

MESTRADO
ECONOMIA E ADMINISTRAÇÃO EMPRESAS

Métodos de Previsão: aplicação prática no volume de negócios de uma empresa de consultoria

Tiago Marques Teixeira da Cunha

M

2022



FACULDADE DE ECONOMIA



Métodos de Previsão: aplicação prática no volume de negócios de uma empresa de consultoria

Tiago Marques Teixeira da Cunha

Dissertação

Mestrado em Economia e Administração de Empresas

Orientado por

Alexandra Patrícia Horta Ramos

2022

Agradecimentos

Finda esta etapa, acredito que a mesma não tenha sido alcançada sem aqueles que me apoiaram e motivaram a tornar este objetivo numa conquista real, dirigindo-me especialmente às pessoas que de alguma forma tornaram possível a realização desta dissertação.

À minha família, em especial aos meus pais, agradeço-lhes o incentivo que desde sempre me deram para alcançar os meus objetivos, proporcionando-me as melhores condições de vida possíveis face todas as adversidades e nunca deixaram que me faltasse nada. Agradeço o apoio incondicional que me deram mesmo nos momentos mais difíceis. A vocês devo quem sou hoje.

À minha namorada, Patrícia, que desde o primeiro momento me marcou com o seu carinho, a sua maneira de ser e pensar. Agradeço por todo o amor, compreensão e sobretudo paciência, que me ajudou a superar as dificuldades encontradas e me deu muita motivação durante o desenvolvimento desta dissertação. És uma das melhores pessoas que o mundo tem a sorte de ter presente.

Aos meus amigos, tenho a agradecer pelo apoio, força e incentivo na conclusão deste percurso. Ao Luís e à Jéssica, obrigado pelo forte incentivo nas horas mais complicadas, por todas as horas de estudo e partilha de conhecimento. Com vocês o mestrado foi incrível.

Agradeço também à minha orientadora, Professora Doutora Alexandra Patrícia Horta Ramos, por toda a sua disponibilidade e transmissão de experiências para a realização deste estudo.

Um agradecimento muito especial ao meu avô paterno, António Cunha, que foi claramente uma das pessoas mais bondosas que passou na terra, uma das que mais me inspirou e ensinou na vida. Esta era a homenagem que me faltava fazer-lhe e, assim, por escrito como gostava digo que nos faz imensa falta.

Resumo

Prever é o ponto de partida para o processo de organização da atividade futura, tomada de decisão e controlo em qualquer organização. A previsão em séries temporais tem sido influenciada principalmente pela econometria e estatística. Em tempos mais recentes, tem-se verificado que a aprendizagem automática, *machine learning*, é uma área que se dedica ao desenvolvimento de algoritmos que aprendem a partir dos dados, mostrando-se muito ativa em modelação e predição.

A análise financeira das empresas é um tópico que se tem tornado cada vez mais importante ao longo dos anos, pelo que conseguir prever o melhor possível um dos dados financeiros mais relevantes como é o Volume de Negócios é algo que as empresas têm em consideração.

Deste modo, o âmbito desta dissertação é analisar o valor do volume de negócios de uma empresa do setor da consultoria empresarial com o objetivo de averiguar qual dos métodos de previsão presentes na literatura, Método ARIMA, Métodos de alisamento exponencial e o Método de Holt-Winters Aditivo, apresenta melhor performance. As previsões destes métodos foram realizadas no software *RStudio*.

Palavras-chave: Métodos de previsão; Volume Negócios; Consultoria Empresarial; ARIMA; ETS; Holt-Winters; RStudio.

Abstract

Forecasting is the starting point for the process of organizing future activity, decision making and control in any organization. Time series forecasting has been influenced mainly by econometrics and statistics. More recently, machine learning has been an area dedicated to the development of algorithms that learn from data, showing itself to be very active in modeling and forecasting.

The financial analysis of companies is a topic that has become increasingly important over the years, so being able to forecast as best as possible one of the most relevant financial data such as Turnover is something that companies take into consideration.

Therefore, the aim of this dissertation is to analyze the value of the turnover of a company in the business consulting sector with the objective of determining which of the forecasting methods present in the literature, ARIMA Method, Exponential Smoothing Methods and Holt-Winters Additive Method, presents the best performance. The forecasts of these methods were performed in the RStudio software.

Keywords: Forecasting methods; Turnover; Business Consulting; ARIMA; ETS; Holt-Winters; RStudio.

Índice

Agradecimentos.....	i
Resumo	iii
Abstract	v
Índice de Figuras.....	x
Acrónimos	xii
Tabela de Símbolos Matemáticos	xiii
1. Introdução	1
1.1 Enquadramento	1
1.2 Objetivos do trabalho	2
1.3 Estrutura do Documento	2
2. Revisão de Literatura.....	3
2.1 Indústrias Criativas.....	3
2.2 Volume de Negócios	3
2.3 Consultoria Empresarial.....	4
2.3.1 Fatores de (In)Sucesso.....	5
2.3.2 Qualidade do Serviço.....	6
2.4 Métodos de Previsão e Análise de Séries Temporais.....	8
2.4.1 Métodos de Previsão	8
2.4.2 Séries Temporais	9
3. Metodologia	12
3.1 Processos Estocásticos.....	12
3.2 Função Autocorrelação	13
3.3 Função Autocorrelação Parcial	14
3.4 Processos Autorregressivos	14
3.5 Média Móvel.....	16

3.6 Processo Autorregressivo e de Média Móvel - ARMA (p, q)	18
3.7 Processo Autorregressivo Integrado de Média Móvel	20
3.7.1 SARIMA	21
3.7.2 Adequação do Modelo	22
3.8 Métodos de Alisamento Exponencial	23
3.8.1 Métodos de Alisamento Exponencial Simples	23
3.8.2 Métodos de Holt	24
3.8.3 Métodos de Holt-Winters	25
3.9 Erros de Previsão	26
3.10. Critério de Seleção do Modelo	27
4. Caso de Estudo	29
4.1 Descrição da Amostra	29
4.2 Modelação e Análise Quantitativa	29
4.3 Recolha e Preparação dos Dados	30
4.4 Tratamento dos Dados	30
4.4.1 Avaliação da Estacionariedade	30
5. Análise dos Resultados	32
5.1 Análise Descritiva da Série	32
5.2 Estimação dos Modelos	36
5.2.1 ARIMA	36
5.2.2 Holt -Winters Aditivo	37
5.2.3 Alisamento Exponencial Simples	39
5.3 Exemplo da aplicação dos métodos	40
5.4 Comparação dos erros dos modelos estimados	44
6. Conclusões	46
7. Referências Bibliográficas	48

8. Anexos	52
-----------------	----

Índice de Figuras

Figura 1 - Representa uma série com tendência e sazonalidade (Chatfield, 2000).....	11
Figura 2 - Etapas da abordagem iterativa da construção de modelos (Murteira et al., 1993)	28
Figura 3 - Série Temporal "Volume Negócios".....	32
Figura 4 - Decomposição da série temporal "Volume Negócios"	33
Figura 5 - Gráfico ACF da série original "Volume Negócios"	34
Figura 6 - Gráfico PACF da série original "Volume Negócios"	34
Figura 7 - Teste ADF para determinar estacionariedade da série "Volume Negócios"	35
Figura 8 - Teste KPSS para determinar estacionariedade da série "Volume Negócios"	35
Figura 9 - Estimação do modelo ARIMA.....	36
Figura 10 - Gráfico, ACF e PACF dos resíduos e Q-Q Plot com as observações dos resíduos	36
Figura 11 - Teste à estacionariedade dos resíduos do modelo SARIMA (0,1,1)(0,0,1)[12] ..	37
Figura 12 - Estimação do Modelo Holt-Winters Aditivo	38
Figura 13 - Gráfico, ACF e PACF dos resíduos e Q-Q Plot com as observações dos resíduos	38
Figura 14 - Teste à estacionariedade dos resíduos do modelo Holt-Winters Aditivo	38
Figura 15 - Estimação do Modelo Alisamento Exponencial Simples.....	39
Figura 16 - Análise dos Resíduos	40
Figura 17 - Teste à estacionariedade dos resíduos do modelo Alisamento Exponencial Simples	40
Figura 18 - Previsão da Série "Volume Negócios" para o último trimestre de 2021 – Modelo SARIMA (4,1,2)(0,0,2)12	41
Figura 19 - Previsão da Série "Volume Negócios" para o último trimestre de 2021 – Modelo Holt-Winters Aditivo	42
Figura 20 - Previsão da Série "Volume Negócios" para o último trimestre de 2021 – Modelo Alisamento Exponencial Simples	42
Figura 21 - Previsão da Série "Volume Negócios" para o primeiro trimestre 2022 – Modelo SARIMA (0,1,1)(0,0,1)[12]	43
Figura 22 - Previsão da Série "Volume Negócios" para o primeiro trimestre 2022 – Modelo Holt-Winters Aditivo	43
Figura 23 - Previsão da Série "Volume Negócios" para o primeiro trimestre 2022 – Modelo	

Alisamento Exponencial Simples	44
Figura 24 - Erros MAE da série “Volume Negócios”	44
Figura 25 - Erros MAPE da série “Volume Negócios”	45

Acrónimos

ACF	Autocorrelation Function
ADF	Augmented Dickley-Fuller
AR	Autoregressive model
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive–moving-average model
ETS	Exponential Smoothing
HW	Holt-Winters
IID	Independentes e identicamente distribuídas (variáveis)
KPSS	Teste Kwiatkowski-Phillips-Schmidt-Shin
MA	Moving-average model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error

Tabela de Símbolos Matemáticos

Notação Matemática	Descrição
ε_t	Erro associado à previsão no instante t
S_t	Componente Sazonal no instante t
T_t	Componente tendência-ciclo no instante t
Y_t	Valor observado no instante t
Y_{t+k}	Valor observado no instante t+k
$\bar{Y}$	Valor médio da série temporal
$\hat{Y}_t$	Previsão da série temporal para o instante t
$\hat{Y}_{t+k}$	Previsão feita no instante t para K períodos em diante

1. Introdução

1.1 Enquadramento

Nos dias de hoje, a previsão ao nível do volume de negócios, ao nível dos stocks, da meteorologia, entre muitas outras coisas, desempenha um papel primordial tanto na vida das empresas, como na vida da população. Ao satisfazer essas necessidades as empresas conseguem regular melhor a sua atividade.

Não obstante, os fenómenos temporais acarretam o fator incerteza e, por isso, o conhecimento exato dos valores no futuro é impossível de se apurar, o que nos leva a fazer previsões. Deste modo, planejar a evolução futura é uma necessidade atual em qualquer setor de atividade. Este planeamento faculta previsões de acontecimentos capazes de afetar as decisões tomadas pelos gestores das empresas. As previsões de vendas são fulcrais para qualquer plano de negócio, pelo que os gestores de vendas têm de elaborar regularmente planos previsionais de vendas dos produtos ou serviços das empresas. Efetivamente, a experiência dos gestores, a consulta de opiniões qualificadas e a utilização de procedimentos estatísticos apropriados são alguns dos procedimentos que evitam desperdícios de recursos e potenciais custos de oportunidade.

O estudo do volume de negócios tem um valor importante para as organizações, na medida em que este é um indicador que impactua na explicação do desempenho financeiro das empresas ao longo dos períodos. O Volume de Negócios é um fator determinante na tomada de decisões, pelo que a sua análise, modelação e previsão é de extrema importância (Books, 2021). Trabalhos sobre a análise de séries temporais têm sido amplamente considerados na literatura, incluindo o estudo de dados financeiros. O foco desta dissertação incide numa empresa de consultoria empresarial específica. A área da consultoria empresarial tem vindo a evoluir ao longo dos últimos anos e é de grande importância, pelo que captou o meu interesse.

1.2 Objetivos do trabalho

A presente dissertação terá por base o estudo da evolução do volume de negócios de uma empresa em particular, do setor da consultoria empresarial. O âmbito deste estudo baseia-se na análise de séries temporais, com o intuito de determinar o(s) modelo(s) que permitem obter a melhor previsão possível para o volume de negócios da empresa em estudo.

Pretende-se demonstrar que a realização de previsões utilizando os modelos adequados pode ser um indicador que as empresas podem utilizar com confiança e de forma consciente com o objetivo de prever o seu futuro económico. Hoje em dia, face ao contexto atual mundial, apesar de se preverem valores de vendas, stocks, custos, entre outros, há sempre fatores que não se podem controlar, como serve de exemplo a pandemia COVID-19, a guerra sucedida na Ucrânia, o aumento do preço dos combustíveis, etc. Estas previsões são indicadores que ajudam na tomada de decisões e não deixam de ser relevantes para as organizações. Os objetivos desta dissertação levantaram várias questões de investigação que se pretendem ver respondidas no final deste trabalho, designadamente: Como se pode prever o volume de negócios para os próximos períodos? Será que a previsão obtida vai de encontro com a realidade? Que modelo é o mais adequado para descrever o volume de negócios e fornece a melhor previsão, de entre alguns modelos considerados nesta dissertação?

1.3 Estrutura do Documento

A presente dissertação encontra-se estruturada do seguinte modo: no primeiro capítulo faz-se a introdução que define a questão de investigação e objetivos a alcançar. De seguida, no segundo capítulo, é realizada a revisão da literatura e apresentados os principais objetivos da previsão e sua importância numa organização empresarial. No capítulo da metodologia serão abordados alguns dos métodos de previsão presentes na literatura e as suas diferentes etapas de previsão. De seguida, no quarto capítulo, apresenta-se o tratamento e caracterização do conjunto de dados abordado neste trabalho. O quinto capítulo corresponde à análise dos resultados alcançados e as respetivas conclusões no sentido de se averiguar e descobrir os melhores métodos de previsão. Finalmente, no último capítulo, encontram-se as respetivas conclusões e respostas às questões de investigação.

2. Revisão de Literatura

2.1 Indústrias Criativas

Segundo Ratalewska (2018), indústrias criativas são indústrias que têm a sua origem na criatividade, na habilidade, nos talentos individuais, potenciando a riqueza e a criação de emprego. O desenvolvimento da informação e da tecnologia impõe que as indústrias desenvolvam ideias e produtos inovadores. Segundo Yum (2016), empresas especializadas em lançar novos produtos sustentados em ideias criativas e na leitura de tendências contemporâneas sobrepõem-se aos seus concorrentes. Efetivamente, tratam-se de indústrias cujo capital humano é crucial para o sucesso, dado que é onde surgem novas ideias e abordagens (Piergiovanni et al., 2012). As indústrias criativas incluem atividades como: publicidade, arquitetura, computadores e videojogos, design, software, rádio e televisão, entre outros (Ratalewska, 2018).

No âmbito desta dissertação, a empresa sujeita a análise, que incorpora a área em estudo, trabalha com softwares sujeitos a desenvolvimentos, isto é, a soluções inovadoras solicitadas pelos clientes e/ou apresentadas pela própria empresa. Para além disso, é uma empresa em crescimento, que procura novos talentos, cujo número de colaboradores ao longo dos anos tem vindo a aumentar, enquadrando-se no setor das indústrias criativas.

2.2 Volume de Negócios

De acordo com a Ordem dos Contabilistas Certificados, o conceito de volume de negócios e de rédito coincidem, visto que são os rendimentos diretamente relacionados com a atividade corrente de uma empresa (Volume de negócios, 2020). Este conceito é importante, dado que revela a produtividade das vendas de um negócio. A informação sobre as receitas operacionais é especialmente valiosa quando segue uma linha de tendência, uma vez que pode revelar picos ou declínios na atividade, que poderiam indicar uma tendência a longo prazo. Books (2021) argumenta que os investidores se focam neste número quando decidem se devem investir num negócio.

O conceito pode ser aperfeiçoado para situações em que as vendas de uma empresa são em

grande parte constituídas por vendas relacionadas com um único contrato ou cliente. Se esta informação puder ser decomposta de forma a separar as receitas de fonte única e todas as outras receitas, pode-se verificar de qual das fontes a empresa está mais dependente (Books, 2021).

O volume de negócios é um dos fenómenos mais importantes para os estudiosos e profissionais da gestão, sendo o principal indicador de cumprimento do objetivo máximo das organizações: obter lucro (Field et al., 2021). Deste modo, um valor de vendas ou prestação de serviços elevado é a principal referência de rendimento de uma organização, pois é a partir desse valor que se começam a deduzir os custos para se detetar um resultado líquido positivo ou negativo. Por exemplo, para solicitar um empréstimo bancário, uma entidade é obrigada a facultar vários dados à instituição financeira. Sendo que, os indicadores de natureza quantitativa mais referenciados são a capacidade de endividamento da empresa e a autonomia financeira. Contudo, os indicadores quantitativos como o volume de negócios, as hipotecas efetuadas, as despesas em recursos humanos, a solvabilidade da empresa e cash-flow gerado, são indicadores que podem ajudar os bancos a tomar uma decisão (Costa, 2003).

A análise financeira é amplamente utilizada nas empresas para determinar os resultados financeiros e o desempenho. Para além disso, desempenha um papel importante ao medir o desempenho empresarial e a gestão financeira das entidades (Tissen et al., 2018).

2.3 Consultoria Empresarial

Devido às grandes mudanças que se fizeram sentir no século passado e ao crescimento económico, Taylor publicou “The principles of scientific management”, com o intuito de colmatar as lacunas de gestão relacionadas com os sistemas industriais. Assim, Taylor debruçou-se sobre a organização do trabalho e da gestão empresarial, concebendo as bases para a ciência de gestão de sistemas de produção industrial. O Taylorismo protagonizou um novo paradigma na gestão empresarial acompanhada da revolução intelectual (Grachev & Rakitsky, 2013). Li and Liu (2005) defendem que a origem da consultoria tem uma profunda relação com o pensamento da gestão científica de Taylor e o seu surgimento é considerado como a amálgama da engenharia, do direito e da contabilidade.

Neto (2004) afirma que consultoria é “o ato de um cliente proporcionar, solicitar, pedir pareceres, opinião, estudos a um especialista contratado para que este auxilie, apoie, oriente o trabalho administrativo”. A consultoria é importante para as empresas porque o seu foco é proporcionar um bom serviço ao cliente. Assim sendo, os consultores encontram caminhos que os gestores e as pessoas envolvidas nas empresas não conseguem.

A consultoria empresarial caracteriza-se pela diversidade em relação ao conteúdo dos serviços prestados a empresas de várias dimensões e em diferentes fases de desenvolvimento. Serve para resolver problemas na esfera de gestão, identificar e explorar novas oportunidades, aprender e implementar mudanças (Ratalewska, 2018). O papel da consultoria tem sido decisivo, afetando os processos empresariais progressivos da organização na era digital (Valebnikova et al., 2019).

A consultoria empresarial pode ser definida segundo dois critérios: a consultoria funcional e institucional. No que concerne ao primeiro critério, a consultoria empresarial define-se por funcional quando é prestado um serviço a um cliente de maior ou menor dimensão e quando existe um aconselhamento em assuntos específicos de cada empresa (Ratalewska, 2018). Segundo o critério institucional, as empresas de consultoria obtêm a sua vantagem competitiva principalmente por terem a capacidade de criar e manter recursos de conhecimento e capital baseado na legitimidade, reputação ou relações com clientes (Reihlen & Nikolova, 2010).

2.3.1 Fatores de (In)Sucesso

De acordo com Veres & Varga-Toldi (2020), os serviços de consultoria são serviços profissionais caracterizados por elevado risco bilateral, relação de confiança, complexidade e assimetria de informação dos participantes. Porém, as expectativas dos gestores, os objetivos dos projetos e os resultados não são por vezes transparentes, pelo contrário, são mal comunicados e não há uma separação de responsabilidade entre os profissionais da empresa de consultoria e os colaboradores da organização. Consequentemente, sem expectativas claras, um compromisso de consultoria não pode ser considerado bem-sucedido e a sua qualidade dificilmente pode ser avaliada. Pretende-se que os consultores tenham valores como a integridade, a honestidade, a comunicação cliente-consultor e a credibilidade.

Uma falha numa destas áreas causaria um resultado e percepção negativa do projeto de consultoria, o que posteriormente afetaria a avaliação pós-compra. Atualmente, o sucesso dos serviços de consultoria ainda é avaliado à luz da relação cliente-consultor e dos resultados do projeto, em vez de se basear na realização de benefícios e indicadores de desempenho. Isto pode dever-se ao facto de ainda não compreendermos claramente quais são os atributos e as dimensões de qualidade percebidos de forma definitiva dos serviços de consultoria empresarial na perspetiva dos clientes. O âmbito do projeto é muitas vezes indefinido, consistindo em esboços de orientações e ideias iniciais em vez de especificações concretas e a medição da qualidade dos projetos de consultoria empresarial continua a ser uma questão espinhosa (Veres & Varga-Toldi, 2020).

2.3.2 Qualidade do Serviço

Desde o início dos anos 80, foram desenvolvidos numerosos modelos para a medição da qualidade de um serviço prestado. As quatro principais abordagens consistiram no Modelo Nórdico ou o seu refinamento, o SERVQUAL, o modelo Multilevel e, finalmente, o modelo Hierárquico (Buttle, 1996).

Veres & Varga-Toldi (2020) referem que ainda não existe um modelo de qualidade de serviços de consultoria geralmente aceite. Este é particularmente o caso dos projetos de consultoria empresarial, onde o âmbito é muitas vezes indefinido. Apesar disso, os estudos sobre a qualidade dos serviços de consultoria aplicam principalmente o modelo SERVQUAL, embora seja amplamente criticado por se concentrar apenas na qualidade funcional e não na medição da qualidade técnica e/ou qualidade dos resultados. O modelo SERVQUAL, formulado em 1985, fornece uma tecnologia para medir e gerir a qualidade do serviço e baseia-se na opinião de que a avaliação do cliente sobre a qualidade do serviço é primordial (Buttle, 1996).

Sendo as tarefas de consultoria baseadas em projetos, as medidas de sucesso e os critérios de avaliação raramente são estabelecidos com antecedência. Na maior parte das vezes, o âmbito de um projeto é móvel e a definição dos requisitos na fase de conceção é quase impossível. Em vez disso, os clientes constroem os seus requisitos e expectativas de qualidade durante o processo de consultoria e veem a qualidade como um valor percebido que deve ser

construído ao longo da colaboração e não como um fenómeno que pode ser descrito de antemão. Grande parte dos projetos de consultoria são excessivamente complexos e existe um grande nível de incerteza. Quanto maior é a incerteza na avaliação da qualidade, melhores são as impressões pessoais da interação cliente-consultor. Os membros da direção baseiam a sua avaliação de qualidade em incentivos intangíveis, tais como memórias e impressões (Veres & Varga-Toldi, 2020).

A avaliação da qualidade dos serviços de consultoria é bidimensional, ou seja, por um lado há uma avaliação consciente e transacional dos resultados ("o que recebemos") executada durante a missão e logo após o encerramento do projeto. Por outro lado, há uma avaliação retrospectiva, informal pós-transacional ("como temos beneficiado"), que não tem um âmbito temporal definido e que é considerada subjetiva (Veres & Varga-Toldi, 2020).

No início do projeto, a organização compradora presta mais atenção ao controlo da cessão do contrato e os clientes estão muito conscientes e rigorosos sobre a forma, a qualidade e o prazo dos resultados do projeto, para cumprir o que foi contratado. À medida que o projeto avança, os clientes prestam menos atenção aos requisitos formais e qualidade das prestações contratadas. Por outro lado, concentram-se na qualidade da relação cliente-consultor tentando perceber se o seu problema está a ser de alguma forma resolvido. A terceira fase é pós-transacional, à medida que os clientes olham para o nível do valor ganho e dos benefícios recebidos, dificilmente são capazes de distinguir o valor do serviço que lhes é prestado e o valor do consultor (Veres & Varga-Toldi, 2020).

2.4 Métodos de Previsão e Análise de Séries Temporais

2.4.1 Métodos de Previsão

Os métodos de previsão são baseados na hipótese de que a evolução futura dos fenómenos analisados mantém características essenciais, a menos que ocorram situações especiais que não tenham sido verificadas no passado (Pater & Cociu, 2014). Por este motivo, os métodos de previsão podem ser amplamente classificados em dois tipos principais: autónomos e híbridos. Os primeiros aplicam uma única técnica de análise de tendências. Os segundos integram mais do que uma técnica autónoma (Debnath & Mourshed, 2018).

Segundo Debnath & Mourshed (2018), na maioria dos casos, o objetivo da hibridização é racionalizar ou fazer previsões fiáveis e produzir uma maior precisão da projeção. O ponto de partida da previsão híbrida é o de que diferentes modelos de previsão são combinados para se obter melhores resultados de previsão do que qualquer outro método individual (Liu et al., 2016).

De acordo com Caniato et al. (2011), os métodos de previsão podem ser classificados em duas famílias: abordagens quantitativas (ou seja, estatísticas) e qualitativas (ou seja, julgamentos). Em termos gerais, as abordagens quantitativas são geralmente consideradas mais precisas quando uma quantidade adequada de dados históricos está disponível e os fenómenos sob investigação são estáveis. Dentro desta abordagem, existem os modelos causais e não causais.

As abordagens qualitativas são preferidas quando não há dados históricos disponíveis, ou quando os fenómenos sob investigação estão a mudar rapidamente, tornando o passado incapaz de explicar o futuro, ou, ainda, no caso de eventos excecionais pontuais (Caniato et al., 2011). De qualquer forma, estes métodos podem não gerar previsões precisas se não for tomado o cuidado suficiente na identificação de informações relevantes e no seu processamento sistemático (Naik, 2004).

Os métodos de previsão compreendem a análise estatística de uma série temporal, a análise de regressão e correlação. A análise da série cronológica é materializada de acordo com algumas funções matemáticas, que são selecionadas de forma a descrever melhor o

desenvolvimento dos fenômenos. Exemplificando uma série temporal: existe uma sequência de valores observados por uma variável Y , por exemplo, o volume de vendas de uma empresa, em diferentes momentos de tempo como $n, n+1, n+2$, etc, ou seja, uma série cronológica associada à variável Y . Para a variável Y , as vendas serão representadas em euros/t.p.u., onde t.p.u. representa a unidade de período de tempo considerada, seja ela o ano, trimestre, mês, semana, dia ou outro (Pater & Cociu, 2014).

2.4.2 Séries Temporais

Uma série cronológica ou temporal Y_t é uma coleção de observações realizadas ao longo do tempo t , ou seja, é um conjunto de observações quantitativas, em diferentes momentos no tempo, sobre uma determinada variável. Representa-se por Y_t o valor de Y no instante t . Quando as observações sucessivas são dependentes, os valores futuros podem ser previstos a partir de observações passadas, ou seja, quando uma série cronológica puder ser prevista com exatidão, diz-se que é determinista. Contudo, a maioria das séries cronológicas são estocásticas na medida em que o futuro é apenas parcialmente determinado por valores passados, de modo que as previsões exatas são impossíveis e devem ser substituídas pela ideia de que os valores futuros têm distribuição de probabilidade, que é condicionada por um conhecimento dos valores passados (Jenkins et al., 2015).

Um método tradicional para representar uma série temporal com sazonalidade tem sido decompor a série em componentes de tendência, sazonalidade, efeito cíclico e de ruído (componente aleatória) (Jenkins et al., 2015).

A tendência (T_t) tem um significado intuitivo e pode descrever-se como o sentido do deslocamento ou o comportamento da série, ou seja, é a inclinação crescente ou decrescente dos valores da série temporal ao longo do tempo, ou, ainda, mudança de nível. Reflete globalmente o sentido do crescimento (ou decrescimento) da série. Ainda que os movimentos possam mudar ao longo do tempo, assume-se que a tendência é suave para proporcionar decomposições robustas. Admite-se que a tendência para ser modelada varia em função do tempo t , ou seja, $T_t = f(t)$. Numa série com tendência, a média $E[Y_t] = T_t$ não é constante, ou seja, a série não é estacionária (Jenkins et al., 2015).

A sazonalidade (S_t), retrata padrões semelhantes ou, por vezes, iguais, dentro de um ciclo

repetitivo e descreve as variações em relação à tendência. São flutuações periódicas da variável com periodicidade fixa de por exemplo um mês, um trimestre ou até um ano (Jenkins et al., 2015).

O efeito cíclico (C_t), reflete movimentos da série que afetam a sua tendência global a médio prazo. São flutuações cíclicas com periodicidade não fixa e períodos mais longos do que um ano. Atendendo a que a componente cíclica tem carácter irregular (de médio a longo prazo), os seus valores futuros dificilmente poderão ser previstos por métodos estatísticos, recorrendo-se muitas vezes a procedimentos subjetivos. Assim, para séries que não sejam longas despreza-se este efeito (Jenkins et al., 2015).

Por último, a componente aleatória (efeito ruído) (ε_t), apresenta variações de carácter imprevisível, ou seja, descreve mudanças aleatórias e irregulares que não podem ser explicadas pela tendência e pela sazonalidade. Portanto, são resíduos de séries temporais após a tendência e a sazonalidade serem removidos. Também se pode denominar de ruído branco, que matematicamente significa ter valor esperado constante ($E[\varepsilon_t] = \mu$), que é usualmente zero, variância constante ($Var[\varepsilon_t] = \sigma^2$) e covariância zero ($Cov[\varepsilon_t, \varepsilon_s] = 0$), ou seja, é um processo estacionário (Jenkins et al., 2015).

Se para o instante t , Y_t representar o valor observado na série, T_t a tendência, S_t a componente sazonal, C_t a componente cíclica e ε_t a componente aleatória, então admite-se que o valor observado em cada período é uma função das quatro componentes: $Y_t = f(T_t, S_t, C_t, \varepsilon_t)$ (Jenkins et al., 2015).

Jara (2011) evidencia que as séries temporais têm propriedades como a estacionariedade, a sazonalidade, ou a linearidade, que determinam o seu comportamento subjacente. Subsiste uma classe de séries intituladas de processos de longa duração, que têm uma dependência temporal entre observações distantes. Isto quer dizer que os valores da série não dependem só de valores passados recentes, mas, também, de observações de períodos de tempo longínquos. A estacionariedade implica que os dados amostrados de diferentes janelas temporais se comportem de forma semelhante, independentemente de onde se encontrem na linha temporal (Jenkins et al., 2015).

As análises de séries temporais são uma ferramenta útil para as empresas, pois podem ajudar a prever valores futuros, dependendo da série que está a ser analisada, por exemplo, no caso desta dissertação, o volume de negócios de uma organização. Para as empresas, essa previsão é sempre importante porque permite ajudar nos processos, nos ideais e nas ações a concretizar. (Pater & Cociu, 2014).

Há representações gráficas que ajudam no estudo das séries como é o caso do cronograma, pois permite recolher elementos sobre o comportamento da série, no que respeita a pontos de viragem, pontos muito altos ou muito baixos, oscilações ou ciclos e movimentos de natureza sazonal (Chatfield, 2000).

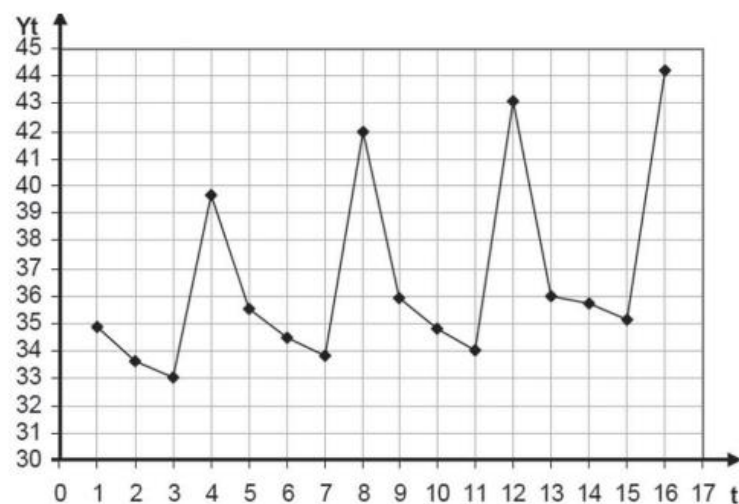


Figura 1 - Representa uma série com tendência e sazonalidade (Chatfield, 2000)

3. Metodologia

No âmbito desta dissertação, propõe-se responder à questão: Que modelo é o mais adequado para descrever o volume de negócios e fornece a melhor previsão, de entre alguns modelos considerados nesta dissertação? Neste capítulo serão apresentadas as diferentes metodologias consideradas para a posterior descrição e análise dos dados estudados nesta dissertação e fornecidos no capítulo 4. Nomeadamente, a metodologia subjacente ao Método ARIMA (casos particulares: AR, MA, ARMA e extensão SARIMA), ao Método de Alisamento Exponencial e ao Método de Holt-Winters Aditivo.

Para além disso, recorreu-se ao software R para estimar cada um dos modelos. O R é uma linguagem e ambiente para análise estatística e gráfica, fornece uma grande variedade de métodos estatísticos (análise de séries cronológicas, modelação linear e não linear, testes estatísticos clássicos, etc.), técnicas gráficas e é altamente extensível (Cryer & Chan, 2008).

3.1 Processos Estocásticos

Pode definir-se um processo estocástico como um conjunto de variáveis aleatórias indexadas a uma variável (geralmente a variável tempo), sendo representado por $X_t: t \in T$, onde T é um conjunto de índices. Uma série temporal ao ser analisada pode ser pensada como uma realização particular, produzida pelo mecanismo probabilístico subjacente, do sistema em estudo. Por outras palavras, ao analisar-se uma série cronológica, pode-se considerá-la como a realização de um processo estocástico (Jenkins et al., 2015).

Estabelecendo o paralelismo com o caso determinístico, onde uma função $f(t)$ toma valores bem definidos ao longo do tempo, um processo estocástico toma valores aleatórios durante o período considerado (Alves & Delgado, 1997).

Uma classe importante dos processos ou modelos estocásticos para descrever séries temporais compreende o que se chama de processos estacionários. Estes assumem que o processo permanece em equilíbrio estatístico com propriedades probabilísticas que não se alteram ao longo do tempo, em particular variando sobre um nível médio fixo constante e com variação constante (Jenkins et al., 2015).

Um processo estocástico é estritamente estacionário se as suas propriedades não forem afetadas por uma mudança de origem temporal, ou seja, se a distribuição conjunta de probabilidade de $\{X_{t_1}, \dots, X_{t_n}\}$ para qualquer conjunto de tempos $\{t_1, \dots, t_n\}$ é a mesma que a associada a n observações $\{X_{t_1+k}, \dots, X_{t_n+k}\}$ em qualquer conjunto de tempos $\{t_1 + k, \dots, t_n + k\}$, qualquer que seja $k \in \mathbb{N}$ (Jenkins et al., 2015).

Estacionariedade estrita significa intuitivamente que os gráficos ao longo de dois intervalos de tempo de igual duração de uma realização da série temporal devem exibir características estatísticas semelhantes (Brockwell & Davis, 1991). Relativamente à estacionariedade fraca, para que um processo seja estritamente estacionário, toda a estrutura de probabilidade deve depender apenas de diferenças temporais. Um requisito menos restritivo, chamado estacionariedade fraca. Ou seja, um processo $\{z_t\}$ é (fracamente) estacionário (de ordem 2), ou estacionário de segunda ordem, se a média $E[z_t] = \mu$ e a variância são constantes fixas para todo o t e as autocovariâncias $cov[z_t, z_{t+k}] = \gamma_k$ dependem apenas do desfasamento k para todo o t . O segundo momento finito, ou seja, $E[z_t^2] < \infty$ e as suas autocovariâncias dependem apenas do intervalo de tempo (lag) k entre as observações (Jenkins et al., 2015).

3.2 Função Autocorrelação

A função autocorrelação, que normalmente se denota como ACF (Autocorrelation Function) serve para determinar a correlação que uma série temporal apresenta com ela própria, desfasada em k períodos e é utilizada como indicador estatístico das características dos dados. É uma ferramenta de diagnóstico muito importante para examinar a dependência (Jenkins et al., 2015).

Assumindo que $X_1, X_2, \dots, X_n$ é uma série observada de um processo estacionário, então

- $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ é um estimador para a média $\mu = \mu_t = E[X_t]$, pois esta é constante.
- Como a autocovariância é constante entre observações com o mesmo desfasamento, então a autocovariância amostral de ordem k , é dada por:

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (1.1)$$

e é um estimador para $\gamma(k) = cov(X_t, X_{t-k})$.

- Da mesma forma, a autocorrelação amostral de ordem k, é dada por:

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \quad (1.2)$$

e é um estimador para $\rho(k) = corr(X_t, X_{t-k})$.

$\hat{\rho}(k)$ refere-se à Função Autocorrelação (ACF).

Caso o processo não seja estacionário, então $\hat{\gamma}(k)$ e $\hat{\rho}(k)$ não são estimadores para $\gamma(t+k, t)$ e $\rho(t+k, t)$ (Jenkins et al., 2015).

3.3 Função Autocorrelação Parcial

Uma vez que para certos modelos de previsão a função de autocorrelação é zero para defasamentos além de q, a autocorrelação da amostra é um bom indicador da ordem do processo. Contudo, a autocorrelação de outro tipo de modelos não se torna zero após um certo número de defasamentos. Portanto, é necessária uma função diferente para ajudar a determinar a ordem dos modelos autoregressivos. Tal função pode ser definida como a correlação entre X_t e X_{t-k} após a eliminação do efeito das variáveis intermédias $X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-k+1}$. Este coeficiente é chamado de autocorrelação parcial (FACP) e será denotado por ϕ_{kk} . Se $\{X_t\}$ é uma série temporal normalmente distribuída, então:

$$\phi_{kk} = corr(X_t, X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}) \quad (1.3)$$

3.4 Processos Autorregressivos

Um processo estacionário $\{X_t, t = 1, 2, 3, \dots\}$ é autoregressivo de ordem p se:

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (1.4)$$

sendo $\mu, \phi_1, \dots, \phi_p$, constantes e $\{\varepsilon_t\}$ um erro aleatório (ruído branco) com média zero e variância σ^2 (variável aleatória independente e identicamente distribuídas (IID)).

$$\varepsilon_t \sim IID(0, \sigma_\varepsilon^2) \quad (1.5)$$

Ou equivalentemente,

$$\phi(B)X_t = \varepsilon_t \quad (1.6)$$

Em que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinómio autoregressivo de ordem p . Sendo B o operador de atraso tal que $X_{t-k} = B^k X_t$, para $k=1, 2, \dots, p$. Geralmente é denotado por AR (p) e não se aplica a séries não estacionárias. A sua FAC tende para zero de forma exponencial ou sinusoidal amortecida (Jenkins et al., 2015).

Na prática, normalmente não se conhece inicialmente a ordem do processo autoregressivo e a ordem tem de ser especificada a partir dos dados. O problema é análogo a decidir sobre o número de variáveis independentes a serem incluídas numa regressão múltipla. A FACP é uma ferramenta que explora o facto de um processo AR(p) ter uma função de autocorrelação que é infinita em extensão, as autocorrelações parciais são zero para além do desfasamento p . Estas podem ser descritas em termos de p funções não nulas das autocorrelações. Num modelo AR (p), o termo ϕ_p (coeficiente de X_{t-p} no modelo AR (p)) mede o excesso da correlação que não é tido em conta num modelo AR ($p-1$). A visualização de um gráfico associado à FACP mostra os valores de ϕ_p obtidos dos modelos AR (p), para $p = 1, 2, 3, \dots$. Para além disso, ajuda na determinação da ordem p de um processo autorregressivo. Por exemplo, se um gráfico da FACP evidenciar valores de ϕ_p significativamente diferentes de zero, apenas para $p=1$, $p=2$ e $p=3$, é provável que se trate de um processo AR (3) (Jenkins et al., 2015).

Exemplo:

Modelo AR (1): $X_t = \phi X_{t-1} + \varepsilon_t$, com $|\phi| < 1$

Note-se que,

$$X_t = \phi X_{t-1} + \varepsilon_t$$

$$X_t = \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$X_t = \phi^2 X_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$$

$$X_t = \phi^2(\phi X_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t$$

$$X_t = \phi^3 X_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$$

$$X_t = \dots$$

$$X_t = \phi^k X_{t-k} + \sum_{i=0}^{k-1} \phi^i \varepsilon_{t-i}$$

Tomando o limite $k \rightarrow \infty$:

$$X_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \quad (1.7)$$

Esta equação decorre do Teorema da Decomposição de Wold, que será oportunamente explanado no subcapítulo 3.6.

Teorema: Os modelos autorregressivos estão associados a processos estacionários, para isso tem de cumprir certos requisitos. Isto é, um processo autorregressivo AR (p) é estacionário se os valores absolutos correspondentes aos zeros $\phi(B)$ estão fora do círculo unitário. Para o caso particular de AR (1) este requisito verifica-se quando $|\phi| < 1$ (Jenkins et al., 2015).

3.5 Média Móvel

Para uma série estacionária e puramente aleatória podem obter-se previsões considerando apenas a média de todas as observações. Isto iria minimizar o erro quadrático médio das previsões. O modelo é denominado de modelo MA (q) e pode ser interpretado como um modelo onde X_t é uma média ponderada dos erros de previsão passados. Seria mais adequado atribuir às observações mais recentes maior importância, o que permitiria detetar possíveis mudanças do seu comportamento. Desta forma, estaria mais sensível às mudanças mais recentes do comportamento da série. Este é o princípio no qual se enquadra a média móvel

aritmética. O modelo da média móvel é definido como uma combinação dos erros aleatórios (ruído branco):

$$\begin{aligned} X_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \mu_t \\ &= \mu_t + \theta(B) \varepsilon_t \end{aligned} \quad (1.8)$$

sendo $\theta_1, \dots, \theta_q$ constantes.

Os modelos de Médias Móveis de ordem q são sempre estacionários e são invertíveis se os zeros do polinómio $\theta(B)$ estiverem fora do círculo unitário (Jenkins et al., 2015).

Este método apresenta as seguintes propriedades:

- A média deste modelo é:

$$E(X_t) = E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \theta_2 E(\varepsilon_{t-2}) + \dots + \theta_q E(\varepsilon_{t-q}) + \mu_t = \mu_t \quad (1.9)$$

- A variância deste modelo é:

$$\text{Var}(X_t) = \sigma^2 (1 + \theta_1^2 + \dots + \theta_q^2) \quad (2.1)$$

A função de autocorrelação do método da média móvel apresenta-se da seguinte forma:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1, & \text{se } h = 0 \\ \frac{-\theta_h \sum_{j=1}^{q-h} \theta_j \theta_{j+h}}{1 + \theta_1^2 + \dots + \theta_q^2}, & \text{se } h = 1, 2, \dots, q \\ 0, & \text{se } h > q \end{cases} \quad (2.2)$$

A função autocorrelação possui uma quebra após o desfasamento q , ou seja, $\rho(h) = 0, h > q$. Esta propriedade é útil para identificar a ordem do modelo. Relativamente à FACP, tem um decaimento exponencial ou sinusoidal para zero (Jenkins et al., 2015).

A invertibilidade é uma condição independente da condição de estacionariedade. Para ilustrar a ideia de invertibilidade considere-se o seguinte caso especial:

$$\tilde{X}_t = (1 - \theta B) \varepsilon_t \quad (2.3)$$

Expressando ε_t em termos do presente e do passado dos $\tilde{X}_t$'s, este modelo torna-se:

$$\begin{aligned}\varepsilon_t &= (1 - \theta B)^{-1} \tilde{X}_t = \\ &= (1 + \theta B + \theta^2 B^2 + \dots + \theta^k B^k)(1 - \theta^{k+1} B^{k+1})^{-1} \tilde{X}_t\end{aligned}\quad (2.4)$$

ou seja,

$$\tilde{X}_t = -\theta \tilde{X}_{t-1} - \theta^2 \tilde{X}_{t-2} + \dots + \varepsilon_t - \theta^{k+1} \varepsilon_{t-k-1} \quad (2.5)$$

Se $|\theta| < 1$, ao deixar k tender para infinito, obtém-se a série infinita:

$$\tilde{X}_t = -\theta \tilde{X}_{t-1} - \theta^2 \tilde{X}_{t-2} - \dots + \varepsilon_t \quad (2.6)$$

Qualquer que seja o valor de θ , $\tilde{X}_t = (1 - \theta B)\varepsilon_t$ define um processo estacionário perfeitamente adequado. Contudo, se $|\theta| \geq 1$, o desvio actual $\tilde{X}_t$ em (2.5) depende de $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots, \tilde{X}_{t-k}$, com pesos que aumentam à medida que k aumenta. Evita-se esta situação exigindo que $|\theta| < 1$. Assim sendo, diz-se que a série é invertível (Jenkins et al., 2015).

3.6 Processo Autorregressivo e de Média Móvel - ARMA (p, q)

Considerando $\{\varepsilon_t, t = 1, 2, \dots\}$ um ruído branco, ou seja, com média 0 e variância σ^2 , um processo ARMA (p, q) define-se como sendo o modelo do tipo (Hamilton, 1994):

$$X_t = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.7)$$

sendo $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ constantes.

No caso de o processo ter média $\mu = E(X_t)$ diferente de zero, o modelo pode ser aplicado para a série com média ajustada, isto é:

$$\begin{aligned}X_t - \mu &= \phi_1 (X_{t-1} - \mu) + \dots + \phi_p (X_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \Leftrightarrow \\ X_t &= \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}\end{aligned}\quad (2.8)$$

em que $\alpha = (1 - \phi_1 - \dots - \phi_p)\mu$ é o intercepto (nome dado à constante de uma regressão).

Os modelos ARMA podem ser escritos em função dos operadores de atraso (B) (recorde que: $BX_t = X_{t-1}$):

$$\phi(B)X_t = \theta(B)\varepsilon_t \quad (2.9)$$

Em que:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.1)$$

é o polinómio autoregressivo e

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (3.2)$$

é o polinómio de média móvel.

Este método conta com as seguintes propriedades:

- O processo é estacionário se os zeros de $\phi(B)$ estão fora do círculo unitário;
- O processo é invertível se os zeros de $\theta(B)$ estão fora do círculo unitário;
- O modelo ARMA (p, 0) equivale ao modelo AR (p) e o modelo ARMA (0, q) equivale ao modelo MA(q);
- Económico, um modelo ARMA por norma requer menos parâmetros que um modelo AR ou MA. Na maior parte das vezes um modelo ARMA (1,1) é suficiente para modelar uma série estacionária;
- Redundância na parametrização, quando os polinómios $\phi(B)$ e $\theta(B)$ partilham o mesmo fator, o modelo pode ser simplificado;

Por exemplo, o modelo ARMA (2,1) dado por:

$$X_t = \frac{5}{6}X_{t-1} - \frac{1}{6}X_{t-2} + \varepsilon_t - \frac{1}{2}\varepsilon_{t-1}, \text{ pode ser escrito como:}$$

$$\left(1 - \frac{5}{6}B + \frac{1}{6}B^2\right)X_t = \left(1 - \frac{1}{2}B\right)\varepsilon_t, \text{ que por sua vez pode ser escrito como:}$$

$$\left(1 - \frac{1}{2}B\right)\left(1 - \frac{1}{3}B\right)X_t = \left(1 - \frac{1}{2}B\right)\varepsilon_t, \text{ assim sendo, pode-se simplificar o termo em comum,}$$

$$\text{resultando no modelo AR (1), dado por } \left(1 - \frac{1}{3}B\right)X_t = \varepsilon_t.$$

- Quando o modelo for estacionário, um modelo ARMA(p,q) pode ser escrito como MA(∞) – Teorema da Decomposição de Wold:

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad (3.3)$$

com $\psi_0 = 1$.

Os coeficientes são dados por:

$$\psi_i = \theta_i + \sum_{j=1}^i \phi_j \psi_{i-j}, \quad j = 1, 2, \dots \quad (3.4)$$

Em que $\phi_i = 0$, para $i > p$ e $\theta_i = 0$ para $i > q$. Esta representação do modelo denomina-se de **forma causal**.

- Se se verificar que um modelo se pode inverter, um modelo ARMA(p,q) pode ser escrito:

$$X_t = \sum_{i=1}^{\infty} \phi_i X_{t-i} + \varepsilon_t \quad (3.5)$$

3.7 Processo Autorregressivo Integrado de Média Móvel

Uma grande parte das séries temporais não são estacionárias. Sendo assim, foram criados os modelos ARIMA. Um modelo ARMA Integrado, ARIMA (p, d, q), consiste em aplicar o modelo ARMA (p, q) na d-ésima diferença da série, onde:

- p é a ordem do modelo autorregressivo;
- d é o grau de diferenciação;
- q é a ordem do modelo de média móvel.

Esta diferença referida anteriormente consiste no cálculo da diferença absoluta (ou percentual) entre uma observação e outra, onde o intervalo entre as observações pode ser selecionado (Jenkins et al., 2015).

Seja $\{X_t\}$ um processo não estacionário. Geralmente, após algumas diferenças (normalmente uma ou duas são suficientes) a série torna-se estacionária. Ou seja, $W_t = \nabla^d X_t$ (d-ésima diferença da série X_t) torna-se estacionária. Assim, o modelo ARIMA (p, d, q) pode ser escrito como o modelo ARMA (p, q) aplicada a série $\{W_t\}$, ou seja,

$$\phi(B)W_t = \theta(B)\varepsilon_t \quad (3.6)$$

Nota: $\nabla X_t = (1 - B)X_t \Rightarrow W_t = \nabla^d X_t = (1 - B)^d X_t$

Assim sendo, o modelo ARIMA (p, d, q) pode ser escrito como:

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (3.7)$$

Propriedades:

- O modelo ARIMA (p, 0, q) é equivalente ao modelo ARMA (p, q);
- Frequentemente, d=1 ou d=2 são suficientes para séries não sazonais (ou com sazonalidade ajustada);
- Um modelo que se torne estacionário após “d” diferenças denomina-se integrado de ordem d.

3.7.1 SARIMA

Para além dos modelos ARIMA tradicionais representados por ARIMA (p,d,q), existem também os modelos ARIMA sazonais, representados por ARIMA (p,d,q)*(P,D,Q)*m. As letras minúsculas têm o mesmo significado que as letras maiúsculas, mas dizem respeito à parte não sazonal, enquanto as maiúsculas à parte sazonal. O m diz respeito ao nº de observações por ano (Jenkins et al., 2015).

A identificação de modelos SARIMA efetua-se a partir da função de correlação da amostra dos dados. Primeiro, encontra-se d e D de modo a fazer as observações diferenciadas. Em seguida, examinam-se as funções de autocorrelação e autocorrelação parcial da variável $\{X_t\}$ com desfasamentos que são múltiplos de m, a fim de se identificar as ordens P e Q no modelo, para que no fim se escolha o modelo ARIMA que melhor se ajusta à série temporal (Brockwell & Davis, 1991).

3.7.2 Adequação do Modelo

Modelo ARMA (p, q):

$$X_t = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.8)$$

em que $\beta = (\alpha, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$ é o vetor de parâmetros e $\{\varepsilon_t\}$ é um ruído branco (média zero e variância σ^2) (Jenkins et al., 2015).

De forma equivalente é possível escrever:

$$X_t = \mu_{t|t-1} + \varepsilon_t \Leftrightarrow$$

$$\Leftrightarrow \mu_{t|t-1} = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \Leftrightarrow \varepsilon_t = X_t - \mu_{t|t-1}$$

onde $\mu_{t|t-1}$ é função dos parâmetros, isto é, $\mu_{t|t-1} = \mu_{t|t-1}(\beta)$.

Em suma, se $X_1, \dots, X_n$ é a série observada e $\hat{\beta}$ é a estimativa de β , os resíduos do modelo são dados por:

$$\varepsilon_t(\hat{\beta}) = X_t - \mu_{t|t-1}(\hat{\beta}) \quad (3.9)$$

para $t = 1, \dots, n$.

Supondo que $\{\varepsilon_t\}$ tem:

- Média zero, $E[\varepsilon_t] = 0, t = 1, 2, 3, \dots$;
- Variância Constante, $Var[\varepsilon_t] = \sigma^2, t = 1, 2, 3, \dots$;
- Covariância nula, $Cov[\varepsilon_t, \varepsilon_{t+h}] = 0, h \neq 0, t = 1, 2, 3, \dots$;
- “Normalidade”.

Um modelo estará bem ajustado se a série de resíduos satisfizer todas as hipóteses levantadas sobre o termo de erro $\{\varepsilon_t\}$. Dos modelos que cumprirem os critérios, deve ser escolhido o mais parcimonioso, ou seja, o que dá o melhor ajuste com o menor grau de complexidade (económico) (Jenkins et al., 2015).

3.8 Métodos de Alisamento Exponencial

Os modelos de alisamento exponencial têm origem em 1944, por Robert Brown, um antigo funcionário das forças armadas dos Estados Unidos. As previsões destes métodos são produzidas através de fórmulas simples tendo em conta a tendência e a sazonalidade. Para além disso, são usadas poucas suposições. Por exemplo, este modelo não pressupõe que a série é estacionária. Na década de 1950 este método de séries temporais contínuas e discretas foi alargado passando a incluir termos para lidar com tendências e sazonalidade (Abraham & Ledolter, 1983).

Há uma variedade de métodos que se enquadram na família de alisamento exponencial, tendo cada um deles a propriedade de que as previsões são combinações ponderadas de observações passadas, tendo as observações recentes relativamente mais peso do que as observações mais antigas. O nome "alisamento exponencial" reflete o facto de que os pesos diminuem exponencialmente à medida que as observações ficam mais antigas (Abraham & Ledolter, 1983).

O método de Holt diferia do de Brown no que diz respeito à suavização da tendência e componentes sazonais. O seu trabalho original foi reproduzido num memorando do ONR (Holt 1957), que foi muito citado, mas não foi publicado até recentemente, quando apareceu no International Journal of Forecasting em 2004. O trabalho de Holt sobre o alisamento exponencial sazonal aditivo e multiplicativo tornou-se bem conhecido através de um artigo do seu estudante Peter Winters (1960), que proporcionou testes empíricos para os métodos de Holt. Como resultado, versões dos métodos de Holt são normalmente chamados métodos de Holt-Winters (Hyndman et al., 2008).

3.8.1 Métodos de Alisamento Exponencial Simples

Para $0 < \alpha < 1$, o método de alisamento exponencial simples, consiste em escrever a previsão y_{t+1} , em função da informação prévia, da seguinte forma (Hyndman et al., 2008):

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \alpha(1 - \alpha)^3 Y_{t-3} + \dots \quad (4.1)$$

Pode-se dizer que o α é o parâmetro de alisamento.

Note-se que:

- I. A previsão de Y_{t+1} é uma ponderação dos pontos anteriores em que os pesos decaem exponencialmente;
- II. Quanto maior for o valor de α , maior é o peso atribuído a observações mais recentes;
- III. Essa previsão não considera tendência nem sazonalidade.

A equação (4.1) pode ser reescrita como:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \sum_{j=0}^{\infty} \alpha (1 - \alpha)^j Y_{t-1-j}, j = i - 1 \quad (4.2)$$

De onde se obtém a relação:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (4.3)$$

3.8.2 Métodos de Holt

Holt, em 1957, estendeu o alisamento exponencial simples ao alisamento exponencial linear para permitir a previsão de dados com tendências. A previsão para o método de alisamento exponencial linear de Holt é encontrada usando duas constantes de alisamento, α e β^* (com valores entre 0 e 1), e três equações:

$$\text{Nível: } L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + B_{t-1}) \quad (4.4)$$

$$\text{Tendência: } B_t = \beta^*(L_t - L_{t-1}) + (1 - \beta^*)B_{t-1} \quad (4.5)$$

$$\text{Previsão: } \hat{Y}_{t+h} = L_t + B_t h \quad (4.6)$$

Aqui L_t denota uma estimativa do nível da série no tempo t , B_t denota uma estimativa da inclinação (ou crescimento) da série no tempo t e h refere-se ao número de previsões, ou seja, o número de passos à frente. Note-se que B_t é uma média ponderada do crescimento anterior B_{t-1} e uma estimativa do crescimento com base na diferença entre os níveis sucessivos. No caso especial em que $\alpha = \beta^*$, o método de Holt é equivalente ao alisamento exponencial duplo de Brown (Brown 1959). Brown usou apenas um argumento para chegar

às suas equações de previsão, já que $(1 - \alpha)$ representa o fator de desconto comum aplicado tanto às componentes de nível como às de tendência (Hyndman et al., 2008).

3.8.3 Métodos de Holt-Winters

Para dados sem tendência ou padrões sazonais, o método de alisamento exponencial simples é o mais apropriado. Para dados com tendência linear, o método linear de Holt é o mais apropriado. Contudo, para dados com padrões sazonais, nenhum dos dois modelos referidos é o mais apropriado. Assim sendo, Holt propôs um método para responder à característica da sazonalidade. O seu método foi estudado por Winters, em 1960, pelo que agora é normalmente conhecido como "método de Holt-Winters". Este método baseia-se em três equações de alisamento - uma para o nível, uma para a tendência e outra para a sazonalidade. É semelhante ao método linear de Holt, com uma equação adicional para lidar com a sazonalidade. Na realidade, existem dois métodos diferentes de Holt-Winters, dependendo se a sazonalidade é modelada de uma forma aditiva ou multiplicativa (Hyndman et al., 2008).

A estimação das componentes da forma aditiva é efetuada com as seguintes equações:

$$\text{Nível: } L_t = \alpha(Y_t - S_{t-m}) + (1 - \alpha)(L_{t-1} - B_{t-1}) \quad (4.7)$$

$$\text{Tendência: } B_t = \beta^*(L_t - L_{t-1}) + (1 - \beta^*)B_{t-1} \quad (4.8)$$

$$\text{Sazonalidade: } S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-m} \quad (4.9)$$

$$\text{Previsão: } \hat{Y}_{t+h} = L_{t-1} + hB_{t-1} + S_{t-m+h} \quad (5.1)$$

Aqui S_t é a componente sazonal da série Y_t , γ é um parâmetro de suavização para termos sazonais que varia entre 0 e 1 (constante de amortecimento do fator sazonal) e m é o número de unidades de tempo em cada ciclo de sazonalidade.

3.9 Erros de Previsão

É importante monitorizar a exatidão dos ajustes a fim de assegurar a utilização racional dos recursos da organização que são investidos no processo de previsão. A tarefa de medir a exatidão dos juízos de valor dos ajustamentos está inseparavelmente ligada à necessidade de escolher uma medida de erro apropriada (Davydenko & Fildes, 2013).

Com o intuito de comparar os diferentes modelos de previsão serão utilizadas duas medidas de erro: MAE e MAPE. O erro de previsão com um passo de antecedência é simplesmente $e_t = y_t - \hat{y}_t$, independentemente de como a previsão foi produzida. Da mesma forma, o erro de previsão com h passos à frente é $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$, que está na mesma escala que os dados. As medidas de exatidão baseadas em e_t são, portanto, dependentes da escala. As duas medidas dependentes da escala mais frequentemente utilizadas são baseadas no erro absoluto ou erros ao quadrado (Hyndman et al., 2008):

$$\text{Erro Absoluto Médio (MAE)} = \text{média} (|e_t|) = \frac{1}{n} * \sum_{t=1}^n |y_t - \hat{y}_t|$$

$$\text{Erro Quadrático Médio (MSE)} = \text{média} (|e_t^2|) = \frac{1}{n} * \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

Ao comparar métodos de previsão numa única série, normalmente usa-se o MAE porque é fácil de compreender e calcular. No entanto, não pode ser utilizado para fazer comparações entre séries, pois não faz sentido comparar a precisão em escalas diferentes (Hyndman et al., 2008).

A percentagem de erro é dada por $\rho_t = 100e_t/y_t$. Os erros percentuais têm a vantagem de serem independentes da escala, pelo que são frequentemente utilizados para comparar o desempenho da previsão entre diferentes conjuntos de dados. A medida mais frequentemente utilizada é:

$$\text{Erro Percentual Absoluto Médio (MAPE)} = \text{média} (|\rho_t|) = \frac{1}{n} * \sum_{t=1}^n (|\frac{y_t - \hat{y}_t}{y_t}|)$$

As medidas baseadas em erros percentuais têm a desvantagem de serem infinitas ou indefinidas se $y_t = 0$ para qualquer t no período de interesse e ter uma distribuição

extremamente enviesada quando qualquer y_t está próximo de zero. Outro problema com erros percentuais frequentemente negligenciado é o de assumirem um zero significativo. Por exemplo, um erro percentual não faz sentido ao medir a precisão das previsões de temperatura na escala Celsius. Têm também a desvantagem de aplicar uma penalização mais pesada aos erros positivos do que aos erros negativos, porque a percentagem de erro não pode exceder 100% para previsões demasiado baixas. Embora não haja um limite superior para as previsões que são demasiado altas. (Hyndman et al., 2008).

3.10. Critério de Seleção do Modelo

Existe sempre a possibilidade de dois ou mais modelos cumprirem todas as hipóteses e de serem adequados para prever a série. Quando assim é, normalmente escolhe-se o modelo com menos parâmetros. Isso é um dos maiores problemas da previsão, a determinação do modelo que melhor se adequa a cada série. Os critérios de seleção fazem um balanço entre a qualidade do ajuste e o grau de complexidade do modelo (Jenkins et al., 2015).

O método Box-Jenkins é um procedimento que tem por base um algoritmo, é iterativo, complexo, estima parâmetros de ponderação apropriados, testa o modelo e repete o ciclo até determinar qual o modelo mais adequado para estimar uma determinada série temporal. Este método utiliza as observações mais recentes como valores iniciais e depois analisa os erros de previsão recentes para determinar os ajustamentos adequados para períodos futuros (Jenkins et al., 2015).

Após a estimação dos parâmetros de um modelo, a metodologia Box-Jenkins exige o diagnóstico deste através, por exemplo, da examinação dos resíduos através da diferença entre os valores reais e os valores estimados. Se os resíduos forem aleatórios, presume-se que o modelo é apropriado. Caso contrário, outro modelo é considerado, os seus parâmetros estimados e os seus resíduos são verificados quanto à aleatoriedade. Outros testes e critérios de informação são considerados para medir a qualidade do ajustamento do modelo considerado. A figura 2 resume os passos do método (Murteira et al., 1993).

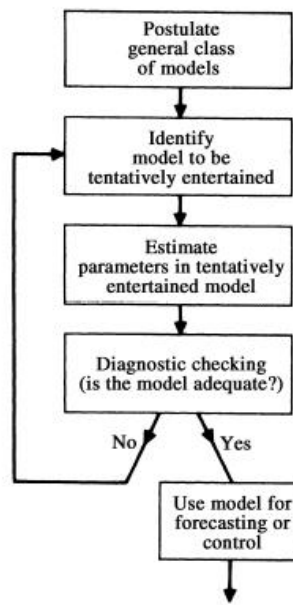


Figura 2 - Etapas da abordagem iterativa da construção de modelos (Murteira et al., 1993)

4. Caso de Estudo

4.1 Descrição da Amostra

Como referido anteriormente, o âmbito da análise recairá num estudo dos dados do volume de negócios facultados por uma empresa de consultoria, que se situa na cidade do Porto e fundada em janeiro de 2007. A empresa conta já com cerca de 40 colaboradores. A sua missão passa por desenvolver soluções tecnológicas de gestão que valorizam os negócios de todas as entidades que acreditam na dinamização da economia nacional. Tem clientes de diversas áreas como a banca, a saúde, as telecomunicações, a gestão de stocks, entre outras.

Estes dados são constituídos pelos valores do volume de negócios mensais desde janeiro de 2007 a dezembro de 2021. Este é um dado financeiro que tem vindo a crescer ao longo da vida da empresa, atingindo em 2021 um valor de faturação de aproximadamente 2 milhões de euros.

4.2 Modelação e Análise Quantitativa

Com o estudo desta dissertação pretende-se efetuar a previsão do volume de negócios da empresa para períodos posteriores, com base nos valores dos anos anteriores. O objetivo é verificar que com um sistema de previsões adequado se torne normal efetuar a previsão do valor de faturação da empresa.

Em termos informáticos, as modelações e as previsões foram efetuadas tendo por base o software R, em conjunto com o ambiente de programação *RStudio*. Neste software é permitido instalar-se packages que consistem num conjunto de funções que quando invocadas executam rotinas de programação (R Core Team, 2014). O package mais usado ao longo desta dissertação foi o “forecast”, pois permite realizar a modelação dos dados utilizando os modelos ARIMA, ETS e Holt-Winters, assim como a sua previsão. Em suma, pretende-se verificar se os modelos são adequados e qual o melhor para estimar a série.

4.3 Recolha e Preparação dos Dados

Devido à falta de dados diários, ao uso recorrente de dados mensais na literatura e com base na disponibilização dos dados por parte da empresa, construiu-se um vetor com 180 observações mensais relativas aos 15 anos. Para prever por exemplo janeiro de 2022, dispõe-se do vetor com 180 linhas relativas aos dados de janeiro de 2007 até dezembro de 2021.

No que concerne à análise descritiva, serão retiradas as últimas observações e serão ajustados os seguintes modelos para cada série:

- Modelo Alisamento Exponencial Simples;
- Modelo Holt-Winters com sazonalidade aditiva;
- Modelo SARIMA.

Para cada modelo ajustado, será realizada a previsão de 1 passo à frente e multi-passo que permite comparar as três observações originais das séries e medir qual o modelo com maior poder preditivo em termos dos erros MAE e MAPE.

4.4 Tratamento dos Dados

4.4.1 Avaliação da Estacionariedade

Com intuito de perceber qual o modelo a ser considerado, realizar-se-á um teste de estacionariedade primeiramente, depois segue-se a análise às funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

Como por hipótese se define que o erro tem média constante, variância constante e que não é autocorrelacionado, é comum ser efetuado um teste de estacionariedade também para os erros, como forma de verificar se o modelo se encontra bem ajustado.

Considere-se 2 tipos de testes para se verificar se a série observada é ou não estacionária:

- Teste Dickey-Fuller Aumentado (ADF);
- Teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS) – geralmente este é o teste efetuado para verificar a estacionariedade dos resíduos.

O objetivo destes testes passa por verificar a existência de raiz unitária (Jenkins et al., 2015).

Considerando a série temporal Y_t :

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad (5.2)$$

correspondente ao modelo AR (1). Neste caso, se $|\phi| < 1$, então Y_t é estacionário. Sendo assim, a hipótese a testar é $H_0: \phi = 1$, ou seja, a série não é estacionária contra a hipótese alternativa $H_1: \phi < 1$ (a série é estacionária) ou, conforme é mais comum, considera-se:

$$\Delta Y_t = (\phi - 1)Y_{t-1} + \varepsilon_t = \delta Y_{t-1} + \varepsilon_t \quad (5.3)$$

Neste caso consideram-se as seguintes hipóteses:

$H_0: \delta = 0$ – Existe pelo menos uma raiz unitária.

$H_1: \delta < 0$ – A série é estacionária.

O teste KPSS tem como hipóteses H_0 a de a série ser estacionária e H_1 a de a série apresentar raiz unitária. Neste teste, as hipóteses diferem dos dois anteriores (Gujarati, 2004).

5. Análise dos Resultados

Neste capítulo apresenta-se a análise dos dados, assim como a aplicação dos métodos de previsão já apresentados: ETS, Holt-Winters e ARIMA. Posteriormente, apresenta-se os valores previstos para um determinado momento. Por fim, apresenta-se uma conclusão onde se faz a análise da previsão obtida.

Para cada método, foi analisado o volume de negócios com início em janeiro de 2007 e fim em dezembro de 2021. Primeiramente, foi feita uma análise descritiva da série de forma a serem identificadas as principais características da série. De seguida, os modelos foram ajustados, foi realizado o seu diagnóstico e finalmente foram comparados em termos de erros. Os erros associados à previsão pelos três métodos encontram-se representados na Figuras 24 e 25.

5.1 Análise Descritiva da Série

Para se seleccionar a melhor metodologia é necessário analisar-se vários métodos, investigando aqueles que melhor se adaptam aos dados existentes. Na Figura 3, é possível visualizar a série original “Volume Negócios”. Da observação do comportamento da série, nota-se uma certa tendência de crescimento, o que poderá indicar não-estacionariedade na média e um aumento da sua variabilidade a partir do ano 2019. Tal sugere que a série será não-estacionária.

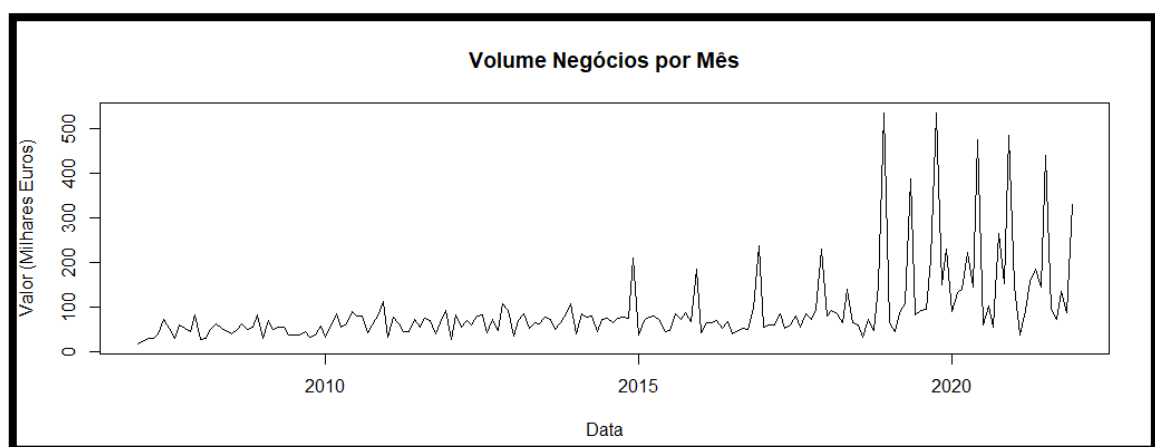


Figura 3 - Série Temporal "Volume Negócios"

Realizou-se uma decomposição para entender que tipo de modelo será adequado para descrever a série, com recurso à função STL do software R. Esta função decompõe a série em duas principais componentes: tendência e sazonalidade e apresenta os resultados graficamente. Para além destas duas componentes a função apresenta ainda os valores observados da série (data) e o que resta da série após ter sido retirada a tendência e a sazonalidade (seasonal).

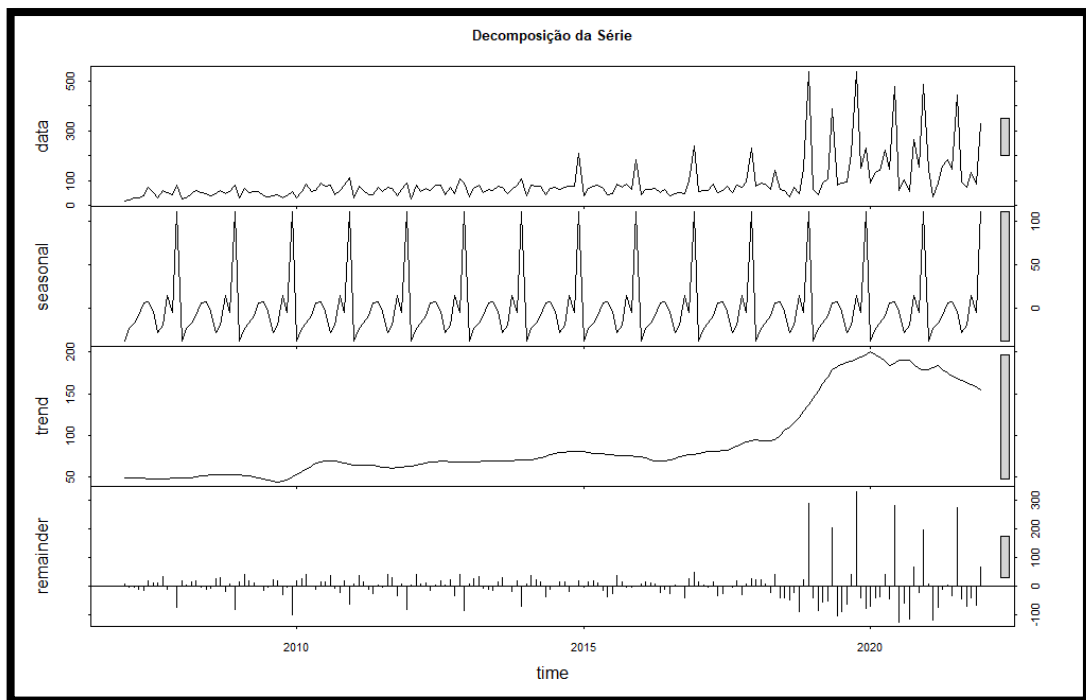


Figura 4 - Decomposição da série temporal "Volume Negócios"

Conclui-se que a série apresenta sazonalidade que se repete a cada doze meses (figura 4). Relativamente à tendência, é possível verificar que esta se encontra presente na série com um declive positivo. Recorrer à decomposição das séries facilita a compreensão do comportamento dos dados, simplificando, desta forma, a seleção do método mais adequado.

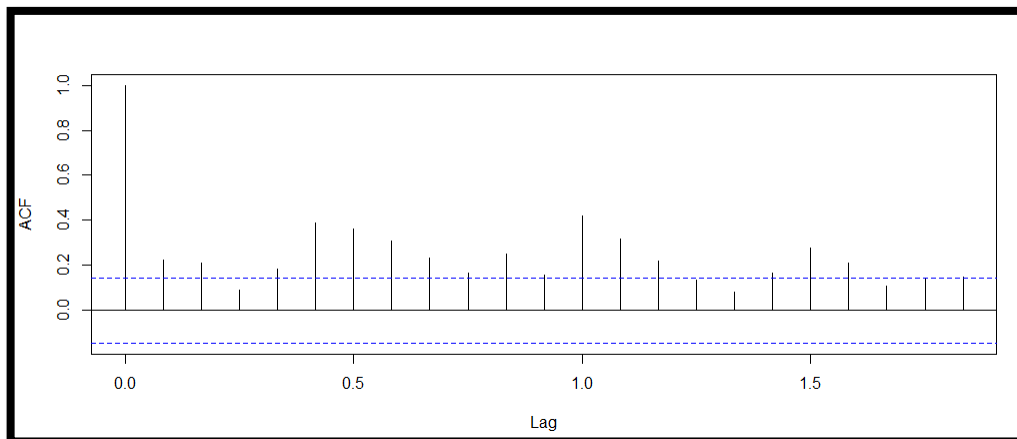


Figura 5 - Gráfico ACF da série original "Volume Negócios"

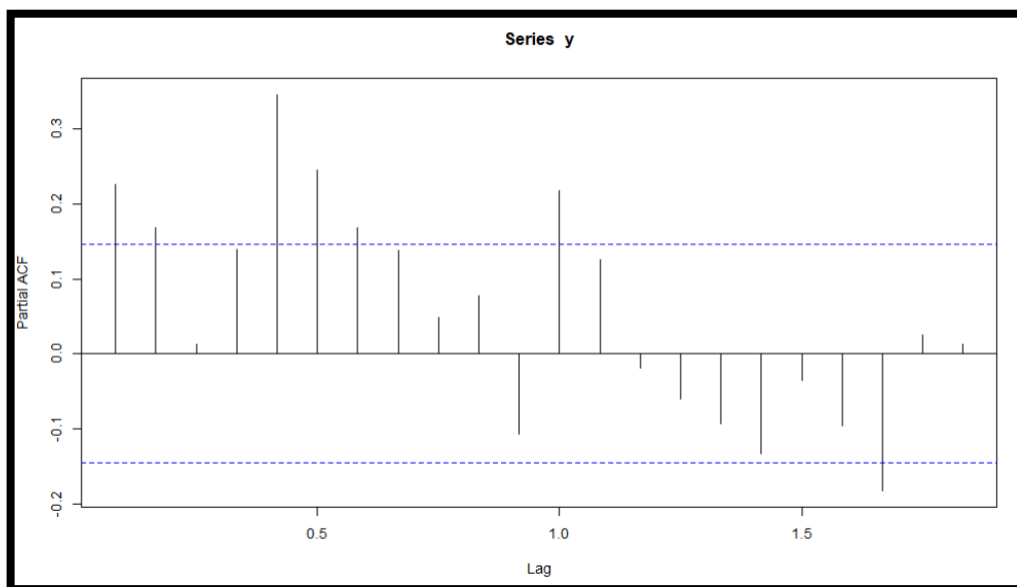


Figura 6 - Gráfico PACF da série original "Volume Negócios"

Com base na ACF e recorrendo ao gráfico da Figura 5, constata-se que os valores não se encontram todos perto dos limites do intervalo de confiança, ou seja, as correlações não são todas próximas de zero, concluindo-se serem significativas. O gráfico da PACF também evidencia que as correlações não são todas próximas de zero, porque algumas delas ultrapassam os limites do intervalo de confiança.

Para se confirmar se a série é ou não estacionária procedeu-se ao teste da estacionariedade da série temporal "Volume Negócios" através das funções `adf.test{tseries}` e `kps.test{tseries}` no *RStudio* que permite realizar os testes de *Dickey-Fuller* Aumentado e o *Kwiatkowski-Phillips-*

Schmidt-Shin.

Com as hipóteses definidas para o teste ADF:

H_0 : a série não é estacionária

H_1 : a série é estacionária

```
Augmented Dickey-Fuller Test  
  
data: y  
Dickey-Fuller = -3.8951, Lag order = 5, p-value = 0.01584  
alternative hypothesis: stationary
```

Figura 7 - Teste ADF para determinar estacionariedade da série "Volume Negócios"

Visto que o $p\text{-value}=0,01584 < 0,05$; rejeitar-se-ia H_0 para $\alpha=0,05$, assim como para qualquer nível de significância superior a 1,6% aproximadamente, o que levaria à rejeição da não estacionariedade da série. Ou seja, apesar da tendência e sazonalidade sugeridas na decomposição da série, tal não é confirmado pelo teste à estacionariedade para estes níveis.

Já para o teste KPSS, com as hipóteses contrárias ao teste ADF, os resultados foram os seguintes:

```
KPSS Test for Level Stationarity  
  
data: y  
KPSS Level = 2.4252, Truncation lag parameter = 4, p-value = 0.01
```

Figura 8 - Teste KPSS para determinar estacionariedade da série "Volume Negócios"

Sendo o $p\text{-value}=0,01 < 0,05$; rejeitar-se-ia H_0 para $\alpha=0,05$, assim como para qualquer nível de significância superior a 1%, o que levaria à rejeição da estacionariedade da série.

5.2 Estimação dos Modelos

5.2.1 ARIMA

Recorrendo à função `auto.arima{tseries}` no *RStudio* foi possível determinar qual o melhor modelo que descreve a série “Volume Negócios”. Através do sistema concluiu-se que o melhor modelo é um SARIMA (0,1,1)(0,0,1)[12].

```
Series: y
ARIMA(0,1,1)(0,0,1)[12] with drift

Coefficients:
      ma1      sma1      drift
    -0.9579  0.2052  0.8183
s.e.    0.0244  0.0660  0.3173

sigma^2 = 5455: log likelihood = -1023.95
AIC=2055.9  AICc=2056.13  BIC=2068.65

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -0.01596745 73.03179 36.90448 -25.35419 41.31356 0.8765771 -0.05807075
```

Figura 9 – Estimação do modelo ARIMA

O critério AIC (Akaike Information Criterion) e o critério BIC (Bayesian Information Criterion) são critérios de avaliação do diagnóstico do modelo. Estes critérios servem para medir a qualidade do ajustamento do modelo e podem ser usados como critérios de seleção do modelo (comparando os valores do AIC e do BIC para os vários modelos). Quanto menor forem o AIC e o BIC melhor é o modelo.

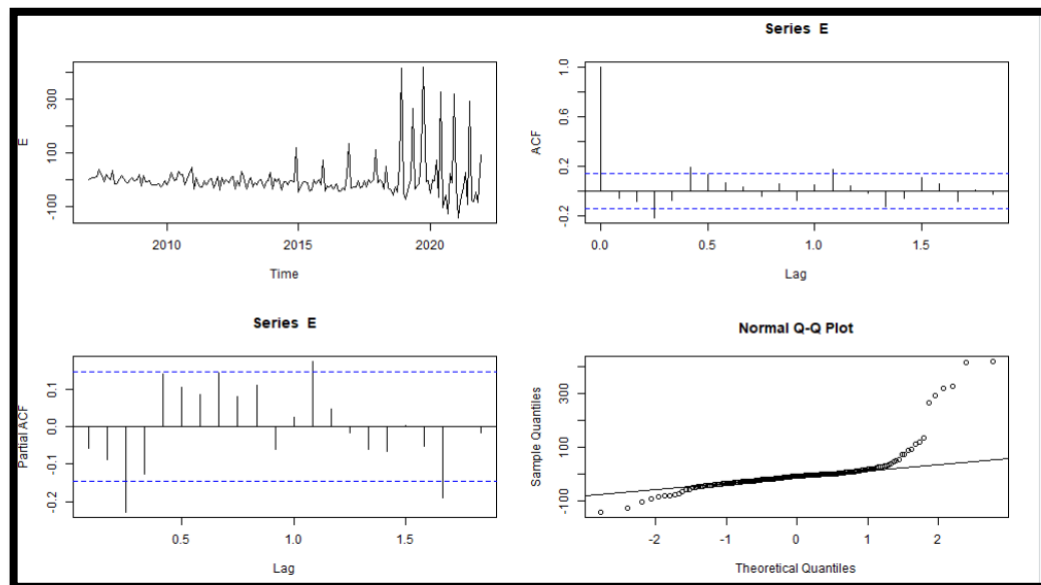


Figura 10 - Gráfico, ACF e PACF dos resíduos e Q-Q Plot com as observações dos resíduos

Os resíduos foram calculados através da função `{auto.arima$residuals}` e denominados por E. Um gráfico Q-Q é um gráfico de dispersão criado para evidenciar dois conjuntos de quantis um contra o outro. No caso dos conjuntos de quantis resultarem da mesma distribuição, os pontos formam uma linha mais ou menos reta. Ao nível dos resíduos do modelo SARIMA (0,1,1)(0,0,1)[12], verifica-se que as caudas fogem um pouco da reta de normalidade devido ao facto de nos últimos anos das observações haver uma variabilidade significativa. Como é possível verificar na imagem acima, a maior parte dos pontos encontram-se em cima da reta da normalidade. Os gráficos Q-Q normais que exibem este comportamento significam que os dados têm valores mais extremos do que seria de esperar caso se resultassem verdadeiramente de uma distribuição Normal.

Foi realizado o teste KPSS aos resíduos deste modelo para analisar a estacionariedade dos mesmos, porque geralmente é o teste utilizado para esta análise. Verifica-se um p-value de 0,1 (10%), ou seja, maior do que 0,05, pelo que se conclui que os resíduos podem ser estacionários, já que não se rejeita a hipótese H_0 de estacionariedade da série de resíduos, o que seria de esperar para se verificar a adequação do modelo estimado.

```

      KPSS Test for Level Stationarity

data: E
KPSS Level = 0.32987, Truncation lag parameter = 4, p-value = 0.1

```

Figura 11 - Teste à estacionariedade dos resíduos do modelo SARIMA (0,1,1)(0,0,1)[12]

5.2.2 Holt -Winters Aditivo

Recorrendo à função `hw{tseries}` no *RStudio* foi possível determinar os parâmetros do modelo Holt-Winters Aditivo para a série “Volume Negócios”.

```

Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
hw(y = y, h = 3, seasonal = "a", initial = "s")

Smoothing parameters:
alpha = 0.0535
beta = 0
gamma = 0.1925

Initial states:
l = 44.6269
b = 0.5254
s = 37.6457 0.6009 6.8121 16.124 -14.3072 9.2532
    27.4305 -4.616 -14.8265 -14.6232 -21.871 -27.6225

sigma: 67.4063
Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 2.747092 67.40631 31.79931 -16.25032 32.81369 0.755316 -0.0175462

```

Figura 12 - Estimação do Modelo Holt-Winters Aditivo

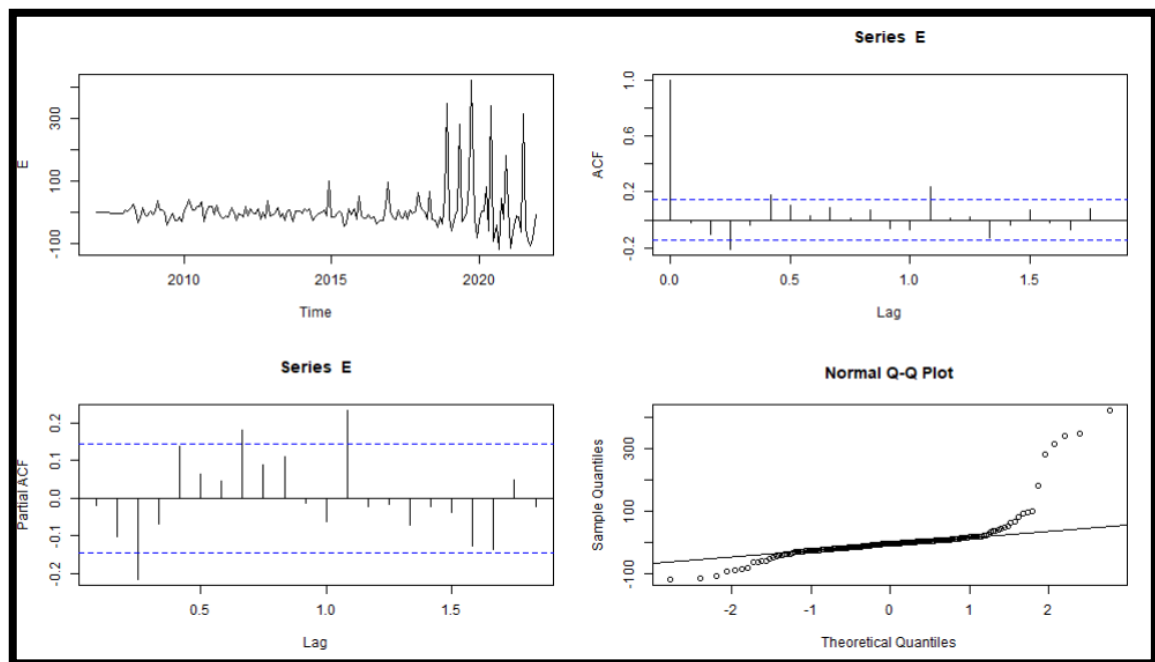


Figura 13 - Gráfico, ACF e PACF dos resíduos e Q-Q Plot com as observações dos resíduos

Os resíduos foram calculados através da função `{hw_model$residuals}` e denominados por E. O teste KPSS revela um p-value de 0,1 (10%), ou seja, maior do que 0,05, pelo que se conclui que os resíduos são estacionários.

KPSS Test for Level Stationarity

```

data: E
KPSS Level = 0.32113, Truncation lag parameter = 4, p-value = 0.1

```

Figura 14 - Teste à estacionariedade dos resíduos do modelo Holt-Winters Aditivo

5.2.3 Alisamento Exponencial Simples

O modelo exponencial simples já se sabe à partida que não poderá ser muito adequado pois este modelo não tem componente de tendência, nem componente sazonal e já se verificou que a série em questão tem tendência e sazonalidade. Logo o modelo exponencial apropriado seria à partida o Holt-Winters. Contudo, o modelo irá ser ajustado aos dados (série) para os resultados serem comparados com os resultados dos outros modelos.

Recorrendo à função `ses{tseries}` no *RStudio* foi possível determinar os parâmetros do modelo Alisamento Exponencial Simples para a série “Volume Negócios”.

```
Forecast method: Simple exponential smoothing
Model Information:
Simple exponential smoothing
Call:
ses(y = y, h = 3)
Smoothing parameters:
alpha = 0.0784
Initial states:
l = 44.2509
sigma: 76.4693
      AIC      AICc      BIC
2500.001 2500.138 2509.580
Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 9.407153 76.04327 37.72422 -14.31848 39.00736 0.896048 -0.06559578
```

Figura 15 - Estimação do Modelo Alisamento Exponencial Simples

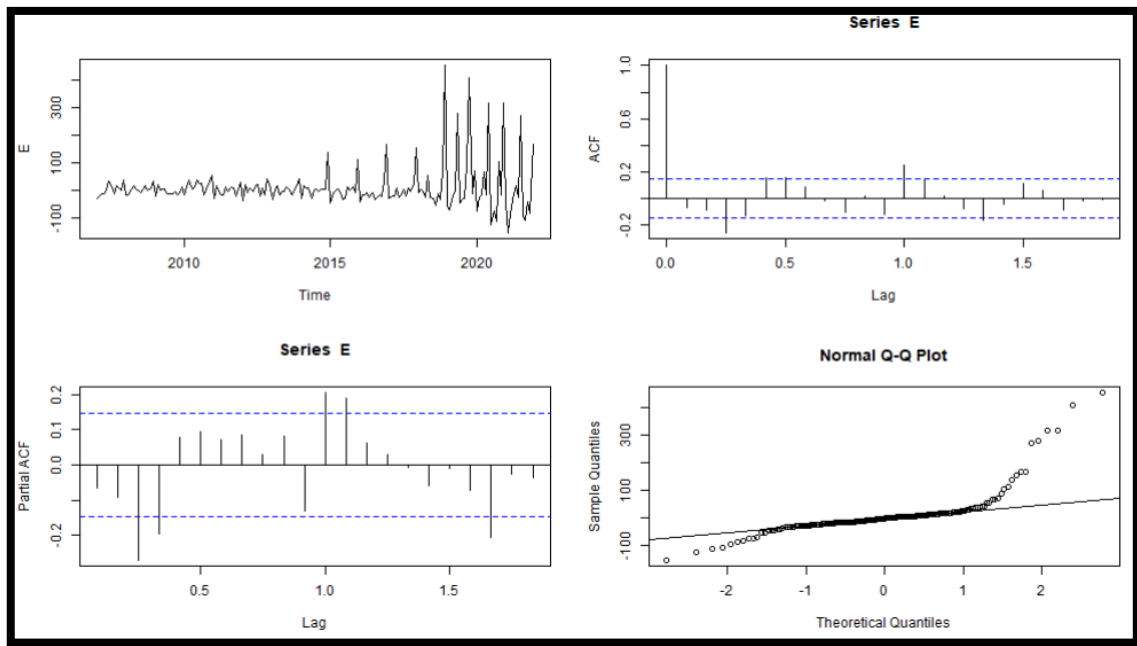


Figura 16 - Análise dos Resíduos

Os resíduos foram calculados através da função `{se_model$residuals}` e denominados por E. O teste KPSS revela um p-value de 0,09174 (9,2 % aproximadamente), ou seja, maior do que 0,05, pelo que se conclui que os resíduos são estacionários.

```
KPSS Test for Level Stationarity

data: E
KPSS Level = 0.36615, Truncation lag parameter = 4, p-value = 0.09174
```

Figura 17 - Teste à estacionariedade dos resíduos do modelo Alisamento Exponencial Simples

5.3 Exemplo da aplicação dos métodos

Nesta secção calculam-se as previsões segundo os modelos estudados (ARIMA, Holt-Winters Aditivo e Alisamento Exponencial Simples) referentes aos meses de outubro, novembro e dezembro de 2021.

Assumiu-se que a série a ser usada para essa previsão iria de janeiro de 2007, mas apenas até setembro de 2021, pelo que recorrendo às funções `auto.arima{tseries}`, `hw{tseries}` e `ses{tseries}` no *RStudio* chegou-se a modelos diferentes dos estimados anteriormente para a série original total, com o objetivo de se prever os meses referidos anteriormente. Para o modelo ARIMA,

o modelo que melhor se enquadra na série de janeiro 2007 a setembro de 2021 é agora um SARIMA $(4,1,2)(0,0,2)[12]$.

Usando estes modelos estimados para a série truncada fez-se a previsão do último trimestre de 2021. Relativamente ao modelo Alisamento Exponencial Simples, o sistema calculou o mesmo valor para os 3 meses. Os pontos azuis das figuras seguintes dizem respeito aos valores previstos para o último trimestre de 2021, a área sombreada escura representa o intervalo de confiança de 95% e a área sombreada clara o intervalo de confiança de 99%.

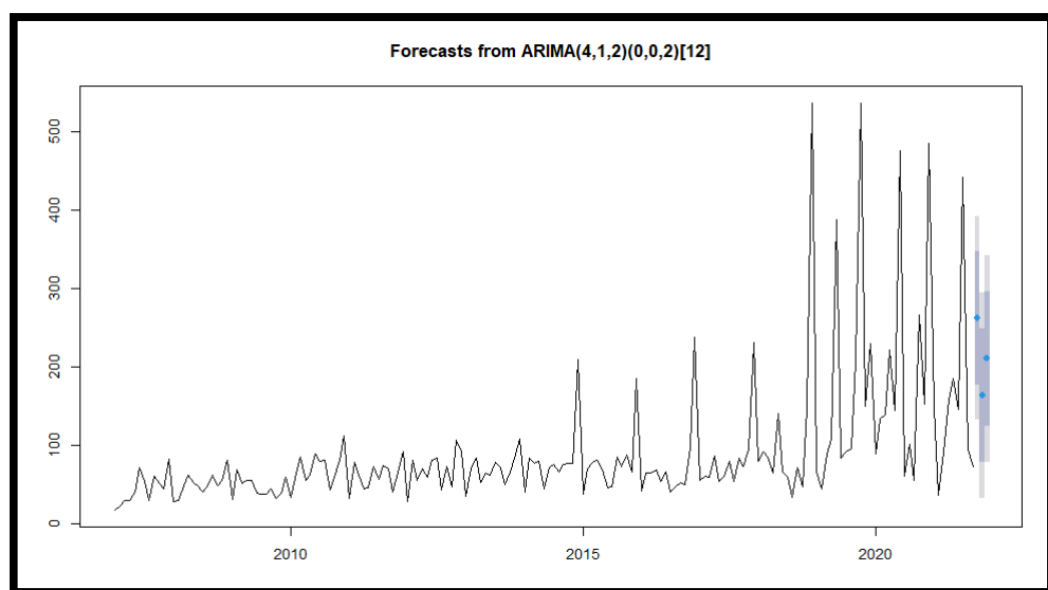


Figura 18 - Previsão da Série "Volume Negócios" para o último trimestre de 2021 – Modelo SARIMA $(4,1,2)(0,0,2)12$

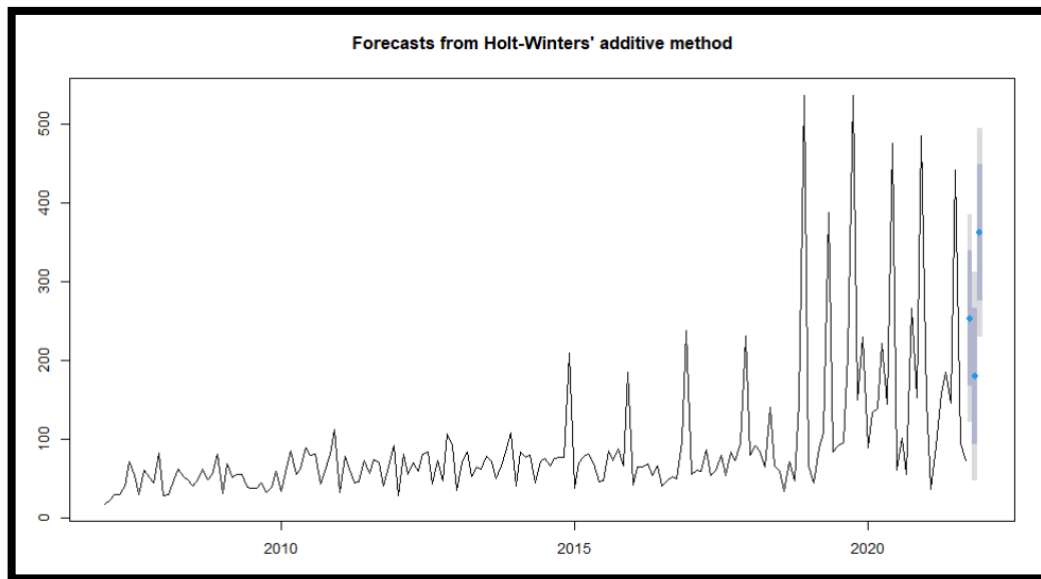


Figura 19 - Previsão da Série "Volume Negócios para o último trimestre de 2021 – Modelo Holt-Winters Aditivo

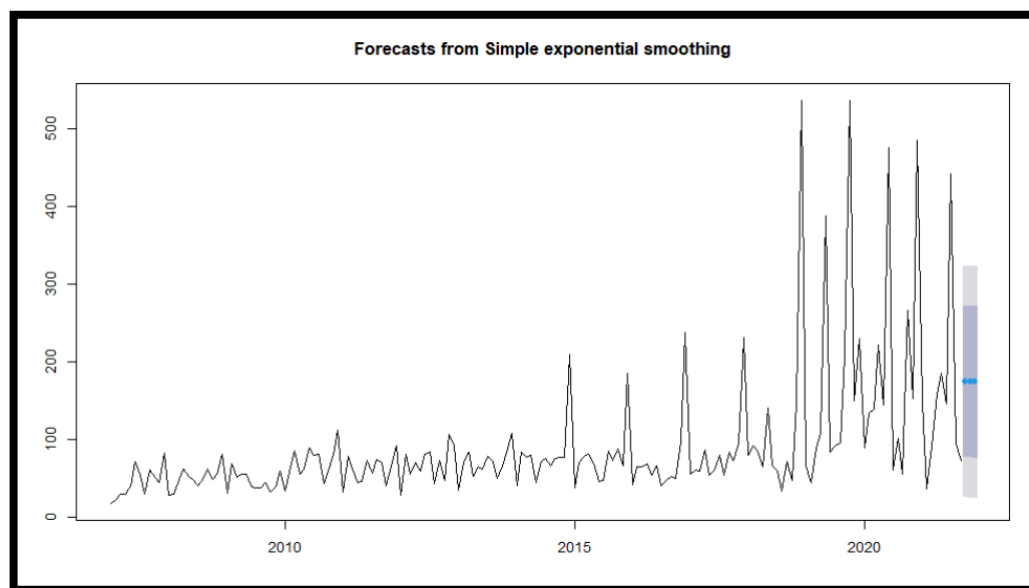


Figura 20 - Previsão da Série "Volume Negócios" para o último trimestre de 2021 – Modelo Alisamento Exponencial Simples

Recorrendo novamente à série original, ou seja, de janeiro 2007 a dezembro 2021, realizou-se a previsão do primeiro trimestre de 2022 para a qual não existe informação sobre os valores reais. Nas Figuras 21, 22 e 23 é possível verificar essa previsão tendo por base os modelos estudados nesta dissertação. Estas são as previsões de acordo com os modelos estimados na secção 5.2. Relativamente ao modelo Alisamento Exponencial Simples, o sistema calculou o mesmo valor para os 3 meses.

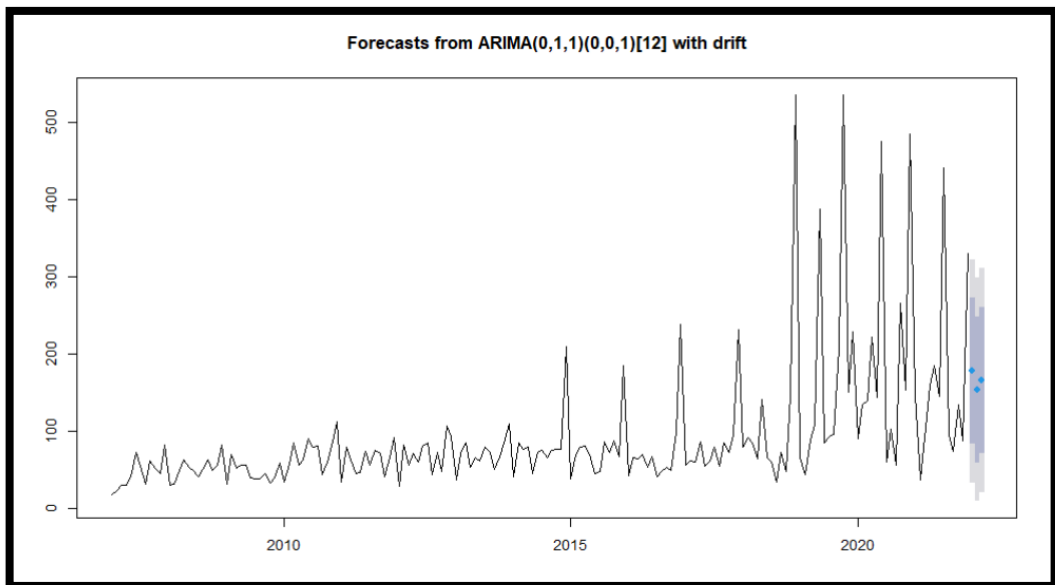


Figura 21 - Previsão da Série "Volume Negócios" para o primeiro trimestre 2022 – Modelo SARIMA $(0,1,1)(0,0,1)[12]$

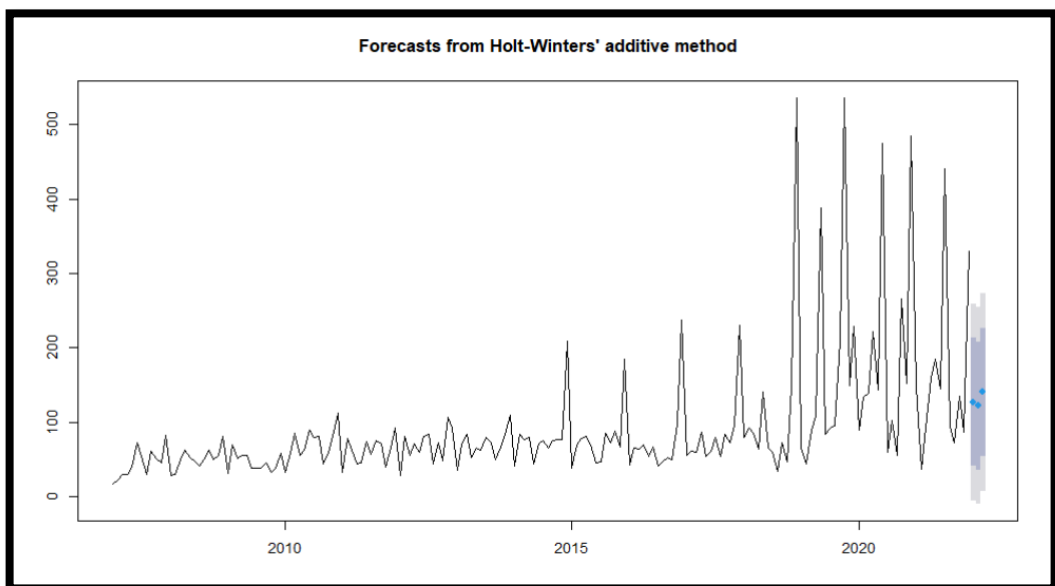


Figura 22 - Previsão da Série "Volume Negócios" para o primeiro trimestre 2022 – Modelo Holt-Winters Aditivo

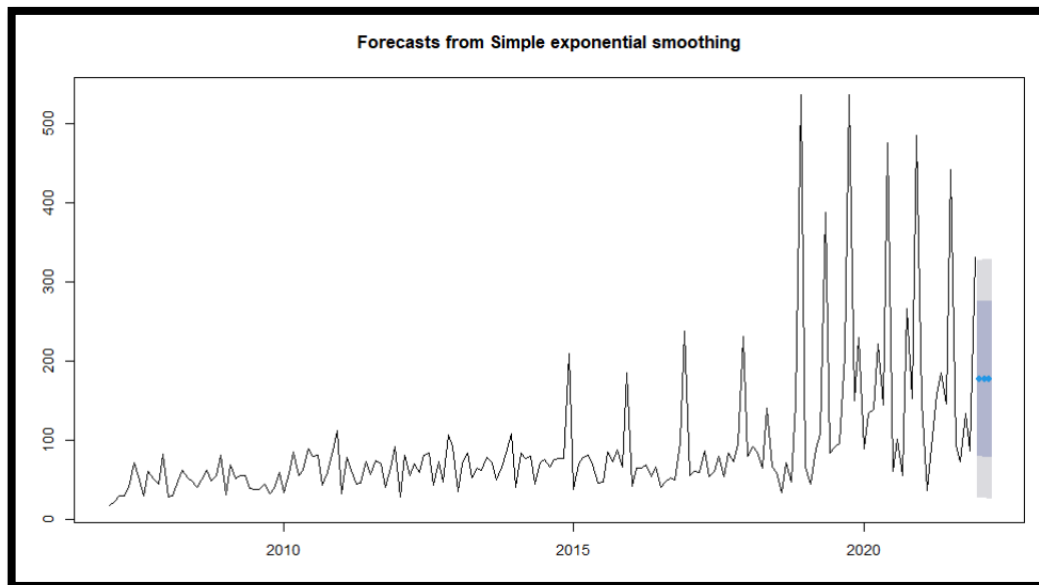


Figura 23 - Previsão da Série "Volume Negócios" para o primeiro trimestre 2022 – Modelo Alisamento Exponencial Simples

5.4 Comparação dos erros dos modelos estimados

Nesta secção são analisados os resultados obtidos quando ajustados os três modelos apresentados no capítulo 3. Os erros MAE e MAPE foram obtidos para cada um dos métodos, com base na função de estimação dos modelos apresentadas nas respectivas secções. Os resultados estão apresentados nas figuras seguintes.

Erro Absoluto Médio

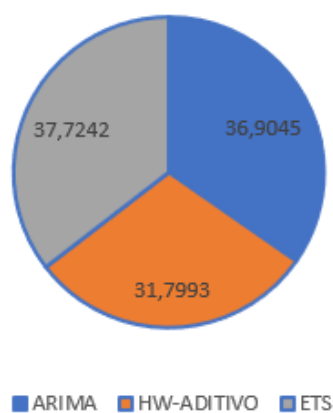


Figura 24 - Erros MAE da série "Volume Negócios"

Erro Percentual Absoluto Médio

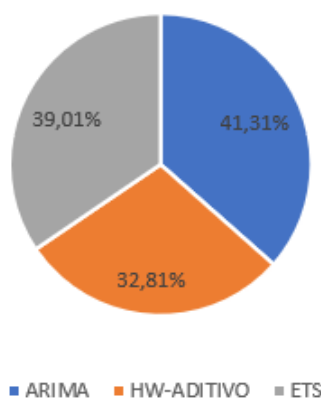


Figura 25 - Erros MAPE da série "Volume Negócios"

Da análise das Figuras 24 e 25 verifica-se que a diferença entre os erros não é muito significativa. Sendo o MAPE uma medida de erro relativa, talvez seja a medida mais indicada para a seleção de um modelo de previsão. Do erro percentual médio absoluto (MAPE), o menor erro médio (32,81) é alcançado pelo método de Holt-Winters Aditivo.

6. Conclusões

As empresas do setor da consultoria empresarial têm vindo a crescer de forma sustentável. No caso da empresa em estudo, é viável realizar o seu trabalho em qualquer lugar, desde que tenha acesso a um ponto de internet. É um fator importante, dado os contextos atuais, porque a empresa não teve de parar o seu trabalho e, consequentemente, não apresentou uma quebra nas vendas, conforme se verificou noutros setores devido aos confinamentos provocados pela COVID-19.

A previsão é muito utilizada nas empresas, pois é uma base de definição de objetivos e adoção de estratégias. É uma maneira de prevenir o futuro. Com a análise efetuada, foi possível concluir que a empresa tem valores muito díspares, muitas vezes em meses seguidos, ou seja, verifica-se uma grande volatilidade.

A previsão do volume de negócios de uma determinada empresa é possível, bastando para isso contar com softwares, como é o caso do *RStudio*, que através das funções e packages disponíveis o permite fazer. Para além disso, os recursos humanos especializados na área também são muito importantes, pois não basta só determinar certos números, mas também é necessário saber analisá-los. É de evidenciar que para tal são necessárias amostras relevantes para que o processo da previsão seja o melhor possível, ou seja, quanto maior for a amostra melhor será a previsão e, possivelmente, menor serão os erros. Visto que os dados obtidos para esta dissertação não eram muitos, os valores previstos não foram tão consistentes como o desejado, isto é, previram-se valores aproximados, bem como valores díspares. Em suma, é uma informação relevante para analisar no curto prazo e quem sabe definir novos objetivos e estratégias de negócio.

Analisaram-se vários modelos para a previsão do volume de negócios e recorreu-se às medidas de erro MAE e MAPE de forma a avaliar o desempenho dos mesmos. Concluindo que, para o conjunto de modelos, o método Holt-Winters Aditivo foi o que conseguiu melhores resultados de acordo com a medida de erro MAPE, pois apresentou a percentagem de erro mais baixa. Ainda assim, devido à notável volatilidade da série é importante realçar que os valores do MAPE, na minha opinião são um pouco elevados. Esta mudança de comportamento da variabilidade da série pode afetar posteriormente a adequação de um

modelo, sendo produzidos valores elevados para os erros. Isto deve-se ao facto de a empresa ter iniciado atividade em 2007 e ao final de alguns anos ter visto os seus valores a crescer e variar de uma forma que até então não tinha acontecido.

7. Referências Bibliográficas

- Abraham, B., & Ledolter, J. (1983). *Statistical Methods for Forecasting*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470316610>
- Alves, R., & Delgado, C. (1997). *Processos estocásticos*. Repositório Faculdade Economia Porto.
- Books, A. C. C. (2021). *Operating revenue definition*. Accounting Tools. Consultado a 27-12-2021 Disponível em <https://www.accountingtools.com/articles/operating-revenue.html>
- Brockwell, P.J., & Davis, R.A. (1991). *Time Series: Theory and Methods*. Springer New York. <https://doi.org/10.1007/978-1-4419-0320-4>
- Brown, R. (1959). *Smoothing, Forecasting, and prediction of Discrete Time series*, Prentice-Hall, New Jersey.
- Buttle, F. A. (1996). SERVQUAL: review, critique, research agenda. *European Journal of Marketing* 30, 1-8.
- Caniato, F., Kalchschmidt, M., & Ronchi, S. (2011). Integrating quantitative and qualitative forecasting approaches: organizational learning in an action research case. *Journal of the Operational Research Society*, 62(3), 413-424. <https://doi.org/10.1057/jors.2010.142>
- Chatfield, C. (2000). *Time-Series Forecasting*. Chapman and Hall/CRC.
- Costa, C. A. (2003). *Métodos Qualitativos na Análise de Risco de Crédito – Confiança e Credibilidade na Relação entre Bancos e Empresas*. Universidade do Minho – Escola de Economia e Gestão
- Cryer, J. D. & Chan, K. (2008). *Time series analysis: With applications in R* (2ª edição). Springer.

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510–522. <https://doi.org/10.1016/j.ijforecast.2012.09.002>

Debnath, K. B., & Mourshed, M. (2018). Forecasting methods in energy planning models. *Renewable & Sustainable Energy Reviews*, 88, 297-325. <https://doi.org/10.1016/j.rser.2018.02.002>

Field, J. G., Bosco, F. A., & Kepes, S. (2021). How robust is our cumulative knowledge on turnover? *Journal of Business and Psychology*, 36(3), 349-365.

Flambó, A. (2010). Métodos de Previsão. *Proelium*, 14. Lisboa, Academia Militar.

Grachev, M., & Rakitsky, B. (2013). Historic horizons of Frederick Taylor's scientific management. *Journal of Management History*.

Gujarati DN (2004). Basic Econometrics, McGraw Hill.

Hamilton, J. (1994). Time series analysis. Princeton: Princeton University Press.

Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). Forecasting with Exponential Smoothing. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-71918-2>

Jara, E. C. (2011). Long memory time series forecasting by using genetic programming. *Genetic Programming and Evolvable Machines*, 12(4), 429-456. <https://doi.org/10.1007/s10710-011-9140-7>

Jenkins, G.M., Reinsel, G.C., Box, G.E. P., & Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control. Wiley & Sons, Incorporated, John.

Li, C. H., & Liu, S. G. (2005). Trajectory of consulting firms and the dynamics of the services: comparative perspectives between USA and China. *Proceedings of ICSSSM '05. International Conference on Services Systems and Services Management*
<https://doi.org/10.1109/icsssm.2005.1500207>

Liu, L., An, Y. J., Fan, Z. Y., Su, W., & Ieee. (2016). Quality evaluation based on multivariate statistical forecasting methods. *25th International Symposium on Industrial Electronics (ISIE)*.
<https://doi.org/10.1109/isie.2016.7745033>

Murteira, B., Müller, D. and Turkman, K. (1993). *Análise de Sucessões Cronológicas*, Lisboa: McGraw-Hill.

Naik, G. (2004). The structural qualitative method: a promising forecasting tool for developing country markets. *International Journal of Forecasting*, 20(3), 475-485.

Neto, A. S. (2004) *A Importância da consultoria na implementação de sistemas ERP*.

Pater, L., & Cociu, N. (2015). Adjustment of time-series using the modified exponential for turnover forecast. *Proceedings of the international conference on numerical analysis and applied mathematics 2014 (icnaam-2014)*. AIP Publishing LLC. <https://doi.org/10.1063/1.4912900>

Piergiovanni, R., Carree, M. A., & Santarelli, E. (2012). Creative industries, new business formation, and regional economic growth. *Small Business Economics*, 39(3), 539-560.
<https://doi.org/10.1007/s11187-011-9329-4>

R Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ratalewska, M. (2018). The Role of Entrepreneurship and Business Consulting in Companies in Creative Industries. *Problemy Zarządzania-Management Issues*, 16(1), 181-191.
<https://doi.org/10.7172/1644-9584.73.11>

Reihlen, M., & Nikolova, N. (2010). Knowledge production in consulting teams. *Scandinavian Journal of Management*, 26(3), 279-289. <https://doi.org/10.1016/j.scaman.2010.05.004>

Tissen, M., Sneidere, R., Univ Latvia, F. B. M., & Econ. (2018). The theoretical and practical analysis of turnover ratios modification for the company's business activity. *10th International Scientific Conference on New Challenges of Economic and Business Development - Productivity and Economic Growth*. Universidade Latvia, Riga, LATVIA.

Valebnikova, O. A., Valebnikova, N. V., & Kalinina, O. V. (2019). Intellectually-Oriented Consulting for Financial Function in the era of Technology and Digitalization. *Proceedings of the European Conference on Knowledge Management*. Universidade Europeia Lisboa, Portugal.

Veres, Z., & Varga-Toldi, K. (2020). ERIP: service quality model of management consulting projects. *Journal of Business & Industrial Marketing*.

Volume de negócios. (2020). OCC - Ordem dos Contabilistas Certificados. Consultado a 13-01-2022 Disponível em <https://www.occ.pt/pt/noticias/volume-de-negocios/>

Yum, S. (2016). The economic growth of creative industries in the Miami metropolitan area. *Creative Industries Journal*, 9(2), 130-145. <https://doi.org/10.1080/17510694.2016.1206358>

8. Anexos

Análise da Série

```
##### VARIÁVEIS #####
X <- valores_tese$`Ano - Mês`
Y <- valores_tese$Valor

##### SERIE ORIGINAL #####

require(graphics)

gnp <- ts(valores_tese$Valor, frequency = 12, start = c(2007, 1), end =
c(2021, 9))
#print( ts(valores_tese$Valor, frequency = 12, start = c(2007, 1), end =
c(2021, 12)))
plot(gnp, xlab = "Data", ylab= "Valor (Milhares Euros)",main = "Volume
Negócios por Mês" )

#####
##### STL #####
#####

library(magrittr)
z <- ts(valores_tese$Valor, frequency = 12, start= c(2007, 1), end=
c(2021, 9)) #estimar 2021
y <- ts(valores_tese$Valor, frequency = 12, start= c(2007, 1), end=
c(2021, 12)) #estimar 2022
stl(x= y, s.window = 'periodic') %>% plot(main='Decomposição da Série')

#####
##### ACF #####
#####

acf(y, main = '')

#####
# augmented Dickey-Fuller test #####
#####

library(tseries)

#perform augmented Dickey-Fuller test
adf.test(y)
```

```
#####  
##### KPSS #####  
#####  
library("fUnitRoots")  
kpss.test(y)  
tsstationary = diff(y, differences=1)  
plot(tsstationary, type = "l")
```

Arima

```
#####  
##### ARIMA #####  
#####  
  
library(forecast)  
auto.arima(z, trace =TRUE) ##### 2021  
  
#fit1:  
fit1<- arima(z, order =c(5, 1, 1), seasonal = c(0, 0, 2), include.mean =  
FALSE) ##### 2021 #####  
  
# residuos  
E <- fit1$residuals ##### 2021 #####  
  
# analise visual  
par(mfrow=c(2,2))  
plot(E)  
acf(E)  
pacf(E)  
qqnorm(E)  
qqline(E)  
  
## testes estatisticos  
#estacionariedade  
kpss.test(E) #H0: serie estacionaria  
  
#####  
##### MAPE #####  
#####  
  
library(readr)  
library(ggplot2)  
library(forecast)  
library(fpp2)  
library(TTR)  
library(dplyr)  
  
dat_real = subset(valores_tese, CLASS == 'REAL')  
dat_teste = subset(valores_tese, CLASS == 'TESTE')  
  
nrow(dat_real); nrow(dat_teste)
```

```

mape <- function(actual,pred){
  mape <- mean(abs((actual - pred)/actual))*100
  return (mape)
}

#Previsão 2021
arima_model <- auto.arima(z)
summary(arima_model)
fore_arima = forecast::forecast(arima_model, h=3)
summary(fore_arima)
#Previsão 2021

plot(fore_arima)
df_arima = as.data.frame(fore_arima)
dat_teste$arima = df_arima$`Point Forecast`
mape(dat_teste$Valor, dat_teste$arima)

#Previsão 2022
arima_model <- auto.arima(y)
summary(arima_model)
fore_arima = forecast::forecast(arima_model, h=3)
summary(fore_arima)
plot(fore_arima)
#Previsão 2022

#####
##### MAE #####
#####

library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)

dat_real = subset(valores_tese, CLASS == 'REAL')
dat_teste = subset(valores_tese, CLASS == 'TESTE')

nrow(dat_real); nrow(dat_teste)

mae <- function(actual,pred){
  mae <- mean(abs((actual - pred)))
  return (mae)}

```

```

#Previsão 2021
arima_model <- auto.arima(z)
summary(arima_model)
fore_arima = forecast::forecast(arima_model, h=3)
summary(fore_arima)
#Previsão 2021

df_arima = as.data.frame(fore_arima)
dat_teste$arima = df_arima$`Point Forecast`
mae(dat_teste$Valor, dat_teste$arima)

#####
##### MSE #####
#####

library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)

dat_real = subset(valores_tese, CLASS == 'REAL')
dat_teste = subset(valores_tese, CLASS == 'TESTE')

nrow(dat_real); nrow(dat_teste)

mse <- function(actual,pred){
  mse <- mean(abs((actual - pred)^2))
  return (mse)}

#Previsão 2022
arima_model <- auto.arima(y)
summary(arima_model)
fore_arima = forecast::forecast(arima_model, h=3)
#Previsão 2022

df_arima = as.data.frame(fore_arima)
dat_teste$arima = df_arima$`Point Forecast`
mse(dat_teste$Valor, dat_teste$arima)

```


Holt-Winters Aditivo

```
#####  
##### HOLT-WINTERS #####  
#####  
  
#####  
##### MAPE #####  
#####  
library(readr)  
library(ggplot2)  
library(forecast)  
library(fpp2)  
library(TTR)  
library(dplyr)  
  
dat_real = subset(valores_tese, CLASS == 'REAL')  
dat_teste = subset(valores_tese, CLASS == 'TESTE')  
  
nrow(dat_real); nrow(dat_teste)  
  
mape <- function(actual,pred){  
  mape <- mean(abs((actual - pred)/actual))*100  
  return (mape)  
}  
  
#Previsão 2021  
hw_model <- hw(z, h = 3, seasonal = "a", initial = "s")  
summary(hw_model)  
plot(hw_model)  
#Previsão 2021  
  
df_holt = as.data.frame(hw_model)  
dat_teste$holt = df_holt$`Point Forecast`  
mape(dat_teste$Valor, dat_teste$holt)  
  
#Previsão 2022  
hw_model <- hw(y, h = 3, seasonal = "a", initial = "s")  
summary(hw_model)  
plot(hw_model)  
#Previsão 2022  
  
#####
```

```
##### MAE #####
#####

library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)

dat_real = subset(valores_tese, CLASS == 'REAL')
dat_teste = subset(valores_tese, CLASS == 'TESTE')

nrow(dat_real); nrow(dat_teste)

mae <- function(actual,pred){
  mae <- mean(abs((actual - pred)))
  return (mae)}

#Previsão 2021
hw_model <- hw(z, h = 3, seasonal = "a", initial = "s")
summary(hw_model)
#plot(hw_model)
#Previsão 2021

df_holt = as.data.frame(hw_model)
dat_teste$holt = df_holt$`Point Forecast`
mae(dat_teste$Valor, dat_teste$holt)

#####
##### MSE #####
#####

library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)

dat_real = subset(valores_tese, CLASS == 'REAL')
dat_teste = subset(valores_tese, CLASS == 'TESTE')

nrow(dat_real); nrow(dat_teste)
```

```

mse <- function(actual,pred){
  mse <- mean(abs((actual - pred)^2))
  return (mse)}

#Previsão 2022
hw_model <- hw(y, h = 3, seasonal = "a", initial = "s")
summary(hw_model)
#plot(hw_model) #outra maneira de plotar a previsão
#Previsão 2022

df_holt = as.data.frame(hw_model)
dat_teste$holt = df_holt$`Point Forecast`
mse(dat_teste$Valor, dat_teste$holt)

```

Alisamento Exponencial Simples

```
#####  
##### ETS #####  
#####  
  
#####  
##### MAPE #####  
#####  
library(readr)  
library(ggplot2)  
library(forecast)  
library(fpp2)  
library(TTR)  
library(dplyr)  
  
dat_real = subset(valores_tese, CLASS == 'REAL')  
dat_teste = subset(valores_tese, CLASS == 'TESTE')  
  
nrow(dat_real); nrow(dat_teste)  
  
mape <- function(actual,pred){  
  mape <- mean(abs((actual - pred)/actual))*100  
  return (mape)  
}  
  
#Previsão 2021  
se_model <- ses(z, h = 3)  
summary(se_model)  
plot(se_model)  
#Previsão 2021  
  
df_fc = as.data.frame(se_model)  
dat_teste$simplexp = df_fc$`Point Forecast`  
mape(dat_teste$Valor, dat_teste$simplexp)  
  
#Previsão 2022  
se_model <- ses(y, h = 3)  
summary(se_model)  
plot(se_model)  
#Previsão 2022  
  
#####  
##### MAE #####
```

```
#####
library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)

dat_real = subset(valores_tese, CLASS == 'REAL')
dat_teste = subset(valores_tese, CLASS == 'TESTE')

nrow(dat_real); nrow(dat_teste)

mae <- function(actual,pred){
  mae <- mean(abs((actual - pred)))
  return (mae)}

#Previsão 2021
se_model <- ses(z, h = 3)
summary(se_model)
#Previsão 2021

df_fc = as.data.frame(se_model)
dat_teste$simplexp = df_fc$`Point Forecast`
mae(dat_teste$Valor, dat_teste$simplexp)

#####
##### MSE #####
#####

library(readr)
library(ggplot2)
library(forecast)
library(fpp2)
library(TTR)
library(dplyr)

dat_real = subset(valores_tese, CLASS == 'REAL')
dat_teste = subset(valores_tese, CLASS == 'TESTE')

nrow(dat_real); nrow(dat_teste)

mse <- function(actual,pred){
```

```

mse <- mean(abs((actual - pred)^2))
return (mse)}

#Previsão 2022
se_model <- ses(y, h = 3)
summary(se_model)
#Previsão 2022

df_fc = as.data.frame(se_model)
dat_teste$simplexp = df_fc$`Point Forecast`
mse(dat_teste$Valor, dat_teste$simplexp)

```