

DM

**Estudo Comparativo
entre as Análises Preditiva e Prescritiva
como Ferramentas de Apoio na Tomada de Decisão
num Sistema de Gestão de Stocks**

DISSERTAÇÃO DE MESTRADO

Viviana Lilian da Silva Gomes

MESTRADO EM MATEMÁTICA, ESTATÍSTICA E APLICAÇÕES



UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

janeiro | 2021

**Estudo Comparativo
entre as Análises Preditiva e Prescritiva
como Ferramentas de Apoio na Tomada de Decisão
num Sistema de Gestão de Stocks**

DISSERTAÇÃO DE MESTRADO

Viviana Lilian da Silva Gomes

MESTRADO EM MATEMÁTICA, ESTATÍSTICA E APLICAÇÕES

ORIENTAÇÃO
Paulo Sérgio Abreu Freitas

Agradecimentos

A realização desta dissertação de mestrado não seria possível sem o apoio de várias pessoas, as quais quero agradecer.

Em primeiro lugar ao Professor Doutor Paulo Sérgio Abreu Freitas pela sua orientação ao longo da dissertação, pelo conhecimento e opiniões que partilhou comigo, pela sua disponibilidade e em geral por todo o seu apoio.

A todos os professores do mestrado em Matemática, Estatística e Aplicações pelos conhecimentos e competências que me transmitiram ao longo deste percurso académico que me ajudaram na realização desta tese.

Aos meus pais pelo apoio incondicional, preocupação, incentivo e ajuda na superação dos obstáculos ao longo do mestrado e dissertação.

Ao meu irmão pela partilha dos seus conhecimentos e a sua pronta disponibilidade por ajudar-me sempre que precisei.

Por último ao Marco Spínola por estar sempre do meu lado, pelo seu apoio, compreensão, partilha e companheirismo.

Resumo

No âmbito da gestão, o processo da tomada de decisão tem vindo a utilizar ferramentas de previsão cada vez mais adequadas. Nos sistemas de gestão de *stocks*, a tomada de decisão baseia-se particularmente nas previsões da procura e nas encomendas dos clientes. As previsões são feitas de forma a diminuir a incerteza e a garantir melhores resultados para a empresa, seja por meio do aumento do nível de serviço ou da redução de custos.

Nesta dissertação, pretende-se apresentar um estudo comparativo entre dois tipos de análises, uma preditiva e outra prescritiva, cujas decisões são avaliadas através de funções de utilidade com custos assimétricos.

Na análise preditiva, as decisões são precedidas pela identificação e estimação de um modelo com boas capacidades preditivas, isto é, baseado em previsões ótimas, em certo sentido. Com base em uma função de utilidade com custos assimétricos, as decisões finais são obtidas a partir de quantis de probabilidade dos erros de previsão.

Na análise prescritiva, o modelo é identificado e estimado em termos da função de utilidade. Neste caso, as estimativas produzidas pelo modelo prescritivo correspondem a quantis de probabilidade da variável em estudo.

Assumindo que a previsão constitui apenas um meio de suporte na tomada de decisão, este estudo pressupõe que previsões ótimas favorecem decisões ótimas e que decisões ideais não exigem previsões ótimas. Para levar a cabo este estudo, é elaborado dois testes estatísticos baseados na ANOVA de medidas repetidas, onde são usados dois tipos de conjuntos de dados, um simulado e outro constituído por séries temporais reais.

Palavras-chaves: gestão de stocks, tomada de decisão, previsão, análise preditiva, análise prescritiva, custos assimétricos.

Abstract

In the process of decision-making regarding management, more adequate forecasting tools are increasingly used. In inventory management systems, decision-making is based particularly on demand forecasts and customer orders. Predictions are made in order to reduce uncertainty, and to ensure better results to the company, either through rising service level or costs reduction.

This dissertation aims to present a comparative study between two types of analyses, one predictive and other prescriptive, where decisions are evaluated through utility functions with asymmetric costs.

In predictive analysis, decisions are preceded by the identification and estimation of a model with good predictive capabilities, that is, based on optimal predictions, at some extent. Based on a utility function with asymmetric costs, the final decisions are obtained from quantiles of the prediction errors.

In prescriptive analysis, the model is identified and estimated in terms of the utility function. In this case, the estimates produced by the prescriptive model correspond to the quantiles of the variable under study.

Assuming that prediction is just a means of supporting decision-making, this study presumes that optimal forecasts favour optimal decisions and that ideal decisions do not require optimal forecasts. To carry out this study, two repeated measures ANOVA are employed, where two types of data sets are considered, one simulated and the other consisting of real time series.

Keywords: inventory management, decision making, forecasting, predictive analysis, prescriptive analysis, asymmetric costs.

Índice

1	INTRODUÇÃO	1
1.1	Objetivos	2
1.2	Revisão da Literatura	2
1.3	Estrutura.....	4
2	GESTÃO DE STOCKS	5
2.1	Custos de stocks.....	6
2.2	Modelos de gestão de stocks.....	7
2.2.1	Modelos determinísticos.....	8
2.2.2	Modelos estocásticos	15
3	PREVISÃO DE SÉRIES TEMPORAIS	25
3.1	A procura	25
3.2	Séries temporais	26
3.3	Previsão.....	29
3.4	Medidas de desempenho	30
3.5	Métodos de alisamento exponencial	34
3.5.1	Alisamento exponencial simples	36
3.5.2	Método de Holt.....	37
3.5.3	Método de Holt-Winters.....	38
4	ANÁLISE PREDITIVA VS. ANÁLISE PRESCRITIVA.....	43
4.1	Motivação e objetivos	43
4.2	Previsões ótimas vs. decisões ótimas.....	44
4.3	ANOVA de medidas repetidas.....	49
4.4	Conjuntos de dados	54
4.4.1	Simulação de componentes de sazonalidade	55
4.4.2	Simulação de componentes de tendência	58
4.4.3	Simulação das séries temporais	62

4.5	Obtenção de resultados e interpretação.....	64
5	CONCLUSÕES E TRABALHO FUTURO	75
	BIBLIOGRAFIA	77
ANEXO A	Gráficos das séries temporais reais.....	81
ANEXO B	Gráficos das séries temporais simuladas	85
ANEXO C	Custos obtidos para as séries temporais simuladas	89
ANEXO D	Custos obtidos para as séries temporais reais.....	91
ANEXO E	Testes de normalidade dos dados originais	93
ANEXO F	Gráficos associados a custos simulados	95

Índice de Figuras

Figura 1 – Modelo da Quantidade Económica de Encomenda [Fonte: Gonçalves, 2012].	9
Figura 2 – Gráfico do custo total em função das quantidades.	11
Figura 3 – Modelo com reposição instantânea e com rutura permitida [Fonte: Gonçalves, 2012].	12
Figura 4 – Modelo com reposição não instantânea e sem rutura permitida [Fonte: Gonçalves, 2012].	13
Figura 5 – Modelo com reposição não instantânea e com rutura permitida (figura adaptada de vários autores).	14
Figura 6 – Sistema de Revisão Contínua [Fonte: Gonçalves, 2012].	17
Figura 7 – Representação gráfica do stock de segurança, nível de serviço e probabilidade de rutura [Fonte: Gonçalves, 2012].	18
Figura 8 – Sistema de Revisão Periódica [Fonte: Gonçalves, 2012].	19
Figura 9 – Exemplo de uma série temporal [Fonte: datamarket.com].	27
Figura 10 – Produção trimestral de gás (milhões de megajoules) na Austrália (janeiro 1956 a dezembro 1994). Os painéis abaixo da série original são as componentes de tendência, de sazonalidade e irregular, que, em conjunto, conformam a série observada [Fonte: datamarket.com].	28
Figura 11 – Diagrama ilustrativo da análise preditiva.	46
Figura 12 – Diagrama ilustrativo da análise prescritiva.	48
Figura 13 – Fatores e respetivos níveis na ANOVA fatorial dupla.	50
Figura 14 – Componentes sazonais simuladas, $S^{(1)}$ e $S^{(2)}$, respetivamente, usando uma composição de funções trigonométricas.	56
Figura 15 – Componentes sazonais simuladas, $S^{(3)}$, $S^{(4)}$ e $S^{(5)}$, respetivamente, usando a série de Mackey-Glass.	58
Figura 16 – Componentes de tendência simuladas, $T^{(1)}$, $T^{(2)}$ e $T^{(3)}$, respetivamente, usando modelos determinísticos.	60
Figura 17 – Componentes de tendência simuladas, $T^{(4)}$, $T^{(5)}$ e $T^{(6)}$, respetivamente, usando modelos de regressão de tendência dinâmica.	62
Figura 18 – Interação dos fatores Paradigma e Horizonte Temporal para os custos reais.	68
Figura 19 – Interação dos fatores Paradigma e Horizonte Temporal para os custos simulados.	72

Índice de Tabelas

Tabela 1 – Disposição das observações de medidas repetidas para o SPSS.	51
Tabela 2 – Condições de simulação das componentes de sazonalidade usando uma composição de funções trigonométricas.	56
Tabela 3 – Condições de simulação das componentes de sazonalidade usando a série caótica de Mackey-Glass.....	58
Tabela 4 – Condições de simulação das componentes de tendência usando modelos determinísticos.	60
Tabela 5 – Condições de simulação das componentes de tendência usando modelos de regressão de tendência dinâmica.	62
Tabela 6 – Teste de normalidade aos custos reais logaritmizados.	65
Tabela 7 – Teste de normalidade aos custos simulados logaritmizados.....	65
Tabela 8 – Teste de esfericidade de Mauchly aos custos reais.....	66
Tabela 9 – Teste de efeitos dentre-sujeitos aos custos reais.....	67
Tabela 10 – Procedimento de Bonferroni aos custos reais.....	69
Tabela 11 – Teste de esfericidade de Mauchly aos custos simulados.	70
Tabela 12 – Teste de efeitos dentre-sujeitos aos custos simulados.	71
Tabela 13 – Procedimento de Bonferroni aos custos simulados.	73

CAPÍTULO 1

1 INTRODUÇÃO

A gestão de stocks, designada também por gestão de inventários, desempenha um papel de importância primordial na gestão das empresas. Os *stocks* permitem não só assegurar o consumo regular de um produto, como também preparar a empresa para possíveis situações de incerteza futura, muitas vezes derivada de níveis de procura superiores ao esperado. Assim, para que os gestores possam otimizar da melhor forma os níveis de existência no inventário, procuram encontrar um nível “ideal” que represente uma solução de compromisso entre a minimização dos custos globais e um alto nível de serviço prestado aos clientes. A concretização destes objetivos é realizada por um modelo de gestão de stocks que sirva o propósito da empresa, sendo aquele dependente da política a considerar e das próprias características da procura.

A procura, como se sabe, é incerta, variando de forma irregular ao longo do tempo. Por este motivo, a procura deve ser considerada como uma variável aleatória ou, mais concretamente, um processo estocástico. Uma vez que os níveis de *stock* estão diretamente relacionados com a procura, a sua influência num sistema de gestão de *stocks* é deveras elevada. Na tentativa de anular ou, no mínimo, reduzir essa incerteza, é fundamental considerar o problema da previsão no apoio à tomada de decisão. Assim, a previsão de séries temporais (ou cronológicas) constitui uma ferramenta matemática de grande importância nos sistemas de gestão de stocks.

Na previsão, o objetivo consiste na estimação de uma sequência de valores da série temporal em estudo para instantes onde não se conhecem observações. Apesar de ser possível obter-se estimativas para instantes passados onde existem observações omissas, vamos nos limitar apenas à obtenção de estimativas para instantes futuros. A abordagem típica ao problema de previsão consiste em escolher um modelo adequado, de entre uma variedade de opções, que satisfaça um ou mais critérios previamente definidos. O modelo escolhido deverá ter a capacidade de reproduzir os vários padrões contidos na série, por exemplo, tendência e sazonalidade, quando existirem.

Na gestão de stocks, a avaliação final de uma decisão, quer seja baseada na previsão da procura, quer em outro tipo de informação (como por exemplo, por pessoas experientes), é tipicamente medida em função de, pelo menos, um dos seguintes critérios:

- Custos monetários decorrentes da definição de níveis de stocks;
- Nível de serviço prestado ao cliente.

1.1 Objetivos

Para avaliar a capacidade de generalização de um modelo de previsão, consideram-se normalmente funções baseadas em erros de previsão. Provavelmente, a função mais usada neste contexto é o erro quadrático médio, que é minimizado em função dos parâmetros do modelo. Para funções de erro desse tipo, os erros cometidos por excesso e por defeito têm o mesmo peso.

Em muitas situações práticas, observa-se que os problemas de previsão contêm erros não simétricos, ou seja, os custos decorrentes de errar por excesso são diferentes dos de errar por defeito, exigindo uma correção ex-post dos valores previstos através de stocks de segurança. Dentro da gestão de stocks, por exemplo, prever a procura de um produto por excesso acarreta custos de armazenamento. No caso de errar por defeito, surgem custos de oportunidade, com eventual perda de confiança por parte dos clientes.

Em situações desse tipo, a previsão pode ser realizada tendo como objetivo, não minimizar uma função de erro tradicional, mas sim uma função de utilidade construída para o problema particular, onde se consideram pesos diferentes quando se prevê por defeito e por excesso.

Desta forma, nesta dissertação, pretende-se fazer um estudo comparativo entre essas duas abordagens (erros simétricos e erros assimétricos) como ferramentas de apoio nas tomadas de decisão no contexto da gestão de stocks.

1.2 Revisão da Literatura

No artigo de Ord *et al.* (1997), os métodos de alisamento exponencial mostraram-se aplicáveis sob condições muito mais gerais do que aquelas que eram tradicionalmente previstas na literatura.

Snyder *et al.* (2002) analisaram mais detalhadamente os métodos de alisamento exponencial no contexto da previsão na gestão de stocks e sugerem que os métodos de alisamento exponencial podem ser estendidos a uma classe mais ampla de modelos em

que as variações de erro, em vez de permanecerem constantes, podem mudar ao longo do tempo.

Têm sido desenvolvidos muitos estudos, pesquisas e artigos sobre como os erros de superestimação e de subestimação têm consequências diferentes e incorrem em custos assimétricos. Por exemplo, Dress *et al.* (2018) mencionam no seu estudo que Granger, em 1969, foi o primeiro autor a sugerir que o que se pretende prever no mundo real, raramente é caracterizado por custos baseados em erros quadráticos simétricos. Posteriormente, alguns autores desenvolveram funções de custo assimétricas e demonstraram o seu potencial para melhorar a qualidade da decisão por meio de experiência empírica (por exemplo, Diebold e Mariano, 1995, e Leitch e Tanner, 1991).

Crone (2003) usou redes neurais artificiais com funções de custo assimétricas na previsão de séries temporais, mais especificamente, na otimização do reabastecimento de máquinas de venda automática. Nesse contexto, a assimetria de custos resultou do facto de que um excedente de mercadorias causa despesas mais baixas em comparação com a rutura de stocks. No mesmo estudo, é referido ainda que considerações semelhantes se aplicam a quase todas as cadeias de suprimentos e enfatizam novamente a importância e a prevalência da assimetria de custos em aplicações do mundo real, conhecido há muito tempo, mas não menos relevante hoje.

No âmbito da justiça criminal, Berk (2011) estudou previsões com funções de custo assimétricas, mostrando que o custo de uma superestimação ou de uma subestimação do número de futuros roubos em uma cidade são diferentes. A superestimação pode ser um desperdício, enquanto que a subestimação pode levar a um aumento do crime. A mensagem de Berk é muito clara, na medida em que, quando os custos dos erros de previsão são assimétricos e é implementada adequadamente uma função de custos assimétrica, as previsões informam com maior precisão as atividades de justiça criminal do que quando é usada uma função de custos simétrica.

Zhao *et al.* (2011) propuseram um método de ajuste estendido, utilizando uma função polinomial de um modelo de regressão para fazer a previsão final sensível ao custo. Os resultados impressionantes da avaliação deste método sugerem que poderia ser de grande ajuda para os analistas que tentam ajustar o seu processo de previsão, a utilização de custos assimétricos.

Mais recentemente, Dress *et al.* (2018) estudaram os preços de contratos de leasing de automóveis com a particularidade de que os erros de previsão acarretam custos diferentes. Os resultados empíricos observados em várias comparações forneceram fortes evidências de que ignorar os custos de erros assimétricos prejudicam a qualidade da decisão. Em

várias comparações experimentais, encontraram que a abordagem assimétrica reduz consistentemente os custos de erro em comparação com a abordagem simétrica.

1.3 Estrutura

A presente dissertação encontra-se dividida em quatro capítulos, sendo que o primeiro engloba a introdução, os objetivos e a revisão da literatura.

O capítulo 2 é dedicado à gestão de stocks, onde se introduzem alguns conceitos relevantes e onde são descritos, até uma certa extensão, os diferentes modelos usados num sistema de gestão de stocks. Apresenta-se a questão dos custos associados às situações de ruptura de stocks e de custos de posse quando há armazenamento em excesso, e tem-se a oportunidade de discutir o problema entre minimizar o custo total e maximizar o nível de serviço. É neste contexto que surge a motivação para o tema da tese, onde são definidos custos assimétricos nos problemas de decisão em sistemas de gestão de stocks.

No capítulo 3, serão abordados os temas relacionados com a previsão de séries temporais, com a introdução de conceitos, descrição de padrões presentes nas séries e descrição de métodos de previsão baseados no alisamento exponencial, assim como as medidas de desempenho, tendo por base erros simétricos.

O capítulo 4 é dedicado ao desenvolvimento e exploração do tema da tese propriamente dito, ou seja, a comparação de dois paradigmas no contexto da previsão e decisão num sistema de gestão de stocks.

Finalmente, reservamos o capítulo 5 para apresentar as conclusões finais do presente trabalho e perspectivas de trabalho futuro.

CAPÍTULO 2

2 GESTÃO DE STOCKS

Seja qual for a dimensão de uma determinada empresa, para poder funcionar, necessita que seja assegurado o abastecimento de todos os seus departamentos. Para assegurar esse abastecimento, será necessário, por um lado, fazer compras, e por outro, constituir stocks de determinados materiais ou produtos que devem estar disponíveis quando na empresa se põe a sua necessidade (Braga, 1991).

O termo stock é utilizado para designar as existências de determinados bens ou artigos de que dispõe uma organização ou indivíduo num dado momento, com vista ao cumprimento dos seus objetivos. Esses bens podem ser matérias-primas, mercadorias, produtos acabados ou mesmo quase acabados, entre outros.

Os stocks servem primeiramente para assegurar o consumo regular de um produto, mesmo nas situações em que a procura é irregular. Têm ainda a vantagem de permitir compras em grandes quantidades, beneficiando a empresa de reduções no preço ou de descontos, evitando a incómoda tarefa de compras ou entregas muito frequentes.

Os stocks são utilizados também para prevenir atrasos por parte dos fornecedores, ou seja, dos tempos de entrega das encomendas, e são muito importantes para proteger o sistema contra a incerteza da procura, podendo ser superior ao esperado, ou quaisquer outros imprevistos que possam surgir no futuro, como por exemplo, greves (Zermati, 1986).

Uma vantagem de ter um sistema de gestão de stocks é que permite melhorar o serviço ao cliente, uma vez que a empresa pode garantir a satisfação do cliente ao ter disponível, nesse mesmo instante, os produtos em stock, o que por outro lado permite também que a empresa possa aumentar as vendas.

Segundo Gonçalves (2012), existem diversos tipos de stocks:

- Stocks em curso de fabrico, constituídos por produtos que ainda estão em fabrico e que já implicaram o consumo de recurso;
- Stocks de lote de fabrico, quando a produção é feita por lotes quer seja por economias de escala ou por imposições tecnológicas;

- Stocks sazonais, constituídos para enfrentar variações de procura ao longo do tempo;
- Stocks de segurança, stocks criados para proteger o sistema de situações de incerteza;
- Outros stocks, por exemplo, para separar as várias fases de fabrico de um produto ou antecipar o aumento de preço de determinado artigo.

Baseado no papel de importância primordial que a gestão de stocks representa na gestão das empresas, é natural que surjam algumas questões, tais como:

- Qual é a quantidade ótima que se deve encomendar?
- Quando se deve encomendar?
- Como controlar os stocks?

As soluções que forem adotadas nestas questões terão um impacto não só nos custos globais do sistema de gestão de stock, como também no nível de serviço prestado ao cliente. Portanto, deverá avaliar-se e aplicar-se uma solução de compromisso entre minimizar esses custos e maximizar o nível de serviço.

O nível de serviço representa a qualidade do serviço prestado e é uma medida complementar da taxa de rutura. A taxa de rutura, por sua vez, é um indicador que representa a capacidade de não satisfazer uma encomenda por meio do inventário existente, e é medida pela percentagem da procura que não é satisfeita. Quanto menor o seu valor, maior é o nível do serviço prestado.

2.1 Custos de stocks

Os stocks representam um investimento significativo para as empresas, e por isso, também a sua gestão é de grande interesse. Como refere Gonçalves (2012), um sistema de stocks suporta três tipos de custos:

- Custos de aprovisionamento;
- Custos de posse de stocks;
- Custos associados à rutura de stocks.

Os custos de aprovisionamento englobam todos os custos inerentes à constituição e renovação do stock, incluindo o preço da compra dos produtos ao fornecedor e os custos associados ao processamento da encomenda, por exemplo, as despesas de deslocação ou

transporte, material de escritório, ou os salários dos funcionários nas diversas áreas da empresa, desde o pessoal afeto diretamente à gestão de stocks, passando pelos gabinetes de contabilidade e de controlo de qualidade.

Os custos de posse de stocks referem-se às despesas associadas à própria existência de stocks. As despesas de armazenamento englobam custos de funcionamento dos armazéns, prémios de seguros, perdas de qualidade, perdas por deterioração, custo de obsolescência, etc. Os encargos financeiros também são despesas associadas à posse de stocks quando há pagamento de juros dos empréstimos que recaem sobre os montantes investidos nos stocks.

A rutura de stocks acontece quando não é possível satisfazer a totalidade da procura por parte dos clientes, isto é, quando esta é superior à quantidade que se encontra em stock.

Quando a procura não é satisfeita, é sabido que existem custos que, apesar de serem difíceis de quantificar, são na generalidade muito elevados. Por um lado, com vista a satisfazer a totalidade da procura, a empresa pode ter de fazer uma encomenda especial, incorrendo, neste caso, num custo adicional de encomenda. Por outro, se a empresa não conseguir satisfazer a procura, além da perda de vendas (custo de oportunidade) e consequente perda de lucros, pode significar também a perda de imagem junto do mercado, perda de clientes e a não fidelização dos mesmos.

2.2 Modelos de gestão de stocks

Existem diversos modelos de gestão de stocks, os quais tentam responder, de forma distinta, às questões do quanto e quando encomendar. Dependendo das características da procura, as empresas devem escolher o modelo de gestão de stocks que melhor satisfaça os seus propósitos.

Numa situação ideal, claramente pouco realista, a procura e o tempo de entrega por parte dos fornecedores são conhecidos e constantes. Este tipo de situação é modelado usando modelos determinísticos. Quando se verificarem flutuações aleatórias na procura ou no tempo de entrega, ou em ambos, devem ser considerados modelos mais realistas, nomeadamente, modelos estocásticos.

2.2.1 Modelos determinísticos

Nos modelos determinísticos, todas as variáveis e parâmetros envolvidos são conhecidos e constantes, isto é, não existe aleatoriedade (Balakrishnan, 2013). No caso mais simples, consideram-se os seguintes pressupostos:

- Apenas um tipo de artigo (com vista à simplicidade);
- Procura determinística, contínua e constante no tempo;
- Horizonte temporal infinito, isto é, o processo continua indefinidamente;
- Não existe qualquer tipo de restrição (quer nas quantidades ou na capacidade de armazenamento, entre outros);
- O prazo de aprovisionamento, ou tempo de entrega (tempo desde o momento que a encomenda é colocada até à sua receção), é constante;
- Os custos não variam com o tempo;
- Reposição instantânea;
- Não são permitidas ruturas de stock;
- Não existem descontos de quantidade.

Quando se fala em reposição instantânea de stock, significa que o fornecimento (ou entrega) é realizado de uma só vez e pela totalidade. No caso em que a reposição é faseada ou gradual, diz-se que não é instantânea. Na prática, uma situação onde a reposição é não instantânea é o caso em que os artigos são produzidos pela própria empresa, a uma taxa fixa por unidade de tempo.

O modelo básico, onde se assume os pressupostos anteriormente descritos, é conhecido como Quantidade Económica de Encomenda (do inglês, *Economic order quantity*). Por vezes, é também designado por Quantidade Económica de Wilson (embora tendo sido originalmente desenvolvido por Harris (1913), foi Wilson (1934) quem o aplicou extensivamente e o divulgou). Este modelo consiste em definir uma quantidade fixa de encomenda, com o objetivo de minimizar o custo total. Este modelo permite resolver um dos maiores problemas da gestão de stocks, que é, quanto encomendar?

O modelo da Quantidade Económica de Encomenda é ilustrado na Figura 1. A quantidade encomendada, Q , depende tanto da procura como dos custos do stock, mas também depende do tempo que demora para obter-se uma nova encomenda. De modo a determinar quando voltar a encomendar, as empresas necessitam de definir um nível mínimo de

stock, conhecido como ponto de encomenda. Assim, quando o nível de existências do stock atingir o nível predeterminado, é feita uma nova encomenda.

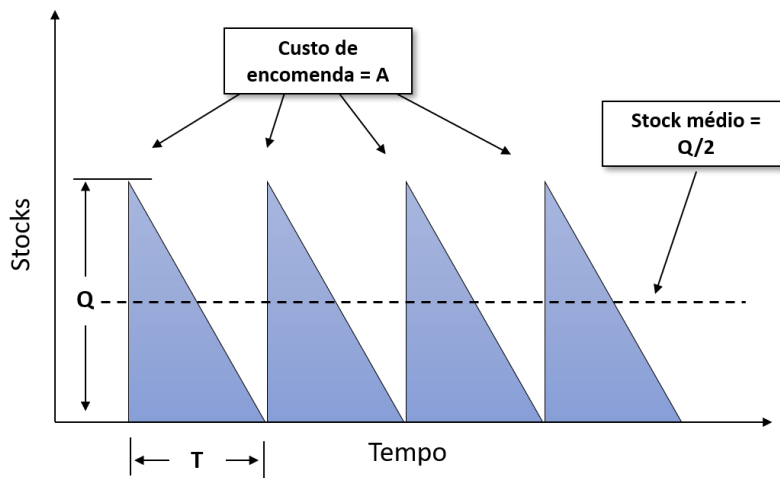


Figura 1 – Modelo da Quantidade Económica de Encomenda [Figura adaptada de Gonçalves, 2012].

Os custos associados a este modelo podem ser agrupados em duas categorias e variam com as quantidades encomendadas e armazenadas (Gonçalves *et al.*, 2008):

- Custos de encomenda – constituídos por duas componentes, uma de custo fixo, independente da quantidade adquirida (por exemplo, custo de transporte, materiais de escritório, despesas gerais associadas à colocação da encomenda, etc.), e outra de custo variável, proporcional à quantidade adquirida;
- Custos de posse – como já foi referido anteriormente, são custos associados à manutenção do stock, como despesas com seguros ou despesas com o local de armazenamento.

Para a definição do custo total e, em particular, das várias componentes de custo envolvidos num sistema de gestão de stocks, vamos usar a seguinte notação:

- Q = quantidade a encomendar por encomenda (em unidades de produto);
- T = duração do ciclo ou período, isto é, intervalo de tempo entre duas entregas ou aprovisionamentos consecutivos (em unidades de tempo);
- A = custo fixo de encomenda (em unidades monetárias);
- c_e = custo unitário do produto (em unidades monetárias por unidade de produto);

- I = quantidade de posse durante um período (em unidades de produto);
- c_p = custo de posse (em unidades monetárias por unidade de produto e unidade de tempo);
- r = taxa de procura (em unidades de produto por unidade de tempo);
- K = custo total (em unidades monetárias por unidade de tempo).

O objetivo é encontrar a quantidade a encomendar que minimize o custo total. O custo total é dado pela soma dos custos de encomenda e custos de posse. De acordo com o modelo básico, os custos associados a uma encomenda são definidos pela seguinte expressão:

$$\text{custos fixos} + \text{custo da encomenda} = A + c_e Q$$

Relativamente aos custos de posse, estes podem ser definidos pelo seguinte:

$$\text{custos de posse por período} = c_p I$$

Juntando estas duas componentes de custo, o custo total associado ao modelo básico vem:

$$K = \frac{\text{custos globais por período}}{\text{duração do ciclo}} = \frac{(A + c_e Q) + c_p I}{T}$$

Uma vez que o modelo básico é determinístico, facilmente verificamos que a quantidade a encomendar é definida pelo produto da taxa de procura e o intervalo de tempo entre duas encomendas, ou seja, $Q = rT$. Por outras palavras, o “quando” define o “quanto”, ou vice-versa, desde que se conheça a taxa de procura. Tem-se a relação $r = \frac{Q}{T}$.

Verificamos também que a quantidade de stock durante um período corresponde à área de um triângulo ilustrado na Figura 1, ou seja, $I = \frac{QT}{2}$. Consequentemente, o custo total resulta em:

$$K(Q) = \underbrace{\left(\frac{Ar}{Q} + c_e r \right)}_{\text{custos de encomenda}} + \underbrace{c_p \frac{Q}{2}}_{\text{custos de posse}} \quad (1)$$

Sem perda de generalidade, para encontrarmos o valor de Q que minimiza os custos globais, é suficiente considerar apenas os termos de K que variam com Q . Assim, os

custos de encomenda relevantes são dados por $\frac{Ar}{Q}$, de onde o custo relevante vem definido por: $K'(Q) = \frac{Ar}{Q} + c_p \frac{Q}{2}$. O ponto mínimo desta função vem $Q^* = \sqrt{\frac{2Ar}{c_p}}$, que corresponde ao ponto de intersecção entre a curva dos custos de encomenda relevantes e a dos custos de posse (ver Figura 2).

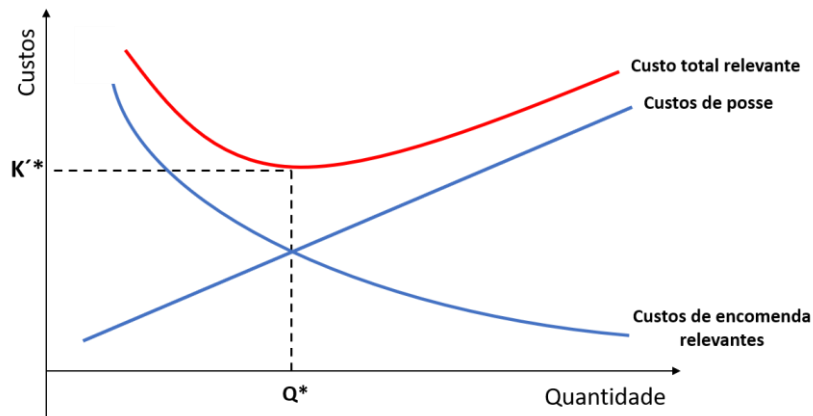


Figura 2 – Gráfico do custo total relevante em função das quantidades [Figura adaptada de vários autores].

Este modelo possui algumas limitações pelo facto de os pressupostos, muitas vezes, não serem realistas do ponto de vista prático. A procura e o tempo de aprovisionamento podem não ser constantes, assim como os custos de posse e de encomenda, podendo haver, por exemplo, descontos para compras de várias unidades. Outra limitação é não ser possível aplicar o método para mais do que um produto.

Uma variante ao modelo básico é permitir situações de rutura, mantendo todos os outros pressupostos inalterados. Esta situação possibilita que as encomendas sejam efetuadas após rutura de stocks e sejam satisfeitas assim que o produto esteja novamente disponível. Isto leva a empresa a suportar custos de rutura resultantes da entrega tardia ao cliente, mas em contrapartida, permite reduzir os custos de posse.

Para esta situação, vamos considerar a seguinte notação adicional:

- S = quantidade de rutura durante um período (em unidades de produto);
- c_r = custo de rutura (em unidades monetárias por unidade de produto e unidade de tempo).

Os custos de rutura, por período, são definidos por $c_r S$. Juntando as três componentes de custo, definidos até o momento, o custo total vem:

$$K(Q, S) = \frac{(A + c_e Q + c_p I) + c_r S}{QT} = \underbrace{\left(\frac{Ar}{Q} + c_e r \right)}_{\text{custos de encomenda}} + \underbrace{c_p \frac{(Q-S)^2}{2Q}}_{\text{custos de posse}} + \underbrace{c_r \frac{S^2}{2Q}}_{\text{custos de rutura}} \quad (2)$$

As variáveis de decisão são a quantidade a encomendar e a quantidade em falta durante um período. Após minimizar K , em função a Q e S , as decisões ótimas são:

$$Q^* = \sqrt{\frac{2Ar}{c_p}} \sqrt{\frac{c_p + c_r}{c_r}} \quad \text{e} \quad S^* = \sqrt{\frac{2Arc_p}{c_r(c_p + c_r)}} \quad (3)$$

Neste modelo, a taxa de procura é também constante, sendo a quantidade encomendada reposta de uma forma instantânea com intervalos de tempo fixos. A duração do ciclo pode ser decomposta em dois intervalos de tempo, um representando o intervalo em que não ocorre rutura, T_1 , e outro o intervalo em que há rutura, T_2 (ver Figura 3).

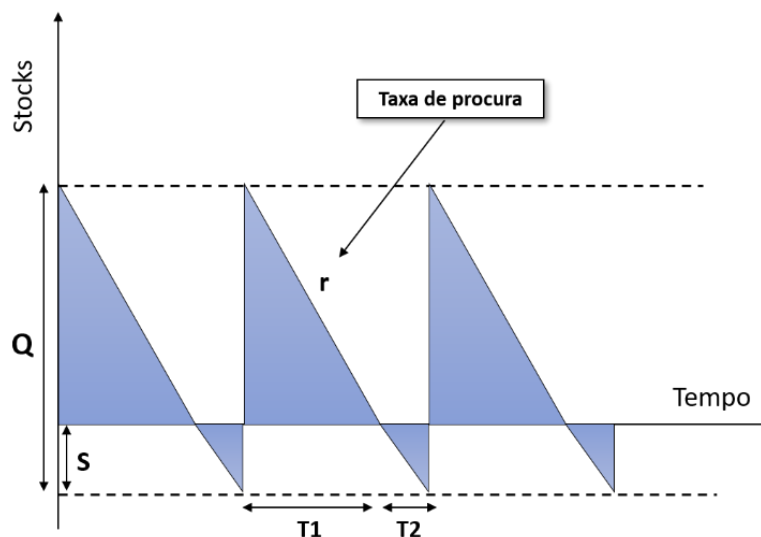


Figura 3 – Modelo com reposição instantânea e com rutura permitida [Figura adaptada de Gonçalves, 2012].

Se considerarmos uma reposição não instantânea e sem rutura permitida, obtemos uma nova variante aos modelos anteriores. Neste caso, assume-se uma taxa de reposição finita, denotada por p ($p > 0$), onde existe um período durante o qual a encomenda é entregue, ou seja, as quantidades encomendadas são entregues de forma faseada. Um exemplo de aplicação deste modelo é quando os produtos são de produção própria da empresa, em vez de serem encomendados. Assim, a taxa de fornecimento corresponde à taxa de produção dos itens. A Figura 4 ilustra esta situação, onde T_1 representa o tempo durante o qual há produção, havendo também consumo, T_2 corresponde ao tempo durante o qual há apenas consumo e I o stock máximo.

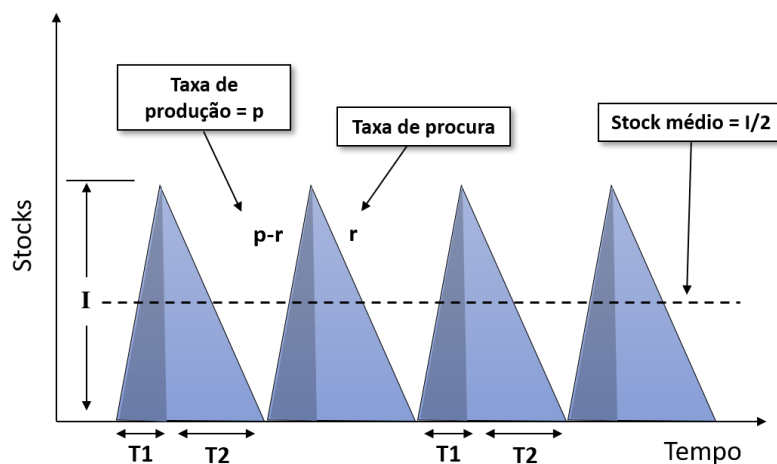


Figura 4 – Modelo com reposição não instantânea e sem rutura permitida [Figura adaptada de Gonçalves, 2012].

A quantidade máxima em stock é dada pela quantidade produzida subtraída do consumo durante o tempo T_1 . Assume-se que a taxa de produção é constante durante os períodos de fornecimento, e que essa taxa é superior à taxa de procura. De modo contrário, não haveria capacidade de satisfazer a procura, estando os níveis de existência dos produtos sempre em rutura.

Neste modelo, não havendo lugar a ruturas de stock, existe apenas uma variável de decisão a ter em conta, isto é, a quantidade a encomendar. Por um lado, a quantidade máxima em stock é dada por $I_{\max} = (p-r)T_1$. Por outro, a quantidade produzida é igual à produção durante o período T_1 , isto é, $Q = pT_1$, sendo, por sua vez, igual ao consumo durante o período T , ou seja, $Q = rT$. O custo total vem:

$$K(Q) = \underbrace{\frac{Ar}{Q} + c_e r}_{\text{custos de encomenda}} + \underbrace{c_p \left(1 - \frac{r}{p}\right) \frac{Q}{2}}_{\text{custos de posse}} \quad (4)$$

A decisão ótima resultante é dada pela expressão:

$$Q^* = \sqrt{\frac{2Ar}{c_p}} \sqrt{\frac{p}{p-r}} \quad (5)$$

Por último, podemos considerar a variante onde são permitidas ruturas de stock, com reposição não instantânea. Este modelo corresponde à variante mais genérica de entre os modelos determinísticos. Assim, considera-se que existe um período durante o qual a encomenda é produzida, T_1 e T_2 , e um período em que há apenas consumo, T_3 e T_4 , havendo nesses períodos dois intervalos de tempo com rutura, T_1 e T_4 , respetivamente, de acordo com a ilustração da Figura 5.

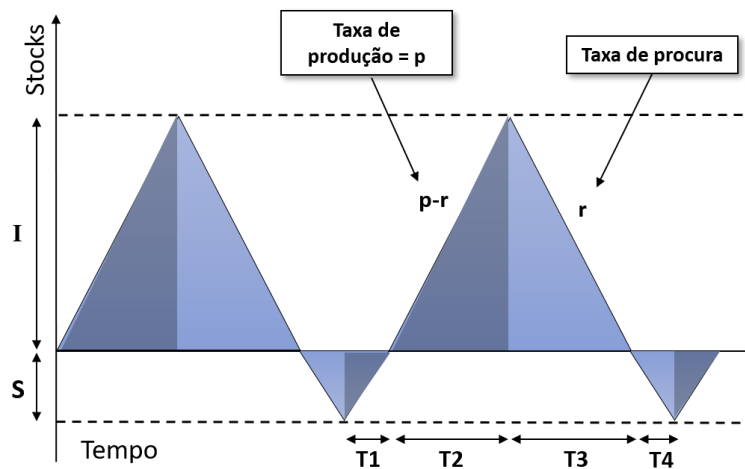


Figura 5 – Modelo com reposição não instantânea e com rutura permitida [Figura adaptada de vários autores].

Neste modelo, assume-se que a taxa de produção é também constante e superior à taxa de procura. As variáveis de decisão são a quantidade a encomendar e a quantidade em falta num período. A expressão que define o custo total é dada por:

$$K(Q, S) = \underbrace{\frac{Ar}{Q} + c_e r}_{\text{custos de encomenda}} + c_p \underbrace{\left[\left(1 - \frac{r}{p}\right) \frac{Q}{2} + \left(\frac{p}{p-r}\right) \frac{S^2}{2Q} - S \right]}_{\text{custos de posse}} + c_r \underbrace{\left(\frac{p}{p-r}\right) \frac{S^2}{2Q}}_{\text{custos de rutura}} \quad (6)$$

As decisões ótimas correspondentes são dadas por:

$$Q^* = \sqrt{\frac{2Ar}{c_p}} \sqrt{\frac{p}{p-r}} \sqrt{\frac{c_p + c_r}{c_r}} \quad \text{e} \quad S^* = \sqrt{\frac{2Arc_p}{c_r(c_p + c_r)}} \sqrt{\frac{p-r}{p}} \quad (7)$$

Observe-se que, as variantes anteriores podem ser consideradas como casos particulares desta última. Por exemplo, se considerarmos a rutura não permitida, que é o mesmo que assumir $c_r = +\infty$, a decisão ótima resultante é dada pela fórmula (5). Se, em contrapartida, a reposição for instantânea, que é o mesmo que considerar $p = +\infty$, a decisão ótima reduz-se às fórmulas (3).

2.2.2 Modelos estocásticos

Como já foi mencionado, os modelos determinísticos são pouco realistas, pois não tomam em consideração o risco e a incerteza da procura, situação recorrente no mundo real. Para esses casos, devem ser aplicados modelos mais adequados e complexos, como é o caso dos modelos estocásticos, onde se sabe que tanto a procura como os tempos de entrega são dinâmicos, variando ao longo do tempo (Renard *et al.*, 2013).

Devido à variação da procura, existe a possibilidade de acontecer ruturas de stocks e, por isso, a empresa tem de definir os chamados stocks de segurança, cuja finalidade é proteger o sistema contra eventuais aumentos da procura acima do previsto, bem como de eventuais atrasos de entrega imputáveis aos fornecedores. Mesmo assim, devido ao carácter inerente à própria imprevisibilidade da realidade, existe sempre a possibilidade da rutura não ser evitada. O nível de proteção da empresa perante estas situações adversas dependerá do valor definido para o stock de segurança.

Para definir o stock de segurança, o nível de serviço é de grande importância. Como já tinha sido referido, o nível de serviço representa a qualidade do serviço prestado ao cliente, e corresponde à probabilidade de a empresa conseguir fornecer, no momento, a quantidade procurada. O nível de serviço é então calculado dividindo a quantidade

entregue num determinado intervalo de tempo pela quantidade encomendada nesse mesmo intervalo de tempo (incluindo ruturas).

Dentro dos modelos estocásticos com horizonte temporal infinito, podemos definir dois tipos básicos de sistemas ou políticas de controlo de stocks:

- Sistema de revisão contínua;
- Sistema de revisão periódica.

O sistema de revisão contínua consiste na verificação assídua das quantidades disponíveis de um determinado produto. Sempre que se verifica uma transação, atualizam-se os níveis de stock. Assim, quando os níveis de existência ficam abaixo de um nível predeterminado, conhecido como ponto de encomenda, é colocado um novo pedido para repor as existências de produtos (Hillier *et al.*, 2001).

Os pressupostos deste sistema são:

- Apenas um tipo de artigo (com vista à simplicidade);
- Horizonte temporal infinito;
- Reposição instantânea;
- Rutura permitida;
- Procura aleatória com distribuição conhecida;
- Tempo de entrega aleatório com distribuição conhecida.

Neste sistema, a quantidade de encomenda é fixa, enquanto que os períodos entre as encomendas são de dimensão variável, uma vez que dependem do momento em que é atingido o ponto de encomenda.

Para caracterizar a política de revisão contínua (ver esquema da Figura 6), vamos considerar a notação seguinte:

- SS = stock de segurança (em unidades de produto);
- M = ponto de encomenda (em unidades de produto);
- R_i = variável aleatória correspondente à taxa de procura num dado instante i (unidades de produto por unidade de tempo), de valor médio ρ e variância σ_R^2 (observe-se que R_i e R_j são independentes para $i \neq j$);

- $L =$ variável aleatória correspondente ao tempo de entrega (em unidades de tempo), de valor médio λ e variância σ_L^2 ;
- $X = \sum_{i=1}^L R_i =$ variável aleatória correspondente à procura durante o tempo de entrega (em unidades de produto), de valor médio μ e variância σ^2 .

Para que a política de revisão contínua fique completamente definida, é indispensável conhecer as características estatísticas da procura durante o tempo de entrega. No caso mais simples, para um tempo de entrega fixo L , a procura durante o tempo de entrega tem valor médio:

$$\mu = E\left(\sum_{i=1}^L R_i\right) = \sum_{i=1}^L E(R_i) = L\rho \quad (8)$$

e variância:

$$\sigma^2 = Var\left(\sum_{i=1}^L R_i\right) = \sum_{i=1}^L Var(R_i) = L\sigma_R^2 \quad (10)$$

No caso geral, para tempos de entrega aleatórios, a variável X é uma variável aleatória composta, onde o valor médio é dado por $\mu = \lambda\rho$ e a variância toma a forma:

$$\sigma^2 = \lambda\sigma_R^2 + \sigma_L^2\rho^2 \quad (9)$$

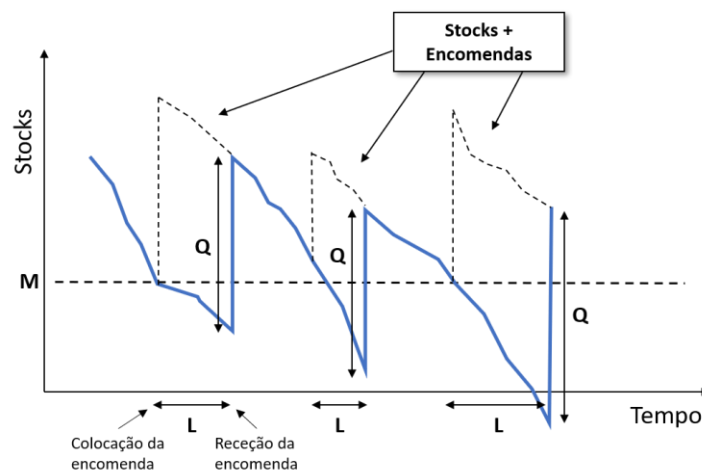


Figura 6 – Sistema de Revisão Contínua [Figura adaptada de Gonçalves, 2012].

O ponto de encomenda é definido como:

$$M = SS + \mu \quad (12)$$

Se a procura durante o tempo de entrega tiver uma distribuição Normal $N(\mu, \sigma)$, o stock de segurança vem expresso como:

$$SS = z_\alpha \sigma \quad (103)$$

onde α é a probabilidade de rutura e z_α o fator de segurança (obtido a partir da tabela da distribuição Normal padrão). A Figura 7 esquematiza estas quantidades usando a distribuição Normal.

Como exemplo, se a probabilidade de acontecer rutura de stocks é de 1%, o nível de serviço será de 99%. Assim, o stock de segurança a constituir pela empresa deverá ser maior, quanto maior for o nível de serviço que pretende prestar aos seus clientes. O ideal seria ter um nível serviço muito próximo de 100%, mas claro que isto implicaria para a empresa ter de suportar custos mais elevados.

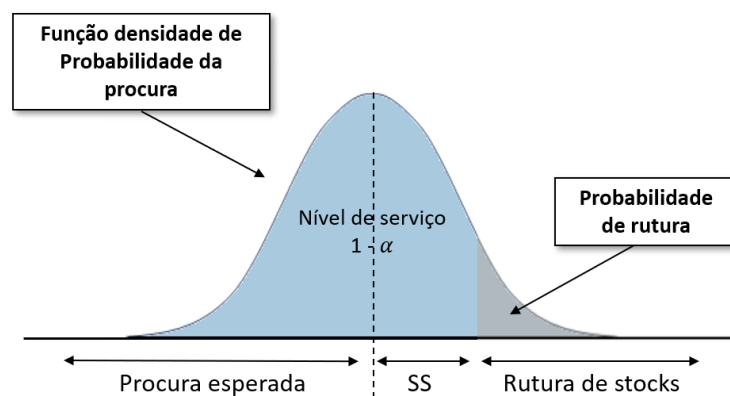


Figura 7 – Representação gráfica do stock de segurança, nível de serviço e probabilidade de rutura [Figura adaptada de Gonçalves, 2012].

Para funcionar corretamente, o sistema de revisão contínua precisa de uma manutenção constante da informação relativa aos stocks, para garantir que a encomenda é colocada quando se atinge o ponto de encomenda, e por forma a evitar possíveis ruturas de stocks. De acordo com Gonçalves (2012), existem situações em que não é económico fazer o

registo de todos os movimentos de stock, contudo, os sistemas informáticos tornam esta tarefa mais fácil e rápida.

Relativamente ao sistema de revisão periódica, as quantidades de stock são verificadas apenas em determinados intervalos de tempo predefinidos (semanal, mensal, etc.). Por outras palavras, ao contrário do que acontece no sistema de revisão contínua, o período entre as encomendas é fixo, sendo as quantidades encomendadas variáveis.

Em cada período de revisão do nível das existências, coloca-se uma nova encomenda. Desta forma, a quantidade a encomendar vai depender do stock atual e do valor máximo de stock predefinido, sendo este o mesmo ao longo do tempo. O stock máximo deve cobrir todas as necessidades até o instante em que é recebida a nova encomenda para o período seguinte. A Figura 8 esquematiza o sistema de revisão periódica (aqui, a letra I representa o stock máximo permitido).

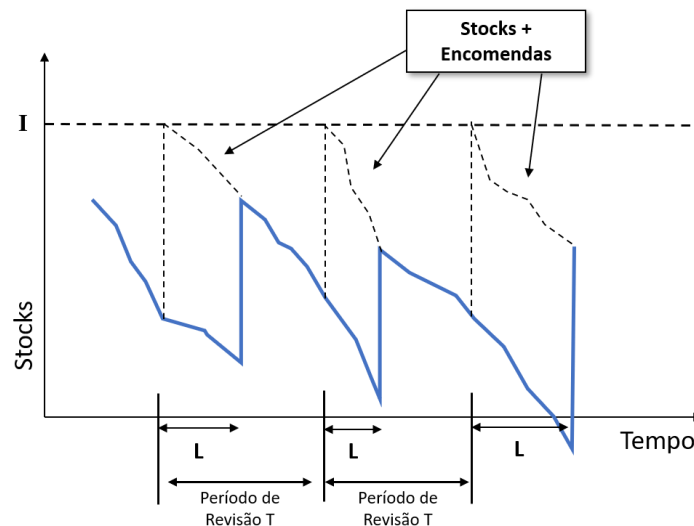


Figura 8 – Sistema de Revisão Periódica [Figura adaptada de Gonçalves, 2012].

Tendo em conta a incerteza da procura e do tempo de entrega, no sistema de revisão periódica, também é constituído um stock de segurança, expresso da mesma forma que (103). Neste caso, e de forma análoga, a procura média durante o tempo de entrega e a respetiva variância são dadas por, respetivamente:

$$\mu = (T + \lambda)\rho \tag{114}$$

$$\sigma^2 = (T + \lambda)\sigma_R^2 + \sigma_L^2\rho^2 \tag{125}$$

onde T corresponde ao período predefinido entre encomendas.

Uma desvantagem do sistema de revisão periódica é o facto de serem necessários níveis de stocks mais elevados, de forma a cobrir a incerteza durante todo o intervalo entre períodos de revisão, originando, por sua vez, custos de manutenção mais elevados para a empresa. Por outro lado, de acordo com Gonçalves (2012), pode ser vantajoso usar este sistema quando existem diversos produtos providos pelo mesmo fornecedor, o que possibilita a encomenda de vários produtos na mesma altura, resultando numa redução de custos de transporte e de custos com o processamento das encomendas.

Dentro dos modelos estocásticos, temos ainda o modelo de período único, que pode ser usado quando um artigo apresenta procura incerta e é encomendado ou produzido somente uma vez (jornais, por exemplo), ou quando são produtos perecíveis, que não podem ser mantidos em stock para serem vendidos no futuro. No final do período, o produto terá sido vendido na totalidade ou, então, a ser possível, o excedente ser vendido com um valor residual.

Os pressupostos deste modelo são:

- Apenas um tipo de artigo (suscetível de perder valor após o período de referência);
- Um único período em consideração;
- A procura total X no período tem distribuição F conhecida;
- No início do período pode já existir em stock uma certa quantidade, I_0 ;
- A quantidade a encomendar, Q_0 , é garantida estar disponível em stock no início do período.

A única decisão a ser tomada neste modelo é a quantidade a encomendar ou produzir no início do período, de forma a minimizar os custos globais. Esta decisão é tomada fazendo uma comparação entre o custo de encomendar uma unidade adicional e o custo de encomendar uma unidade a menos.

Para caracterizar esta política de período único, vamos considerar a seguinte notação, com alguns pressupostos:

- X = variável aleatória correspondente à procura durante o período;
- $Q = I_0 + Q_0$ quantidade de produto total disponível no início do período;
- A = custo fixo de encomenda (podem ser nulos ou não);

- c_e = custo unitário de aquisição (não dependente da quantidade a encomendar);
- c_p = custo unitário de posse (só contabilizável para a quantidade em excesso);
- p_v = preço unitário de venda durante o período;
- p_r = preço residual unitário pós-período (“salvados”, “saldos”, etc.);
- Tipicamente, assume-se que $0 < p_r < c_e < p_v$.

No final do período, se acontecer $X > Q$, irão faltar unidades de produto para cobrir o total da procura, ocorrendo o chamado *understocking*. Se, pelo contrário, acontecer $X < Q$, irão existir unidades de produto em excesso, ocorrendo *overstocking*.

Assim, podemos definir os seguintes custos:

- Custo por unidade de produto em falta no fim do período:

$$c_u = p_v - c_e \quad (16)$$

Este custo representa a perda de lucro pelas unidades que poderiam ter sido vendidas se estivessem disponíveis.

- Custo por unidade de produto em excesso no fim do período:

$$c_o = c_e - p_r + c_p \quad (137)$$

Este custo representa a perda de capital pelas unidades adicionais que não foram vendidas.

O problema de decisão da política de período único corresponde a encontrar o melhor valor de Q que minimize os custos globais esperados no período, ou seja:

$$\hat{K}(Q) = E[K(Q, X)] \quad (148)$$

Vamos considerar o caso particular em que, no início do período, não existe nenhuma quantidade de produto em stock ($I_0 = 0$), e que, sem perda de generalidade, o custo fixo de encomenda é nulo ($A = 0$). Supondo que se encomendava Q e se verificava uma procura x , os custos globais vêm:

$$K(Q, x) = c_u \times \max \{x - Q, 0\} + c_o \times \max \{Q - x, 0\} \quad (159)$$

Para o caso de variáveis aleatórias contínuas X , os custos esperados tomam a forma:

$$E[K(Q, X)] = c_u \int_Q^{+\infty} (x - Q) f(x) dx + c_o \int_0^Q (Q - x) f(x) dx \quad (20)$$

onde f denota a função densidade de probabilidade da procura.

Derivar a expressão de \hat{K} em ordem a Q e igualar a zero, obtém-se:

$$-c_u (1 - F(Q)) + c_o F(Q) = 0 \quad (216)$$

ou seja, a decisão ótima é tal que:

$$F(Q^*) = \frac{c_u}{c_u + c_o} \quad (172)$$

No caso de já existir alguma quantidade de produto em stock ($I_0 > 0$), a quantidade ótima a encomendar vem:

$$Q_0^* = \max \{Q^* - I_0, 0\} \quad (183)$$

Podemos analisar este problema de outra forma. Por um lado, o custo esperado associado ao *understocking* é dado por:

$$c_u \times P(X > Q) = c_u (1 - F(Q)) \quad (194)$$

Por outro, o custo esperado de ocorrer *overstocking* é:

$$c_o \times P(X < Q) = c_o F(Q) \quad (205)$$

Por outras palavras, a situação ideal de encomenda corresponde ao equilíbrio entre os custos esperados de *understocking* e *overstocking*.

Em suma, a quantidade ótima Q^* corresponde a um quantil de probabilidade definido em função dos custos de ocorrer *understocking* e *overstocking*. Esta decisão ótima é a base que consideramos no nosso trabalho, onde se define uma função de utilidade (ver secção 4.2) com base em custos assimétricos.

O custo de subestimar a procura é, em geral, mais difícil de determinar do que o custo de sobrestimar. Isto deve-se ao facto de, quando há ruturas de stocks, o custo poder incluir uma perda de fidelidade do cliente, uma vez que este não conseguiu adquirir o produto naquele instante.

Neste modelo, dá-se a possibilidade então de avaliar qual seria o melhor cenário para a empresa, através de uma análise de custos de rutura de stocks (*understocking*) versus custos de armazenamento em excesso (*overstocking*). Assim, a decisão ótima dependerá da quantidade que minimiza os custos totais, como foi referido anteriormente.

Em todos os modelos vistos nesta secção, pretendeu-se encontrar a quantidade ótima a encomendar que minimize o custo total. Contudo, se a política da empresa for oferecer um elevado nível de serviço aos seus clientes, esses custos, irão certamente aumentar.

De acordo com Pozo (2002), o aumento dos custos de stock tende a ser exponencial em relação à proximidade de um nível de serviço de 100%, levando os custos a valores quase insuportáveis para a empresa. Apesar disto, é importante avaliar se esse aumento dos custos poderia ser compensado pelo aumento da procura resultante de um nível de serviço mais elevado. Sistemas de gestão de stocks eficientes têm um impacto muito positivo na empresa, na medida em que uma sucessão de decisões ótimas, no que diz respeito às quantidades encomendadas, influenciando, por sua vez, tanto os custos de stocks como o nível de serviço escolhido, podem evitar ruturas e perdas substanciais para a empresa.

CAPÍTULO 3

3 PREVISÃO DE SÉRIES TEMPORAIS

3.1 A procura

Como foi visto anteriormente, a incerteza associada à tomada de decisão na gestão de stocks, resulta do facto de tanto a procura como o prazo de entrega, serem variáveis aleatórias. No caso da procura, ela é afetada por diferentes fatores. Gonçalves (2012) refere alguns desses fatores que podem contribuir para a incerteza:

- Fatores externos à empresa:
 - Os ciclos de negócio;
 - O ciclo de vida dos produtos;
 - Concorrência;
 - Época do ano (sazonalidade).
- Fatores internos à empresa:
 - Preços;
 - Qualidade;
 - Publicidade;
 - Reputação do produto, da marca e da empresa.

Com vista a diminuir o fator de incerteza na procura, podem fazer-se estimações usando diferentes modelos e/ou métodos de previsão. Na gestão de stocks, as previsões são um pré-requisito na generalidade das decisões baseadas no planeamento. Assim, a qualidade de uma previsão é avaliada considerando a sua capacidade de melhorar a decisão resultante.

Quando se fala em previsão, o objetivo consiste na geração de uma sequência de valores da série para instantes onde não se conhecem observações. Pelo facto de os valores resultantes da previsão não serem valores reais, mas sim estimativas, podem verificar-se dois tipos de erros associados:

- Errar por defeito – acontece quando a procura prevista é inferior à procura observada, resultando em custos de ruturas de stocks;

- Errar por excesso – acontece quando a procura prevista é superior à procura observada, resultando em aumentos de custos de posse.

Em geral, as rupturas de stocks têm consequências mais graves do que os excessos de stocks, daí que a solução de compromisso que normalmente se utiliza, como foi visto, seja a de ter um stock de segurança, destinado a proteger o sistema contra os custos associados aos erros de previsão, nos eventuais aumentos da procura acima do previsto, bem como de eventuais atrasos de entrega imputáveis aos fornecedores.

Existem também casos em que os custos de posse são muito elevados e, portanto, pode ser preferível para a empresa incorrer com custos associados a rupturas de stocks. Tudo isto influencia na hora da tomada da decisão sobre as quantidades a encomendar ou a frequência das encomendas.

No caso de existirem dados históricos suficientes da procura, deve utilizar-se um modelo de previsão quantitativo, como é o caso dos modelos de séries temporais, que têm como objetivo, identificar e caracterizar padrões presentes nos dados para posteriormente extrapolar esses padrões para o futuro.

3.2 Séries temporais

De acordo com Granger *et al.* (1986), uma série temporal é uma sequência de observações ordenadas por um parâmetro de tempo (intervalos de tempo), e que pode ser medida contínua ou discretamente. Neste trabalho, apenas se irá trabalhar com dados discretos e com intervalos de tempo regulares, isto é, intervalos medidos diariamente, semanalmente, mensalmente, trimestralmente, ou com outra periodicidade. Um exemplo de uma série temporal deste tipo, observada trimestralmente, é apresentado na Figura 9.

Numa série temporal, podem ser reconhecidos até quatro tipos de padrões diferentes, e no caso da gestão de stocks, mais propriamente quando falamos da procura de um produto, temos:

- Tendência (T) – verifica-se quando existe um aumento ou uma diminuição do nível médio da procura ao longo do tempo, podendo ser também horizontal. A tendência descreve o comportamento mais notório da série (crescente, decrescente ou constante) durante o período considerado.
- Sazonalidade (S) – representa uma flutuação da procura acima ou abaixo do nível médio e que se repete consistentemente ao longo do tempo. Acontece

quando a série é influenciada por fatores sazonais, isto é, a procura de produtos de acordo com a estação do ano, com o mês ou o dia da semana. Um exemplo de sazonalidade pode ser a procura por gelados, uma vez que se verifica, todos os anos, um aumento durante os meses de verão.

- Variações cíclicas (C) – representam também flutuações de subidas e descidas da procura, mas estas repetem-se com uma frequência superior a um ano e sem um período fixo. Este movimento é normalmente atribuído aos ciclos de negócio. De acordo com Makridakis *et al.* (1998), a principal distinção entre um padrão sazonal e um padrão cíclico é que o primeiro tem duração constante e se repete regularmente, enquanto o segundo varia em comprimento. Além disso, a duração média de um ciclo é geralmente mais longa que a da sazonalidade e a magnitude de um ciclo é geralmente mais variável do que a da sazonalidade.
- Variações irregulares (I) – representam flutuações de natureza aleatória de curto prazo que não seguem nenhum padrão, decorrentes de uma variedade de fatores, contendo também a parte aleatória da variável. Podem representar o que não se consegue integrar nas componentes anteriores.

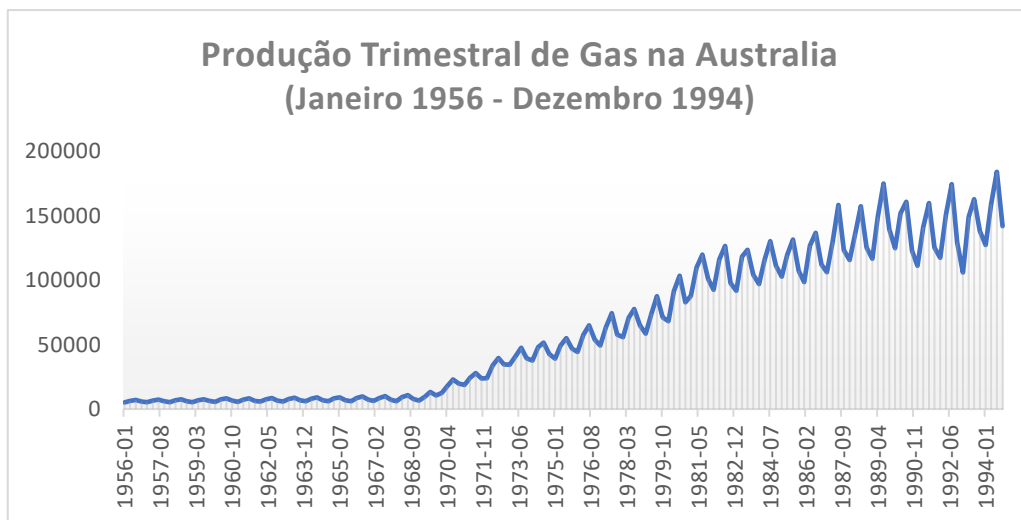


Figura 9 – Exemplo de uma série temporal [Fonte: datamarket.com].

De acordo com a decomposição clássica, uma série temporal pode ser decomposta nestas componentes básicas, o que permite escrevê-la como função das mesmas, ou seja:

$$Y = f(T, C, S, I) \quad (216)$$

onde Y representa a série temporal e f o modelo funcional que define a forma como estas componentes se combinam entre si para gerarem a série. A Figura 10 ilustra as várias componentes de uma série real, nomeadamente, as componentes de tendência, de sazonalidade e irregular.

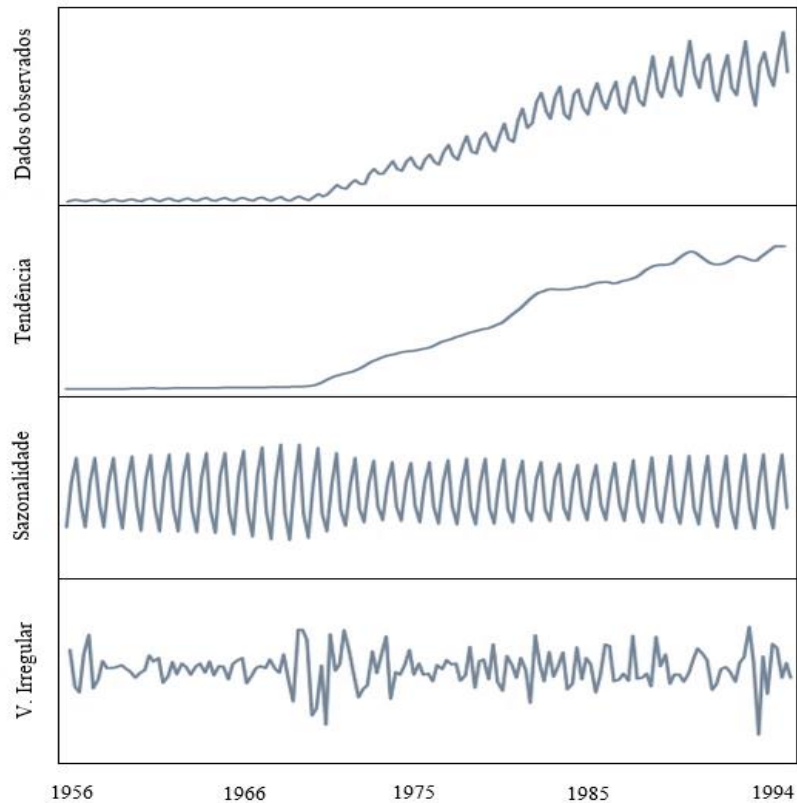


Figura 10 – Produção trimestral de gás (milhões de megajoules) na Austrália (janeiro 1956 a dezembro 1994). Os painéis abaixo da série original são as componentes de tendência, de sazonalidade e irregular, que, em conjunto, conformam a série observada [Fonte: datamarket.com].

Considerando o efeito das quatro componentes, podemos identificar diferentes modelos de decomposição para as séries temporais, a depender da forma como as mesmas são combinadas entre si. Assim, temos:

- Modelo multiplicativo:

$$Y = T \times C \times S \times I \quad (227)$$

Este modelo é apropriado para situações onde a amplitude da variação sazonal é proporcional ao ciclo da série.

- Modelo aditivo:

$$Y = T + C + S + I \quad (238)$$

Este modelo é mais adequado quando a amplitude da variação sazonal é independente do termo $T + C$.

- Modelos mistos, por exemplo:

$$Y = (T + C) \times S + I \quad (249)$$

Este modelo não é mais do que uma combinação de termos aditivos e multiplicativos, definida em concordância com a situação identificada em cada caso.

3.3 Previsão

Um dos objetivos mais importantes no estudo das séries temporais é a previsão de valores futuros. Podemos identificar dois tipos de previsão:

- Previsão para “dentro da amostra” (do inglês, *in-sample*): é aquela que, dadas as estimativas dos parâmetros, produz resultados para o intervalo temporal da amostra da série temporal;
- Previsão para “fora da amostra” (do inglês, *out-of-sample*): corresponde ao valor futuro obtido da variável aleatória para uma data “fora” da janela temporal da amostra.

O cálculo de previsões requer um conjunto de informação. Por exemplo, em modelos de séries temporais, é usado tipicamente o histórico das observações da série. Seja y_k , para cada instante k , o valor observado nesse instante. Se k representar o instante de tempo corrente e $k + 1$ o instante futuro seguinte, as observações y_1, y_2, \dots, y_k representam a informação histórica disponível, incluindo a do instante presente. Para produzir uma previsão futura com horizonte temporal 1 (isto é, 1-passo-à-frente) para o instante $k + 1$, usa-se a informação disponível até o instante k . Neste sentido, as previsões para instantes futuros são condicionadas pela informação histórica disponível.

De um modo geral, a variável sujeita à previsão, digamos Y_k (para cada k entre 1 e n), é uma variável aleatória. Esta variável pode ser completamente caracterizada pela função densidade de probabilidade. No contexto da previsão, a modelação de uma série temporal como uma sequência de variáveis aleatórias é baseada nas distribuições condicionais $Y_k | Y_{k-1}, Y_{k-2}, \dots, Y_1$. Estas distribuições condicionais dão a solução mais completa do problema de previsão: dadas as observações das variáveis aleatórias Y_1, Y_2, \dots, Y_{k-1} , que asserções podemos fazer sobre Y_k ?

A distribuição conjunta de Y_1, Y_2, \dots, Y_n pode ser fatorizada da seguinte forma:

$$f_{1:n}(y_1, y_2, \dots, y_n) = f_1(y_1) f_{2|1}(y_2 | y_1) \cdots f_{n|n-1}(y_n | y_{n-1}, y_{n-2}, \dots, y_1) \quad (30)$$

o que significa que a função de verossimilhança é determinada por estas distribuições condicionais. A distribuição condicional pode ser definida por:

- Média condicional:

$$\mu_k = E(Y_k | Y_{k-1} = y_{k-1}, Y_{k-2} = y_{k-2}, \dots, Y_1 = y_1) \quad (31)$$

- Variância condicional:

$$\sigma_k^2 = \text{Var}(Y_k | Y_{k-1} = y_{k-1}, Y_{k-2} = y_{k-2}, \dots, Y_1 = y_1) \quad (252)$$

- Forma da distribuição condicional.

Na prática, caracterizar a sequência Y_1, Y_2, \dots, Y_n é muito difícil, sendo, portanto, usual obter-se estimativas pontuais (as previsões) e, quando possível, intervalos de confiança. Deste modo, a média condicional é a melhor estimativa para a previsão e o desvio padrão condicional mede até que ponto acreditamos que o valor real pode diferir da previsão. A forma da distribuição condicional, muitas vezes assumida como a distribuição Normal, permite fazer asserções probabilísticas acerca do valor real.

3.4 Medidas de desempenho

Como foi referido anteriormente, as previsões são por natureza incertas. No entanto, algumas são melhores que outras, devendo-se escolher o modelo ou método que seja mais adequado. Com isto surge uma questão fundamental:

Como medir a adequação de um método de previsão específico para um determinado conjunto de dados? Isto é, quão bem o modelo de previsão é capaz de reproduzir os dados que já são conhecidos?

Geralmente, essa capacidade é avaliada usando medidas de desempenho, baseadas nos erros de previsão produzidos pelo modelo. Assim, para responder à questão acima, deve-se calcular primeiro os erros de previsão e, posteriormente, avaliar a capacidade preditiva do modelo de previsão.

O erro de previsão obtém-se subtraindo ao valor observado, o valor de previsão naquele mesmo instante de tempo:

$$e_k = y_k - \hat{y}_k \quad (33)$$

onde, y_k é a observação no instante de tempo k e \hat{y}_k é a previsão nesse mesmo instante.

Usualmente, para efeitos de previsão, a notação \hat{y}_k é usada para representar uma previsão (ou estimativa preditiva) 1-passo-à-frente da observação y_k , fazendo uso das observações disponíveis até o instante $k-1$, ou seja, com origem de previsão $k-1$ e horizonte temporal igual a 1. Portanto, essa notação visa à simplificação de $\hat{y}_{k|k-1}$. Da mesma forma, e_k é uma notação abreviada de $e_{k|k-1}$, representando o erro de previsão 1-passo-à-frente.

No caso geral, em que se consideram horizontes temporais h quaisquer, com $h \geq 1$, usa-se a notação $\hat{y}_{k+h|k}$ para representar uma previsão h -passos-à-frente da observação y_{k+h} , com origem k e horizonte temporal h . Correspondentemente, os erros de previsão h -passos-à-frente são expressos como:

$$e_{k|k-h} = y_k - \hat{y}_{k|k-h} \quad (264)$$

O responsável pela decisão necessita de uma medida de desempenho que sirva de base de comparação e que seja utilizada para avaliar a qualidade das previsões, sendo que o modelo escolhido, de entre os candidatos, é aquele que minimiza essa medida de erro.

Na prática, é comum utilizar algumas das medidas de desempenho indicadas a seguir.

- **Erro Médio (ME, do inglês, *Mean Error*)**

$$ME = \frac{1}{n} \sum_{k=1}^n e_k \quad (275)$$

Esta medida apenas indica a direção do erro, isto é, a presença de um desvio sistemático por defeito ou por excesso. Dado que os erros positivos e negativos tendem a compensar-se, não consegue indicar o tamanho do erro, ou seja, não consegue diferenciar entre modelos de previsão mais precisos e menos precisos. Um ME igual a zero não significa que as previsões são perfeitas, mas apenas indica que as previsões estão no alvo adequado (Adhikari e Agrawal, 2013). Assim, é desejável que o ME esteja o mais próximo possível de zero.

- **Erro Absoluto Médio (MAE, do inglês, *Mean Absolut Error*)**

$$MAE = \frac{1}{n} \sum_{k=1}^n |e_k| \quad (286)$$

Esta medida quantifica o desvio médio absoluto das estimativas relativamente aos valores observados e mostra a magnitude do erro geral ocorrido devido à previsão. Ao contrário do ME, esta medida evita que os erros se cancelem entre si, uma vez que usa os seus valores absolutos. No entanto, o MAE não fornece nenhuma ideia sobre a direção dos erros, isto é, se o erro foi por defeito ou por excesso.

Para uma boa previsão, o MAE obtido deve ser o menor possível. Da mesma maneira com o ME, esta medida depende da ordem de grandeza dos valores e é influenciada pela transformação de dados.

- **Erro Quadrático Médio (MSE, do inglês, *Mean Squared Error*)**

$$MSE = \frac{1}{n} \sum_{k=1}^n e_k^2 \quad (297)$$

É uma medida do desvio quadrado médio dos valores previstos, muito usada pela sua facilidade de aplicação. Tem uma ideia similar ao MAE, onde os erros são todos positivos ou nulos, uma vez que se trabalha com o quadrado de cada um deles, portanto, fornece uma ideia geral do erro ocorrido durante a previsão, mas não sobre a direção do mesmo.

Refira-se que, no MSE, o erro total de previsão é muito afetado ou influenciado por erros individuais grandes, ou seja, erros de grande magnitude têm muito mais influência do que erros de pequena magnitude.

- **Raiz do Erro Quadrático Médio (RMSE, do inglês, *Root Mean Squared Error*)**

$$RMSE = \sqrt{MSE} \quad (308)$$

Esta medida não é mais do que a raiz quadrada do MSE calculado e, portanto, tem as propriedades do MSE. A única diferença é que o valor vem na mesma ordem de grandeza que os erros.

- **Erro Percentual Absoluto Médio (MAPE, do inglês, *Mean Absolute Percentage Error*)**

$$MAPE = \frac{100}{n} \sum_{k=1}^n \left| \frac{e_k}{y_k} \right| \quad (319)$$

Esta medida corresponde à percentagem do erro absoluto médio ocorrido. Uma vez que representa uma medida de erro relativa, é muito utilizada para comparar diferentes conjuntos de dados e tem a vantagem de não depender da ordem de grandeza dos valores. Esta medida não mostra a direção do erro, pois os erros com sinais opostos não se cancelam. Caso existirem valores observados muito próximos de zero, esta medida não deve ser aplicada.

- **Erro Percentual Médio (MPE, do inglês, *Mean Percentage Error*)**

$$MPE = \frac{100}{n} \sum_{k=1}^n \frac{e_k}{y_k} \quad (40)$$

Representa a percentagem do erro médio ocorrido durante a previsão e possui propriedades semelhantes ao MAPE, com a diferença de que o MPE mostra a direção do erro ocorrido e onde os erros positivos e negativos se compensam entre si.

É importante salientar que, nos problemas de gestão de stocks, as funções de erro do tipo indicado acima assumem que os erros cometidos por defeito e por excesso têm o mesmo

peso (pesos simétricos), ou seja, o mesmo grau de importância, sendo necessário realizar uma correção ex-post dos valores previstos.

De acordo com Makridakis *et al.* (1998), estas medidas não fornecem, por si só, uma boa base de comparação quanto aos ganhos de precisão obtidos com a aplicação de um modelo ou método de previsão específico. Uma boa base para fazer a comparação entre as diferentes medidas é definir alguns métodos de previsão muito simples, adiante designados por *naïve* (ingênuos).

Ao aplicarmos métodos *naïve*, permite-se fazer uma comparação entre o desempenho desses métodos com modelos ou métodos mais sofisticados.

De entre os métodos *naïve*, podemos destacar dois:

- **Naïve 1** – É o método de previsão *naïve* mais simples, onde a última observação disponível é usada como previsão para o próximo instante, com horizonte temporal h -passos-à-frente:

$$\hat{y}_{k+h|k} = y_k \quad (41)$$

- **Naïve 2** – É parecido ao *naïve 1*, com a diferença de que este tem em conta a tendência presente na série:

$$\hat{y}_{k+h|k} = y_k + (y_k - y_{k|k-h}) \quad (42)$$

3.5 Métodos de alisamento exponencial

Quando são necessárias previsões para centenas ou até milhares de artigos, como é o caso de muitos sistemas de gestão de stock, os métodos de alisamento são um dos poucos métodos de previsão suficientemente rápidos para uma implementação aceitável, além de serem simples de aplicar e apresentarem baixos custos computacionais (Makridakis *et al.*, 1998). Por isso, são frequentemente usados como técnicas de previsão na gestão de stocks e usados para estimar consumos.

Muitos métodos de previsão são baseados na ideia de que, quando um padrão subjacente existe numa série, esse padrão possa ser distinguido da aleatoriedade alisando os valores

passados. O efeito desse alisamento tem por finalidade eliminar a aleatoriedade, de modo que o padrão possa ser projetado no futuro e seja usado como previsão.

Se a média for usada como uma ferramenta de previsão, então, como em todos os métodos de previsão, o seu uso requer um conhecimento das condições que determinam a sua adequação. Para a média, a condição é que os dados sejam estacionários. De acordo com Yaffee e McGee (2002), as séries estacionárias são caracterizadas por uma espécie de equilíbrio estatístico em torno de um nível médio constante, bem como uma variância constante em torno desse nível médio ao longo do tempo. Assim, se uma série temporal for gerada por um processo constante sujeito a um erro aleatório, a média é uma estatística útil e pode ser usada como previsão para o período seguinte.

No entanto, se a série temporal apresentar uma tendência (seja numa direção ascendente ou descendente), ou um efeito sazonal (por exemplo, uma forte procura de chocolates e amêndoas na época da Páscoa), ou uma tendência e um efeito sazonal em conjunto, a média simples não é capaz de capturar os padrões nos dados. Neste caso, o melhor será considerar métodos de alisamento que buscam melhorar a média e a previsão para o período seguinte.

São vários os métodos de alisamento, mas os que ganham especial relevo neste trabalho são os métodos de alisamento exponencial. Uma característica comum a todos eles, é o facto de se atribuir um maior peso às observações mais recentes. Os métodos mais sofisticados procuram isolar as tendências ou a sazonalidade das variações irregulares, para poderem ser incorporadas na previsão.

O método de alisamento exponencial foi originalmente desenvolvido por Robert G. Brown (1956), para a previsão extrapolativa a partir de dados, tendo surgido vários trabalhos sobre o assunto (por exemplo, Holt, 1957, Brown, 1959, Winters, 1960, Brown, 1963, e Chatfield, 1978). Estes métodos usam médias ponderadas dos dados anteriores para fazer previsão, onde os efeitos das observações são exponencialmente decrescentes com a antiguidade das mesmas, isto é, quanto mais recentes forem as observações, maior a sua influência nas previsões.

Podemos enumerar essencialmente os métodos mais conhecidos:

- O alisamento exponencial simples;
- O método de Holt;
- O método de Holt-Winters.

Estes métodos têm por base o modelo de decomposição clássica (216), diferenciando-se entre si nas suposições relativamente aos padrões presentes na série.

3.5.1 Alisamento exponencial simples

O alisamento exponencial simples (AES) é mais adequado para séries temporais sem tendência ou sazonalidade. É um método fácil de aplicar, uma vez que só são necessárias a previsão mais recente, a observação real mais recente e uma constante de alisamento. Pelo facto de utilizar fórmulas recursivas, permite atualizar o valor seguinte de uma forma rápida e simples.

O modelo subjacente ao AES é formado por um único parâmetro: o nível a_k . O método é constituído por duas equações, uma dedicada à estimação do nível e outra ao cálculo de previsões. A equação que permite estimar o nível é representada por uma equação de alisamento recursiva, dada por:

$$a_k = \alpha y_k + (1 - \alpha) a_{k-1} \quad (323)$$

onde α é a constante de alisamento do nível (variando entre 0 e 1). Esta constante representa o peso que se atribui à observação mais recente. A escolha do seu valor pode fazer a diferença entre uma previsão imprecisa e uma previsão mais precisa. Uma maneira de escolher α é considerar aquele que minimiza a medida de desempenho pré-definida, como por exemplo o MSE, que é usado frequentemente.

Quando α toma um valor próximo de 1, o AES aproxima-se do método *naïve* 1, onde a previsão é quase toda ela baseada na última observação disponível. Neste caso, diz-se que o método possui pouca memória, no sentido de reter pouca informação sobre as observações mais antigas. Quando α toma um valor próximo de 0, inclui pouco ajustamento, onde o valor da nova previsão é muito próximo da anterior. Neste caso, as observações mais antigas passam a influenciar em demasia a estimação do novo valor. Posto isto, podemos afirmar que, quanto maior o valor da constante de alisamento, mais irregular é a componente de nível, uma vez que esta copia em demasia o comportamento irregular da série.

As previsões h -passos-à-frente obtêm-se da seguinte forma:

$$\hat{y}_{k+h|k} = a_k \quad (334)$$

Como todos os métodos recursivos, o alisamento exponencial simples necessita de uma inicialização. A forma mais simples e usual de inicializar este método é considerar o valor inicial do nível igual à primeira observação, $a_0 = y_1$.

3.5.2 Método de Holt

O método de Holt foi originalmente introduzido por Holt (1957) e corresponde a uma extensão do método de alisamento exponencial simples para séries com tendência. O modelo subjacente a este método é formado por dois parâmetros: o nível a_k e o declive b_k . O declive não é mais do que uma estimativa da inclinação da série em cada instante k . O nível e o declive constituem a componente de tendência do modelo. As equações que permitem estimar recursivamente o nível e o declive são:

$$\begin{cases} a_k = \alpha y_k + (1-\alpha)(a_{k-1} + b_{k-1}) \\ b_k = \beta (a_k - a_{k-1}) + (1-\beta)b_{k-1} \end{cases} \quad (345)$$

onde, β é a constante de alisamento associada ao declive (variando entre 0 e 1). De forma análoga, os valores de α e β são escolhidos como sendo aqueles que minimizam a medida de desempenho. Esta escolha é tipicamente realizada através de uma pesquisa de valores pré-definidos numa grelha (Cortez, 2010).

A equação que permite calcular previsões h -passos-à-frente é definida por:

$$\hat{y}_{k+h|k} = a_k + h \cdot b_k \quad (356)$$

Uma forma muito simples de inicializar o método de Holt é considerar o nível igual à primeira observação, $a_0 = y_1$, e o declive como sendo a diferença entre as duas primeiras observações, $b_0 = y_2 - y_1$.

3.5.3 Método de Holt-Winters

Os estudos realizados por Holt (1957) e Winters (1960) permitiram estender o método de Holt por forma a capturar a sazonalidade. O modelo subjacente ao método de Holt-Winters é constituído por 3 parâmetros: o nível a_k , o declive b_k e a sazonalidade c_k , cada um possuindo uma constante de alisamento a si associada, respetivamente, α , β e γ (todas variando entre 0 e 1). Vamos usar a letra m para denotar o período da sazonalidade, isto é, o número observações em cada ciclo. Por exemplo, para uma série com sazonalidade anual, se as observações forem registadas mensalmente, o período é $m = 12$, e se os registos forem trimestrais, o período vem $m = 4$.

No método de Holt-Winters, podemos distinguir dois modelos diferentes, a depender da forma como a componente sazonal se combina com a de tendência. Quando as variações sazonais são aproximadamente constantes ao longo da série, não dependendo do seu nível, o modelo é aditivo. Pelo contrário, quando as variações sazonais mudam proporcionalmente ao nível da série, o modelo é multiplicativo.

Com o modelo aditivo, a componente sazonal é expressa em termos absolutos na mesma escala que as observações. Consequentemente, na equação do parâmetro de nível, a série é ajustada sazonalmente subtraindo a componente sazonal. Com o modelo multiplicativo, a componente sazonal é expressa em termos relativos (percentagens) e a série é ajustada sazonalmente dividindo-a pela componente sazonal.

As equações que definem o método de Holt-Winters, para cada um dos modelos, são dadas de acordo com o seguinte.

Modelo aditivo:

$$\begin{cases} a_k = \alpha(y_k - c_{k-m}) + (1-\alpha)(a_{k-1} + b_{k-1}) \\ b_k = \beta(a_k - a_{k-1}) + (1-\beta)b_{k-1} \\ c_k = \gamma(y_k - a_k) + (1-\gamma)c_{k-m} \end{cases} \quad (367)$$

A equação que permite calcular previsões h -passos-à-frente é definida por:

$$\hat{y}_{k+h|k} = (a_k + h \cdot b_k) + c_{k+(h-1) \bmod m+1-m} \quad (378)$$

Modelo multiplicativo:

$$\begin{cases} a_k = \alpha (y_k / c_{k-m}) + (1 - \alpha) (a_{k-1} + b_{k-1}) \\ b_k = \beta (a_k - a_{k-1}) + (1 - \beta) b_{k-1} \\ c_k = \gamma (y_k / a_k) + (1 - \gamma) c_{k-m} \end{cases} \quad (389)$$

A equação que permite calcular previsões h -passos-à-frente é dada por:

$$\hat{y}_{k+h|k} = (a_k + h \cdot b_k) c_{k+(h-1) \bmod m+1-m} \quad (5039)$$

As constantes de alisamento, α , β e γ , pelo facto de serem parâmetros associados ao método e não ao modelo subjacente, são muitas vezes chamadas de hiperparâmetros, designação essa que passaremos a utilizar neste trabalho. Para a sua identificação, à semelhança dos métodos de alisamento exponencial anteriores, pode ser realizada uma otimização por pesquisa de valores numa grelha, seleccionando aqueles que minimizam o MSE ou outra medida de desempenho pré-definida. No entanto, à medida que aumenta o número de hiperparâmetros a identificar, mais árdua e morosa se torna a pesquisa de valores numa grelha. Além do mais, tem a desvantagem de considerar apenas um conjunto de valores discreto e limitado, ficando excluídos potenciais valores que poderiam originar melhorias substanciais no desempenho do método.

O método de Holt-Winters é o método de previsão escolhido para a realização deste trabalho. Por forma a tirar o máximo partido das suas capacidades, não só não foi considerada a pesquisa de valores numa grelha, como também foi usada uma versão alternativa e melhorada no que diz respeito à estimação da componente de tendência, nomeadamente, tendência amortecida, que passamos a apresentar seguidamente.

A versão original do método de Holt-Winters, em termos de previsão, projeta linearmente a tendência usando a expressão $a_k + h \times b_k$. No entanto, de acordo com alguns trabalhos empíricos realizados sobre os métodos de Holt e Holt-Winters, entre os quais destacamos Armstrong (2006) e Gardner e McKenzie (2011), é amplamente recomendada uma versão

de tendência amortecida (do inglês, *damped trend*), abordagem alternativa que surgiu na sequência dos trabalhos de Gardner e McKenzie (1985, 1988 e 1989). Esta versão visa suavizar a tendência existente em um período de dados.

Modelo multiplicativo com tendência amortecida:

As equações que definem o método de Holt-Winters com tendência amortecida, incluindo o cálculo de previsões, são dadas a seguir. Indicam-se apenas as equações para o modelo multiplicativo, já que é aquele que é usado no nosso estudo, uma vez que é o mais comum na prática e que se adequa melhor nas muitas situações reais.

$$\begin{cases} a_k = \alpha (y_k / c_{k-m}) + (1 - \alpha)(a_{k-1} + \phi b_{k-1}) \\ b_k = \beta (a_k - a_{k-1}) + (1 - \beta)\phi b_{k-1} \\ c_k = \gamma (y_k / a_k) + (1 - \gamma)c_{k-m} \end{cases} \quad (51)$$

$$\hat{y}_{k+h|k} = \left[a_k + (\phi + \phi^2 + \dots + \phi^h) b_k \right] c_{k+(h-1) \bmod m+1-m} \quad (402)$$

onde ϕ ($\phi > 0$) é a constante de amortecimento. Se $\phi = 1$, o método é idêntico ao método de Holt-Winters original. Para valores entre 0 e inferiores a 1, a tendência é amortecida de modo a se aproximar a uma constante para algum instante no futuro. Na prática, o valor de ϕ raramente é inferior a 0.8, uma vez que o efeito de amortecimento é demasiado forte para valores inferiores (Hyndmann e Athanasopoulos, 2018).

O estudo realizado por Gardner e Acar (2019), onde refere que a otimização dos hiperparâmetros traz melhorias significativas na precisão da previsão, independentemente do critério usado, é o mote que nos leva a substituir a pesquisa dos seus valores numa grelha por métodos de otimização mais precisos, para além das razões já mencionadas acima. Neste trabalho, para a identificação dos quatro hiperparâmetros do método de Holt-Winters com tendência amortecida, foi usado um algoritmo de otimização estocástica – o *simulated annealing* (Goffe *et al.*, 1994), onde se apresenta uma breve descrição no Anexo G.

Inicialização:

Relativamente à inicialização dos parâmetros de nível, de declive e dos valores sazonais, existem várias sugestões na literatura. Aqui, apresentamos dois desses métodos:

- 1) O primeiro é uma proposta de Makridakis *et al.* (1998), onde a inicialização é obtida a partir dos primeiros dois ciclos sazonais de observações. As primeiras estimativas para o nível e declive correspondem ao instante $k = m$. O nível é definido como a média das primeiras m observações, ou seja, um ciclo sazonal:

$$a_m = \frac{y_1 + y_2 + \dots + y_m}{m} \quad (413)$$

O declive é obtido a partir da média dos declives para cada ciclo:

$$b_m = \frac{(y_{m+1} + y_{m+2} + \dots + y_{m+m}) - (y_1 + y_2 + \dots + y_m)}{m^2} \quad (424)$$

Quanto aos primeiros m valores sazonais, para o modelo multiplicativo, estes são obtidos do seguinte modo:

$$c_i = \frac{y_i}{a_m}, \quad i = 1, \dots, m \quad (435)$$

Uma das desvantagens deste método é que só permite obter previsões a partir do instante $m + h$.

- 2) O segundo é baseado numa decomposição proposto por Hyndmann *et al.* (2008), constituindo um método preferível ao anterior. Para tal, são aplicados sucessivamente os seguintes passos apenas aos primeiros dois ou três ciclos completos de observações:
- (i) Primeiro, calculam-se médias móveis centradas de comprimento m às primeiras observações, caso o período seja ímpar. Se o período for par, as médias móveis centradas de comprimento m precedem o cálculo de médias móveis centradas de comprimento 2.
 - (ii) Seguidamente, para o modelo multiplicativo, divide-se a série original pela tendência alisada obtida anteriormente, resultando numa série sem tendência. Os valores sazonais iniciais são obtidos fazendo as médias dos valores da tendência alisada pertencentes ao mesmo período do ciclo. Por exemplo, para dados mensais, o valor sazonal para o mês de janeiro corresponde à média dos valores de janeiro, etc.
 - (iii) Depois, divide-se a série original pelos valores sazonais iniciais, de modo a se obter uma série sazonalmente ajustada, isto é, sem componente sazonal.

- (iv) Finalmente, ajusta-se uma reta de regressão à série sazonalmente ajustada. Os valores iniciais para o nível e o declive correspondem aos parâmetros da reta obtida.

No final da inicialização, tem-se $a_0, b_0, c_{-m+1}, c_{-m+2}, \dots, c_0$.

Neste trabalho, para o cálculo dos valores iniciais dos parâmetros associados ao método de Holt-Winters, versão da tendência amortecida, foi considerado o segundo método de inicialização (para a obtenção dos resultados, ver secção 4.5), uma vez que se considera que é uma variante melhorada do primeiro método que tem sido alvo de alguma crítica na literatura.

CAPÍTULO 4

4 ANÁLISE PREDITIVA VS. ANÁLISE PRESCRITIVA

Este é o capítulo central da dissertação, onde é realizado um estudo comparativo entre dois tipos de análises, uma baseada em previsões ótimas – análise preditiva – e outra em decisões ótimas – análise prescritiva. A comparação foi realizada com recurso à análise de variância (ANOVA) de medidas repetidas.

Para levar a cabo este estudo, foram usados dois tipos de conjuntos, um simulado e outro obtido a partir de observações reais (ver Anexo A). No cálculo de previsões, usou-se o método de Holt-Winters com tendência amortecida, cujo modelo subjacente é o de sazonalidade multiplicativa. Na análise preditiva, para fins de comparação e avaliação da capacidade preditiva do modelo, foi usada a medida de desempenho MSE. As previsões foram posteriormente ajustadas com vista a obter-se valores prescritos (as decisões), a partir de quantis de probabilidade definidos pela assimetria dos custos originados entre errar por defeito e errar por excesso. Na análise prescritiva, para fins de comparação e avaliação da capacidade prescritiva do modelo, foi usada uma função de custos assimétrica (assimetria essa análoga à usada na análise preditiva), não sendo necessário, neste caso, proceder-se a qualquer tipo de correção ex-post.

As várias secções a seguir apresentadas têm por objetivo dar uma descrição de todos os procedimentos considerados no estudo comparativo realizado.

4.1 Motivação e objetivos

Num sistema de gestão de stocks, o processo da tomada de decisões baseia-se particularmente na previsão da procura e nas encomendas dos clientes. As previsões de procuras futuras são produzidas de forma a escolher um nível de stock que vá de encontro aos objetivos traçados pela empresa, equilibrando os custos de posse com os custos de rutura de stocks. Diz-se que o referido sistema é eficiente quando os objetivos são atingidos. De acordo com Crone (2003), embora a amplitude dos custos aumente geralmente com a magnitude numérica dos erros de previsão, os custos decorrentes de previsão por excesso e por defeito não são frequentemente simétricos nem quadráticos. Para refletir essa característica, é recomendável a utilização de uma função de utilidade,

baseada em custos assimétricos, como função objetivo, visando a obtenção de previsões melhoradas e, conseqüentemente, de um nível de stock mais eficiente.

É intuitivo e relativamente fácil de reconhecer que as boas decisões são favorecidas por boas previsões, mas pode já não ser tão evidente que as boas decisões não requerem necessariamente boas previsões. Daqui, podem-se identificar dois paradigmas:

- 1) O papel que o modelo desempenha como modelo preditivo: durante o processo de identificação e estimação, dá-se ênfase à obtenção de boas previsões (ou previsões ótimas); a análise preditiva assenta-se neste paradigma;
- 2) O papel que o modelo desempenha como modelo prescritivo: aqui, mais do que a obtenção de boas previsões, o problema centra-se na obtenção de decisões ótimas que promovam bons resultados, estando diretamente relacionados com a tomada de decisões; este paradigma constitui a base da análise prescritiva.

O confronto destes paradigmas teve surgimento, em parte, no trabalho de Freitas (2008), no contexto da mistura de modelos em problemas de previsão e decisão. Na literatura, é rara ou praticamente inexistente essa comparação. Tipicamente, apenas uma das abordagens é explorada e aplicada no problema particularizado. Por exemplo, no contexto da gestão de stocks, sabendo que é fundamental considerar custos assimétricos e, por sua vez, funções assimétricas, opta-se cada vez mais pelo papel prescritivo do modelo global. Acredita-se que, independentemente do caminho escolhido, o mesmo depende tanto do problema particular como do horizonte temporal que se considera para a previsão ou decisão, entre outros hipotéticos fatores.

A partir da discussão dos paradigmas apresentados acima, o presente trabalho centralizou-se no estudo comparativo entre as análises preditiva e prescritiva, avaliando-se custos de decisão através do uso de funções de utilidade baseadas em custos assimétricos. Um dos fatores considerados no estudo comparativo, para além dos paradigmas, foi o de tentar perceber até que ponto o horizonte temporal afeta os resultados.

4.2 Previsões ótimas vs. decisões ótimas

Para se compreender melhor os procedimentos realizados no estudo comparativo, há que distinguir primeiro decisão ótima de previsão ótima. Para tal, vamos designar por P1 o procedimento que caracteriza o primeiro paradigma, ou seja, aquele onde as decisões são precedidas de previsões, sendo estas calculadas a partir de um modelo otimizado em função da sua capacidade preditiva. Vamos designar por P2 o procedimento

que caracteriza o segundo paradigma, isto é, aquele onde as decisões são obtidas a partir de um modelo otimizado em função da sua capacidade prescritiva. De forma sucinta, temos:

- 1) P1 – as decisões são precedidas pela identificação e estimação de um modelo com boas capacidades preditivas (representa a análise preditiva);
- 2) P2 – a identificação e estimação do modelo são realizadas com o objetivo de se obter boas decisões (representa a análise prescritiva).

Enquanto que, na maioria das vezes, P1 é talvez a abordagem mais comumente usada e menos complexa de se realizar (por exemplo, para modelos paramétricos lineares, a medida de desempenho MSE permite considerar métodos de estimação baseados nos mínimos quadrados), P2 pode não ser exequível em todas as situações (uma vez que as funções de custos assimétricas não são diferenciáveis e, portanto, exigindo recorrer-se a métodos de otimização não lineares, o que dificulta a estimação do modelo subjacente). Não obstante, com vista a concretizar mais facilmente o objetivo proposto neste trabalho, a classe de modelos ou métodos que se usou aqui corresponde aos métodos de alisamento exponencial, onde a exequibilidade dos procedimentos anteriormente descritos são relativamente simples de pôr em prática.

As previsões ótimas são aquelas que levam à minimização de uma medida de desempenho, definida com base em erros de previsão sem assimetria, como por exemplo, MSE. Após a identificação de um modelo com boas capacidades preditivas, as decisões são obtidas pelo cálculo de quantis a partir da sequência de erros de previsão. Essas decisões dependem dos pesos que se atribuem a errar por excesso e errar por defeito.

A depender do horizonte temporal considerado, a medida de desempenho utilizada na análise preditiva foi o MSE (secção 3.4), com vista à sua minimização:

$$J(\alpha, \beta, \gamma, \phi) = \frac{1}{n - n_0} \sum_{k=n_0+1}^n e_{k|k-h}^2 \quad (446)$$

onde α , β , γ e ϕ são os hiperparâmetros do método de Holt-Winters (secção 3.5.3), com tendência amortecida, a otimizar, e h o horizonte temporal da previsão pré-definido. O valor n_0 representa o número de observações iniciais a ignorar no cálculo da medida de desempenho, devido à inicialização do próprio método e tempo necessário para um

ajustamento inicial adequado dos parâmetros de nível, declive e sazonalidade. A Figura 11 ilustra, de forma sucinta, a análise preditiva.

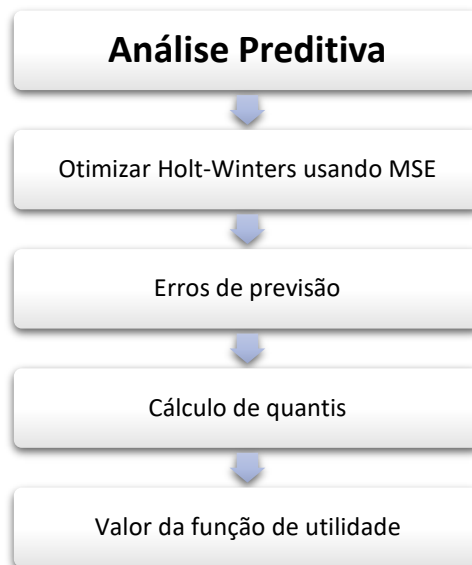


Figura 11 – Diagrama ilustrativo da análise preditiva.

Com vista a clarificar melhor o cálculo de decisões a partir de previsões ótimas, sejam Y_k a variável aleatória que define a quantidade procurada de um tipo de artigo, no instante de tempo k , e y_k a procura real nesse mesmo instante. Se u e v (ambos positivos) são os pesos que refletem o grau de assimetria entre errar por defeito e errar por excesso, respetivamente, o valor de probabilidade acumulada, dado por:

$$P(Y_k \leq y_k) = \frac{u}{u + v} \quad (457)$$

permite definir o quantil de probabilidade Q_k , de ordem $u/(u + v)$, a partir da fórmula:

$$F(Q_k) = \frac{u}{u + v} \quad (468)$$

onde Q_k é a quantidade a encomendar para o instante k . Sem perda de generalidade, os valores de u e v foram definidos no intervalo real $]0,1[$, de modo que $u + v = 1$. A

probabilidade dada por (457) pode ser vista como uma probabilidade condicional de Y_k ser inferior a y_k , dados os valores da procura nos instantes passados, representados por y_1, y_2, \dots, y_{k-h} (tendo em conta o horizonte temporal da previsão). Portanto, podemos assumir que:

$$P(Y_k \leq y_k | Y_{k-h} = y_{k-h}, Y_{k-h-1} = y_{k-h-1}, \dots, Y_1 = y_1) = \frac{u}{u+v} \quad (479)$$

Isto permite interpretar as previsões ótimas, obtidas pelo método de Holt-Winters, como médias condicionais (ver Secção 0):

$$\hat{y}_{k|k-h} = E(Y_k | Y_{k-h} = y_{k-h}, Y_{k-h-1} = y_{k-h-1}, \dots, Y_1 = y_1) \quad (60)$$

de modo que as decisões correspondem a quantis de probabilidade de ordem $u/(u+v)$ das previsões. Portanto, se $\hat{y}_{k|k-h}$ corresponde à previsão h -passos-à-frente da procura no instante k , a decisão $q_{k|k-h}$ correspondente é calculada do seguinte modo:

$$q_{k|k-h} = \hat{y}_{k|k-h} + \tilde{e}_{k|k-h} \quad (61)$$

onde $\tilde{e}_{k|k-h}$ é o quantil de probabilidade de ordem $u/(u+v)$, calculado a partir dos erros de previsão h -passos-à-frente $e_{j|j-h}$, para $j \leq k$.

Um dos problemas associados aos métodos de alisamento exponencial, é que os erros de previsão não são garantidos terem média nula, nem tão pouco terem distribuição Normal. Na verdade, esta é uma situação comum na generalidade dos modelos de séries temporais. No estudo comparativo, para o cálculo de quantis a partir dos erros de previsão, foi considerado o seguinte procedimento: para cada horizonte temporal h e em cada instante k , foram realizados dois testes de normalidade, o teste de Shapiro-Wilk (Shapiro e Wilk, 1965) e o teste D'Agostino-Pearson (D'Agostino, 1970), tendo-se agido da seguinte forma:

- Sempre que os erros de previsão, em ambos os testes, provinham de uma distribuição Normal, os quantis foram obtidos a partir dessa distribuição assumindo, como parâmetros, a média e o desvio padrão dos erros;
- Caso contrário, os quantis foram determinados a partir da distribuição empírica dos erros de previsão.

Independentemente da situação considerada, foi sempre garantido um número mínimo de 50 erros de previsão em todas as séries temporais do estudo.

As decisões ótimas são aquelas que levam à minimização da função de utilidade, sendo esta tipicamente baseada em custos assimétricos, refletindo os pesos atribuídos quando se erra por defeito e quando se erra por excesso. As decisões finais correspondem às previsões produzidas pelo próprio modelo considerado. A Figura 12 ilustra, de forma sucinta, a análise prescritiva.

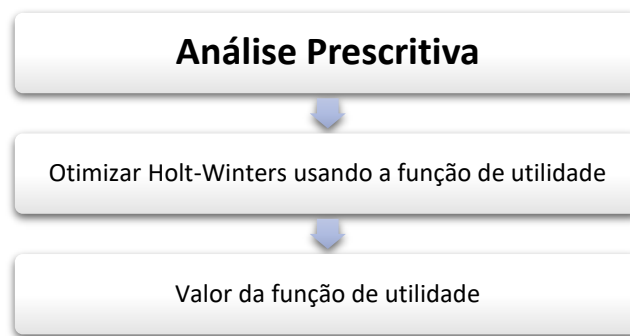


Figura 12 – Diagrama ilustrativo da análise prescritiva.

A função de custos assimétrica utilizada na análise prescritiva é definida por:

$$D(\alpha, \beta, \gamma, \phi) = \frac{1}{n - n_0} \sum_{k=n_0+1}^n d_{k|k-h} \quad (482)$$

onde:

$$d_{k|k-h} = \begin{cases} ue_{k|k-h}, & e_{k|k-h} \geq 0 \\ -ve_{k|k-h}, & e_{k|k-h} < 0 \end{cases} \quad \text{e} \quad e_{k|k-h} = y_k - q_{k|k-h} \quad (493)$$

com $q_{k|k-h}$ a denotar as decisões ótimas obtidas diretamente pelo método de Holt-Winters, e u e v os pesos assimétricos atribuídos aos diferentes tipos de erros. No caso de os pesos serem simétricos, isto é, $u = v$, o problema de minimização definido pela função de utilidade é equivalente ao de minimização da medida de desempenho MAE.

Na aplicação da ANOVA, foram considerados os valores da função utilidade dadas pela fórmula (482), obtidos a partir das sequências $\{q_{k|k-h} : k > h\}$, quer as obtidas pela análise prescritiva descrita acima, como as da análise preditiva (61).

4.3 ANOVA de medidas repetidas

O conceito de análise de variância (ANOVA, vem do inglês “Analysis of Variance”) foi estabelecido pelo geneticista e estatístico britânico Sir Ronald Fisher, em 1918, e publicado formalmente em seu livro "Statistical Methods for Research Workers", em 1925. A técnica foi desenvolvida para fornecer procedimentos estatísticos de um teste de significância das médias dos diferentes grupos (ou tratamentos).

De acordo com Neideen *et al* (2007), os testes paramétricos são mais robustos que os testes não paramétricos, e geralmente, requerem menos dados para obter uma conclusão mais forte. No entanto, exigem premissas mais rígidas, isto é, um conjunto de pressupostos. Quando esses pressupostos são violados, os resultados da análise podem ficar comprometidos, correndo o risco de não serem os mais acertados ou estarem completamente errados. Neste estudo, optou-se por aplicar um teste paramétrico.

A análise de variância é uma ferramenta estatística usada para detetar diferenças entre as médias dos grupos experimentais, através de duas hipóteses, a hipótese nula e a hipótese alternativa. Na terminologia ANOVA, as variáveis independentes são chamadas de fatores e os grupos dentro de cada fator são chamados de níveis (Sawyer, 2009).

Quando se quer estudar o efeito ou influência de mais do que um fator sobre a variável de interesse, utiliza-se a análise de variância múltipla. Cada fator pode ter mais do que um nível. Quando se quer estudar as combinações de todos os níveis de todos os fatores, diz-se estar perante uma análise de variância fatorial (Sawyer, 2009).

No estudo comparativo entre as análises preditiva e prescritiva, houve o interesse estudar o efeito de dois fatores sobre a variável de interesse. Esta variável, adiante denotada por D , é representada por valores de custo obtidos a partir da função de custos assimétrica (482). Os fatores considerados foram o paradigma, contendo dois níveis, P1 (previsões

ótimas) e P2 (decisões ótimas), e o horizonte temporal, constituído por três níveis, CP (curto-prazo), MP (médio-prazo) e (LP) longo-prazo (ver Figura 13), definidos seguidamente.

Para representar cada um dos níveis do horizonte temporal, precisamos de definir *a priori* 3 horizontes temporais que distingam o curto, o médio e o longo prazos. Para o horizonte temporal de curto prazo, a escolha natural é 1-passo-à-frente. Relativamente ao horizonte temporal de médio prazo, foi definido um período correspondente à metade de um ciclo sazonal. Para um horizonte temporal de longo prazo, deverá se escolhido um período não inferior ao comprimento do ciclo sazonal. A característica que achamos importante reter aqui é que a distância entre os horizontes temporais escolhidos não seja demasiado curta relativamente ao comprimento do ciclo. Portanto, isto permite considerar a ANOVA de efeitos fixos para o nosso estudo. Em suma, temos o seguinte:

- CP (curto-prazo): $h_{CP} = 1$;
- MP (médio-prazo): $h_{MP} = \lfloor m/2 \rfloor + 1$, onde $\lfloor \cdot \rfloor$ denota a parte inteira do valor numérico e m o comprimento do ciclo sazonal;
- LP (longo-prazo): $h_{LP} = m + 1$.

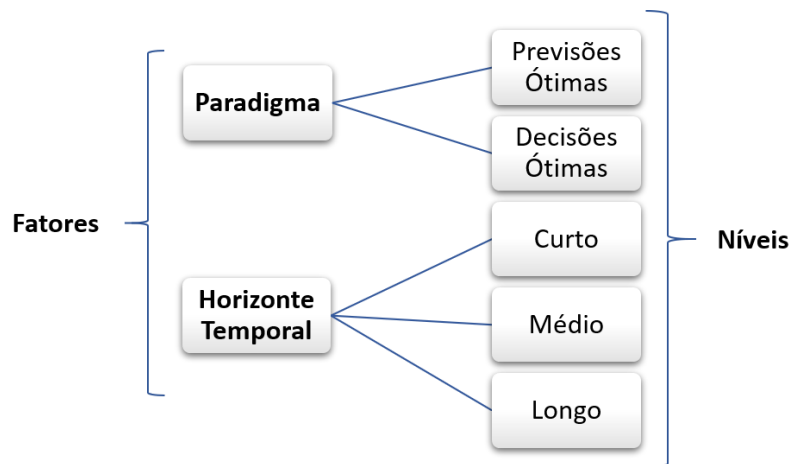


Figura 13 – Fatores e respetivos níveis na ANOVA fatorial dupla.

Existem, portanto, $2 \times 3 = 6$ combinações possíveis dos níveis dos fatores. Deve-se ter em conta que estes 2 fatores podem interagir entre si, onde a variação que ocorre na variável resposta produzida por uma alteração de um nível de um dos fatores, pode variar consoante os níveis dos restantes fatores.

Na elaboração do estudo, foi implementada a ideia de que os mesmos indivíduos (adiante designados por unidades experimentais), são testados em diferentes condições experimentais. Por outras palavras, a ANOVA foi planeada de modo que cada unidade experimental (série temporal) foi medida várias vezes, ou seja, sujeita a cada combinação de um dos paradigmas e um dos horizontes temporais. Assim, considerou-se a ANOVA de medidas repetidas.

A realização de uma ANOVA de medidas repetidas permite tirar partido de algumas suas vantagens, nomeadamente, no que diz respeito à redução da variabilidade dos dados e, principalmente, do número total de unidades experimentais considerado. Note-se, contudo, que as medidas repetidas constituem também uma desvantagem, na medida em que as observações, para as diferentes condições a que foram sujeitas, estão relacionadas. Como haveremos de referir mais à frente, esta desvantagem impõe que as variâncias das diferenças entre todos os pares de medidas sejam similares. A violação desta condição tem o potencial de aumentar o chamado falso positivo (erro do tipo I), ou seja, rejeitar a hipótese nula quando ela é verdadeira.

Considerando os 2 fatores estudados, paradigma e horizonte temporal, as observações da variável de interesse são indexadas por 3 índices, D_{ijk} , onde i representa o nível do fator paradigma, j representa o nível do fator horizonte temporal, e k representa a posição dentro do grupo ij . Uma vez que foi usado o SPSS para a realização da ANOVA de medidas repetidas, considerou-se a disposição dos dados esquematizada na Tabela 1.

Tabela 1 – Disposição das observações de medidas repetidas para o SPSS.

Sujeitos	P1_CP	P1_MP	P1_LP	P2_CP	P2_MP	P2_LP
Série 1	D_{111}	D_{121}	D_{131}	D_{211}	D_{221}	D_{231}
Série 2	D_{112}	D_{122}	D_{132}	D_{212}	D_{222}	D_{232}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Série K	D_{11K}	D_{12K}	D_{13K}	D_{21K}	D_{22K}	D_{23K}

O modelo matemático geral que descreve a relação entre a resposta e os grupos da ANOVA fatorial dupla considerada no estudo é dado por:

$$D_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad \begin{cases} i = 1, 2 \\ j = 1, 2, 3 \\ k = 1, \dots, K \end{cases} \quad (504)$$

onde:

- D_{ijk} representa a k -ésima observação do grupo ij (K denota o número de observações dentro de cada grupo ij ; no nosso caso concreto, o número de séries temporais);
- μ é a média de todas as observações;
- τ_i representa o i -ésimo efeito do fator paradigma;
- β_j representa o j -ésimo efeito do fator horizonte temporal;
- γ_{ij} representa o efeito da interação entre o nível i do fator paradigma e o nível j do fator horizonte temporal;
- ε_{ijk} representa o erro aleatório de cada observação; assume-se que estes erros são independentes entre si, normalmente distribuídos, de média nula e variância constante σ_ε^2 .

Quando existe interação entre os dois fatores, o efeito de um deles depende dos níveis do outro. Assim, o efeito de uma interação significativa pode sobrepor-se ao efeito individual de cada um dos fatores. Por este motivo, através dos testes de hipóteses, primeiro avalia-se se existe interação entre os fatores, e caso o efeito da interação não seja significativo, avaliam-se os efeitos individuais. Deve-se ter em atenção que a existência de interação entre os fatores pode levar a que os testes relativos aos fatores individuais não tenham significado.

Para o modelo considerado, é exigido um conjunto de pressupostos e hipóteses a estudar. Começando pelas hipóteses, tem-se:

- 1) Efeito principal do fator paradigma:

$$H_0 : \tau_1 = \tau_2 = 0 \quad \text{vs.} \quad H_1 : \tau_i \neq 0, \text{ pelo menos para algum } i$$

- 2) Efeito principal do fator horizonte temporal:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0, \text{ pelo menos para algum } j$$

- 3) Interação entre os fatores paradigma e horizonte temporal:

$$H_0 : \gamma_{11} = \gamma_{12} = \dots = \gamma_{23} = 0 \quad \text{vs.} \quad H_1 : \gamma_{ij} \neq 0, \text{ pelo menos para um par } i, j$$

As hipóteses nulas afirmam que as médias da população para cada nível da variável independente são iguais. Pelo contrário, as hipóteses alternativas afirmam que as médias da população diferem (Sawyer, 2009).

A estatística de teste utilizada na ANOVA é a estatística F, que recebeu o seu nome em homenagem a Sir Ronald Fisher, e consiste numa razão de duas variâncias, permitindo avaliar o efeito da interação e os efeitos principais. Ao rejeitar a hipótese nula, afirma-se que existem diferenças significativas em algum lugar entre as médias (em pelo menos um dos fatores), mas não indica onde está a diferença.

Para determinar quais os pares de níveis que apresentam diferenças significativas entre si (dois a dois), podem utilizar-se uma variedade de métodos de comparação múltipla. Neste estudo, tendo em conta que foi utilizado o SPSS para a realização da ANOVA, e dado que não existe uma facilidade própria para produzir testes post-hoc para medidas repetidas, foi considerada a correção ou procedimento de Bonferroni.

Relativamente aos pressupostos assumidos no modelo, estes incluem o seguinte (Sawyer 2009):

- 1) Cada grupo de observações deve provir de uma distribuição Normal;
- 2) As variâncias das diferenças entre todos os pares de medidas devem ser similares. Este pressuposto é conhecido por esfericidade e constitui um dos mais importantes pressupostos na ANOVA de medidas repetidas. A sua violação faz aumentar o erro do tipo I, fazendo com que o teste aceite mais frequentemente os falsos positivos.

Quando os dados originais não seguem a curva em forma de sino, podemos transformar esses dados para torná-los com uma distribuição o mais normal possível, de modo que os resultados da análise estatística desses dados se tornem mais válidos. De entre os vários métodos de transformação de dados existentes, consideramos aqui a transformação logarítmica de base 10, em que se substitui cada variável x pelo seu $\log(x)$, com vista a reduzir ou remover a distorção dos dados originais.

No estudo, foi aplicado o teste de normalidade de Shapiro-Wilk aos dados logaritmizados. A hipótese nula consiste em assumir que os dados provêm de uma distribuição Normal, enquanto que, na hipótese alternativa, assume-se que os mesmos não seguem uma

distribuição Normal. Refira-se que, na ANOVA de medidas repetidas, a violação do pressuposto de normalidade não é tão severa como o da esfericidade.

Para verificar a hipótese de esfericidade, foi aplicado o teste de Mauchly (Mauchly, 1940). No caso de os dados não verificarem o pressuposto da esfericidade, a situação pode ser solucionada com recurso a uma correção, visando a obtenção de valores críticos mais válidos da estatística F. Este tipo de correções tem o intuito de estimar o grau com que a esfericidade é violada. Ao ser aplicado um fator de correção ao número de graus de liberdade da distribuição F, fazendo diminuir esse número, permite que a estatística F seja mais conservativa. Este grau é usualmente denotado por ε . Quando $\varepsilon = 1$, diz-se que a esfericidade é verificada na sua plenitude. Quanto menor é o valor de ε , maior é o grau de violação da esfericidade (Sawyer, 2009).

No SPSS, podem ser considerados, pelo menos, dois tipos de correções sempre que o teste de esfericidade de Mauchly for violado, nomeadamente:

- Correção de Huynh-Feldt, se $\varepsilon \geq 0.75$;
- Correção de Greenhouse-Geisser, caso $\varepsilon < 0.75$.

4.4 Conjuntos de dados

Como já foi mencionado anteriormente, com o objetivo de comparar as análises preditiva e prescritiva, foram considerados dois conjuntos de dados, um constituído por séries simuladas e outro por séries reais. Com o intuito das séries simuladas serem adequadas para a aplicação do método de Holt-Winters (sem perda de generalidade, foi considerado apenas a variante do modelo multiplicativo), foi utilizado o modelo de decomposição seguinte como modelo gerador:

$$Y = T \times S + I \quad (515)$$

onde T e S representam as componentes de tendência e sazonalidade, respetivamente, combinadas de acordo com o modelo multiplicativo, e I a componente de ruído aditivo.

De modo a ser gerado um conjunto de séries relativamente diverso, foram considerados vários processos na geração de cada uma das componentes do modelo. Uma das vantagens de simular dados desta forma é podermos controlar as características de cada uma das componentes contidas na série. Para tornar o estudo mais completo, os

procedimentos realizados com as séries simuladas foram depois replicados com os dados reais (ver gráficos apresentados nos Anexos A e B).

4.4.1 Simulação de componentes de sazonalidade

Para a geração de componentes de sazonalidade, foram considerados dois procedimentos distintos: o primeiro usa uma composição de funções trigonométricas e, o segundo, utiliza a série caótica de Mackey-Glass, a qual será apresentada mais tarde.

Funções trigonométricas

As funções trigonométricas têm a vantagem de permitir a geração de séries com quaisquer periodicidades. Usando uma composição de funções trigonométricas, foram consideradas as duas fórmulas indicadas a seguir para se gerarem componentes sazonais, $S^{(1)}$ e $S^{(2)}$, ambas de periodicidade 12 (simulando, por exemplo, ciclos anuais para registos mensais):

$$s_k^{(1)} = a_k + \eta_k^{(1)}, \quad \eta_k^{(1)} \sim N(0, \sigma_\eta^{(1)}) \quad (526)$$

$$s_k^{(2)} = b_k + \eta_k^{(2)}, \quad \eta_k^{(2)} \sim N(0, \sigma_\eta^{(2)}) \quad (537)$$

onde:

$$a_k = A_1 \cos\left(\frac{2\pi k}{12} + w_1\right) + A_2 \cos\left(\frac{2\pi k}{2} + w_2\right) \quad (548)$$

$$b_k = B_1 \cos\left(\frac{2\pi k}{12} + w_3\right) + B_2 \cos\left(\frac{2\pi k}{6} + w_4\right) + B_3 \cos\left(\frac{2\pi k}{3} + w_5\right) \quad (559)$$

Todas as constantes envolvidas nas expressões de a_k e b_k foram geradas de acordo com a distribuição Uniforme contínua. Os valores A_i ($i=1,2$) e B_j ($j=1,2,3$) representam as amplitudes e w_l ($l=1,2,\dots,5$) partes das fases das ondas harmónicas correspondentes. Nas expressões das componentes sazonais, estão presentes componentes de ruído de muito baixa magnitude, com distribuição Normal (média nula e desvio padrão até 1% da amplitude máxima da sazonalidade), de modo a evitar um comportamento demasiado regular das componentes resultantes.

A Tabela 2 resume as condições de simulação das componentes de sazonalidade usando funções trigonométricas. A Figura 14 ilustra duas componentes sazonais obtidas usando as equações (526) e (537), respectivamente.

Tabela 2 – Condições de simulação das componentes de sazonalidade usando uma composição de funções trigonométricas.

Constantes
$A_1, B_1 \sim U(5,10)$
$A_2 \sim U(0,5)$
$B_2, B_3 \sim U(0,10)$
$w_1, w_2, \dots, w_5 \sim U(0, 2\pi)$

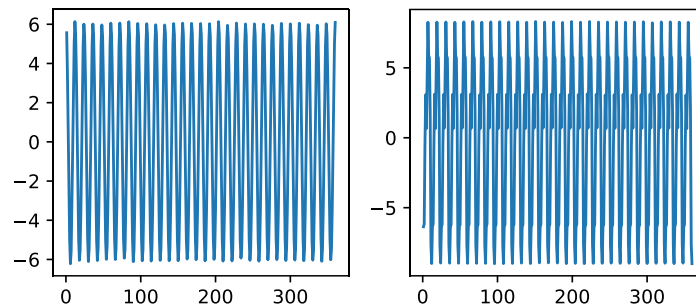


Figura 14 – Componentes sazonais simuladas, $S^{(1)}$ e $S^{(2)}$, respectivamente, usando uma composição de funções trigonométricas.

Série caótica de Mackey-Glass

Esta série tem servido de motivação para muitos estudos sobre a previsão de séries temporais difíceis, já que constitui uma série caótica possuindo muitas não linearidades. Apesar de, à partida, constituir uma série difícil e não adequada para a aplicação do método de Holt-Winters, é possível, perante certas condições, obter-se casos particulares onde a série resultante é periódica, perfazendo as condições mínimas para o pretendido.

Originalmente, a série de Mackey-Glass é definida por uma equação diferencial com atraso no tempo (Mackey e Glass, 1977). Para poder ser usada no contexto da previsão

de séries temporais regularmente espaçadas, é necessário proceder-se à sua discretização, o que poderá ser feito através da seguinte equação recursiva:

$$x_{k+1} = \frac{ax_{k-d}}{1+x_{k-d}^c} + bx_k \quad (70)$$

onde a , b , c e d constituem os parâmetros da equação. Tipicamente, a e b são números reais e c e d são números inteiros positivos.

Para a geração de componentes de sazonalidade usando o segundo procedimento, foram consideradas três subséries da equação (70) do seguinte modo:

$$s_k^{(3)} = R_1 x_k + \eta_k^{(3)}, \quad \eta_k^{(3)} \sim N(0, \sigma_\eta^{(3)}) \quad (71)$$

$$s_k^{(4)} = R_2 x_k + \eta_k^{(4)}, \quad \eta_k^{(4)} \sim N(0, \sigma_\eta^{(4)}) \quad (562)$$

$$s_k^{(5)} = R_3 x_{2k} + \eta_k^{(5)}, \quad \eta_k^{(5)} \sim N(0, \sigma_\eta^{(5)}) \quad (573)$$

As constantes R_i ($i=1,2,3$), geradas a partir de uma distribuição Uniforme contínua, representam fatores de escala. Em todos os casos, de modo a introduzir algum grau de aleatoriedade, foram adicionadas componentes de ruído aditivo de muito baixa magnitude (de forma análoga ao realizado na composição de funções trigonométricas). Apesar das componentes sazonais não possuírem um comportamento perfeitamente regular (à parte do ruído introduzido), muito por causa do elevado grau de não linearidade presente na série original, os picos mais altos ocorrem nos instantes espaçados de 25 em 25, 39 em 39 e 40 em 40, respetivamente, correspondendo aos períodos das componentes sazonais respetivas.

A Tabela 3 sumariza as condições de simulação das componentes de sazonalidade obtidas a partir da série caótica de Mackey-Glass. A Figura 15 ilustra três componentes sazonais obtidas usando as equações (71)-(573), respetivamente.

Tabela 3 – Condições de simulação das componentes de sazonalidade usando a série caótica de Mackey-Glass.

Constantes
$S^{(3)} : a = 0.6, b = 0.85, c = 10, d = 9$
$S^{(4)} : a = 0.65, b = 0.8, c = 10, d = 7$
$S^{(5)} : a = 0.2, b = 0.9, c = 10, d = 13$
$R_1, R_2, R_3 \sim U(10, 20)$

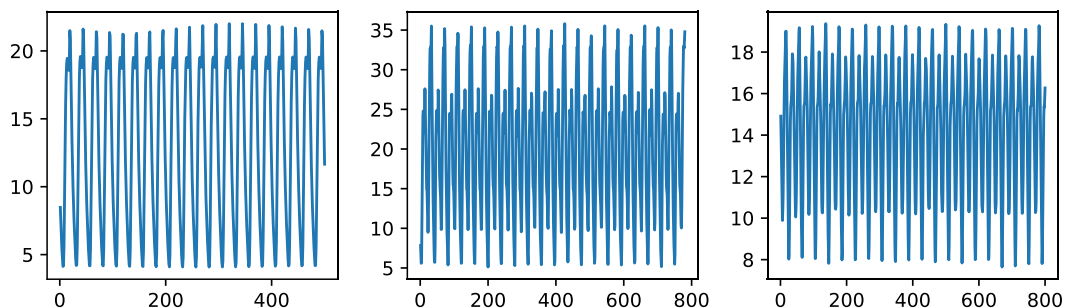


Figura 15 – Componentes sazonais simuladas, $S^{(3)}$, $S^{(4)}$ e $S^{(5)}$, respectivamente, usando a série de Mackey-Glass.

4.4.2 Simulação de componentes de tendência

Tendo em conta que as componentes sazonais simuladas não são perfeitamente regulares, constituindo, por si só, fatores que dificultam a previsão, optou-se por gerar componentes de tendência desprovidas do mesmo efeito de ruído. No total, foram considerados seis procedimentos distintos para se gerarem componentes de tendência.

Os três primeiros procedimentos basearam-se em modelos determinísticos: a reta de regressão, ondas sinusoidais de período extremamente longo e a função logística. Os três últimos procedimentos basearam-se em modelos de regressão de tendência dinâmica, descritos posteriormente.

Reta de regressão

Em termos de tendência, a reta de regressão é um dos modelos mais simples que se pode considerar. A equação usada, neste caso, foi a seguinte:

$$t_k^{(1)} = 200 + \frac{R}{N}k \quad (584)$$

onde N é o número de observações simuladas e R é um número aleatório gerado a partir de uma distribuição Uniforme contínua, dando informação sobre a taxa de variação da tendência. As condições de simulação do R estão representadas na Tabela 4.

Ondas sinusoidais com período extremamente longo

Apesar das funções circulares serem mais apropriadas para gerarem efeitos periódicos, elas poderão ser usadas como modelos geradores de efeitos de tendência, desde que seja considerado um período extremamente longo relativamente à dimensão da série. Desse modo, perde-se a noção de que o efeito sinusoidal resultante é de facto derivado de uma componente periódica, sendo, portanto, caracterizado como um padrão de tendência.

A seguinte equação foi usada para gerar o segundo tipo de componente de tendência:

$$t_k^{(2)} = 200 + A \cos\left(\frac{2\pi k}{P_1} + w_1\right) + B \cos\left(\frac{2\pi k}{P_2} + w_2\right) \quad (595)$$

onde A e B são números aleatórios gerados a partir de uma distribuição Uniforme contínua, os quais limitam a variação máxima e mínima dos níveis da tendência (i.e., fatores de escala), e P_1 e P_2 são números aleatórios gerados também a partir de uma distribuição Uniforme contínua, representando os períodos longos das ondas harmónicas (consultar a Tabela 4 para ver as condições de simulação).

Função logística

Como é sabido, a função logística encontra aplicações nas mais diversas áreas do conhecimento humano, passando pela demografia, uma das primeiras suas aplicações, e também pela economia, área afim onde se enquadra a aplicação do presente estudo, e

razão principal de considerá-la como um dos modelos geradores de tendência. Assim, a terceira componente de tendência foi obtida usando a equação:

$$t_k^{(3)} = 200 + \frac{L}{1 + \exp\left[-C\left(\frac{k}{N} - 0.5\right)\right]} \quad (606)$$

onde C contém informação sobre a taxa de crescimento da logística e L o nível máximo da curva. A Tabela 4 sumariza as condições de simulação associadas aos procedimentos determinísticos usados na obtenção de componentes de tendência. Para efeitos de comparação visual, a Figura 16 ilustra três componentes de tendência simuladas por estes três primeiros procedimentos.

Tabela 4 – Condições de simulação das componentes de tendência usando modelos determinísticos.

Constantes
$R, L \sim U(20, 80)$
$A, B \sim U(10, 40)$
$P_1, P_2 \sim U(N, 2N)$
$w_1, w_2 \sim U(0, 2\pi)$
$C \sim U(5, 15)$

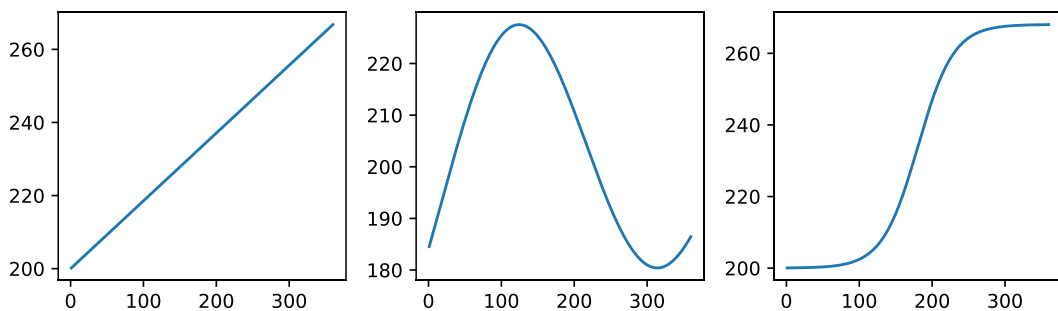


Figura 16 – Componentes de tendência simuladas, $T^{(1)}$, $T^{(2)}$ e $T^{(3)}$, respectivamente, usando modelos determinísticos.

Modelos de regressão de tendência dinâmica

Para contrapor as componentes determinísticas obtidas anteriormente, foram geradas mais três componentes de tendência, mas agora usando modelos de regressão de tendência dinâmica, englobados na chamada família do Passeio Aleatório Generalizado (do inglês, *Generalized Random Walk*; Young, 1994).

A evolução de um processo de Passeio Aleatório Generalizado pode ser descrita através do seguinte modelo de tendência:

$$\begin{cases} a_k = \beta a_{k-1} + \gamma b_{k-1} + \eta_k \\ b_k = \alpha b_{k-1} + \xi_k \end{cases} \quad (617)$$

com $0 < \alpha, \beta, \gamma \leq 1$, onde a_k e b_k registam, respetivamente, as mudanças de nível e o declive. As componentes η_k e ξ_k constituem ruído aditivo, tipicamente com distribuição Normal, de média nula e variâncias constantes σ_η^2 e σ_ξ^2 , respetivamente. A partir deste processo genérico, foram considerados três casos particulares, um para cada tipo de tendência simulada:

- Tendência Amortecida (do inglês, *Damped Trend*):

$$t_k^{(4)} = T_0^{(4)} + a_k, \text{ com } 0 < \alpha < 1, \beta = \gamma = 1, \sigma_\eta \neq 0, \sigma_\xi \neq 0$$

- Tendência Linear Local (do inglês, *Local Linear Trend*):

$$t_k^{(5)} = T_0^{(5)} + a_k, \text{ com } \alpha = \beta = \gamma = 1, \sigma_\eta = 0, \sigma_\xi \neq 0$$

- Passeio Aleatório Alisado (do inglês, *Smoothed Random Walk*), por vezes, chamado de processo Autoregressivo Integrado de 1ª Ordem (do inglês, *First-Order Integrated Autoregressive process*; Rodrigues, 1996):

$$t_k^{(6)} = T_0^{(6)} + a_k, \text{ com } 0 < \alpha < 1, \beta = \gamma = 1, \sigma_\eta = 0, \sigma_\xi \neq 0$$

Quanto mais próximo α está da unidade, para os processos de Tendência Amortecida e Passeio Aleatório Alisado, mais regular é a tendência resultante. As constantes $T_0^{(i)}$ ($i = 4, 5, 6$) representam valores iniciais, e visam a obtenção de observações estritamente positivas e não próximas de zero.

A Tabela 5 sintetiza as condições de simulação das componentes de tendência usando modelos de regressão de tendência dinâmica. A Figura 17 ilustra o resultado que se pode obter para estas três componentes, todas com 360 valores simulados.

Note-se o potencial elevado grau de complexidade que estas componentes podem constituir para o método de Holt-Winters, pois este não é favorecido perante tais comportamentos estocásticos. A ideia que se pretendeu implementar aqui foi a simulação de potenciais situações reais onde, perante as mesmas, a previsão tende a perder precisão principalmente para horizontes temporais mais longos.

Tabela 5 – Condições de simulação das componentes de tendência usando modelos de regressão de tendência dinâmica.

Constantes
$T^{(4)} : \sigma_\eta, \sigma_\xi \sim U(0.01, 0.1); \alpha \sim U(0.975, 0.999)$
$T^{(5)} : \sigma_\xi \sim U(0.01, 0.05)$
$T^{(6)} : \sigma_\xi \sim U(0.05, 0.1); \alpha \sim U(0.975, 0.999)$
$T_0^{(i)} = 200 - \min\{a_k\}; i = 4, 5, 6$

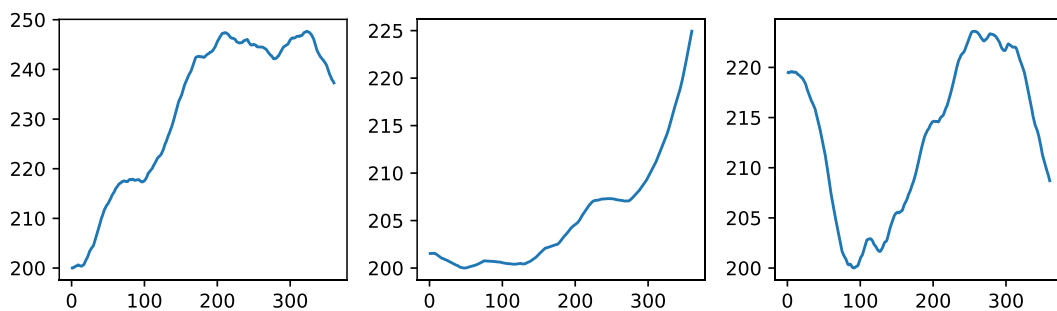


Figura 17 – Componentes de tendência simuladas, $T^{(4)}$, $T^{(5)}$ e $T^{(6)}$, respetivamente, usando modelos de regressão de tendência dinâmica.

4.4.3 Simulação das séries temporais

No total, foram simuladas 30 séries temporais, cada uma resultante da combinação de um par $(t_k^{(i)}, s_k^{(j)})$, com $i = 1, \dots, 6$ e $j = 1, \dots, 5$, de acordo com o modelo de decomposição clássica supracitado (515). Mais concretamente, temos:

$$y_k^{(ij)} = t_k^{(i)} \times s_k^{(j)} + \varepsilon_k^{(ij)} \quad (628)$$

Portanto, por um lado, para o mesmo i , cada tipo de componente de tendência $t_k^{(i)}$ aparece 5 vezes representado no conjunto simulado. Por outro, de forma análoga, cada tipo de componente de sazonalidade $s_k^{(j)}$ aparece 6 vezes representado no conjunto simulado. Com vista a promover uma maior diversidade no conjunto simulado, todo o número aleatório envolvido nos procedimentos de simulação, tanto nas componentes de tendência como nas de sazonalidade, foi gerado tantas vezes quanto as necessárias, usando a distribuição pré-definida correspondente. Assim, é garantido que, para o mesmo tipo de componente de tendência $t_k^{(i)}$, quando combinado com dois tipos de componentes de sazonalidade $s_k^{(j)}$ e $s_k^{(l)}$ distintos ($j \neq l$), as componentes de tendência resultantes nos pares $(t_k^{(i)}, s_k^{(j)})$ e $(t_k^{(i)}, s_k^{(l)})$ não são constituídas exatamente pelos mesmos valores. O mesmo se passa para cada tipo de componente de sazonalidade $s_k^{(j)}$, quando combinado com dois tipos de componentes de tendência $t_k^{(i)}$ e $t_k^{(l)}$ distintos ($i \neq l$).

Para que as séries simuladas pudessem ser obtidas usando o modelo (515), definido por uma sazonalidade multiplicativa, foi necessário proceder a um pré-processamento das componentes de sazonalidade antes destas serem combinadas com as componentes de tendência. Isto deve-se ao facto de as mesmas representarem percentagens (índices sazonais) e, como tal, não conterem valores negativos. Assim, para cada componente sazonal $S^{(j)}$, quando combinada com uma componente de tendência $T^{(i)}$, realizou-se sucessivamente os seguintes passos:

- 1) Reposicionamento relativamente à tendência, com vista à obtenção de valores estritamente positivos e não próximos de zero:

$$s_k^{(j)} = s_k^{(j)} - \bar{s}^{(j)} + \min\{t_k^{(i)}\} \quad (639)$$

A escrita \bar{s} denota a média dos valores s_k ($k = 1, 2, \dots, N$).

- 2) Normalização, com vista à obtenção de índices sazonais: a sequência $s_k^{(j)}$ resultante do passo anterior foi agrupada em grupos de m valores (onde m representa o comprimento do ciclo sazonal), tendo-se realizado a seguinte transformação em cada grupo:

$$s_l := m \times \frac{s_l}{\sum_{j=1}^m s_j}, \quad l = 1, \dots, m \quad (640)$$

Esta transformação garante que, dentro de cada ciclo sazonal, a soma dos índices sazonais seja igual ao comprimento do ciclo.

Para cada série simulada, o ruído $\varepsilon_k^{(ij)}$, adicionado após a transformação das componentes de sazonalidade, foi gerado a partir da distribuição normal $N(0, \sigma_\varepsilon^{(ij)})$, onde $\sigma_\varepsilon^{(ij)}$ define o grau de irregularidade aditiva que se queira impor. Na simulação, considerou-se o ruído gerado pela distribuição Normal padrão (média nula e desvio padrão igual a 1). Como já foi referido, as séries simuladas podem ser vistas no Anexo B.

4.5 Obtenção de resultados e interpretação

Nesta secção, são apresentados os resultados obtidos com recurso aos testes estatísticos através do programa IBM SPSS Statistics (versão 23), tendo em conta as hipóteses nulas e alternativas descritas no capítulo anterior. Os custos simulados e reais obtidos no decorrer deste estudo (recorrendo-se ao programa Python (versão 3.7) para a otimização e estimação do método de Holt-Winters, assim como a implementação dos paradigmas) podem ser consultados nos Anexos C e D, respetivamente. No Anexo G, apresentamos os algoritmos, na linguagem de programação Python, dos procedimentos realizados para a obtenção dos vários custos simulados e reais (Anexos C e D).

O primeiro passo nesta análise foi a aplicação do teste de Shapiro-Wilk para verificar a normalidade dos dados, tanto nos custos reais como nos custos simulados. Conforme se observa nas tabelas do Anexo E, a maioria dos dados violam o pressuposto de normalidade (o nível de significância, representado na coluna “Sig.”, é inferior a 0.05), pelo que optamos por aplicar uma transformação logarítmica aos dados originais.

Em ambas as Tabelas 6 e 7 encontram-se os resultados do teste de Shapiro-Wilk aplicado aos dados logaritmizados relativos às séries reais e simuladas, respetivamente, nas quais se pode constatar que a maioria dos níveis de significância são superiores a 0.05, o que implica a não rejeição da hipótese nula. Deste modo, podemos assumir que os dados provêm de uma distribuição Normal.

No caso da Tabela 7, a interação de fatores P2_LP é o único caso em que não se verifica a normalidade através do teste de Shapiro-Wilk, apesar desta violação não ser muito

severa. Ao observar os gráficos do Anexo F, poderemos assumir que este também segue uma distribuição aproximadamente normal.

Tabela 6 – Teste de normalidade aos custos reais logaritmizados.

Testes de Normalidade						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
P1_CP	,151	24	,166	,954	24	,325
P1_MP	,066	24	,200 [*]	,993	24	1,000
P1_LP	,087	24	,200 [*]	,986	24	,977
P2_CP	,100	24	,200 [*]	,971	24	,698
P2_MP	,136	24	,200 [*]	,971	24	,690
P2_LP	,093	24	,200 [*]	,984	24	,960

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Tabela 7 – Teste de normalidade aos custos simulados logaritmizados.

Testes de Normalidade						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
P1_CP	,100	30	,200 [*]	,957	30	,263
P1_MP	,104	30	,200 [*]	,978	30	,767
P1_LP	,149	30	,086	,964	30	,398
P2_CP	,100	30	,200 [*]	,974	30	,656
P2_MP	,146	30	,104	,936	30	,071
P2_LP	,159	30	,051	,921	30	,028

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Uma vez verificado o pressuposto de normalidade, podemos agora verificar se o pressuposto de esfericidade foi violado, através do teste de Mauchly, como foi referido anteriormente. Na interpretação da tabela do teste de esfericidade de Mauchly, basta analisar os fatores com mais de dois níveis.

Na Tabela 8, encontram-se os resultados referentes aos custos reais logaritmizados. O resultado do teste indica que a esfericidade foi violada para o Horizonte Temporal e a interação deste com o Paradigma, pelo facto de os níveis de significância serem inferiores

a 0.05 (sig = 0.0, para o Horizonte Temporal, e sig = 0.026, para a interação entre os 2 fatores).

Tabela 8 – Teste de esfericidade de Mauchly aos custos reais logaritmizados.

Teste de esfericidade de Mauchly^a

Medida: MEASURE_1

Efeito dentre-sujeitos	W de Mauchly	Aprox. Qui-quadrado	gl	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Limite inferior
Paradigma	1,000	,000	0	.	1,000	1,000	1,000
Horizonte_Temporal	,352	22,997	2	,000	,607	,622	,500
Paradigma *							
Horizonte_Temporal	,716	7,337	2	,026	,779	,825	,500

Testa a hipótese nula para a qual a matriz de covariâncias de erro das variáveis transformadas ortonormalizadas é proporcional em relação a uma matriz identidade.

a. Design: Intercepto

Design Dentre-Sujeitos: Paradigma + Horizonte_Temporal + Paradigma * Horizonte_Temporal

b. Pode ser usado para ajustar os graus de liberdade dos testes de significância dentro da média. Os testes corrigidos são exibidos na tabela Testes de efeitos dentre-sujeitos.

Nos casos em que a esfericidade foi violada, foi necessário aplicar as correções, como foi indicado no capítulo anterior, e, para isso, deve-se tomar em conta o valor do parâmetro épsilon (grau com que o pressuposto de esfericidade é violado) para cada um dos métodos de correção, que é visto também na Tabela 8.

Para o fator Horizonte Temporal, o valor de épsilon é inferior a 0.75 e, portanto, consideramos a correção de Greenhouse-Geisser. Para o caso da interação entre os dois fatores, uma vez que o valor de épsilon é superior a 0.75, consideramos a correção de Huynh-Feldt. Só por questões de registo, o valor de épsilon associado ao fator Paradigma é igual a 1 em ambas as colunas das correções (Greenhouse-Geisser e Huynh-Feldt), pelo que a esfericidade é verificada na sua plenitude, não sendo necessário qualquer correção.

Seguidamente, segue-se um dos passos mais importantes da ANOVA de medidas repetidas, que é o estudo dos efeitos dentre-sujeitos, cujos resultados são apresentados na Tabela 9 (para os custos reais). Para o fator Paradigma, o nível de significância foi de 0.624, pelo que se conclui que os dados não fornecem evidência suficiente para rejeitar a hipótese nula. Relativamente ao fator Horizonte Temporal e a interação deste com o Paradigma, os níveis de significância foram de 0.000 (com a correção de Greenhouse-

Geisser) e 0.001 (com a correção de Huynh-Feldt), respetivamente, de modo que se rejeita a hipótese nula em ambos estes casos.

Tabela 9 – Teste de efeitos dentre-sujeitos aos custos reais logaritmizados.

Testes de efeitos dentre-sujeitos

Medida: MEASURE_1

Origem		Tipo III Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Paradigma	Esfericidade considerada	,004	1	,004	,246	,624
	Greenhouse-Geisser	,004	1,000	,004	,246	,624
	Huynh-Feldt	,004	1,000	,004	,246	,624
	Limite inferior	,004	1,000	,004	,246	,624
Erro(Paradigma)	Esfericidade considerada	,390	23	,017		
	Greenhouse-Geisser	,390	23,000	,017		
	Huynh-Feldt	,390	23,000	,017		
	Limite inferior	,390	23,000	,017		
Horizonte_Temporal	Esfericidade considerada	,455	2	,227	41,208	,000
	Greenhouse-Geisser	,455	1,213	,375	41,208	,000
	Huynh-Feldt	,455	1,245	,365	41,208	,000
	Limite inferior	,455	1,000	,455	41,208	,000
Erro(Horizonte_Temporal)	Esfericidade considerada	,254	46	,006		
	Greenhouse-Geisser	,254	27,905	,009		
	Huynh-Feldt	,254	28,629	,009		
	Limite inferior	,254	23,000	,011		
Paradigma * Horizonte_Temporal	Esfericidade considerada	,074	2	,037	9,441	,000
	Greenhouse-Geisser	,074	1,558	,048	9,441	,001
	Huynh-Feldt	,074	1,651	,045	9,441	,001
	Limite inferior	,074	1,000	,074	9,441	,005
Erro(Paradigma*Horizonte_Temporal)	Esfericidade considerada	,181	46	,004		
	Greenhouse-Geisser	,181	35,837	,005		
	Huynh-Feldt	,181	37,967	,005		
	Limite inferior	,181	23,000	,008		

Com estes resultados, podemos concluir que, para um nível de 95% de confiança, não há diferenças significativas entre os diferentes paradigmas. Em contrapartida, as diferenças entre os diferentes horizontes temporais e as diferenças entre as combinações dos 2 fatores são significativas, com níveis de significância iguais a 0.000 e 0.001, respetivamente.

Para interpretar melhor a interação dos fatores, podemos usar o gráfico da Figura 18. Uma primeira observação é que, independentemente do paradigma, os custos aumentam com o horizonte temporal. No entanto, o tipo de paradigma parece influenciar a forma como esses custos variam com o horizonte temporal. Enquanto que, para um horizonte temporal de curto prazo, os custos aumentam do paradigma P1 para o paradigma P2, para horizontes temporais mais longos, passa-se exatamente o inverso, ou seja, os custos diminuem do paradigma P1 para o paradigma P2.

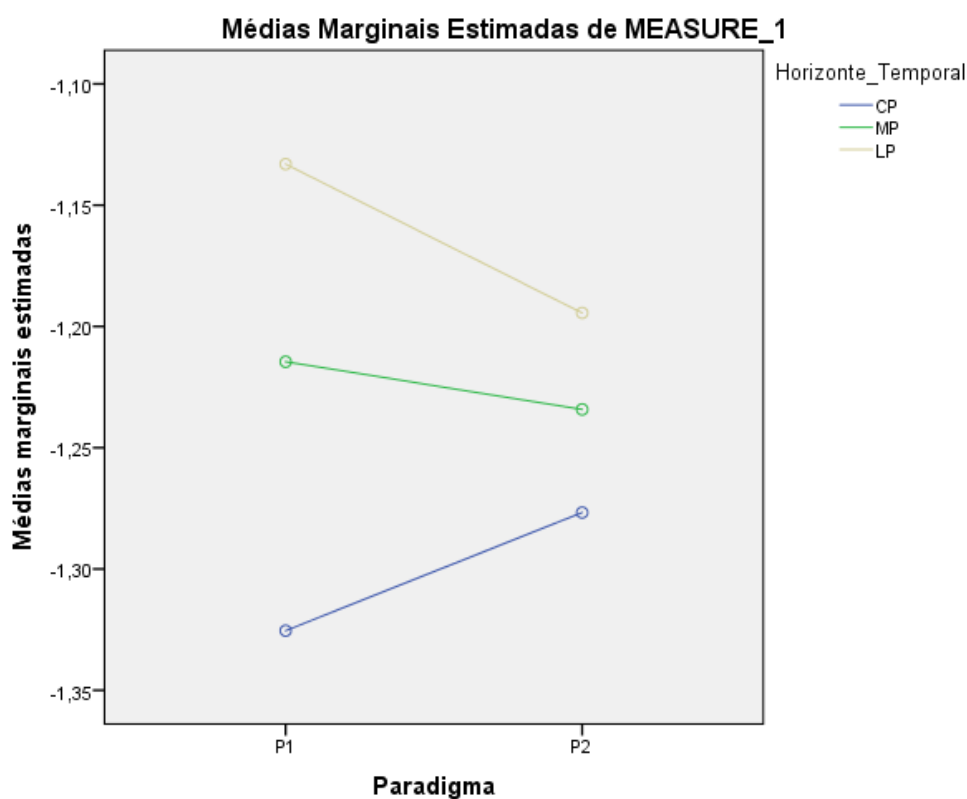


Figura 18 – Interação dos fatores Paradigma e Horizonte Temporal para os custos reais logaritmizados.

Uma vez que existem diferenças no fator Horizonte Temporal, realizou-se uma comparação através do procedimento de Bonferroni, para determinar quais os pares de níveis desse fator que apresentam diferenças significativas entre si. Os resultados são apresentados na tabela 10.

Tabela 10 – Procedimento de Bonferroni aos custos reais logaritmizados.

Comparações por Método Pairwise

Medida: MEASURE_1

(I) Horizonte_ Temporal	(J) Horizonte_ Temporal	Diferença média (I-J)	Erro Erro	Sig. ^b	95% Intervalo de Confiança para Diferença ^b	
					Limite inferior	Limite superior
CP	MP	-,077*	,017	,000	-,119	-,034
	LP	-,137*	,019	,000	-,187	-,088
MP	CP	,077*	,017	,000	,034	,119
	LP	-,061*	,007	,000	-,079	-,042
LP	CP	,137*	,019	,000	,088	,187
	MP	,061*	,007	,000	,042	,079

Baseado em médias marginais estimadas

*. A diferença média é significativa no nível ,05.

b. Ajustamento para diversas comparações: Bonferroni.

No SPSS, as diferenças médias estatisticamente significativas são marcadas com um asterisco (*), o que acontece, como regra geral, quando o nível de significância é inferior a 0.05. Assim, de acordo com os resultados obtidos, podemos afirmar que existem diferenças estatisticamente significativas entre todos os pares de níveis dos diferentes tipos de horizontes temporais.

Na Tabela 11, encontram-se os resultados do teste de esfericidade de Mauchly referentes aos custos simulados. Similarmente ao que aconteceu nos resultados dos custos reais, o teste indica que a esfericidade foi violada para o fator Horizonte Temporal e para a interação deste com o Paradigma. Como, em ambos os casos, os valores de épsilon são inferiores a 0.75, consideramos a correção de Greenhouse-Geisser em ambas as situações.

Tabela 11 – Teste de esfericidade de Mauchly aos custos simulados logaritimizados.

Teste de esfericidade de Mauchly^a

Medida: MEASURE_1

Efeito dentre-sujeitos	W de Mauchly	Aprox. Qui-quadrado	gl	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Limite inferior
Paradigma	1,000	,000	0	.	1,000	1,000	1,000
Horizonte_Temporal	,096	65,508	2	,000	,525	,528	,500
Paradigma * Horizonte_Temporal	,320	31,944	2	,000	,595	,606	,500

Testa a hipótese nula para a qual a matriz de covariâncias de erro das variáveis transformadas ortonormalizadas é proporcional em relação a uma matriz identidade.

a. Design: Intercepto

Design Dentre-Sujeitos: Paradigma + Horizonte_Temporal + Paradigma * Horizonte_Temporal

b. Pode ser usado para ajustar os graus de liberdade dos testes de significância dentro da média. Os testes corrigidos são exibidos na tabela Testes de efeitos dentre-sujeitos.

Seguidamente, realizou-se o estudo dos efeitos dentre-sujeitos (desta feita, com os custos simulados) usando as correções correspondentes, cujos resultados podem ser vistos na Tabela 12. Como os níveis de significância foram de 0.002 para o fator Paradigma, 0.000 para o fator Horizonte Temporal (com a correção de Greenhouse-Geisser) e 0.036 para a interação destes fatores (também com a correção de Greenhouse-Geisser), a hipótese nula foi rejeitada em todos estes casos. Deste modo, podemos dizer que existem diferenças para os diferentes níveis dos fatores.

Com estes resultados, podemos concluir que, para um nível de 95% de confiança, há diferenças significativas entre os diferentes paradigmas, entre os diferentes horizontes temporais e entre as combinações dos dois fatores. Apesar de haver indicação de que existe interação dos fatores, para o nível de confiança dado, essa interação não é tão forte como a do caso dos custos associados às séries reais. De acordo com a Figura 19, essa interação aparenta ser suave, onde as linhas associadas aos horizontes temporais de curto e médio prazos são sensivelmente paralelas e com uma inclinação ligeiramente superior à inclinação da linha associada ao horizonte temporal de longo prazo.

Tabela 12 – Teste de efeitos dentre-sujeitos aos custos simulados logaritmizados.

Testes de efeitos dentre-sujeitos

Medida: MEASURE_1

Origem		Tipo III Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Paradigma	Esfericidade considerada	,362	1	,362	11,134	,002
	Greenhouse-Geisser	,362	1,000	,362	11,134	,002
	Huynh-Feldt	,362	1,000	,362	11,134	,002
	Limite inferior	,362	1,000	,362	11,134	,002
Erro(Paradigma)	Esfericidade considerada	,943	29	,033		
	Greenhouse-Geisser	,943	29,000	,033		
	Huynh-Feldt	,943	29,000	,033		
	Limite inferior	,943	29,000	,033		
Horizonte_Temporal	Esfericidade considerada	1,628	2	,814	36,317	,000
	Greenhouse-Geisser	1,628	1,051	1,549	36,317	,000
	Huynh-Feldt	1,628	1,056	1,541	36,317	,000
	Limite inferior	1,628	1,000	1,628	36,317	,000
Erro(Horizonte_Temporal)	Esfericidade considerada	1,300	58	,022		
	Greenhouse-Geisser	1,300	30,468	,043		
	Huynh-Feldt	1,300	30,628	,042		
	Limite inferior	1,300	29,000	,045		
Paradigma * Horizonte_Temporal	Esfericidade considerada	,027	2	,014	4,440	,016
	Greenhouse-Geisser	,027	1,190	,023	4,440	,036
	Huynh-Feldt	,027	1,212	,022	4,440	,036
	Limite inferior	,027	1,000	,027	4,440	,044
Erro(Paradigma*Horizonte_Temporal)	Esfericidade considerada	,177	58	,003		
	Greenhouse-Geisser	,177	34,514	,005		
	Huynh-Feldt	,177	35,147	,005		
	Limite inferior	,177	29,000	,006		

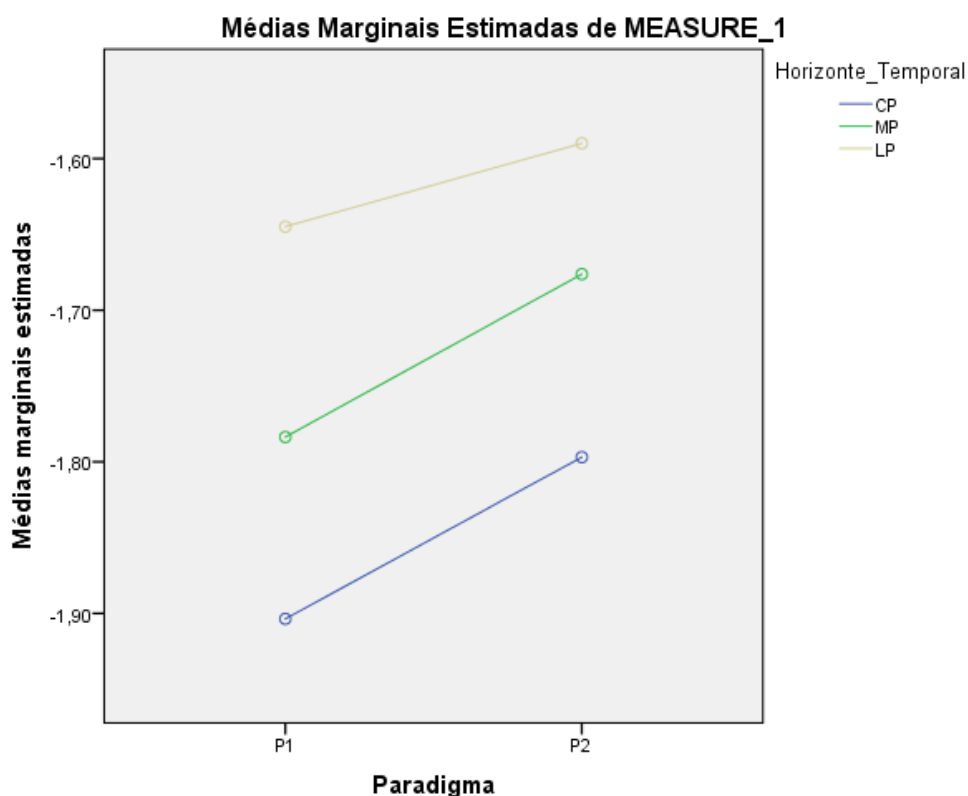


Figura 19 – Interação dos fatores Paradigma e Horizonte Temporal para os custos simulados logaritmizados.

Apesar de não ser muito evidente na Figura 19, poderá fazer sentido conjecturar que a interação resultante dos dois fatores em estudo tende para uma situação análoga à verificada com os dados reais, ou seja, que o paradigma P2 tem o potencial de diminuir a variabilidade das médias dos custos associadas aos diferentes horizontes temporais, quando comparada com o que ocorre no paradigma P1. No limite, para horizontes temporais cada vez maiores, o paradigma P2 poderá ser preferível ao paradigma P1. Contudo, seriam necessários mais estudos para averiguar tal situação e se existem outros fatores em causa que possam influenciar as conclusões.

No fator Paradigma verificou-se que existem diferenças significativas entre os diferentes níveis, mas uma vez que só existem 2 níveis, não é necessário aplicar o procedimento de Bonferroni. No caso do fator Horizonte Temporal, como existem 3 níveis diferentes, foi efetuada a comparação através do procedimento de Bonferroni para determinar quais os

pares de níveis que apresentam diferenças significativas. Os resultados estão apresentados na Tabela 13.

Tabela 13 – Procedimento de Bonferroni aos custos simulados logaritimizados.

Comparações por Método Pairwise

Medida: MEASURE_1

(I) Horizonte_ Temporal	(J) Horizonte_ Temporal	Diferença média (I-J)	Erro Erro	Sig. ^b	95% Intervalo de Confiança para Diferença ^b	
					Limite inferior	Limite superior
CP	MP	-,120*	,023	,000	-,180	-,061
	LP	-,233*	,038	,000	-,329	-,137
MP	CP	,120*	,023	,000	,061	,180
	LP	-,113*	,016	,000	-,153	-,072
LP	CP	,233*	,038	,000	,137	,329
	MP	,113*	,016	,000	,072	,153

Baseado em médias marginais estimadas

*. A diferença média é significativa no nível ,05.

b. Ajustamento para diversas comparações: Bonferroni.

Assim como se verificou nos resultados dos custos reais, vemos que existem diferenças estatisticamente significativas entre todos os pares de níveis dos diferentes horizontes temporais.

CAPÍTULO 5

5 CONCLUSÕES E TRABALHO FUTURO

Esta dissertação destinou-se a apresentar um estudo comparativo entre dois tipos de análises, uma preditiva e outra prescritiva, como ferramentas de apoio nas tomadas de decisão no contexto da gestão de stocks e previsão.

Provavelmente, a função mais usada na análise preditiva, é o erro quadrático médio, onde os erros cometidos por excesso e por defeito têm o mesmo peso. No entanto, em muitas situações práticas, observa-se que os problemas de previsão contêm erros não simétricos, ou seja, os custos decorrentes de errar por excesso são diferentes dos de errar por defeito.

Neste sentido, realizou-se primeiramente uma revisão da literatura, na qual se verificou que não existem praticamente estudos que façam a comparação entre estes paradigmas, apesar de haver diversas análises sobre previsões com funções de custo assimétrica, mostrando que os custos de uma superestimação e de uma subestimação são diferentes.

A análise comparativa apoiou-se num conjunto de dados reais e simulados, onde analisamos situações práticas e reais de vários países, através de séries temporais reais e situações simuladas, com a intenção de obter resultados mais abrangentes, e de um conjunto de dados mais diverso.

Os resultados obtidos para os custos reais e simulados são bastantes análogos. Apesar de, nos custos reais, se concluir que não há diferenças significativas entre os diferentes paradigmas, verificou-se que o tipo de paradigma parece influenciar a forma como esses custos variam com o horizonte temporal. Nos custos simulados, os resultados indicaram que existem diferenças significativas entre os diferentes paradigmas, entre os diferentes horizontes temporais e entre as combinações dos dois fatores.

Desta forma, consideramos que os objetivos desta dissertação foram cumpridos, pelo facto de, para além da comparação entre os diferentes paradigmas, foi possível analisar o efeito do horizonte temporal nos resultados.

Consideramos também que esta dissertação contribui para o conhecimento de duas abordagens na tomada de decisão, no âmbito da gestão de stocks, baseadas na previsão de séries temporais, mas de igual maneira, poderá contribuir para outras áreas.

No âmbito da previsão, o método escolhido foi o método de Holt-Winters por ser eficiente, de fácil entendimento e de simples aplicação, mas sabemos que existem outros métodos de previsão que poderiam ser implementados em futuras investigações. Como já referimos, pode-se aplicar este estudo comparativo noutras áreas, utilizando também amostras mais amplas, ou até outras variáveis que não foram consideradas neste estudo.

Por fim, estamos convencidos de que este é um tema em que há ainda muito por percorrer e que, portanto, é um bom campo de trabalho para outros investigadores.

BIBLIOGRAFIA

- Adhikari, R. e Agrawal, R. K. *An Introductory Study on Time Series Modeling and Forecasting*. Bergisch Gladbach: LAP Lambert Academic Publishing, 2013.
- Armstrong, J. S. “Findings from evidence-based forecasting: Methods for reducing forecast error.” *International Journal of Forecasting*, 22, 2006: 583-598.
- Balakrishnan, N., Render, B. e Stair, R. M. “Inventory Control Models.” Cap. 12 em *Managerial Decision Modeling With Spreadsheets*. Pearson Education Limited, 2013.
- Berk, R. “Asymmetric loss functions for forecasting in criminal justice settings.” *Journal of Quantitative Criminology*, 27(1), 2011: 107-123.
- Braga, M. *Gestão do Aproveitamento*. Lisboa: Editorial Presença, 1991.
- Brown, R. G. *Exponential Smoothing for Predicting Demand*. Cambridge, Massachusetts: Arthur D. Little Inc., 1956.
- . *Smoothing Forecasting and Prediction of Discrete Time Series*. Englewood Cliffs, NJ: Prentice-Hall, 1963.
- Chatfield, C. “The Holt-Winters forecasting procedure.” *Journal of the Royal Statistical Society. Series C*, 27(3), 1978: 264-279.
- Cortez, P. “Sensitivity Analysis for Time Lag Selection to Forecast Seasonal Time Series Using Neural Networks and Support Vector Machine.” *The 2010 International Joint Conference on Neural Networks*. Barcelona, Spain, 2010. 1-8.
- Crone, S. F. “Artificial neural networks for time series prediction - A novel approach to inventory management using asymmetric cost functions.” *International Conference on Artificial Intelligence*. Las Vegas, Nevada, USA: CSREA Press, Athens, 2003. 193-199.
- D'Agostino, R. B. “Transformation to normality of the null distribution of g_1 .” *Biometrika*, 57(3), 1970: 679-681.
- Diebold, F. X. e Mariano, R. S. “Comparing predictive accuracy.” *Journal of Business & Economic Statistics*, 13, 1995: 253-263.

- Dress, K., Lessmann, S. e Mettenheim, H.-J. von. "Residual value forecasting using asymmetric cost functions." *International Journal of Forecasting*, 34(4), 2018: 551-565.
- Freitas, P. S. A. *The Combination of Neural Estimates in Prediction and Decision Problems*. Tese de Doutoramento, Lisboa: Faculdade de Ciências da Universidade de Lisboa (FCUL), 2008.
- Gardner Jr., E. S. e Acar, Y. "Fitting the damped trend method of exponential smoothing." *Journal of the Operational Research Society*, 70(6), 2019: 926-930.
- Gardner Jr., E. S. e McKenzie, E. "Forecasting trends in time series." *Management Science*, 31, 1985: 1237-1246.
- Gardner Jr., E. S. e McKenzie, E. "Model identification in exponential smoothing." *Journal of the Operational Research Society*, 39, 1988: 863-867.
- Gardner Jr., E. S. e McKenzie, E. "Seasonal exponential smoothing with damped trends." *Management Science*, 35, 1989: 372-376.
- Gardner Jr., E. S. e McKenzie, E. "Why the damped trend works." *Journal of the Operational Research Society*, 62, 2011: 1177-1180.
- Goffe, W. L., Ferrier, G. D. e Rogers, J. "Global optimization of statistical functions with simulated annealing." *Journal of Econometrics*, 60(1-2), 1994: 65-99.
- Gonçalves, E. e Lopes, N. *Séries Temporais: Modelações Lineares e Não Lineares*. Lisboa: Sociedade Portuguesa de Estatística, 2008.
- Gonçalves, J. F. *Gestão de Aprovisionamentos*. Porto: Publindústria - Edições Técnicas, 2012.
- Granger, C. W. J. e Newbold, P. *Forecasting Economic Time Series*. 2ª Ed. Academic Press, 1986.
- Harris, F. W. "How many parts to make at once." *Factory, the Magazine of Management*, 10(152), 1913: 135-152.
- Hillier, F. e Lieberman, G. *Introduction to Operations Research*. 2ª Ed. New York: McGraw Hill, 2001.

- Holt, C. E. *Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52)*. Pittsburgh USA: Carnegie Institute of Technology, 1957.
- Hyndmann, R. J. e Athanasopoulos, G. *Forecasting: Principles and Practice*. 2ª Ed. Melbourne, Australia: OTexts, 2018.
- Hyndmann, R. J., Koehler, A. B., Ord, J. K. e Snyder, R. D. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, 2008.
- Leitch, G. e Tanner, J. E. “Economic forecast evaluation: profits versus the conventional error measures.” *The American Economic Review*, 81, 1991: 580-590.
- Mackey, M. C. e Glass, L. “Oscillation and chaos in physiological control systems.” *Science*, 197(4300), 1977: 287-289.
- Makridakis, S., Wheelwright, S. C. e Hyndman, R. J. *Forecasting: Methods and Applications*. 3ª Ed. New Jersey: John Wiley & Sons, Inc., 1998.
- Mauchly, J. W. “Significance test for sphericity of a normal n-variate distribution.” *The Annals of Mathematical Statistics*, 11(2), 1940: 204-209.
- Neideen, T., e Brasel K. “Understanding Statistical Tests.” *Journal os Surgical Education*, 64 (2), 2007: 93-96.
- Ord, K., Koehler, A. e Snyder, R. D. “Estimation and prediction for a class of dynamic nonlinear statistical models.” *Journal of the American Statistical Association*, 92(440), 1997: 1621-1629.
- Pozo, H. *Administração de Recursos Materiais e Patrimoniais*. São Paulo: Atlas, 2002.
- Renard, P., Alcolea, A. e Ginsbourger, D. “Stochastic versus deterministic approaches.” Cap. 8 em *Environmental Modelling: Finding Simplicity in Complexity*., de J. e Mulligan, M. Wainwright, 133-149. 2ª Ed. John Wiley & Sons, Inc., 2013.
- Rodrigues, A. J. *Dynamic Regression and Supervised Learning Methods in Time Series Modelling and Forecasting*. Tese de Doutorado, Lancaster: Lancaster University, 1996.
- Sawyer, S. P. “Analysis of Variance: The Fundamental.” *The Journal of Manual & Manipulative Therapy*, 17 (2), 2009: 27-38.

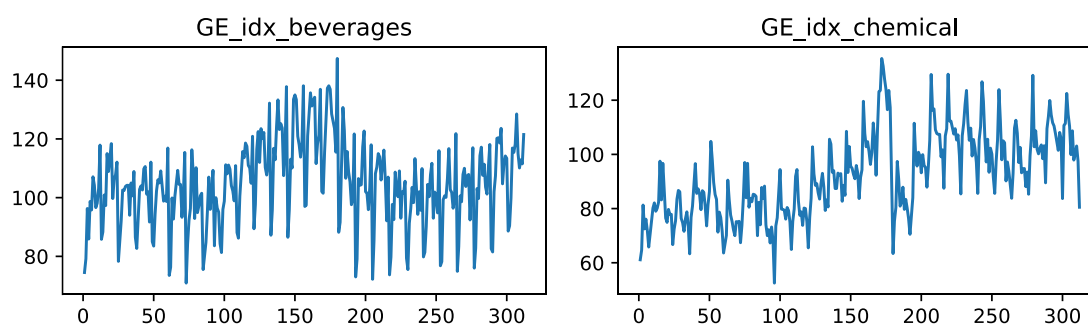
- Shapiro, S. S. e Wilk, M. “An analysis of variance test for normality (complete samples).” *Biometrika*, 52, 1965: 591-611.
- Snyder, R. D., Koehler, A. B. e Ord, J. K. “Forecasting for inventory control with exponential smoothing.” *International Journal of Forecasting*, 18, 2002: 5-18.
- Wilson, R. H. “A scientific routine for stock control.” *Harvard Business Review*, 13, 1934: 116-128.
- Winters, P. R. “Forecasting sales by exponentially weighted moving averages.” *Management Science*, 6, 1960: 324-342.
- Yaffee, R. A. e McGee, M. *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*. Amsterdam: Academic Press, 2002.
- Young, P. C. “Time variable parameters and trend estimation in nonstationary economic time series.” *Journal of Forecasting*, 13, 1994: 179-210.
- Zermati, P. *A Gestão de Stocks*. Tradução: Garcia, M. H. Lisboa: Editorial Presença, 1986.
- Zhao, H., Sinha, A. e Bansal, G. “An extended tuning method for cost-sensitive regression and forecasting.” *Decision Support Systems*, 51(3), 2011: 372-383.

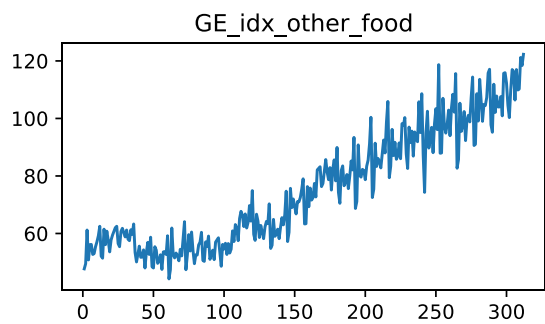
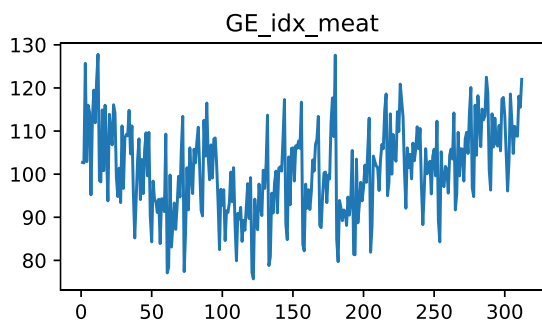
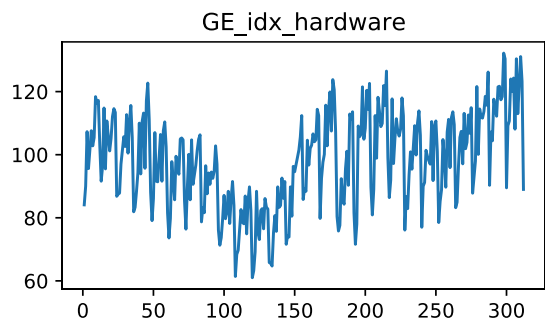
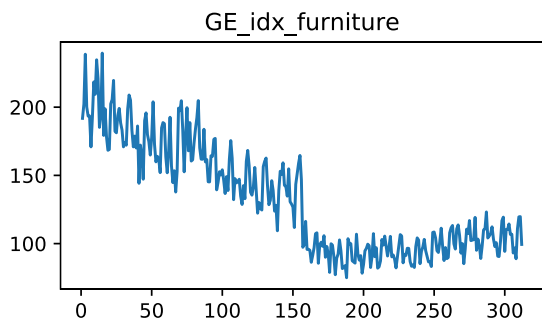
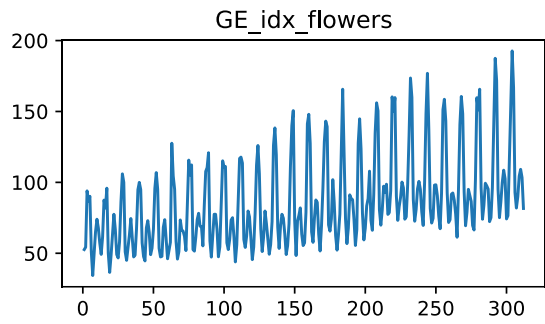
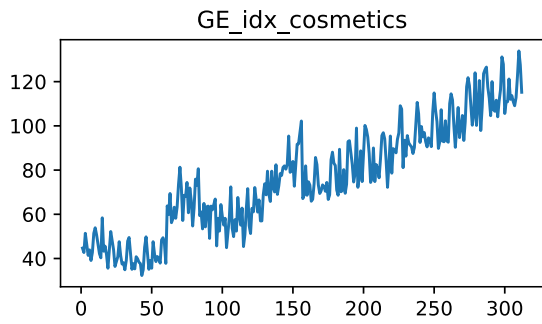
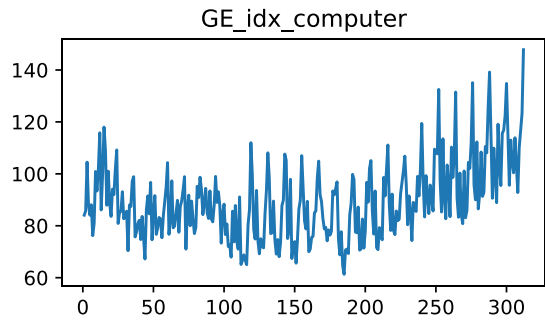
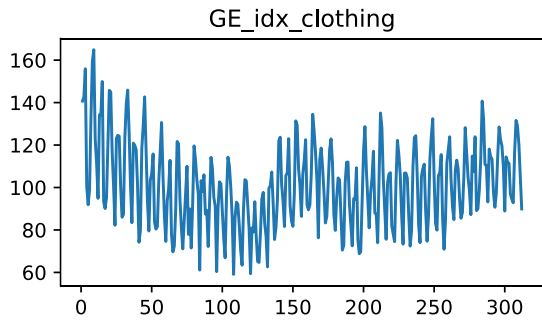
ANEXO A Gráficos das séries temporais reais

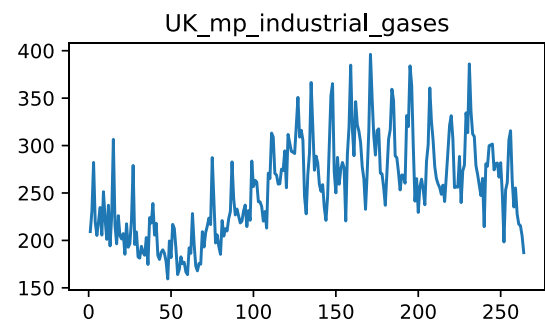
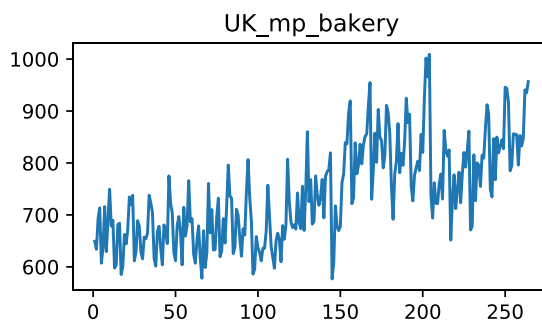
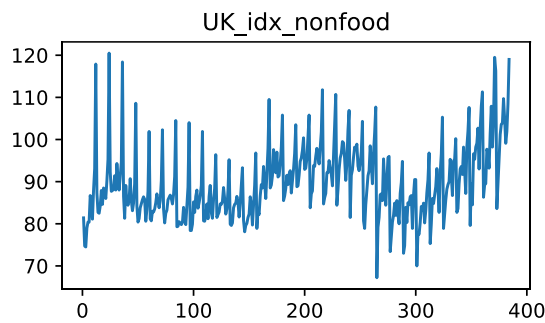
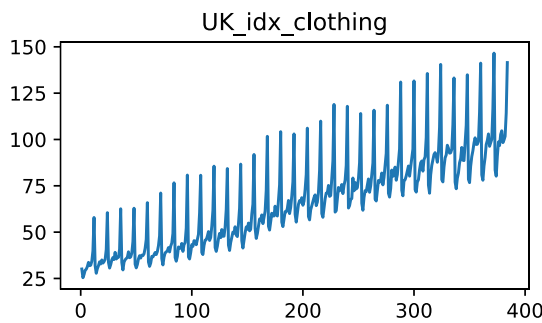
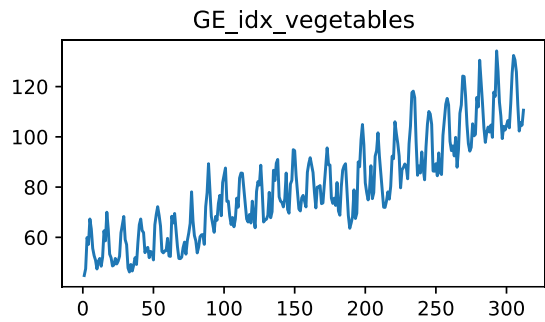
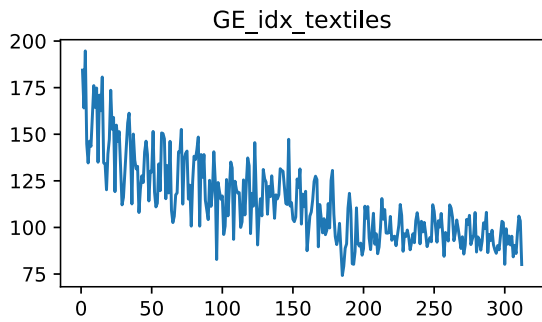
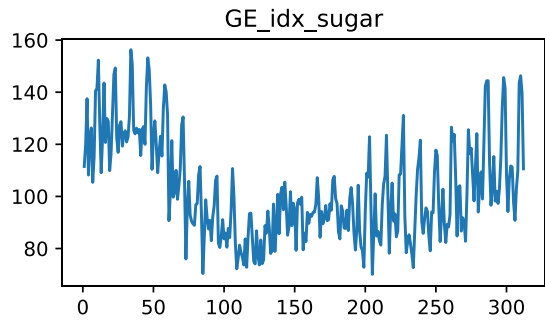
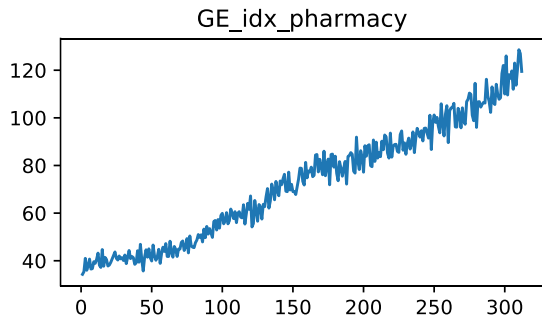
As várias séries reais consideradas no estudo foram obtidas a partir das seguintes fontes:

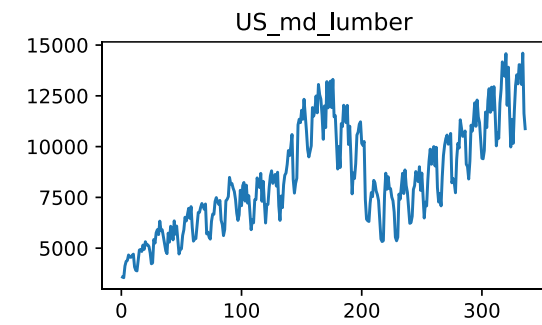
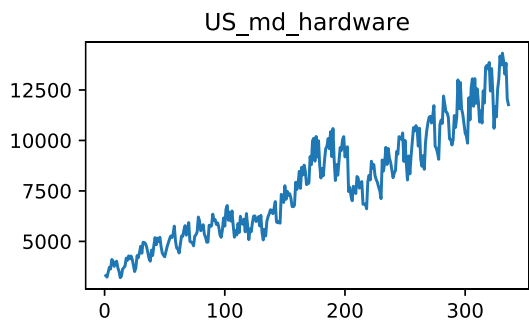
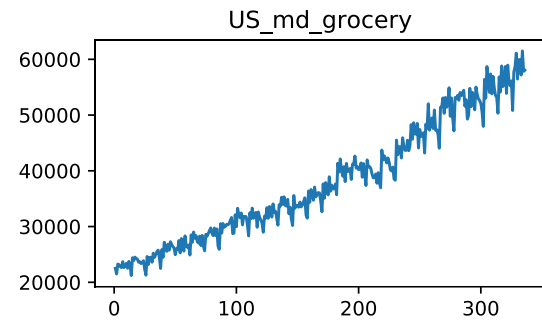
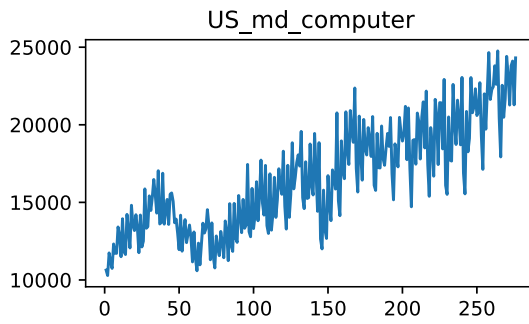
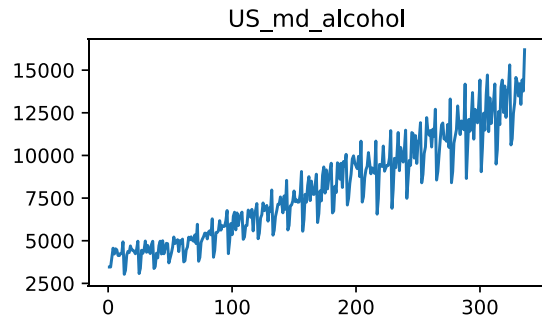
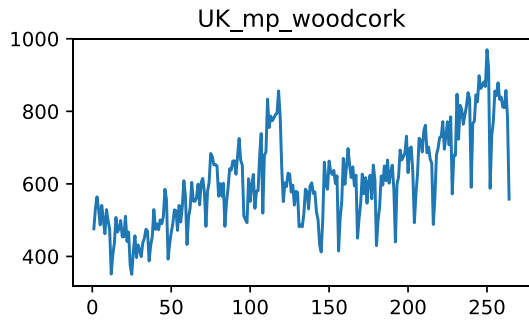
- United States Census Bureau, Monthly Retail Trade:
Acedido em 9 de novembro de 2020, em
<https://www.census.gov/econ/currentdata/dbsearch?program=MARTS&startYear=1992&endYear=2020&categories=44X72&dataType=SM&geoLevel=US&adjusted=1¬Adjusted=1&errorData=0>
- Office for National Statistics, Time Series Explorer:
Acedido em 14 de novembro de 2020, em
<https://www.ons.gov.uk/timeseriestool?uri=timeseriestool&uri=timeseriestool&uri=timeseriestool&uri=timeseriestool&uri=timeseriestool&uri=timeseriestool&uri=timeseriestool&topic=/businessindustryandtrade/retailindustry&page=7>
- Destatis Statistisches Bundesamt Genesis-Online, Database of the Federal Statistical Office of Germany:
Acedido em 14 de novembro de 2020, em
<https://www-genesis.destatis.de/genesis/online?operation=statistic&levelindex=0&levelid=1605354481363&code=45211#abreadcrumb>

Nos gráficos representados a seguir, os títulos apresentam três tipos de informação, nomeadamente: o país (GE=Alemanha, UK=Reino Unido, US=Estados Unidos da América), a unidade (idx=índices, mp=milhões de libras, md=milhões de dólares) e a variável de interesse.

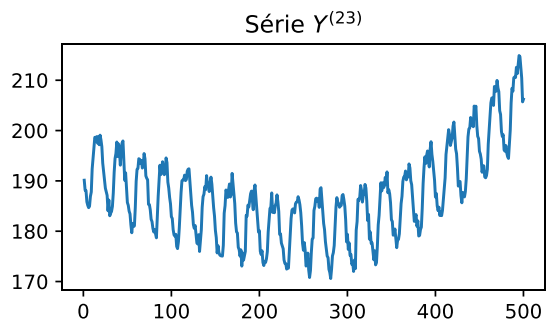
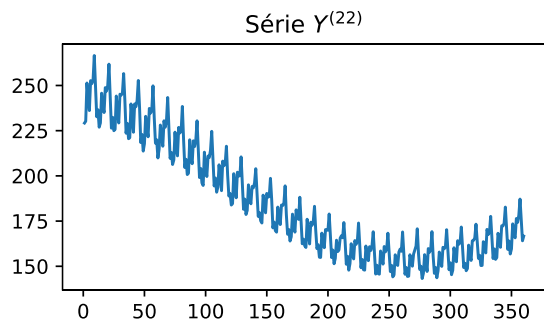
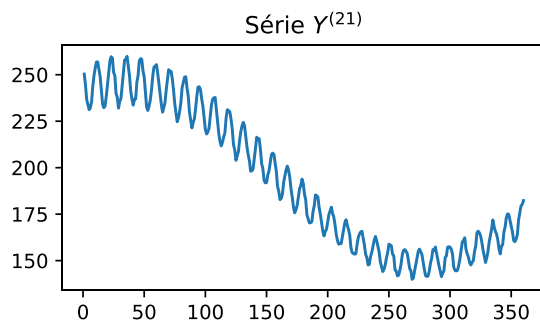
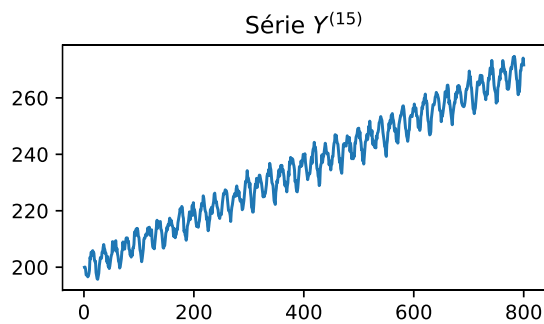
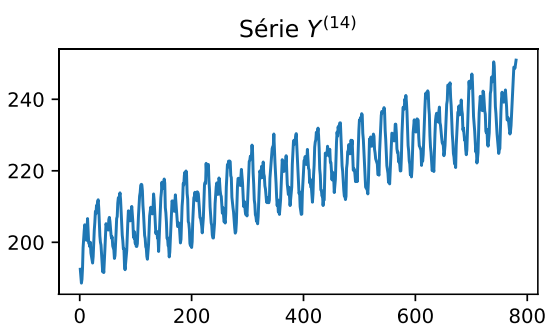
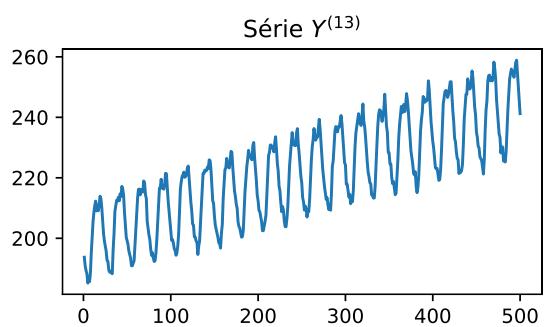
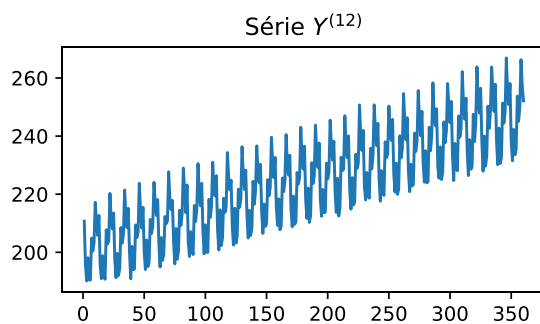
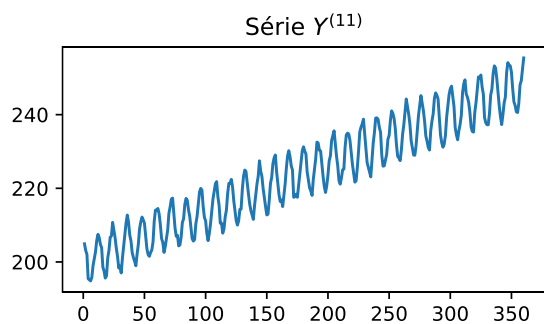


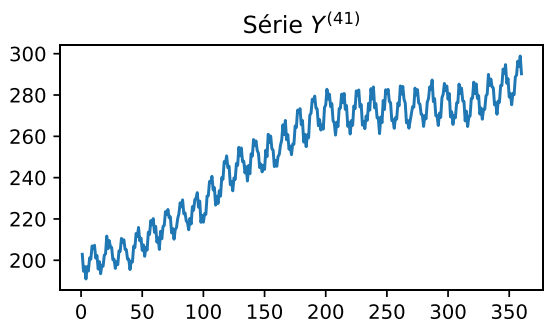
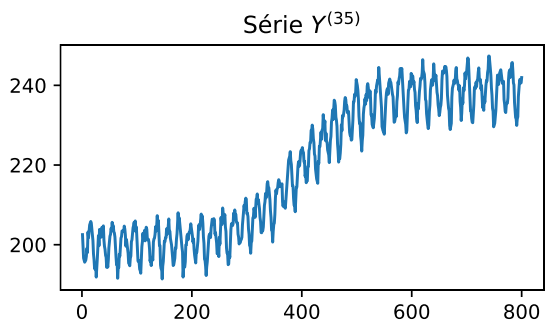
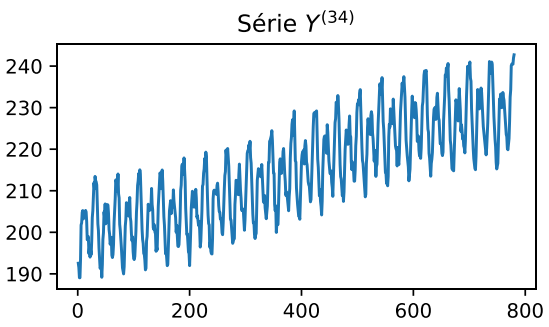
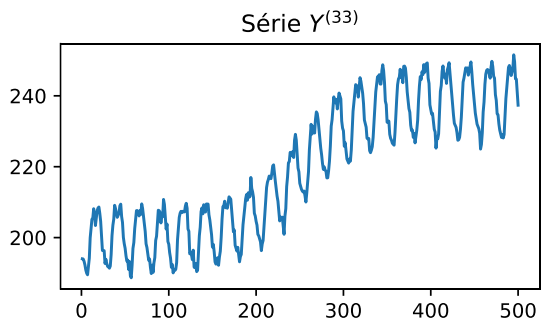
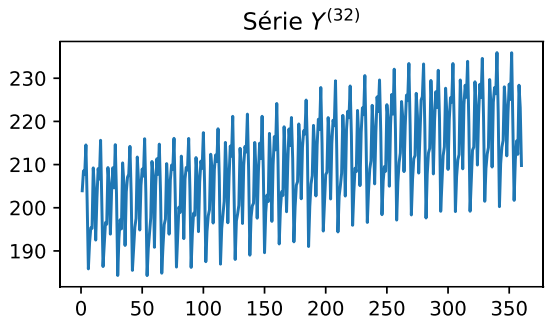
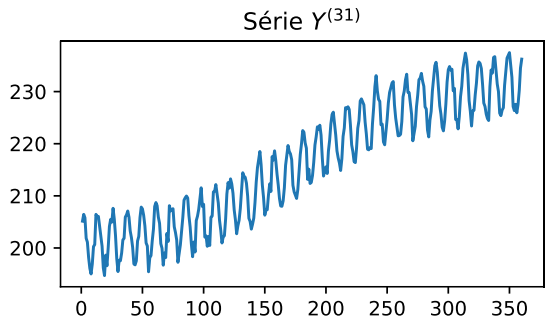
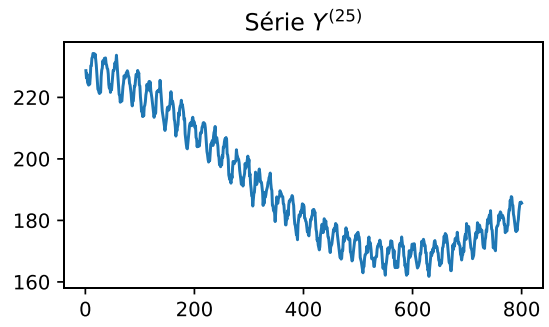
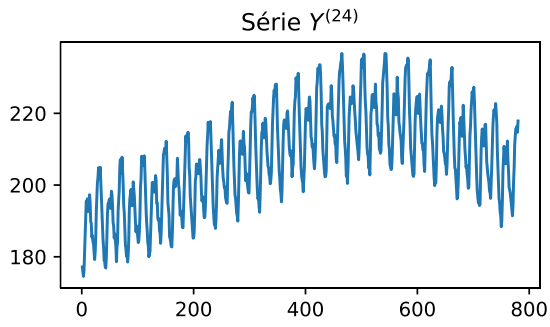


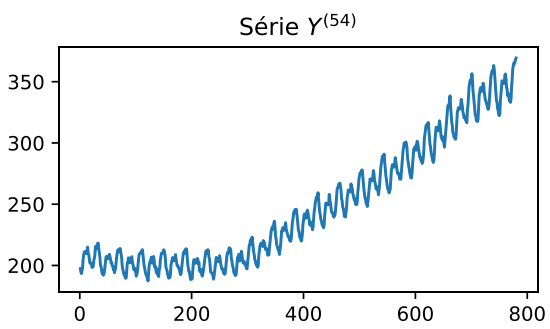
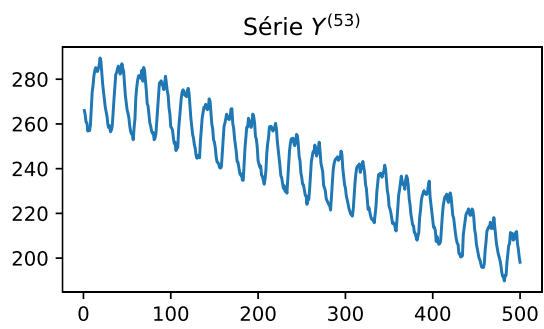
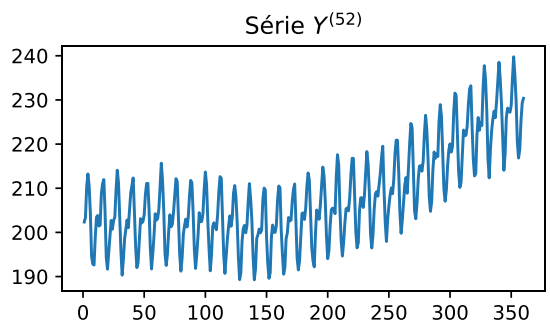
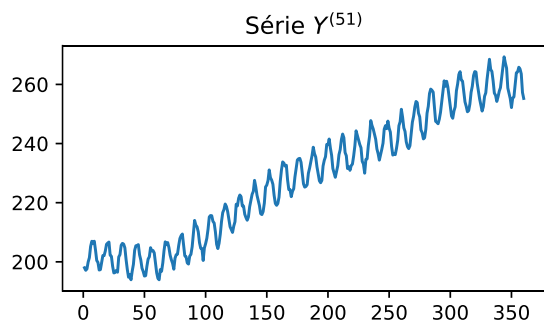
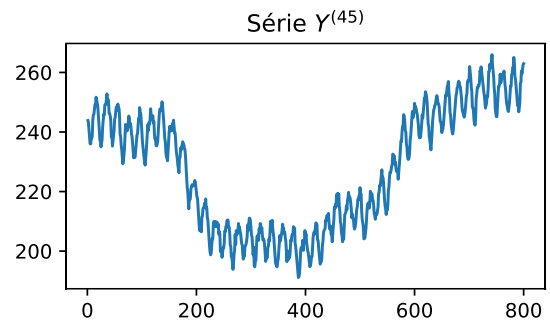
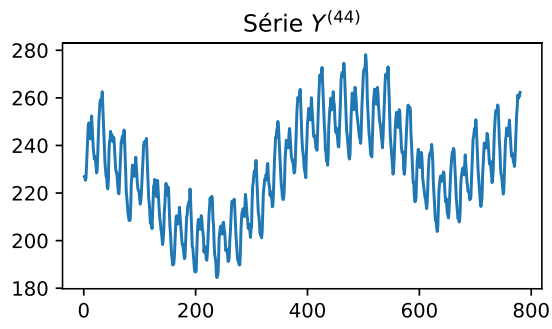
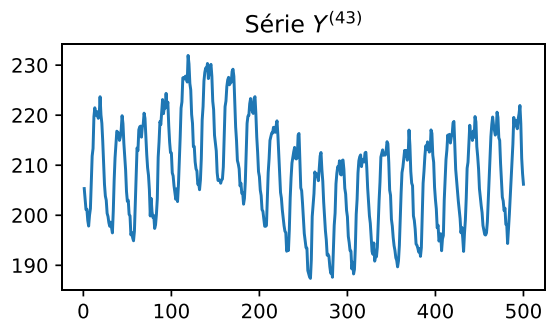
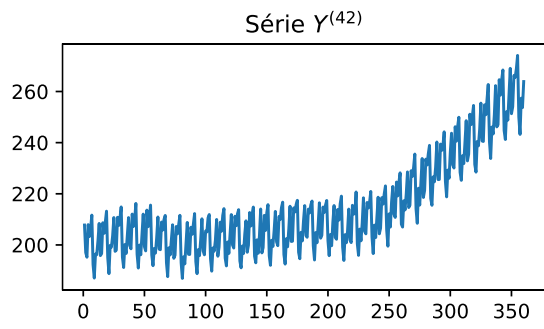


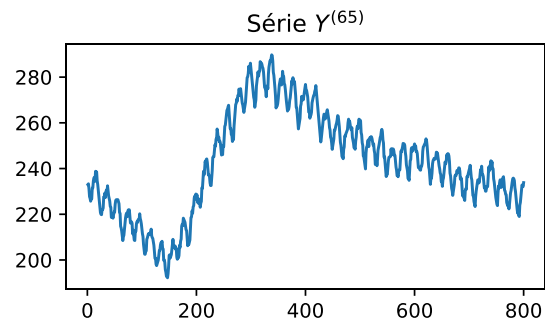
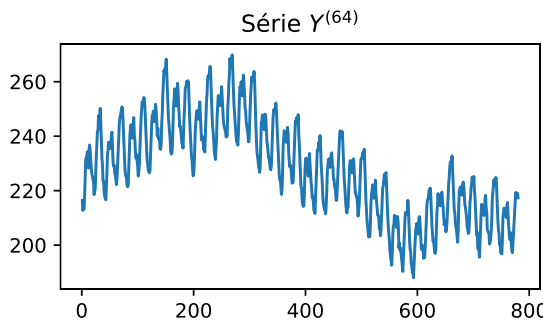
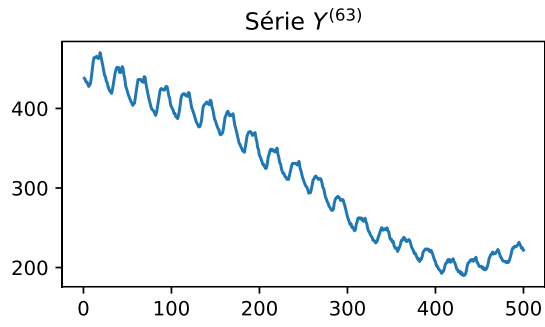
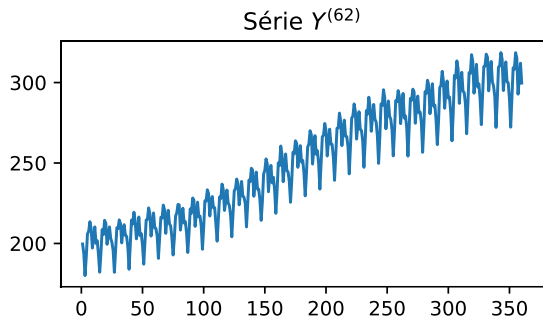
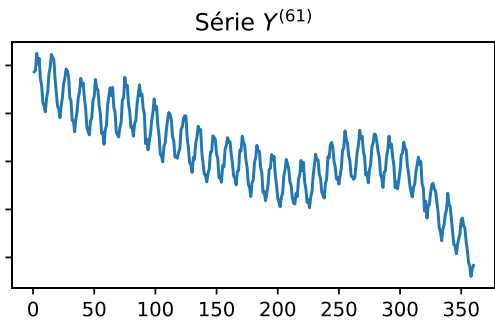
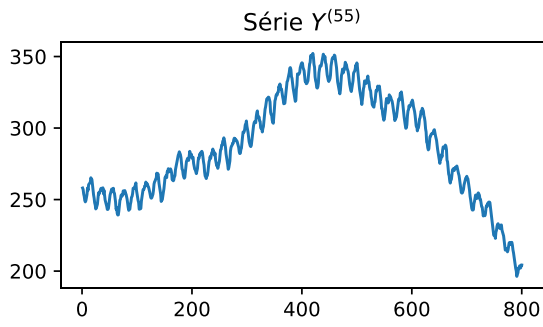


ANEXO B Gráficos das séries temporais simuladas









ANEXO C Custos obtidos para as séries temporais simuladas

	P1_CP	P1_MP	P1_LP	P2_CP	P2_MP	P2_LP
1	0.011411	0.012034	0.014085	0.011334	0.011543	0.011751
2	0.012315	0.010658	0.014135	0.009123	0.008889	0.009602
3	0.012122	0.012077	0.012951	0.012156	0.012485	0.012091
4	0.020812	0.020835	0.019693	0.02106	0.02121	0.020559
5	0.011469	0.011463	0.011491	0.01154	0.011626	0.011594
6	0.00522	0.006069	0.006375	0.005806	0.011621	0.018515
7	0.006553	0.006698	0.007512	0.013805	0.017197	0.021523
8	0.024074	0.024739	0.031733	0.041194	0.047839	0.054937
9	0.023895	0.024696	0.027216	0.024773	0.024994	0.025177
10	0.010831	0.013317	0.015716	0.02499	0.030187	0.040376
11	0.017536	0.016929	0.017579	0.017975	0.020041	0.019089
12	0.017513	0.02073	0.018119	0.017817	0.018304	0.018316
13	0.011574	0.014269	0.019576	0.013619	0.018996	0.022021
14	0.026526	0.025489	0.03967	0.026145	0.026576	0.026682
15	0.015091	0.017463	0.027374	0.01784	0.02263	0.024995
16	0.008565	0.011548	0.017376	0.008642	0.011613	0.01629
17	0.012089	0.014246	0.017499	0.013091	0.013811	0.016321
18	0.022371	0.030441	0.043102	0.049996	0.061302	0.074055
19	0.019295	0.032879	0.060804	0.045014	0.076767	0.125029
20	0.014814	0.034482	0.075192	0.030298	0.07286	0.128001
21	0.010541	0.013219	0.019306	0.010458	0.012829	0.016054
22	0.017266	0.018966	0.022813	0.027981	0.030018	0.027666
23	0.009213	0.011344	0.023183	0.009126	0.010425	0.011977
24	0.00773	0.011561	0.016773	0.008219	0.012487	0.017359
25	0.009782	0.021476	0.042877	0.011877	0.030012	0.050762
26	0.013529	0.021273	0.032828	0.023168	0.03148	0.046007
27	0.005808	0.007621	0.010184	0.006215	0.008081	0.0108
28	0.003192	0.007601	0.015173	0.005168	0.010975	0.02014
29	0.021362	0.044687	0.075245	0.024865	0.049729	0.069989
30	0.014056	0.051457	0.104674	0.026222	0.04711	0.052263

ANEXO D Custos obtidos para as séries temporais reais

	P1_CP	P1_MP	P1_LP	P2_CP	P2_MP	P2_LP
1	0.073145	0.075387	0.083656	0.068907	0.07185	0.080765
2	0.06007	0.160106	0.159857	0.064479	0.102432	0.118351
3	0.051899	0.064233	0.082001	0.084398	0.083589	0.091818
4	0.059939	0.074685	0.097281	0.058163	0.062529	0.065391
5	0.040347	0.041122	0.05698	0.047439	0.051651	0.061169
6	0.057041	0.052491	0.066735	0.057737	0.055907	0.058698
7	0.03652	0.044121	0.044455	0.052752	0.05044	0.052155
8	0.074931	0.090704	0.103771	0.108654	0.107929	0.118337
9	0.087299	0.096125	0.108621	0.153811	0.153002	0.154335
10	0.041787	0.040461	0.045162	0.039041	0.041501	0.041529
11	0.025711	0.028661	0.036155	0.024445	0.024996	0.029955
12	0.069105	0.100994	0.150364	0.073882	0.076446	0.074524
13	0.062719	0.060403	0.06305	0.057118	0.057391	0.05756
14	0.036916	0.048646	0.056562	0.043894	0.049175	0.049214
15	0.01511	0.017974	0.017519	0.015019	0.016841	0.018655
16	0.060327	0.080979	0.0977	0.056755	0.073977	0.086052
17	0.0839	0.117433	0.131949	0.13562	0.166969	0.200525
18	0.099054	0.136074	0.14966	0.12566	0.129761	0.13378
19	0.05743	0.230457	0.257428	0.073045	0.066665	0.08775
20	0.022603	0.023694	0.032206	0.021936	0.022785	0.023537
21	0.04606	0.061373	0.129784	0.049014	0.057213	0.057252
22	0.02158	0.025245	0.032664	0.020115	0.025231	0.027075
23	0.028598	0.036698	0.042402	0.026945	0.03462	0.044282
24	0.042883	0.059357	0.071493	0.039666	0.056415	0.068568

ANEXO E Testes de normalidade dos dados originais

Testes de Normalidade dos custos Reais

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
P1_curto	,085	24	,200*	,978	24	,854
P1_medio	,159	24	,118	,859	24	,003
P1_longo	,120	24	,200*	,897	24	,019
P2_curto	,171	24	,067	,900	24	,021
P2_medio	,168	24	,078	,900	24	,022
P2_longo	,143	24	,200*	,907	24	,030

*. Este é um limite inferior da significância verdadeira.

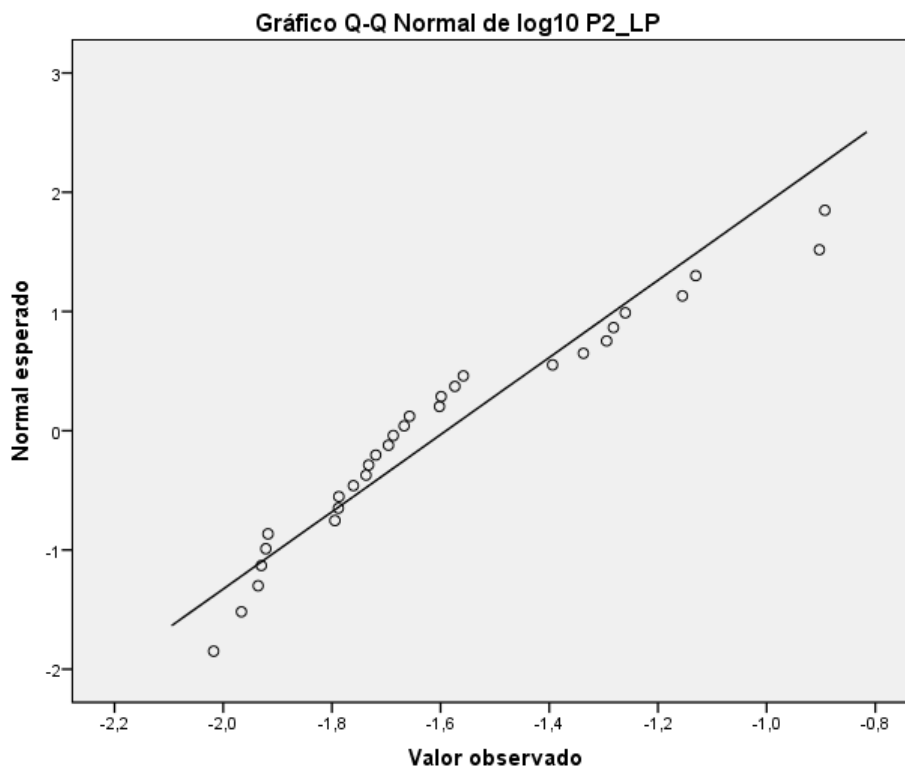
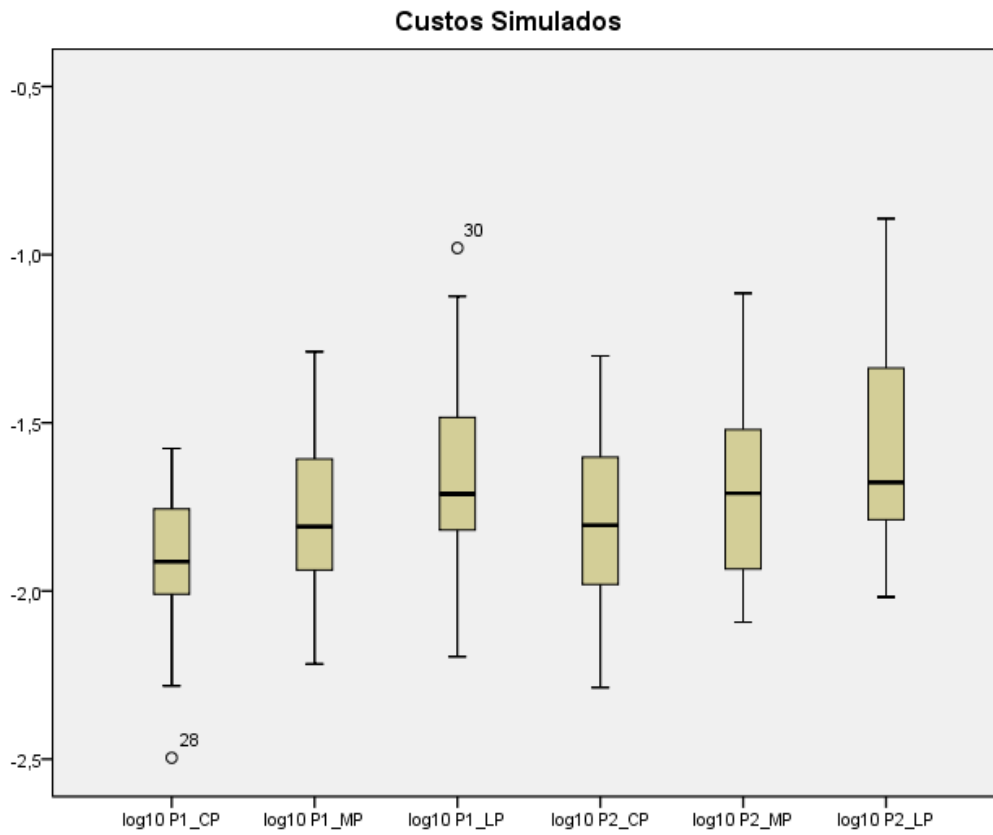
a. Correlação de Significância de Lilliefors

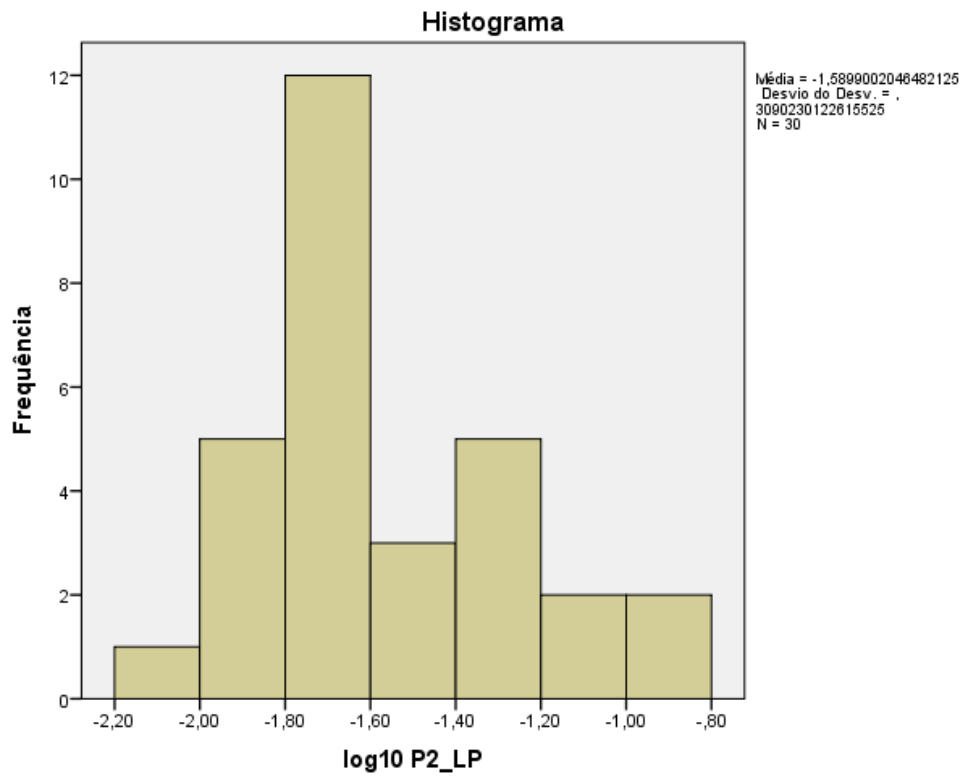
Testes de Normalidade dos custos Simulados

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
P1_curto	,135	30	,169	,968	30	,494
P1_medio	,167	30	,032	,875	30	,002
P1_longo	,228	30	,000	,770	30	,000
P2_curto	,173	30	,023	,889	30	,005
P2_medio	,189	30	,008	,810	30	,000
P2_longo	,282	30	,000	,719	30	,000

a. Correlação de Significância de Lilliefors

ANEXO F Gráficos associados a custos simulados





ANEXO G Scripts em Python

```
import math
import numpy as np
import scipy.stats as ss

def mse(Y, Yhat, *args):
    '''
    Erro quadrático médio

    Input:
        Y = (Y(n0+1),...,Y(n)) série temporal (n-n0 observações)
        Yhat = (Yhat(n0+1),...,Yhat(n)) previsões h-passos-à-frente (n-n0 estimativas)
    Output:
        perf = valor da medida de desempenho na previsão
    '''
    E = Y-Yhat
    return np.dot(E,E)/len(E)

def funcao_utilidade(Y,Q,u,*args):
    '''
    Função de custos assimétrica

    Input:
        Y = (Y(n0+1),...,Y(n)) série temporal (n-n0 observações)
        Q = (Q(n0+1),...,Q(n)) sequência de decisões h-passos-à-frente (n-n0 estimativas)
        u = peso (0<u<1) associado aos erros de sinal positivo (representando a assimetria
entre erros de sinal positivo e negativo)
        Por exemplo: se cu = custo por unidade de produto em falta (erro de sinal
positivo) em cada período, e co = custo por unidade de produto em excesso (erro de sinal
negativo) em cada período, então, u = cu/(cu+co)
    Output:
        D = (pseudo) custo total na decisão; para saber o custo real, basta calcular
D*(cu+co)
    '''
    if isinstance(u,tuple):
        u = u[0]
    E = Y-Q
    I = E >= 0.0
    J = np.logical_not(I)
    E[I] = u*E[I] # erros de sinal positivo
    E[J] = (u-1.0)*E[J] # erros de sinal negativo
    return np.mean(E)

def holt_winters(Y,h,hp,param0):
    '''
    Método de Holt-Winters

    Input:
        Y = (Y(1),...,Y(n)) série temporal (n observações)
```

```

    h = horizonte temporal das previsões
    hp = (alfa,beta,gama,phi) constantes de alisamento
    param0 = (a0,b0,c0) valores iniciais para o nível, declive e índices sazonais,
com c0 = (c0(1),...,c0(m))
Output:
    Yhat = (Yhat(1+h),...,Y(n),...,Y(n+h)) previsões h-passos-à-frente (n estimativas)
    ...
    alfa,beta,gama,phi = hp
    phi_h = np.sum(phi**np.arange(1,h+1))
    a0,b0,c0 = param0; c1 = np.copy(c0)
    n,m = len(Y),len(c0)
    Yhat = np.empty(n+h)
    for k in range(n):
        a1 = alfa*(Y[k]/c1[k%m]) + (1-alfa)*(a0 + phi*b0) # nível
        b1 = beta*(a1-a0) + (1-beta)*phi*b0 # declive
        c1[k%m] = gama*(Y[k]/a1) + (1-gama)*c1[k%m] # índices sazonais
        a0 = a1; b0 = b1
        Yhat[k+h] = (a1 + phi_h*b1) * c1[(k+h)%m] # previsões h-passos-à-frente
    return Yhat

def init_holt_winters(Y,m):
    ...
    Calcula valores iniciais para o método de Holt-Winters (2º método)

    Input:
        Y = (Y(1),...,Y(m),...,Y(q*m)) observações durante os primeiros q (>=2) ciclos sazonais
        m = comprimento do ciclo sazonal
    Output:
        param0 = (a0,b0,c0) valores iniciais para o nível, declive e índices sazonais,
com c0 = (c0(1),...,c0(m))
    ...
    n,d = len(Y),m//2
    # (i) médias móveis centradas
    T = np.empty(n)
    if m%2: # m ímpar
        for k in range(d,n-d):
            T[k] = np.mean(Y[k-d:k+d])
    else: # m par
        for k in range(d,n-d):
            T[k] = 0.5*(np.mean(Y[k-d:k+d-1])+np.mean(Y[k-d+1:k+d]))
    # (ii) índices sazonais
    S = Y/T # série sem tendência
    c0 = np.empty(m)
    for k in range(d,n-d):
        c0[k%m] = np.mean(S[k:n-d:m])
    # (iii) ajustamento sazonal
    for k in range(0,n,m):
        mk = min(m,n-k)
        T[k:k+mk] = Y[k:k+mk]/c0[:mk] # série sem sazonalidade

```

```

# (iv) regressão linear
p = np.polyfit(np.linspace(1.0,n,n),T,1)
return (p[1],p[0],c0)

def pesquisa_em_grelha(Y,h,grid_hp,param0,n0,funcao,*args):
    '''
    Realiza uma pesquisa de valores pré-definidos numa grelha (adaptada para o método de
    Holt-Winters).
    Estes valores são depois usados para a inicialização do simulated annealing.

    Input:
        Y = (Y(1),...,Y(n)) série temporal (n observações)
        h = horizonte temporal das previsões
        grid_hp = (galfa,gbeta,ggama,gphi) grelha de valores para as constantes de
    alisamento
        param0 = (a0,b0,c0) valores iniciais para o nível, declive e índices sazonais,
    com c0 = (c0(1),...,c0(m))
        n0 = número de observações iniciais ignoradas na avaliação do método de Holt-
    Winters
        funcao = rotina/função que avalia o método de Holt-Winters
        args = argumentos adicionais para a rotina funcao

    Output:
        best_hp = (alfa,beta,gama,phi) melhores valores encontrados para as constantes de
    alisamento
    '''
    n = len(Y)
    galfa,gbeta,ggama,gphi = grid_hp
    emin = np.inf
    print("Pesquisa em grelha:")
    for ip in range(len(gphi)):
        for ia in range(len(galfa)):
            for ib in range(len(gbeta)):
                for ig in range(len(ggama)):
                    hp = (galfa[ia],gbeta[ib],ggama[ig],gphi[ip])
                    # avaliação do método de Holt-Winters
                    Yhat = holt_winters(Y,h,hp,param0)
                    perf = funcao(Y[n0:n],Yhat[n0:n],args)
                    if (perf < emin):
                        emin = perf
                        best_hp = hp
                print("{} / {} : best_hp= {} ; emin= {}".format(ip+1,len(gphi),best_hp,emin))
    return best_hp

def simulated_annealing(Y,h,hp0,param0,n0,funcao,*args):
    '''
    Algoritmo de otimização estocástica (adaptado para o método de Holt-Winters)

    Input:
        Y = (Y(1),...,Y(n)) série temporal (n observações)
        h = horizonte temporal das previsões

```

```

    hp0 = (alfa,beta,gama,phi) valores iniciais para as constantes de alisamento
    param0 = (a0,b0,c0) valores iniciais para o nível, declive e índices sazonais,
com c0 = (c0(1),...,c0(m))
    n0 = número de observações iniciais ignoradas na avaliação do método de Holt-
Winters
    funcao = rotina/função que avalia o método de Holt-Winters
    args = argumentos adicionais para a rotina funcao
Output:
    best_hp = (alfa,beta,gama,phi) melhores valores encontrados para as constantes de
alisamento
'''
n = len(Y)
T0 = 800.0 # temperatura inicial (tipicamente, muito alta)
T1 = 1.0 # temperatura final (congelamento)
Trate = 0.85 # taxa de redução da temperatura
N = 10000 # número de tentativas para estabilizar a solução
# uma solução inicial
S0 = np.asarray(hp0)
# limites do espaço de procura para as constantes de alisamento
bounds_min = np.asarray((0.0001,0.0001,0.0001,0.80))
bounds_max = np.asarray((0.9999,0.9999,0.9999,1.05))
# raio de vizinhança de S0 para a obtenção de uma nova solução
radius = np.asarray((0.01,0.01,0.01,0.001))
# avaliação do método de Holt-Winters para a solução corrente
Yhat = holt_winters(Y,h,hp0,param0)
perf0 = funcao(Y[n0:n],Yhat[n0:n],args)
# melhor solução encontrada até o momento
best_S = S0
best_perf = perf0
min_perf = perf0
print("Simulated annealing:")
while T0 > T1: # repetir até atingir o congelamento
    for j in range(N):
        # nova solução é escolhida aleatoriamente de uma vizinhança de S0
        S1 = S0 + (2*np.random.rand(4)-1)*radius
        # averigua se a nova solução está dentro dos limites do espaço de procura
        I = S1 < bounds_min
        J = S1 > bounds_max
        S1[I] = bounds_min[I]
        S1[J] = bounds_max[J]
        # avaliação do método de Holt-Winters para a nova solução
        Yhat = holt_winters(Y,h,S1,param0)
        perf1 = funcao(Y[n0:n],Yhat[n0:n],args)
        # verifica se houve melhorias
        if perf1 < perf0: # movimento downhill
            update = True
        else: # movimento uphill (aceita uma pior solução com probabilidade P)
            P = math.exp(-(perf1-perf0)/T0)
            update = np.random.rand() < P
        if update: # atualiza a solução corrente

```

```

        S0 = S1
        perf0 = perf1
        if perf0 < min_perf:
            min_perf = perf0
            best_S = S0
    # averigua se deve reduzir a temperatura
    if min_perf < best_perf: # temperatura é mantida (solução ainda é instável)
        best_perf = min_perf
        print("T= {} ; best_S= {} ; best_perf= {}".format(T0,best_S,best_perf))
    else: # temperatura é reduzida (porque não houve melhorias)
        print("T= {} ; perf_atual= {}".format(T0,perf0))
        T0 = T0*Trate
    return best_S

```

def analise_preditiva(Y,m,h,u):

```

'''
Calcula o custo de decisão usando o paradigma P1

Input:
    Y = (Y(1),...,Y(n)) série temporal (n observações)
    m = comprimento do ciclo sazonal
    h = horizonte temporal das previsões
    u = peso (0<u<1) associado aos erros de sinal positivo, representando a assimetria
entre erros de sinal positivo e negativo (na tese, foi considerado u=0.9)

Output:
    D = custo de decisão
'''
# normalização da série temporal
Y = Y/np.std(Y)
n = len(Y)
nE = 50 # número mínimo de erros para o cálculo de quantis
n0 = 6*m+nE # número de observações iniciais ignoradas na avaliação do método de Holt-
Winters (idealmente, 6*m >> h)
param0 = init_holt_winters(Y[:2*m],m)
# Otimizar Holt-Winters usando MSE
# -----
# valores iniciais para o simulated annealing
grid = np.linspace(0.1,0.9,9)
grid_hp = (grid,grid,grid,(1.0,))
hp = pesquisa_em_grelha(Y,h,grid_hp,param0,n0,mse)
# otimização das constantes de alisamento
hp = simulated_annealing(Y,h,hp,param0,n0,mse)
# previsões ótimas
Yhat = holt_winters(Y,h,hp,param0)
# Erros de previsão
# -----
E = Y-Yhat[:n] # existem apenas n-h erros disponíveis, mas os primeiros serão ignorados
# Cálculo de quantis
# -----
Qe = np.empty(n)

```

```

for k in range(n0-h,n-h):
    data = E[n0-h-nE:k] # informação disponível até o instante k
    stat1,p1 = ss.shapiro(data) # teste de Shapiro-Wilk
    stat2,p2 = ss.normaltest(data) # teste de D'Agostino K^2
    if p1 > 0.05 and p2 > 0.05: # quantis a partir da distribuição Normal
        mu,sigma = np.mean(data),np.std(data,ddof=1)
        z = ss.norm.ppf(u)
        Qe[k+h] = mu+z*sigma
    else: # quantis empíricos
        Qe[k+h] = np.quantile(data,u)
# Valor da função de utilidade
# -----
Q = Yhat[:n] + Qe # decisões baseadas em previsões
D = funcao_utilidade(Y[n0:n],Q[n0:n],u)
return D

def analise_prescritiva(Y,m,h,u):
    ...
    Calcula o custo de decisão usando o paradigma P2

    Input:
        Y = (Y(1),...,Y(n)) série temporal (n observações)
        m = comprimento do ciclo sazonal
        h = horizonte temporal das previsões
        u = peso (0<u<1) associado aos erros de sinal positivo, representando a assimetria
entre erros de sinal positivo e negativo (na tese, foi considerado u=0.9)
    Output:
        D = custo de decisão
    ...
    # normalização da série temporal
    Y = Y/np.std(Y)
    n = len(Y)
    n0 = 6*m+50 # número de observações iniciais ignoradas na avaliação do método de Holt-
Winters (idealmente, 6*m >> h)
    param0 = init_holt_winters(Y[:2*m],m)
    # Otimizar Holt-Winters usando a função de utilidade
    # -----
    # valores iniciais para o simulated annealing
    grid = np.linspace(0.1,0.9,9)
    grid_hp = (grid,grid,grid,(1.0,))
    hp = pesquisa_em_grelha(Y,h,grid_hp,param0,n0,funcao_utilidade,u)
    # otimização das constantes de alisamento
    hp = simulated_annealing(Y,h,hp,param0,n0,funcao_utilidade,u)
    # decisões ótimas
    Q = holt_winters(Y,h,hp,param0)
    # Valor da função de utilidade
    # -----
    D = funcao_utilidade(Y[n0:n],Q[n0:n],u)
    return D

```