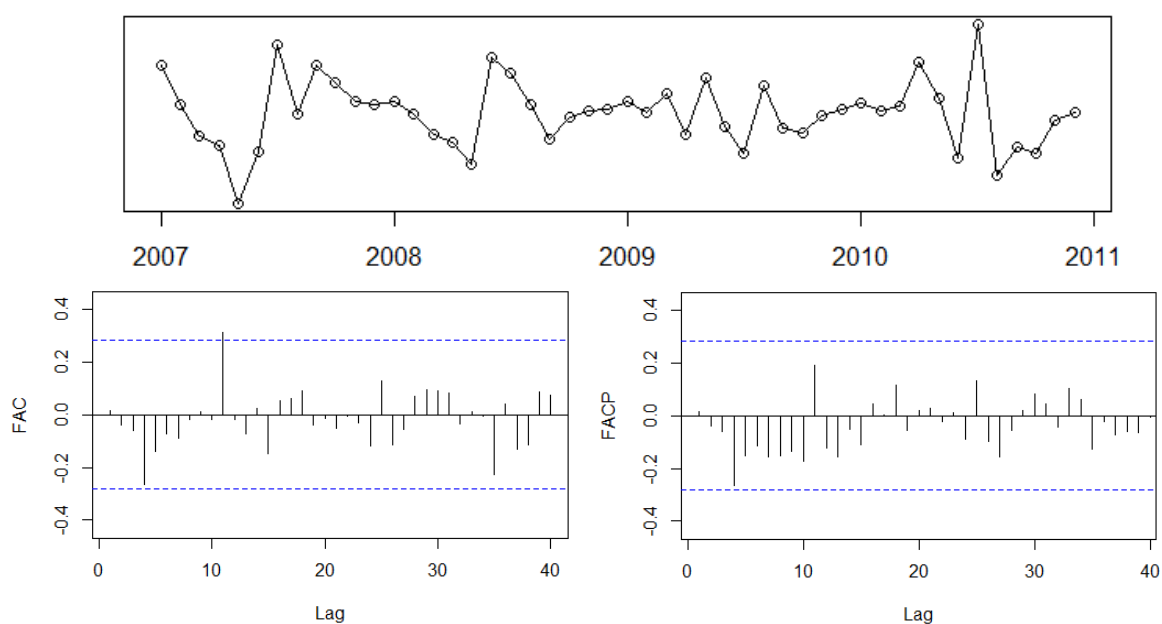


Figura 14: Cronograma, FAC e FACP dos resíduos do modelo ETS(A,A,A) para a Sandália.

Mês de 2011	Erro de previsão	Erro percentual
Março	-907	-107.3%
Abril	2722	33.7%
Mai	2496	17.5%
Junho	907	5.4%
Julho	5166	23.6%
Agosto	548	5.0%
Setembro	-820	-31.6%
	EAM	EPAM
Estação (Mar-Set)	1937.9	32.03%

Tabela 13: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sandália.

No quadro seguinte encontram-se as estimativas dos parâmetros de alisamento do modelo ETS (A,A,A), estimado utilizando agora todos os dados históricos, e as estimativas dos valores iniciais utilizados.

Na Figura 15 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Sandália, (2) dos valores ajustados pelo modelo ETS(A,A,A), e (3) das previsões de vendas da categoria Sandália para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo ETS aos dados históricos e o modo

como este captou a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se no aumento rápido da amplitude dos intervalos de previsão à medida que o horizonte temporal da previsão aumenta.

	$\alpha$	$\beta$	$\gamma$				
Parâmetros de alisamento	0.76	1e-04	1e-04				
Estados iniciais	$l_0$	$b_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
	3366.39	119.11	-5363.09	-5202.35	-5029.43	-2746.81	4483.87
	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
	12018.74	10188.84	6646.10	345.35	-4085.03	-5583.82	-5672.39

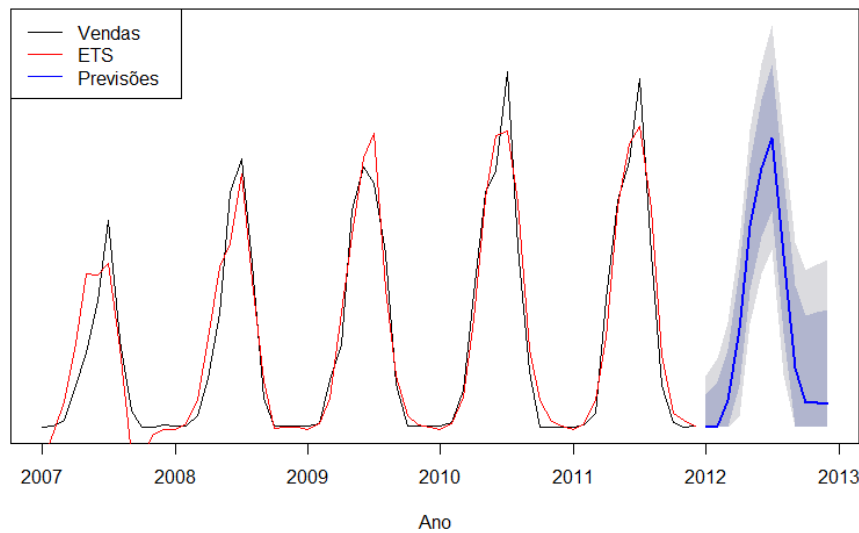


Figura 15: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Sandália.

#### 4.1.5 Categoria Sapato

Na Tabela 14 encontram-se as medidas de avaliação dos erros de previsão dos modelos ETS selecionados dentro e fora da amostra para a série de vendas da categoria Sapato. Como se pode observar na Tabela 14, para os dados *raw* foi selecionado o modelo de espaço de estado com erros multiplicativos, sem tendência e sazonalidade multiplicativa – ETS(M,N,M). Para os dados transformados logaritmicamente foi selecionado o modelo de espaço de estado com erros aditivos, tendência aditiva e sazonalidade aditiva – ETS(A,A,A). Atente-se que nenhum dos modelos passou o teste de *Ljung-Box*, significando isto que nenhum dos modelos ETS estimados, quer para os dados *raw* quer para os dados transformados logaritmicamente, passou esse teste, tendo sido selecionado o modelo que apresentou o menor valor da REQM para o conjunto de teste. Sendo assim, selecionou-se o modelo ETS(A,A,A) para a previsão da série de vendas da categoria Sapato, visto que de entre os dois modelos é o que apresenta melhor

desempenho de previsão. Saliente-se que na previsão com modelação de espaço de estado, apenas na série desta categoria o modelo selecionado corresponde a dados transformados logaritmicamente. Acrescente-se que, para além da REQM, todas as restantes medidas de avaliação dos erros de previsão fora da amostra apresentam o menor valor no modelo selecionado. Excetuando o EM e a REQM, todas as restantes medidas de avaliação dos erros de previsão dentro da amostra apresentam o menor valor no modelo selecionado. Observe-se o EPAM fora da amostra da série do Sapato, 45.59% é superior ao EPAM fora da amostra da série da Sabrina, 38.25%, mas é inferior ao EPAM fora da amostra da série da Bota, 69.22%, ao EPAM fora da amostra da série do Botim, 119.12% e ao EPAM fora da amostra da série da Sandália, 12889.20%. Repare-se que o modelo ETS(A,A,A) está de acordo com a análise exploratória da série apresentada na Secção 3.2.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
ETS(M,N,M)	-2.61	726.96	589.89	-1.45%	12.04%	0.45	0.01*
Log ETS(A,A,A)	12.2	732.39	574.23	-1.03%	11.63%	0.44	0*

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ETS(M,N,M)	-1350.68	1483.71	1350.68	-45.59%	45.59%
Log ETS(A,A,A)	-1213.34	1357.60	1213.34	-41.13%	41.13%

Tabela 14: Medidas de avaliação dos erros de previsão dos modelos ETS dentro e fora da amostra para a série de vendas da categoria Sapato.

A Figura 16 mostra o cronograma, a FAC e FACP dos resíduos do modelo ETS(A,A,A). Atente-se que este modelo não passou o teste de *Ljung-Box*, e apesar do cronograma mostrar um comportamento semelhante a um ruído branco, a FAC e a FACP mostram correlações significativas no *lag* 24 e a FACP ainda também no *lag* 16. Contudo, este foi de entre todos os modelos estimados, aquele que apresentou melhor desempenho global.

A Tabela 15 mostra medidas de avaliação dos erros de previsão do modelo ETS(A,A,A) para os meses do ano de 2011. O erro percentual mensal varia entre -5.9% (mês de janeiro) e -65.9% (mês de junho).

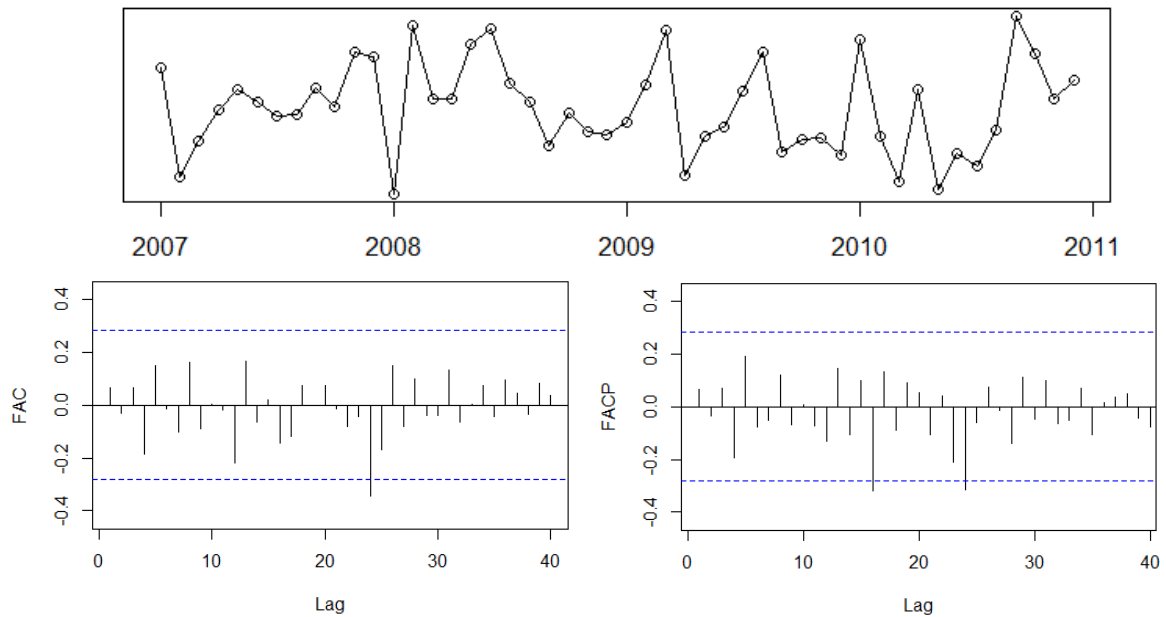


Figura 16: Cronograma, FAC e FACP dos resíduos do modelo ETS(A,A,A) para o Sapato.

Mês de 2011	Erro de previsão	Erro percentual
Janeiro	-157	-5.9%
Fevereiro	-207	-6.1%
Março	-1559	-41.2%
Abril	-1510	-38.1%
Maio	-1639	-55.1%
Junho	-1227	-65.9%
Julho	-1346	-65.0%
Agosto	-1227	-45.3%
Setembro	-1426	-31.9%
Outubro	-2485	-64.3%
Novembro	-1037	-47.7%
Dezembro	-740	-27.0%

Tabela 15: Medidas de avaliação dos erros de previsão dos meses de 2011 da categoria Sapato.

No quadro seguinte encontram-se as estimativas dos parâmetros de alisamento do modelo ETS (A,A,A), estimado utilizando agora todos os dados históricos, e as estimativas dos valores iniciais utilizados.

	$\alpha$	$\beta$	$\gamma$				
Parâmetros de alisamento	0.56	1e-04	3e-04				
Estados iniciais	$l_0$	$b_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$
	8.40	5.9e-03	-0.10	-0.23	0.45	0.40	-0.04
Estados iniciais	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
	-0.23	-0.34	0.05	0.24	0.24	-0.08	-0.36

Na Figura 17 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Sapato, (2) dos valores ajustados pelo modelo ETS(A,A,A), e (3) das previsões de vendas da categoria Sapato para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo ETS aos dados históricos e o modo como este captou razoavelmente a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se no aumento da amplitude dos intervalos de previsão à medida que o horizonte temporal da previsão aumenta.

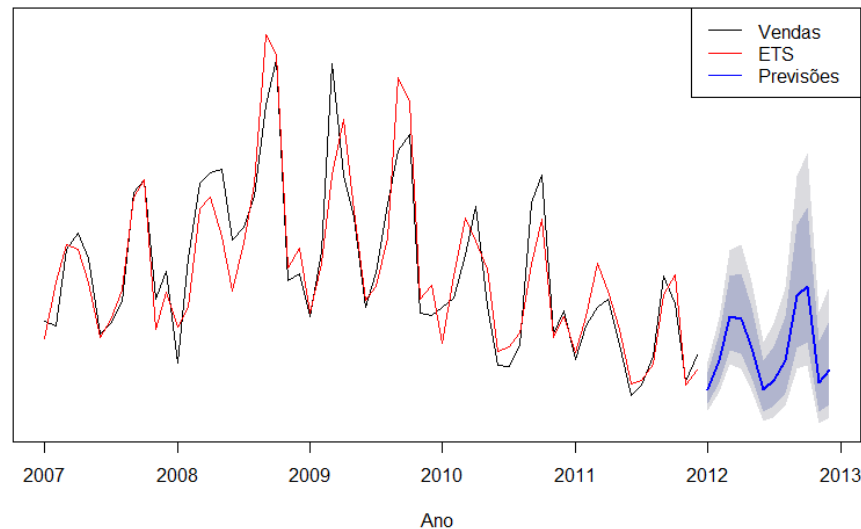


Figura 17: Previsões de vendas do modelo ETS para o ano de 2012 da categoria Sapato.

## 4.2 Modelação e previsão ARIMA

### 4.2.1 Categoria Bota

Na Tabela 16 encontram-se as medidas de avaliação dos erros de previsão dos modelos ARIMA selecionados dentro (conjunto de treino) e fora da amostra (conjunto de teste) para a série de vendas da categoria Bota. Para cada tipo de dados (dados *raw*, dados com diferenciação simples, dados com diferenciação sazonal, dados com diferenciação simples e sazonal, dados transformados logaritmicamente, dados transformados logaritmicamente com diferenciação simples, dados transformados logaritmicamente com diferenciação sazonal, e dados transformados logaritmicamente com diferenciação simples e sazonal) foi selecionado o modelo ARIMA que: (1) apresentou o menor valor da REQM para o conjunto de teste e que (2) passou o teste de *Ljung-Box* com um nível de significância de 5%. As medidas de avaliação apresentadas na Tabela 16 encontram-se definidas na Secção 2.2. Os modelos encontram-se ordenadas pelo valor da REQM para o conjunto de teste, do melhor para o pior desempenho. Como se pode observar na Tabela 16, o modelo que apresentou melhor desempenho de

previsão foi o modelo ARIMA  $(2,1,5) \times (1,0,1)_{12}$  para os dados transformados logaritmicamente (com diferenciação simples), consequentemente este é o modelo ARIMA selecionado para a previsão da série de vendas da categoria Bota. Atente-se que todos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. Excetuando a REQM, todas as medidas de avaliação dos erros de previsão fora da amostra apresentam um valor superior no modelo selecionado, o que é habitual, como já foi referido. Todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam um valor superior no modelo selecionado.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
Log ARIMA $(2,1,5) \times (1,0,1)_{12}$	-29.11	1215.36	759.72	-16.62%	36.24%	0.32	0.88
ARIMA $(2,1,2) \times (2,1,2)_{12}$	-0.83	740.59	453.79	20.11%	48.55%	0.19	0.83
Log ARIMA $(5,1,1) \times (2,1,2)_{12}$	397.11	1061.82	608.77	-7.43%	25.08%	0.26	1.00
ARIMA $(3,1,5) \times (2,0,0)_{12}$	-109.81	1019.78	753.30	-108.89%	208.67%	0.32	0.42
ARIMA $(0,0,0) \times (1,1,2)_{12}$	-146.82	951.62	556.38	-20.86%	33.86%	0.24	0.75
Log ARIMA $(5,0,1) \times (0,0,1)_{12}$	-121.87	2546.64	1366.86	-23.87%	53.21%	0.58	0.73
ARIMA $(5,0,1) \times (1,0,2)_{12}$	-116.51	1051.82	815.35	-115.17%	183.10%	0.35	0.26
Log ARIMA $(1,0,2) \times (1,1,2)_{12}$	232.78	960.50	561.49	-17.17%	31.85%	0.24	1.00

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
Log ARIMA $(2,1,5) \times (1,0,1)_{12}$	-266.93	753.73	485.84	-109.48%	124.30%
ARIMA $(2,1,2) \times (2,1,2)_{12}$	25.34	813.66	588.14	-34.72%	119.98%
Log ARIMA $(5,1,1) \times (2,1,2)_{12}$	55.26	831.47	494.91	-130.97%	150.39%
ARIMA $(3,1,5) \times (2,0,0)_{12}$	-212.22	871.07	364.68	31.96%	48.93%
ARIMA $(0,0,0) \times (1,1,2)_{12}$	-83.02	904.68	592.33	-140.80%	193.84%
Log ARIMA $(5,0,1) \times (0,0,1)_{12}$	-80.57	927.83	686.99	-266.51%	282.46%
ARIMA $(5,0,1) \times (1,0,2)_{12}$	-424.04	937.69	460.82	-169.43%	187.25%
Log ARIMA $(1,0,2) \times (1,1,2)_{12}$	-351.60	1101.60	590.00	-191.98%	202.96%

Tabela 16: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Bota.

Repare-se que o modelo ARIMA  $(2,1,5) \times (1,0,1)_{12}$  com dados transformados logaritmicamente está de acordo com a análise exploratória da série apresentada na Secção 3.2. A Figura 18 mostra o cronograma, a FAC e FACP dos resíduos do modelo ARIMA  $(2,1,5) \times (1,0,1)_{12}$ . O cronograma mostra que os resíduos apresentam um comportamento relativamente semelhante a um ruído branco. Apesar dos valores dos resíduos relativos aos meses junho e julho de 2010 serem um pouco mais elevados, este facto não compromete o diagnóstico já que a FAC e a FACP não contêm correlações significativas.

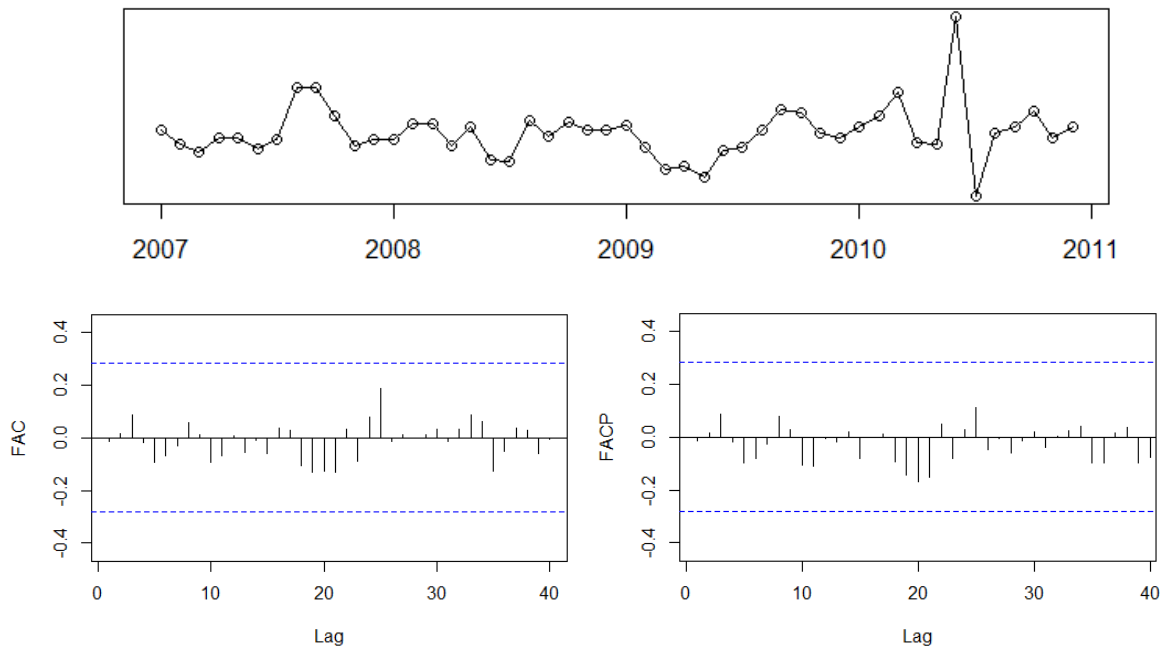


Figura 18: Cronograma, FAC e FACP dos resíduos do modelo ARIMA  $(2,1,5) \times (1,0,1)_{12}$  para a Bota.

Mês de 2011	Erro de previsão	Erro percentual
Janeiro	-833	-7.1%
Fevereiro	746	11.9%
Março	15	0.9%
Setembro	327	23.9%
Outubro	-1298	-38.8%
Novembro	-67	-0.8%
Dezembro	-1913	-14.7%
	EAM	EPAM
Estações (Jan-Mar;Set-Dez)	742.68	14.00%

Tabela 17: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Bota.

A Tabela 17 mostra medidas de avaliação dos erros de previsão do modelo  $ARIMA(2,1,5) \times (1,0,1)_{12}$  apenas para os meses da estação de inverno do ano de 2011. O erro de previsão foi calculado de acordo com a Equação 9 e o erro percentual foi calculado de acordo com a Equação 11, ambas apresentadas na secção 2.2. O erro percentual mensal varia entre -0.8% (mês de novembro) e -38.8% (mês de outubro).

Selecionado o modelo  $ARIMA(2,1,5) \times (1,0,1)_{12}$  para a caracterização da série de vendas da categoria Bota, este foi novamente estimado utilizando agora todos os dados históricos (60 observações) e a sua equação é:

$$\begin{aligned} (1 - 1.7327B + 0.9955B^2)(1 - 0.9368B^{12})(1 - B)\log(Y_t) = \\ = (1 - 2.4485B + 2.0479B^2 - 0.4024B^3 - 0.2003B^4 + 0.0338B^5)(1 - 0.6974B^{12})\epsilon_t \end{aligned}$$

Na Figura 19 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Bota, (2) dos valores ajustados pelo modelo  $ARIMA(2,1,5) \times (1,0,1)_{12}$ , e (3) das previsões de vendas da categoria Bota para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo  $ARIMA$  aos dados históricos e o modo como este captou a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se que a amplitude dos intervalos de previsão aumenta à medida que o horizonte temporal da previsão aumenta, mas é inferior fora dos meses da estação de inverno.

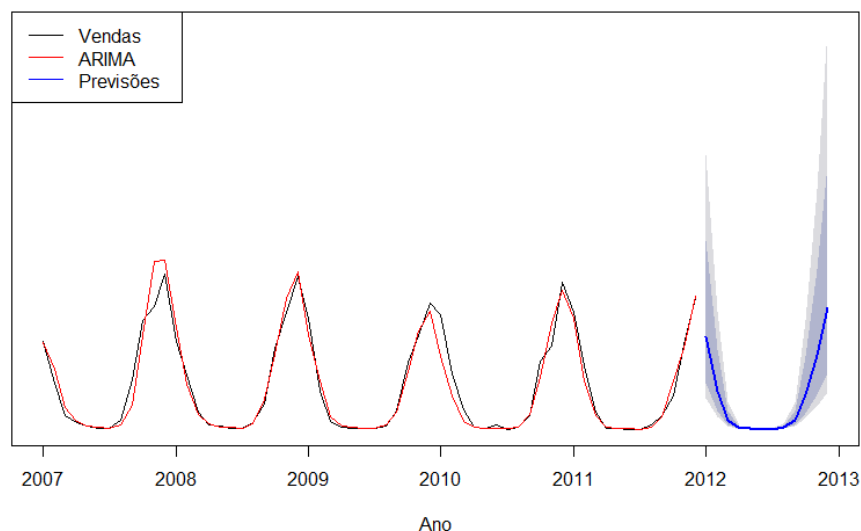


Figura 19: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Bota.

### 4.2.2 Categoria Botim

Como se pode observar na Tabela 18, o modelo que apresentou melhor desempenho de previsão foi o modelo ARIMA  $(5,0,2) \times (0,1,0)_{12}$  para os dados transformados logaritmicamente (com diferenciação sazonal), consequentemente este é o modelo ARIMA selecionado para a previsão da série de vendas da categoria Botim. Atente-se que todos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. Excetuando o EPM e o EPAM, todas as medidas de avaliação dos erros de previsão fora da amostra apresentam um valor inferior no modelo selecionado.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						<i>Ljung-Box</i> Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
Log ARIMA $(5,0,2) \times (0,1,0)_{12}$	40.83	391.67	226.98	-7.55%	31.04%	0.45	0.96
ARIMA $(0,0,2) \times (0,1,1)_{12}$	92.24	335.42	199.86	-14.14%	59.88%	0.39	0.66
ARIMA $(5,1,3) \times (0,0,0)_{12}$	75.28	477.43	320.28	-66.00%	123.16%	0.63	0.11
ARIMA $(4,0,2) \times (1,0,1)_{12}$	28.67	324.86	223.31	-32.59%	60.20%	0.44	0.76
Log ARIMA $(1,1,2) \times (0,1,2)_{12}$	78.67	312.30	188.95	-2.69%	25.49%	0.37	0.97
Log ARIMA $(0,1,4) \times (1,0,0)_{12}$	145.75	468.71	275.92	-11.86%	42.44%	0.54	0.49
ARIMA $(5,1,5) \times (0,1,0)_{12}$	50.45	282.34	172.59	6.70%	50.38%	0.34	1.00
Log ARIMA $(2,0,0) \times (1,0,0)_{12}$	148.38	407.21	266.00	-15.90%	44.27%	0.52	0.54

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
Log ARIMA $(5,0,2) \times (0,1,0)_{12}$	19.56	347.12	278.54	-186.83%	199.31%
ARIMA $(0,0,2) \times (0,1,1)_{12}$	-79.62	428.60	346.18	-224.30%	236.83%
ARIMA $(5,1,3) \times (0,0,0)_{12}$	-154.55	468.94	354.82	-47.31%	84.67%
ARIMA $(4,0,2) \times (1,0,1)_{12}$	-247.45	502.34	413.81	-183.29%	189.30%
Log ARIMA $(1,1,2) \times (0,1,2)_{12}$	-33.94	513.47	434.53	-269.93%	288.93%

Log ARIMA $(0,1,4) \times (1,0,0)_{12}$	-356.20	527.33	454.85	-411.30%	413.44%
ARIMA $(5,1,5) \times (0,1,0)_{12}$	-638.95	769.33	663.80	-497.33%	497.87%
Log ARIMA $(2,0,0) \times (1,0,0)_{12}$	280.87	898.87	656.33	-256.78%	283.44%

Tabela 18: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Botim.

Todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam um valor superior no modelo seleccionado.

Repare-se que o modelo ARIMA  $(5,0,2) \times (0,1,0)_{12}$  com dados transformados logaritmicamente está de acordo com a análise exploratória da série apresentada na Secção 3.2. A Figura 20 mostra o cronograma, a FAC e FACP dos resíduos do modelo ARIMA  $(5,0,2) \times (0,1,0)_{12}$ . O cronograma mostra que os resíduos apresentam um comportamento relativamente semelhante a um ruído branco. Apesar dos valores dos resíduos relativos aos meses junho e julho de 2010 serem um pouco mais elevados, este facto não compromete o diagnóstico já que a FAC e a FACP não contêm correlações significativas.

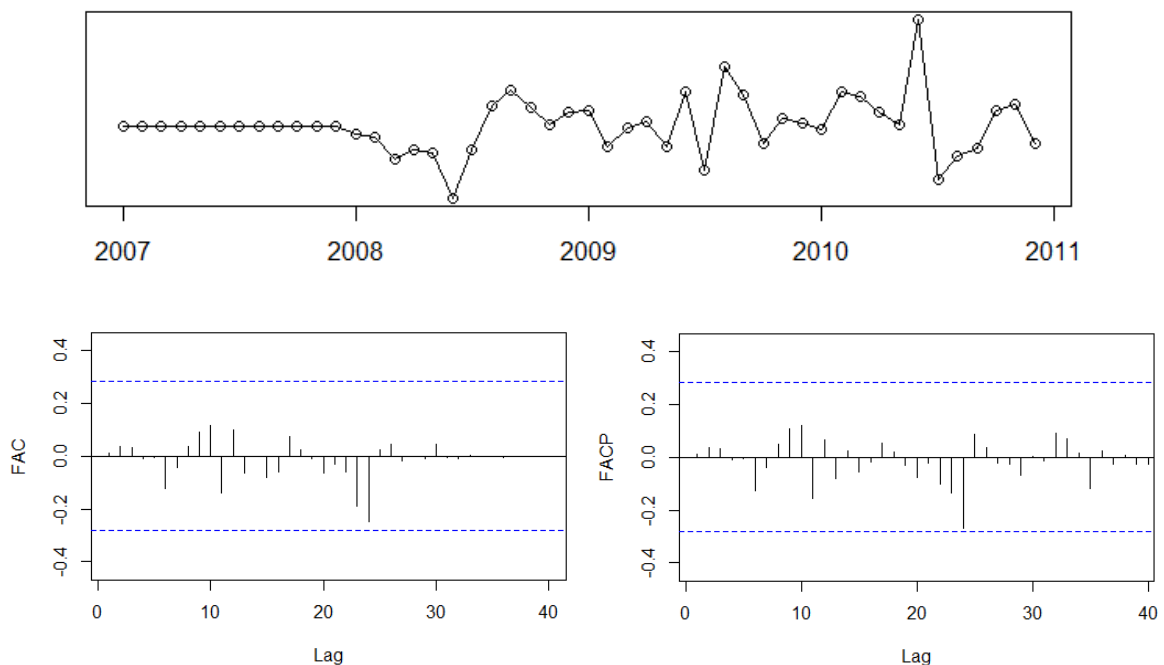


Figura 20: Cronograma, FAC e FACP dos resíduos do modelo ARIMA  $(5,0,2) \times (0,1,0)_{12}$  para o Botim.

A Tabela 19 mostra medidas de avaliação dos erros de previsão do modelo  $ARIMA(5,0,2) \times (0,1,0)_{12}$  apenas para os meses da estação de inverno do ano de 2011. O erro percentual mensal varia entre 0.1% (mês de janeiro) e -63.2% (mês de março).

Atente-se que o EPAM relativo aos meses da estação de inverno da série do Botim, 16.52% é ligeiramente superior ao EPAM relativo aos meses da estação de inverno da série da Bota, 14.00%.

Mês de 2011	Erro de previsão	Erro percentual
Janeiro	3	0.1%
Fevereiro	45	2.3%
Março	-462	-63.2%
Setembro	467	20.3%
Outubro	-258	-6.6%
Novembro	728	15.8%
Dezembro	349	7.2%
	EAM	EPAM
Estações (Jan-Mar;Set-Dez)	330.45	16.52%

Tabela 19: Medidas de avaliação dos erros de previsão para os meses da estação de inverno do ano de 2011 da categoria Botim.

O modelo  $ARIMA(5,0,2) \times (0,1,0)_{12}$  para a série de vendas da categoria Botim, estimado utilizando agora todos os dados históricos tem equação:

$$\begin{aligned} (1 + 0.9952B - 0.2543B^2 - 0.6953B^3 - 0.0066B^4 + 0.0708B^5)(1 - B^{12}) \log(Y_t) = \\ = (1 + 1.7136B + 1.0000B^2) \varepsilon_t \end{aligned}$$

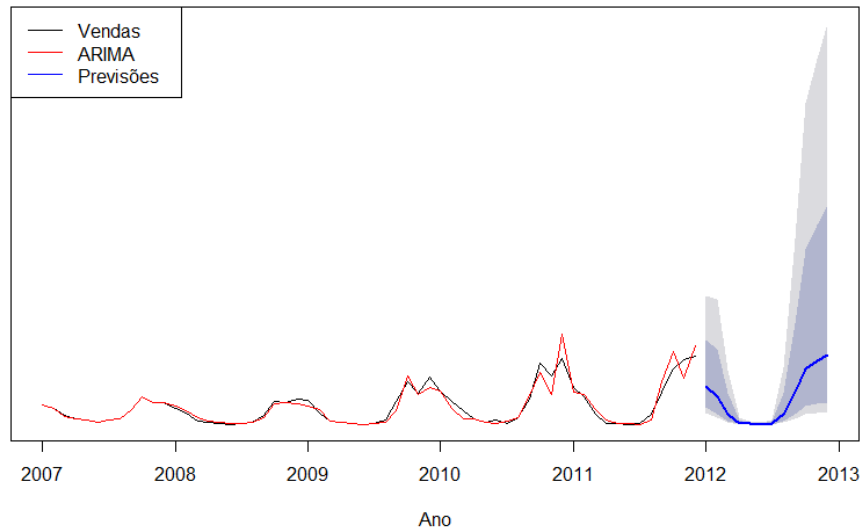


Figura 21: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Botim.

Na Figura 21 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Botim, (2) dos valores ajustados pelo modelo  $ARIMA(5,0,2) \times (0,1,0)_{12}$ , e (3) das previsões de vendas da categoria Botim para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a capacidade de ajuste do modelo ARIMA aos dados históricos e o modo como este captou a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se que a amplitude dos intervalos de previsão aumenta à medida que o horizonte temporal da previsão aumenta, mas é inferior fora dos meses da estação de inverno.

### 4.2.3 Categoria Sabrina

Como se pode observar na Tabela 20, o modelo que apresentou melhor desempenho de previsão foi o modelo  $ARIMA(5,0,5) \times (2,1,0)_{12}$  para os dados *raw* (com diferenciação sazonal), consequentemente este é o modelo ARIMA selecionado para a previsão da série de vendas da categoria Sabrina. Atente-se que todos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. Excetuando a REQM, todas as medidas de avaliação dos erros de previsão fora da amostra apresentam um valor superior no modelo selecionado. Todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam um valor superior no modelo selecionado. Repare-se que o modelo  $ARIMA(5,0,5) \times (2,1,0)_{12}$  com dados *raw* está de acordo com a análise exploratória da série apresentada na Secção 3.2.

A Figura 22 mostra o cronograma, a FAC e FACP dos resíduos do modelo ARIMA  $(5,0,5) \times (2,1,0)_{12}$ . O cronograma mostra que os resíduos apresentam um comportamento relativamente semelhante a um ruído branco. Apesar dos valores dos resíduos relativos aos meses maio e junho de 2008 serem um pouco mais elevados, este facto não compromete o diagnóstico já que a FAC e a FACP não contêm correlações significativas.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						Ljung-Box Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
ARIMA $(5,0,5) \times (2,1,0)_{12}$	-51.83	264.79	173.94	-13.24%	33.91%	0.62	0.86
ARIMA $(1,0,0) \times (2,0,1)_{12}$	-21.48	344.77	240.13	-20.09%	36.67%	0.86	0.97
Log ARIMA $(0,0,2) \times (1,1,0)_{12}$	-90.38	231.58	148.63	-21.79%	29.42%	0.53	0.22
ARIMA $(1,1,1) \times (2,0,1)_{12}$	-38.63	344.04	244.51	-20.03%	37.08%	0.88	0.93
Log ARIMA $(0,0,2) \times (2,0,0)_{12}$	64.94	402.59	229.18	-13.29%	32.25%	0.82	0.12
Log ARIMA $(2,1,5) \times (0,1,0)_{12}$	-2.09	208.43	137.07	-2.70%	22.04%	0.49	0.92
ARIMA $(2,1,2) \times (1,1,0)_{12}$	5.97	323.40	202.78	-2.92%	36.63%	0.73	0.89
Log ARIMA $(5,1,5) \times (2,0,0)_{12}$	8.50	257.38	175.61	-6.78%	22.53%	0.63	0.33

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ARIMA $(5,0,5) \times (2,1,0)_{12}$	84.87	240.51	187.01	12.90%	30.70%
ARIMA $(1,0,0) \times (2,0,1)_{12}$	17.19	242.44	197.62	-11.98%	33.47%
Log ARIMA $(0,0,2) \times (1,1,0)_{12}$	115.31	261.08	196.84	7.39%	28.66%
ARIMA $(1,1,1) \times (2,0,1)_{12}$	111.36	268.83	195.78	4.99%	27.44%
Log ARIMA $(0,0,2) \times (2,0,0)_{12}$	105.11	272.53	198.55	0.98%	28.91%
Log ARIMA $(2,1,5) \times (0,1,0)_{12}$	64.70	279.73	214.05	0.56%	31.83%
ARIMA $(2,1,2) \times (1,1,0)_{12}$	192.84	321.65	233.61	22.83%	33.40%
Log ARIMA $(5,1,5) \times (2,0,0)_{12}$	272.37	408.46	304.96	24.76%	36.72%

Tabela 20: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Sabrina.

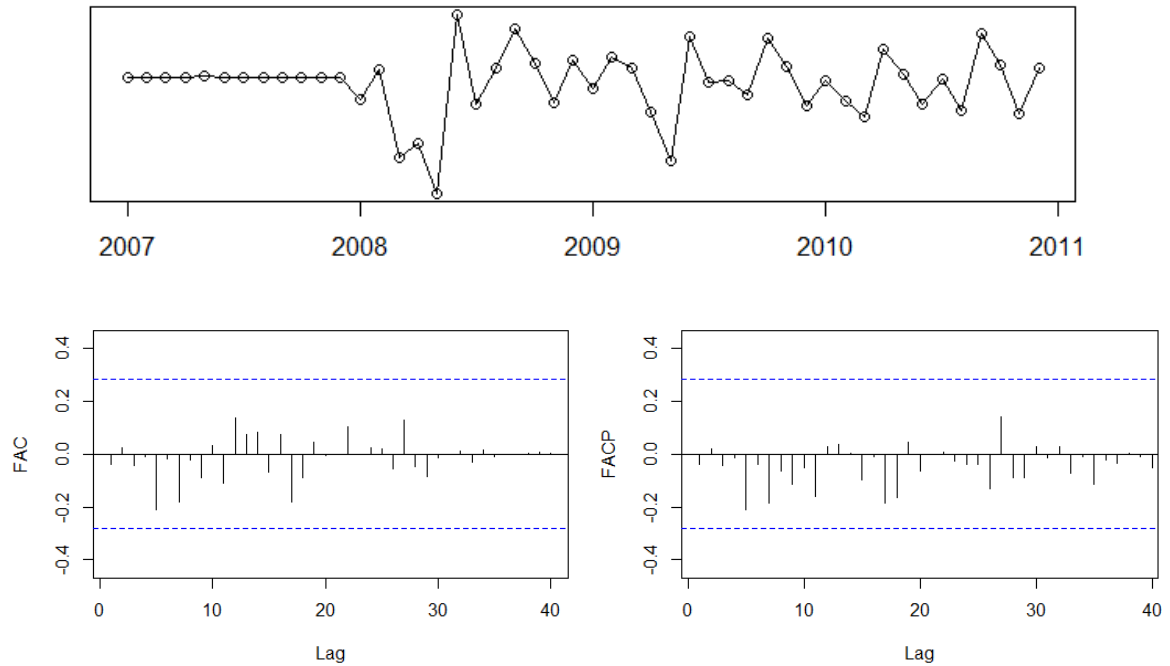


Figura 22: Cronograma, FAC e FACP dos resíduos do modelo  $ARIMA(5,0,5) \times (2,1,0)_{12}$  para a Sabrina.

A Tabela 21 mostra medidas de avaliação dos erros de previsão do modelo  $ARIMA(5,0,5) \times (2,1,0)_{12}$  apenas para os meses da estação de verão do ano de 2011. O erro percentual mensal varia entre -2.7% (mês de agosto) e -59.9% (mês de julho).

Mês de 2011	Erro de previsão	Erro percentual
Março	-124	-12.1%
Abril	134	8.5%
Maio	-117	-12.7%
Junho	-91	-19.6%
Julho	-263	-59.9%
Agosto	-19	-2.7%
Setembro	250	26.7%
	EAM	EPAM
Estação (Mar-Set)	142.37	20.28%

Tabela 21: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sabrina.

O modelo  $ARIMA(5,0,5) \times (2,1,0)_{12}$  para a série de vendas da categoria Sabrina, estimado utilizando agora todos os dados históricos tem equação:

$$(1+1557B+0.3406B^2-0.1591B^3-0.0148B^4-0.4590B^5)(1+0.5151B^{12}-0.1359B^{24})(1-B^{12})Y_t = (1+1.0652B+1.4308B^2+0.8260B^3+0.6067B^4-0.2901B^5)\varepsilon_t$$

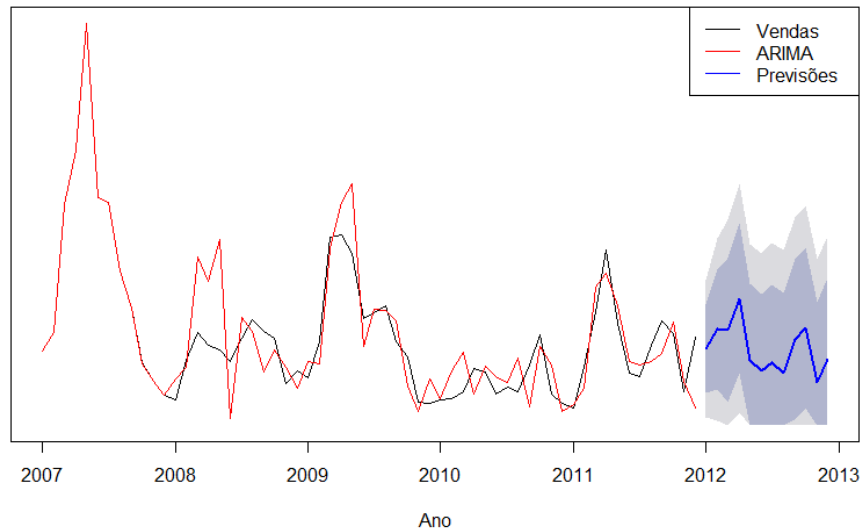


Figura 23: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Sabrina.

Na Figura 23 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Sabrina, (2) dos valores ajustados pelo modelo  $ARIMA(5,0,5) \times (2,1,0)_{12}$ , e (3) das previsões de vendas da categoria Sabrina para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro). Saliente-se a dificuldade de ajuste do modelo ARIMA aos dados históricos. Observe-se que as previsões, como é esperado, tentam manter o padrão de comportamento da série. Repare-se que curiosamente a amplitude dos intervalos de previsão mantém-se razoavelmente constante à medida que o horizonte temporal da previsão aumenta.

#### 4.2.4 Categoria Sandália

Como se pode observar na Tabela 22, o modelo que apresentou melhor desempenho de previsão foi o modelo  $ARIMA(3,0,1) \times (1,1,0)_{12}$  para os dados *raw* (com diferenciação sazonal), consequentemente este é o modelo ARIMA selecionado para a previsão da série de vendas da categoria Sandália.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						Ljung-Box Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
$ARIMA(3,0,1) \times (1,1,0)_{12}$	393.22	1782.79	1005.61	-663.55%	1359.57%	0.34	0.67
Log $ARIMA(0,0,3) \times (0,1,0)_{12}$	442.67	1937.27	971.58	-38.93%	75.18%	0.33	0.92

ARIMA $(0,1,0) \times (0,1,0)_{12}$	-1.57	2206.47	1196.30	-350.21%	406.10%	0.41	0.33
ARIMA $(3,0,3) \times (1,0,0)_{12}$	362.41	2029.58	1350.97	-310.83%	2906.64%	0.46	0.39
Log ARIMA $(4,0,0) \times (1,0,0)_{12}$	1167.13	2537.48	1376.52	-62.86%	112.09%	0.47	0.21
ARIMA $(5,1,3) \times (1,0,0)_{12}$	100.67	1818.99	1266.89	782.16%	2673.12%	0.43	0.29
Log ARIMA $(5,1,2) \times (2,0,0)_{12}$	1019.99	2327.13	1261.33	-39.23%	89.85%	0.43	0.47
Log ARIMA $(1,1,3) \times (0,1,0)_{12}$	-770.17	3208.71	1396.37	-72.58%	91.96%	0.48	0.96

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
ARIMA $(3,0,1) \times (1,1,0)_{12}$	-132.88	461.46	350.20	-40.64%	86.76%
Log ARIMA $(0,0,3) \times (0,1,0)_{12}$	-266.97	559.11	409.30	-44.89%	75.23%
ARIMA $(0,1,0) \times (0,1,0)_{12}$	-246.08	562.85	401.58	9.12%	58.14%
ARIMA $(3,0,3) \times (1,0,0)_{12}$	-18.85	904.45	717.53	-3552.79%	3573.33%
Log ARIMA $(4,0,0) \times (1,0,0)_{12}$	93.35	1097.54	677.27	-54.80%	87.01%
ARIMA $(5,1,3) \times (1,0,0)_{12}$	-1164.15	1415.56	1212.46	-8357.72%	8374.60%
Log ARIMA $(5,1,2) \times (2,0,0)_{12}$	-1027.06	2392.99	1598.98	-143.06%	173.53%
Log ARIMA $(1,1,3) \times (0,1,0)_{12}$	4642.95	7286.72	4643.08	55.63%	61.79%

Tabela 22: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Sandália.

Atente-se que todos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. Excetuando a REQM e o EAM, todas as medidas de avaliação dos erros de previsão fora da amostra apresentam um valor superior no modelo selecionado. Excetuando a REQM todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam um valor superior no modelo selecionado.

Repare-se que o modelo ARIMA  $(3,0,1) \times (1,1,0)_{12}$  com dados *raw* está de acordo com a análise exploratória da série apresentada na Secção 3.2. A Figura 24 mostra o cronograma, a FAC e FACP dos resíduos do modelo ARIMA  $(3,0,1) \times (1,1,0)_{12}$ . O cronograma mostra que os resíduos apresentam um comportamento semelhante a um ruído branco, como é esperado. A FAC e a FACP mostram que não existe correlação entre os resíduos, como é desejável.

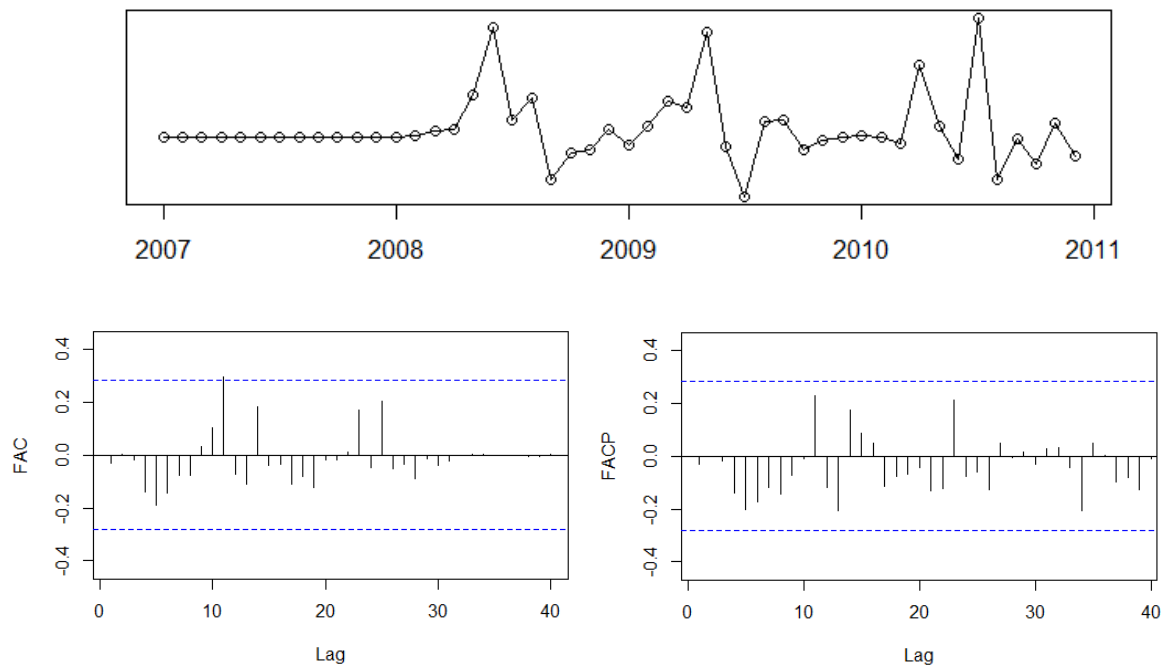


Figura 24: Cronograma, FAC e FACP dos resíduos do modelo  $ARIMA(3,0,1) \times (1,1,0)_{12}$  para a Sandália.

A Tabela 23 mostra medidas de avaliação dos erros de previsão do modelo  $ARIMA(3,0,1) \times (1,1,0)_{12}$  apenas para os meses da estação de verão do ano de 2011. O erro percentual mensal varia entre 0.4% (mês de maio) e -106.6% (mês de março).

Atente-se que o EPAM relativo aos meses da estação de verão da série da Sabrina, 20.28% é equivalente ao EPAM relativo aos meses da estação de verão da série da Sandália, 20.70%.

Mês de 2011	Erro de previsão	Erro percentual
Março	-901	-106.6%
Abril	-246	-3.0%
Maio	58	0.4%
Junho	285	1.7%
Julho	741	3.4%
Agosto	-487	-4.5%
Setembro	-655	-25.3%
	EAM	EPAM
Estação (Mar-Set)	481.75	20.70%

Tabela 23: Medidas de avaliação dos erros de previsão para os meses da estação de verão do ano de 2011 da categoria Sandália.

O modelo  $ARIMA(3,0,1) \times (1,1,0)_{12}$  para a série de vendas da categoria Sandália, estimado utilizando agora todos os dados históricos tem equação:

$$(1 + 0.5341B - 0.4912B^2 - 0.2193B^3)(1 + 0.1011B^{12})(1 - B^{12})Y_t = (1 + 0.9676B)\epsilon_t$$

Na Figura 25 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Sandália, (2) dos valores ajustados pelo modelo  $ARIMA(3,0,1) \times (1,1,0)_{12}$ , e (3) das previsões de vendas da categoria Sandália para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro).

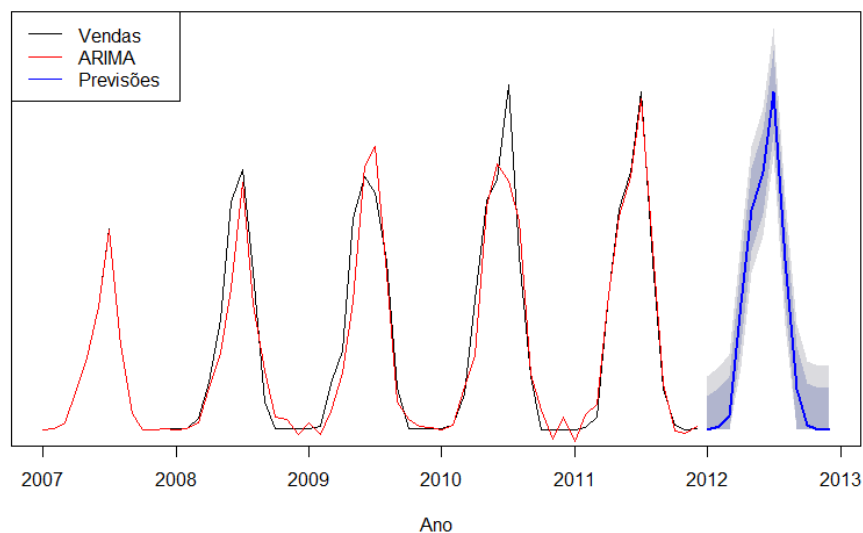


Figura 25: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Sandália.

Saliente-se a capacidade de ajuste do modelo ARIMA aos dados históricos e o modo como este captou a tendência e a sazonalidade. Observe-se que as previsões, como é esperado, mantêm o padrão de comportamento da série. Repare-se que a amplitude dos intervalos de previsão mantém-se razoavelmente constante à medida que o horizonte temporal da previsão aumenta, mas é superior fora dos meses da estação de verão.

#### 4.2.5 Categoria Sapato

Como se pode observar na Tabela 24, o modelo que apresentou melhor desempenho de previsão foi o modelo  $ARIMA(0,1,3) \times (1,0,0)_{12}$  para os dados transformados

logaritmicamente (com diferenciação simples), consequentemente este é o modelo ARIMA selecionado para a previsão da série de vendas da categoria Sapato. Atente-se que todos os modelos passaram o teste estatístico de diagnóstico de resíduos de *Ljung-Box*. Excetuando a REQM e o EAM, todas as medidas de avaliação dos erros de previsão fora da amostra apresentam um valor superior no modelo selecionado. Todas as medidas de avaliação dos erros de previsão dentro da amostra apresentam um valor superior no modelo selecionado.

MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						Ljung-Box Valor-p
	EM	REQM	EAM	EPM	EPAM	EEAM	
Log ARIMA $(0,1,3) \times (1,0,0)_{12}$	56.07	1043.50	874.00	-2.63%	18.17%	0.67	0.11
ARIMA $(3,1,4) \times (0,0,2)_{12}$	-10.83	927.21	733.66	-3.65%	16.28%	0.56	0.07
Log ARIMA $(3,1,5) \times (1,1,0)_{12}$	-18.81	808.18	552.73	-1.47%	11.07%	0.42	0.64
ARIMA $(2,1,4) \times (0,1,0)_{12}$	-26.85	908.05	631.37	-1.46%	12.99%	0.48	0.53
ARIMA $(5,0,0) \times (0,1,0)_{12}$	-54.70	931.14	675.79	-2.88%	14.09%	0.52	0.42
ARIMA $(5,0,0) \times (0,0,2)_{12}$	32.60	944.50	806.42	-3.53%	17.63%	0.62	0.13
Log ARIMA $(5,0,0) \times (1,0,0)_{12}$	130.94	932.99	782.30	-1.50%	16.63%	0.60	0.39
Log ARIMA $(5,0,0) \times (0,1,0)_{12}$	-13.57	945.51	694.41	-2.16%	14.17%	0.53	0.52

MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)				
	EM	REQM	EAM	EPM	EPAM
Log ARIMA $(0,1,3) \times (1,0,0)_{12}$	-533.03	609.70	533.03	-19.29%	19.29%
ARIMA $(3,1,4) \times (0,0,2)_{12}$	-495.37	663.68	552.47	-19.02%	20.50%
Log ARIMA $(3,1,5) \times (1,1,0)_{12}$	-638.12	777.92	638.12	-20.59%	20.59%
ARIMA $(2,1,4) \times (0,1,0)_{12}$	-385.20	782.92	581.72	-9.43%	17.84%
ARIMA $(5,0,0) \times (0,1,0)_{12}$	-558.02	950.83	657.59	-15.74%	19.32%
ARIMA $(5,0,0) \times (0,0,2)_{12}$	-828.82	970.89	838.53	-29.36%	29.72%
Log ARIMA $(5,0,0) \times (1,0,0)_{12}$	-935.76	1029.99	935.76	-33.52%	33.52%
Log ARIMA $(5,0,0) \times (0,1,0)_{12}$	-927.38	1138.71	927.38	-29.20%	29.20%

Tabela 24: Medidas de avaliação dos erros de previsão dos modelos ARIMA dentro e fora da amostra para a série de vendas da categoria Sapato.

Repare-se que o modelo ARIMA  $(0,1,3) \times (1,0,0)_{12}$  com dados transformados logaritmicamente está de acordo com a análise exploratória da série apresentada na Secção 3.2. A Figura 26 mostra o cronograma, a FAC e FACP dos resíduos do modelo ARIMA  $(0,1,3) \times (1,0,0)_{12}$ .

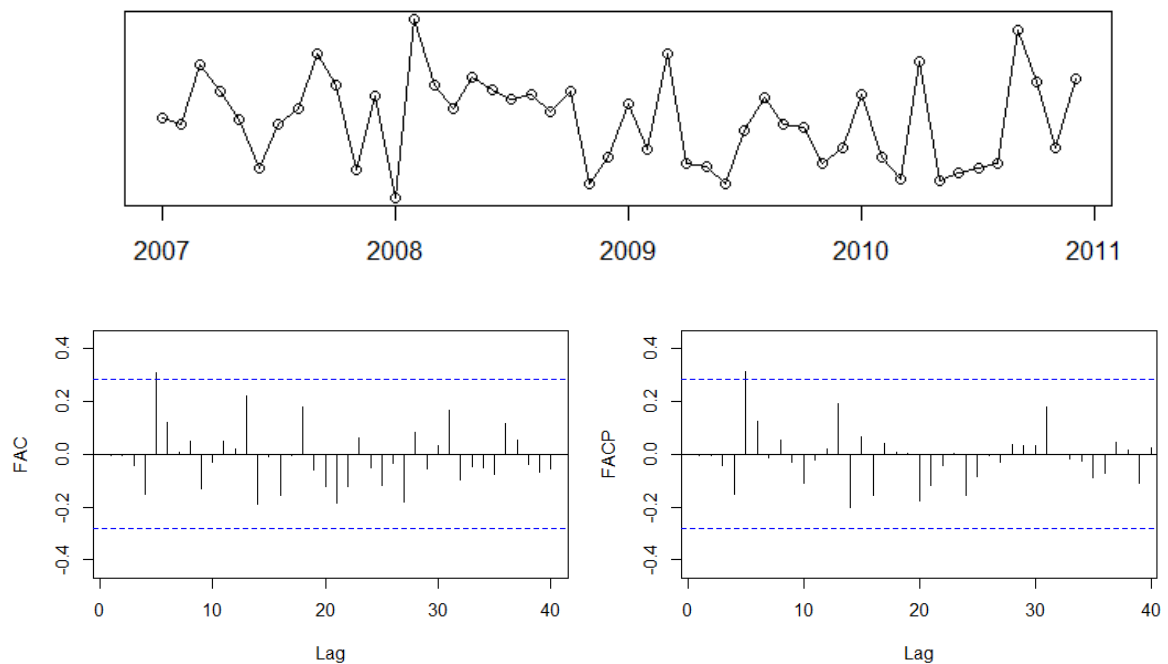


Figura 26: Cronograma, FAC e FACP dos resíduos do modelo  $ARIMA(0,1,3) \times (1,0,0)_{12}$  para o Sapato.

O cronograma mostra que os resíduos apresentam um comportamento semelhante a um ruído branco, como é esperado. A FAC e a FACP mostram que não existe correlação entre os resíduos, como é desejável. A Tabela 25 mostra medidas de avaliação dos erros de previsão do modelo  $ARIMA(0,1,3) \times (1,0,0)_{12}$  para os meses do ano de 2011. O erro percentual mensal varia entre -4.6% (mês de agosto) e -37.7% (mês de novembro).

Mês de 2011	Erro de previsão	Erro percentual
Janeiro	-838	-31.6%
Fevereiro	-218	-6.4%
Março	-258	-6.8%
Abril	-668	-16.9%
Maio	-405	-13.6%
Junho	-678	-36.4%
Julho	-456	-22.0%
Agosto	-125	-4.6%
Setembro	-210	-4.7%
Outubro	-1138	-29.5%
Novembro	-820	-37.7%
Dezembro	-580	-21.2%

Tabela 25: Medidas de avaliação dos erros de previsão para os meses do ano de 2011 da categoria Sapato.

O modelo  $ARIMA(0,1,3) \times (1,0,0)_{12}$  para a série de vendas da categoria Botim, estimado utilizando agora todos os dados históricos tem equação:

$$(1 - 0.7300B^{12})(1 - B)\log(Y_t) = (1 - 0.4594B + 0.0189B^2 - 0.2829B^3)\varepsilon_t$$

Na Figura 27 mostra-se o gráfico conjunto: (1) da série de vendas da categoria Sapato, (2) dos valores ajustados pelo modelo  $ARIMA(0,1,3) \times (1,0,0)_{12}$ , e (3) das previsões de vendas da categoria Sapato para o ano de 2012 e respetivos intervalos de previsão de 80% (azul escuro) e 95% (azul claro).

Saliente-se a relativa capacidade de ajuste do modelo ARIMA aos dados históricos. Observe-se que as previsões mantêm o padrão de comportamento da série, com tendência de descida.

Repare-se que a amplitude dos intervalos de previsão mantém-se razoavelmente constante à medida que o horizonte temporal da previsão aumenta

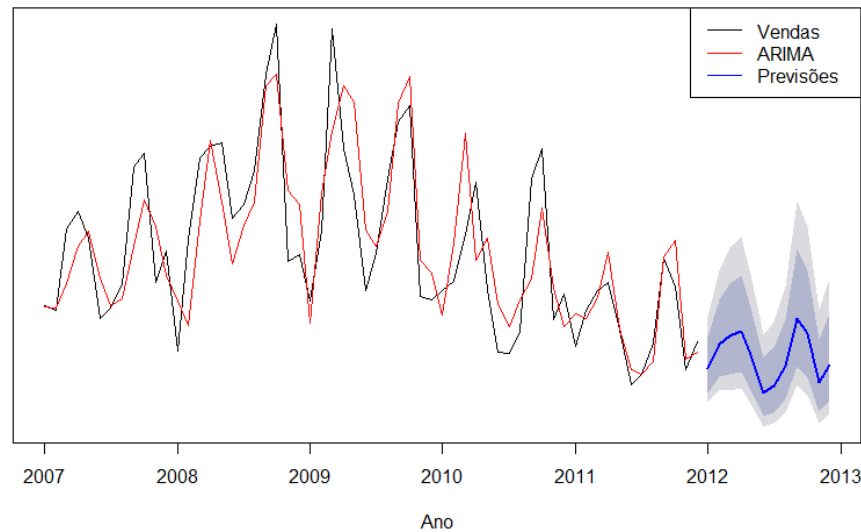


Figura 27: Previsões de vendas do modelo ARIMA para o ano de 2012 da categoria Sapato.

### 4.3-Comparação de resultados

A Tabela 26 mostra as medidas de avaliação dos erros de previsão para o conjunto de treino e para o conjunto de teste dos modelos ETS e ARIMA seleccionados para todas as séries de vendas (Godfrey, 1978). Nesta tabela mostram-se também o EAM e o EPAM apenas para os meses das respetivas estações (verão ou inverno) do ano de 2011, estando referenciados respetivamente por EAMs e EPAMs.

A partir desta tabela observamos que, excetuando a série do Botim, para todas as restantes séries os modelos ARIMA têm um melhor desempenho de previsão que os modelos de espaço de estado, quando avaliados pelas medidas REQM, EAM e EPAM.

Na série de vendas da categoria Bota, os valores da REQM e do EAM relativos ao modelo ARIMA são respetivamente 55% e 57% inferiores aos valores da REQM e do EAM relativos ao modelo de espaço de estado. O valor do EPAM relativo ao modelo de espaço de estado é inferior ao valor do EPAM do modelo ARIMA; contudo o valor do EPAMs relativo ao modelo ARIMA é 70% inferior ao EPAMs relativo ao modelo de espaço de estado.

Na série de vendas da categoria Sabrina, os valores da REQM, do EAM e do EPAM relativos ao modelo ARIMA são respetivamente 21%, 17% e 20% inferiores aos valores da REQM, do EAM e do EPAM relativos ao modelo de espaço de estado.

## Modelação e Previsão de Vendas no Setor do Retalho de Calçado

Na série de vendas da categoria Sandália, os valores da REQM, do EAM e do EPAM relativos ao modelo ARIMA são respetivamente 77%, 76% e 99% inferiores aos valores da REQM, do EAM e do EPAM relativos ao modelo de espaço de estado.

Na série de vendas da categoria Sapato, os valores da REQM, do EAM e do EPAM relativos ao modelo ARIMA são respetivamente 55%, 56% e 53% inferiores aos valores da REQM, do EAM e do EPAM relativos ao modelo de espaço de estado.

Série	MODELO	Medidas de avaliação dos erros de previsão dentro da amostra (2007-2010)						Ljung-Box Valor-p
		EM	REQM	EAM	EPM	EPAM	EEAM	
Bota	ETS (M,A <sub>d</sub> ,A)	-74.11	1024.02	678.32	-8.39%	98.34%	0.29	0.12
	Log ARIMA (2,1,5)×(1,0,1) <sub>12</sub>	-29.11	1215.36	759.72	-16.62%	36.24%	0.32	0.88
Botim	ETS (M,A <sub>d</sub> ,M)	4.93	427.68	279.69	-25.13%	49.01%	0.55	0.06
	Log ARIMA (5,0,2)×(0,1,0) <sub>12</sub>	40.83	391.67	226.98	-7.55%	31.04%	0.45	0.96
Sabrina	ETS (M,N,A)	-29.86	380.08	263.18	-11.98%	34.90%	0.94	0.54
	ARIMA (5,0,5)×(2,1,0) <sub>12</sub>	-51.83	264.79	173.94	-13.24%	33.91%	0.62	0.86
Sandália	ETS (A,A,A)	-103.63	1802.00	1346.83	466.94%	2638.46%	0.46	0.02*
	ARIMA (3,0,1)×(1,1,0) <sub>12</sub>	393.22	1782.79	1005.61	-663.55%	1359.57%	0.34	0.67
Sapato	Log ETS (A,A,A)	12.2	732.39	574.23	-1.03%	11.63%	0.44	0*
	Log ARIMA (0,1,3)×(1,0,0) <sub>12</sub>	56.07	1043.50	874.00	-2.63%	18.17%	0.67	0.11

Série	MODELO	Medidas de avaliação dos erros de previsão fora da amostra (2011)						
		EM	REQM	EAM	EPM	EPAM	EAMs	EPAMs
Bota	ETS (M,A <sub>d</sub> ,A)	-26.15	1682.70	1132.33	35.99%	69.22%	1859.28	47.23%
	Log ARIMA (2,1,5)×(1,0,1) <sub>12</sub>	-266.93	753.73	485.84	-109.48%	124.30%	742.68	14.00%
Botim	ETS (M,A <sub>d</sub> ,M)	-52.59	309.32	226.79	-110.87%	119.12%	295.97	10.27%
	Log ARIMA (5,0,2)×(0,1,0) <sub>12</sub>	19.56	347.12	278.54	-186.83%	199.31%	330.45	16.52%
Sabrina	ETS (M,N,A)	51.71	304.31	224.55	-5.64%	38.25%	231.3	33.95%
	ARIMA (5,0,5)×(2,1,0) <sub>12</sub>	84.87	240.51	187.01	12.90%	30.70%	142.37	20.28%
Sandália	ETS (A,A,A)	505.78	2015.27	1486.89	-12841.70%	12889.20%	1937.9	32.03%
	ARIMA (3,0,1)×(1,1,0) <sub>12</sub>	-132.88	461.46	350.20	-40.64%	86.76%	481.75	20.70%
Sapato	Log ETS (A,A,A)	-1213.34	1357.60	1213.34	-41.13%	41.13%		
	Log ARIMA (0,1,3)×(1,0,0) <sub>12</sub>	-533.03	609.70	533.03	-19.29%	19.29%		

Tabela 26: Medidas de avaliação dos erros de previsão para o conjunto de teste dos modelos ETS e ARIMA selecionados para todas as séries de vendas.

Na série de vendas da categoria Botim, os valores da REQM, do EAM e do EPAM relativos ao modelo ARIMA são respetivamente 11%, 19% e 40% superiores aos valores da REQM, do EAM e do EPAM relativos ao modelo de espaço de estado.

Uma outra observação interessante é que para todas as categorias em ambas as abordagens o valor do EPAMs é sempre inferior ao valor do EPAM o que demonstra a dificuldade dos modelos responderem às alterações de magnitude dos valores das séries temporais.

Como esperado, um modelo que se ajusta bem aos dados históricos não é necessariamente o melhor modelo de previsão e o inverso também é verdadeiro. Este facto é notório quando analisamos os valores das medidas de avaliação dos erros de previsão para o conjunto de treino. Contudo, como já foi referido, é muito frequente este tipo de resultados contraditórios quando comparamos as medidas de avaliação de desempenho de vários modelos candidatos.

## 5. CONCLUSÕES

O objetivo principal deste trabalho consistiu na comparação do desempenho de duas metodologias de previsão na análise de séries de vendas do setor do retalho de calçado.

Os dados analisados consistiram em cinco séries temporais relativas às cinco principais categorias de calçado comercializadas pela empresa Foreva: a Bota, o Botim, a Sabrina, a Sandália, e o Sapato.

Dado que só a partir do ano de 2007 é que a Foreva estabilizou a sua rede de lojas optou-se por considerar as séries de vendas destas categorias de calçado a partir de 1 de janeiro de 2007. Para este estudo foram disponibilizados pela empresa dados diários das vendas destas categorias até dezembro de 2011.

Desde logo foi solicitada pela Foreva uma previsão anual das vendas de cada uma das categorias de calçado para o ano seguinte, tendo a empresa como objetivo a utilização dessas previsões para a especificação do número pares de calçado de cada uma das categorias a adquirir para comercialização.

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

Para melhor comparar o desempenho de previsão de ambas as abordagens, decidimos considerar todos os modelos de espaço de estado possíveis e todos os modelos ARIMA, onde os parâmetros variaram razoavelmente.

Em ambas as abordagens foram considerados dois tipos de dados: dados raw e dados transformados logaritmicamente.

Para cada abordagem foi selecionado o modelo que apresentou o menor valor de REQM na previsão do conjunto de teste e passou o teste estatístico de *Ljung-Box*.

Os resultados mostram que, de um modo geral, os modelos ARIMA têm melhor desempenho que os modelos de espaço de estado na previsão de séries de vendas do setor do retalho de calçado.

## 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, 3<sup>rd</sup> ed., Prentice-Hall.
- BOWERMAN, B., O'CONNELL, R. e KOEHLER, A. (2005). Forecasting, Time Series and Regression, 4<sup>a</sup> edição, Duxbury, Thomson Brooks/Cole, Belmont.
- BROCKWELL, P. J. e DAVIS, R. A. (1991). Time Series: Theory and Methods. 2<sup>nd</sup> ed., Springer, New York.
- BROWN, R (1956). "Exponential smoothing for predicting demand", 10<sup>th</sup> 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<sup>a</sup> 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", 5<sup>th</sup> Time Series Workshop, Arrábida, Portugal.
- CRYER, J. D., CHAN, K-S (2008). Time series analysis with applications in R, 2<sup>nd</sup> 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/>.
- JOHNSON. R, E. M., KURTZ, D. e SCHEVING, E. (1994). *Sales Management - Concepts, Practices and Cases*, McGraw-Hill, Singapore International Editions.
- LAWRENCE, 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*, 4<sup>th</sup> ed., McGraw-Hill, Boston.
- R *Software*. Free software environment for statistical computing and graphics. <http://www.r-project.org/> (acedido em junho 2012).
- 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*. 2<sup>nd</sup> 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.