



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO

MESTRADO EM MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO

TRABALHO DE PROJETO

**MODELO DE PREVISÃO DE COMISSÕES DE CRÉDITO PARA
PARTICULARES**

MIGUEL ÂNGELO LOPES FERREIRA

ORIENTAÇÃO:

PROFESSOR DOUTOR JORGE CAIADO

**DOCUMENTO ESPECIALMENTE ELABORADO PARA OBTENÇÃO DE GRAU
DE MESTRE**

OUTUBRO - 2023

Glossário

ARIMA – *Autoregressive Integrated Moving Average*

ARIMAX – *Autoregressive Integrated Moving Average Exogenous*

VAR – *Vector Autoregressive*

EAM – Erro Absoluto Médio

EPAM – Erro Percentual Absoluto Médio

EQM – Erro Quadrático Médio

REQM – Raiz Erro Quadrático Médio

Com_CH – Comissões de Crédito Habitação

Com_CP - Comissões de Crédito Pessoal

Com_Auto - Comissões de Crédito Automóvel

Com_CC – Comissões de Cartões de Crédito

Com_DL – Comissões de Descobertos e Limites

Com_CH_Var – Comissões de Crédito Habitação em taxa de variação mensal

Com_CP_Var - Comissões de Crédito Pessoal em taxa de variação mensal

Com_Auto_Var - Comissões de Crédito Automóvel em taxa de variação mensal

Com_CC_Var – Comissões de Cartões de Crédito em taxa de variação mensal

Com_DL_Var – Comissões de Descobertos e Limites em taxa de variação mensal

FAC – Função de autocorrelação

FACP – Função de autocorrelação parcial

Resumo

Dado que as comissões bancárias são uma das principais fontes de rentabilidade de qualquer instituição financeira em Portugal, a tentativa de previsão destas componentes acaba por se tornar fulcral no desenvolvimento de campanhas de promoção a certos produtos de crédito. Tendo em conta a estrutura da carteira de créditos concedidos da Caixa Geral de Depósitos, uma vez que é a instituição financeira com mais crédito concedido a particulares em especial o crédito habitação, o estudo da evolução das comissões de créditos concedidos a particulares acaba por ser um trabalho de enorme relevância para o sistema financeiro português, dada a grande representatividade da amostra em questão.

Inicialmente neste trabalho, são realizados testes de estacionariedade das variáveis a serem estudadas, bem como testes de cointegração e de causalidade sendo estes requisitos para a aplicação dos modelos econométricos de séries temporais em estudo.

Depois de preparadas as variáveis, iniciou-se a modelação das comissões usando os métodos de Alisamento Exponencial e os modelos ARIMA, ARIMAX, VAR e VEC e foram comparados os erros de previsão de cada um deles para assim concluir qual o modelo que melhor se ajusta à comissão em questão. Adicionalmente foram realizados testes de avaliação do diagnóstico a cada um dos modelos estimados como ausência de autocorrelação dos resíduos, normalidade dos resíduos, multicolineariedade e significância dos parâmetros. Tendo em conta os resultados obtidos foram escolhidos os modelos que apresentavam os melhores resultados de indicadores de performance e que passavam na maioria dos testes realizados.

Os resultados obtidos apontaram para um melhor ajustamento da maioria das comissões nos modelos ARIMA e ARIMAX, no entanto, existiram duas comissões onde os métodos de Alisamento Exponencial e o modelo VAR foram os que apresentaram melhores resultados. Por fim, foi realizada a previsão *out-sample* que verificou um crescimento mensal em algumas das comissões estudadas e a manutenção da média dos valores registados nos dois anos seguintes à amostra disponível.

Palavras-Chave: Comissões de crédito, ARIMA, ARIMAX, VAR, VEC, Alisamento Exponencial.

Abstract

Given that bank commissions are one of the main sources of profitability for any financial institution in Portugal, trying to predict these components becomes crucial when developing campaigns to promote certain credit products. Bearing in mind the structure of Caixa Geral de Depósitos' loan portfolio, since it is the financial institution with the largest number of loans to private individuals, especially housing loans, studying the evolution of commissions on loans to private individuals turns out to be an extremely important study for the Portuguese financial system, given the large representativeness of the sample in question.

Initially in this work, stationarity tests are carried out on the variables to be studied, as well as cointegration and causality tests, which are requirements for the application of the time series econometric models under study.

Once the variables had been prepared, the commissions were modelled using the Exponential Smoothing methods and the ARIMA, ARIMAX, VAR and VEC models, and the forecast errors of each were compared to conclude which model best fits the commission in question. In addition, diagnostic tests were carried out on each of the estimated models, such as the absence of autocorrelation of the residuals, normality of the residuals, multicollinearity, and significance of the parameters. Considering the results obtained, the models with the best performance indicator results and which passed most of the tests were chosen.

The results showed that the ARIMA and ARIMAX models were the best fit for most of the commissions, but there were two commissions where the Exponential Smoothing method and the VAR model were the best fit. Finally, an out-sample forecast was carried out which found monthly growth in some of the commissions studied and the average values recorded in the two years following the sample available were maintained.

Keywords: Credit commissions, ARIMA, ARIMAX, VAR, VEC, Exponential Smoothing.

Índice

Glossário	iii
Resumo.....	iv
Abstract	v
Índice.....	vi
Lista de Gráficos	viii
Lista de Tabelas.....	viii
Agradecimentos.....	ix
1 Introdução	1
2 Breve Revisão de Literatura.....	3
3 Análise Descritiva	6
4 Metodologia	7
4.1. Modelos Univariados	7
4.1.1. Métodos de Alisamento Exponencial.....	8
4.1.2. <i>Autoregressive Integrated Moving Average (ARIMA)</i>	10
4.2. Modelos Multivariados	11
4.2.1. <i>Autoregressive Integrated Moving Average Exogenous (ARIMAX)</i>	11
4.2.2. Modelo de Vetor Autorregressivo (VAR).....	12
4.2.3. <i>Vector Error Correction</i>	12
4.2.4. Variáveis explicativas	13
5 Testes a realizar.....	14
5.1. Testes a modelos univariados.....	14
5.1.1. Teste <i>Dickey-Fuller</i> Aumentado (ADF).....	14
5.1.2. Significância dos parâmetros	15

5.1.3.	Teste de <i>Ljung-Box</i>	15
5.1.4.	Teste Normalidade dos Resíduos <i>Shapiro-Wilk Test</i>	16
5.2.	Testes a modelos multivariados	17
5.2.1.	Teste de multicolineariedade (VIF)	17
5.2.2.	Teste de Cointegração	18
5.2.3.	Teste de Causalidade de <i>Granger</i>	18
5.3.	Indicadores de Seleção de Modelos	19
5.3.1.	Critério de <i>Akaike</i> (AIC)	19
5.3.2.	Erro Quadrático Médio.....	20
5.3.3.	Erro Absoluto Médio.....	20
5.3.4.	Erro Percentual Absoluto Médio.....	20
6	Resultados	21
6.1.	Estacionariedade.....	21
6.2.	Modelos Univariados	22
6.2.1.	Alisamento Exponencial	22
6.2.2.	ARIMA	23
6.3.	Modelos Multivariados	24
6.3.1.	ARIMAX.....	24
6.3.2.	Modelo de Vetor Autorregressivo.....	25
6.3.3.	<i>Vector Error Correction</i>	27
6.4.	Comparação de resultados.....	28
6.5.	Previsão	29
7	Conclusão.....	31
8	Referências Bibliográficas	33
9	Anexo	35
9.1.	Evolução das comissões na Caixa Geral de Depósitos	35
9.2.	Evolução das comissões de crédito para particulares.....	36

Lista de Gráficos

Gráfico 1 - Evolução das Comissões Recebidas entre 2015 e 2022 – CGD.....	35
Gráfico 2 - Peso das Comissões de Crédito Concedidos a Particulares sobre o total de Comissões Recebidas.....	35
Gráfico 3 – Evolução das comissões de Crédito Habitação.....	36
Gráfico 4 – Evolução das comissões de Crédito Pessoal.....	36
Gráfico 5 – Evolução das comissões de Crédito Automóvel.....	37
Gráfico 6 – Evolução das comissões de Cartões de Crédito.....	37
Gráfico 7 – Evolução das comissões de Descobertos e Limites.....	38

Lista de Tabelas

Tabela 1 – Análise descritiva das comissões escolhidas	7
Tabela 2 - Resumo da Estacionariedade das comissões de crédito.....	21
Tabela 3 – Melhores modelos de alisamento exponencial escolhidos para cada comissão.....	22
Tabela 4 – Melhores modelos ARIMA escolhidos para cada comissão	23
Tabela 5 – Resultados dos testes realizados nos modelos ARIMA.....	23
Tabela 6 – Melhores modelos ARIMAX escolhidos para cada comissão	24
Tabela 7 - Resultados dos testes realizados nos modelos ARIMAX	25
Tabela 8 - Melhores modelos VAR escolhidos para cada comissão.....	26
Tabela 9 - Resultados dos testes realizados nos modelos VAR	26
Tabela 10 - Melhores modelos VEC escolhidos para cada comissão	27
Tabela 11 - Resultados dos testes realizados nos modelos VEC.....	28
Tabela 12 - Modelos finais	29

Agradecimentos

Ao professor Jorge Caiado por todo o apoio prestado, bem como pelas sugestões valiosas que se tornaram fundamentais para o desenvolvimento deste trabalho.

À equipa de Desenvolvimento de Modelos da Caixa Geral de Depósitos, pela generosa partilha de dados e pela colaboração que enriqueceu significativamente a análise realizada nesta investigação.

A todos os amigos que me apoiaram ao longo do processo de desenvolvimento e escrita deste trabalho.

Por fim, um especial agradecimento aos meus avós, pai e irmã pelo apoio incondicional ao longo de todo o percurso académico.

1 | Introdução

Atualmente, as comissões bancárias são uma das principais fontes de rentabilidade de qualquer instituição financeira ou intermediário financeiro português. Estas comissões tratam-se de valores adicionais a cobrar ao cliente devido à prestação de um serviço por parte da entidade bancária. Por exemplo, quando é concedido ao cliente um crédito habitação, o valor a pagar não irá apenas ser o valor da prestação acordada, comissões de avaliação do imóvel, comissões de preparação da documentação contratual ou até mesmo comissão de processamento da prestação paga mensalmente, são exemplos de comissões atualmente cobradas no sistema financeiro português que os clientes têm de ter em conta.

Apesar das instituições financeiras oferecerem muitos produtos em que não existe necessariamente o pagamento de um serviço como depósitos à ordem, a maioria do lucro destas instituições é proveniente da concessão de créditos. Estes créditos têm de ser ajustados ao segmento do cliente. Seja ele um particular, uma empresa inserida no segmento de pequeno negócio ou pequena e média empresa, sendo que são estas os segmentos de empresas mais predominantes em Portugal. Desta forma, as comissões cobradas no processo de concessão ou de acompanhamento do crédito, tornam-se num fator muito importante de controlar, bem como tentar prever, uma vez que representa um valor muito importante no ativo da instituição bancária.

Analisando a evolução do total de comissões bancárias recebidas ao longo de 2015 até 2022, é possível verificar uma tendência constante no seu valor, que apesar de algumas flutuações, tende sempre a rondar os 50 milhões de euros. É também possível verificar uma tendência constante entre 2015 e meados de 2021, apesar do pico em dezembro de 2017 onde o valor das comissões recebidas atingiu cerca de 62,5 milhões de euros. Posteriormente a esta data, do início de 2021 até ao final de 2022 é possível evidenciar uma tendência crescente do valor das comissões recebidas, sendo que o valor médio neste período foi de 53 milhões de euros, contrastando com 46,6 milhões de euros no restante período de observação (Anexo I).

O crédito a particulares é um dos principais produtos onde as instituições financeiras procuram ser mais competitivas e adquirir mais clientes, não só pela volumetria das receitas, mas também pela capacidade de fidelização do cliente à instituição, o que pode proporcionar mais oportunidades de negócio com o mesmo. Assim, tendo em linha de conta a importância da concessão de crédito a particulares e a importância que estes têm na receita da instituição financeira, torna-se necessário tentar prever estas componentes não só para existir um controlo do que este segmento de crédito pode gerar, mas também para auxiliar a tomada de decisão de futuras decisões de investimento estratégico.

Ao analisar os diferentes créditos concedidos pela Caixa Geral de Depósitos, foi decidido modelar as seguintes variáveis em nível e em variação mensal:

- Comissões Crédito Habitação;
- Comissões Crédito Pessoal;
- Comissões Crédito Automóvel;
- Comissões de Cartões de Crédito;
- Comissões de Descobertos e Limites.

Analisando a evolução do peso das comissões de crédito concedidos a particulares sobre o total de comissões recebidas ao longo de 2015 até 2022, é possível verificar uma tendência constante no seu peso, que apesar do seu pico em junho de 2017 atingindo aproximadamente 19%, em média o seu valor foi rondando os 14,5%.

Tendo em conta a volumetria dos produtos concedidos, este valor acaba por ser muito significativo e representativo do grande investimento que a Caixa Geral de Depósitos faz em créditos a particulares. Apesar deste grande investimento e da tendência praticamente constante entre 2015 e final de 2020, de 2021 até ao final de 2022 foi registada uma tendência de decréscimo do peso nas comissões referentes a créditos concedidos a particulares, atingindo um mínimo histórico em dezembro de 2022, onde o total de comissões recebidas de créditos concedidos a particulares representou apenas 8,71% do total de comissões recebidas pela Caixa Geral de Depósitos. De forma a prever estas componentes, no presente trabalho final de mestrado foram estudados diferentes modelos de previsão como o alisamento exponencial, ARIMA, ARIMAX, VAR e VEC bem como os seus erros de previsão. Para cada tipo de modelo e de comissão irão ser

comparados os diferentes indicadores de performance e irá ser selecionado o melhor modelo para cada comissão de crédito. Por fim, para estes modelos finais, irão ser estimados os seus valores futuros para os próximos dois anos e analisados os seus resultados.

As fontes estatísticas do valor das comissões cobradas em créditos presentes neste trabalho tiveram como origem a Caixa Geral de Depósitos. As variáveis exógenas utilizadas nos modelos multivariados tiveram como fonte o Banco de Portugal, Eurostat e o Banco Central Europeu.

Tendo em conta a grande volumetria de créditos concedidos a particulares pela Caixa Geral de Depósitos, em particular créditos habitação, o presente trabalho procura estudar de uma maneira representativa o sistema bancário português e a importância destes créditos para as instituições financeiras.

2 | Breve Revisão de Literatura

Prever componentes bancárias tais como comissões de crédito habitação, pode se tornar numa tarefa bastante complexa, uma vez que estas comissões podem ser influenciadas por imensas variáveis que são alheias às instituições financeiras e que são impossíveis de controlar. Tais como taxas de juro, rendimentos disponíveis das famílias ou até mesmo a taxa de desemprego. Ainda assim, a utilização de modelos de previsão univariados como o modelo *Autoregressive Integrated Moving Average* (ARIMA) ou o método de Alisamento Exponencial ou modelos de previsão multivariados como o modelo *Autoregressive Integrated Moving Average Exogenous* (ARIMAX), bem como as suas versões sazonais (SARIMA e SARIMAX) ou o modelo *Vector Autoregression* (VAR), podem ser capazes de prever com robustez estas componentes.

Segundo Meyler, numa tentativa de prever a evolução do Produto Interno Bruto Irlandês, os modelos univariados como o ARIMA, foram surpreendentemente mais robustos quando comparados com alternativas multivariadas como o VAR. Ao deflacionar o PIB, ou seja, retirar o efeito da inflação nesta componente, o autor conseguiu obter melhores resultados do que utilizando o modelo VAR utilizando as duas séries (Meyler et al., 1998).

Em contra partida, ao estudar a diferença entre o ARIMA e ARIMAX na previsão da procura de roupas para crianças muçulmanas, (Anggraeni et al., 2015) obteve melhores resultados no modelo multivariado quando comparado com o modelo univariado (Anggraeni et al., 2015). A comparação de indicadores de performance como o AIC, MAPE e RSME foram as principais evidências que levaram ao autor concluir que nesta série, a utilização do ARIMAX era mais adequada e precisa.

Numa tentativa de tentar prever dados do sector de Perfuração e Exploração de Petróleo e Refinarias da Bolsa Nacional e da Bolsa de Valores de Bombaim, Kumar utilizou não só o ARIMA, mas também a sua fusão com dois algoritmos de seleção de parâmetros como a *differential evolution* e a *Artificial Bee Colony*. Esta junção de métodos acabou por ser bastante benéfica para a sua previsão uma vez que os indicadores de performance mostraram sinais mais positivos para esta versão modificada do ARIMA, do que a sua versão simples (Kumar et al., 2022).

Apesar do ARIMA ser uma técnica de previsão reconhecida mundialmente pela sua boa capacidade de previsão, ao compará-la com técnicas mais avançadas de previsão pode acabar por ficar um pouco atrás do desejado. Num estudo realizado por Ayodele Adebisi, Aderemi Adewumi Charles Ayo (2014), a metodologia de previsão ARIMA, foi comparada com um modelo de redes neurais na tentativa de prever o valor da bolsa de valores de Nova Iorque (NYSE). Apesar de ambos terem obtido bons resultados de previsão, o modelo mais sofisticado de redes neurais acabou por obter melhores resultados quando comparados com as previsões do ARIMA (Adebisi et al., 2014), mostrando assim a fragilidade que estes modelos mais tradicionais podem ter quando comparados com modelos mais sofisticados.

A previsão de cotações de um índice ou de uma empresa sempre foi um dos principais desafios que os investigadores da área de previsão enfrentaram. Numa tentativa de tentar prever as ações da Infosys, SUN Farmacêutica e do Banco ICICI, um conjunto de investigadores da *Praxis Business School* comparou modelos como o alisamento exponencial *Holt Winters*, ARIMA, modelos de *machine learning* como o *Random Forest* e o MARS (Multivariate Adaptive Regression Splines), e modelos de *deep learning* como RNN e LSTM (Chatterjee et al., 2021). Ao comparar esta variada seleção de modelos, os autores concluíram que o modelo MARS, um modelo que funciona dividindo as variáveis

iniciais em várias funções passo a passo, acabou por ser o melhor modelo, quando comparado o RMSE de todos os modelos em cada uma das empresas. O que mostrou na generalidade menor performance, foi como seria de esperar o método de Alisamento Exponencial *Holt Winters*.

Mais recentemente, um conjunto de investigadores da universidade de Londres (Bellotti et al., 2021), tentou prever taxas de recuperação de *Non-Performing Loans*. Para este efeito, utilizaram variadas técnicas de regressões bem como de algoritmos de *machine learning*, concluindo que algoritmos como *Cubist*, *boosted trees* e *random forest* foram aqueles que obtiveram melhor performance quando comparados com as outras metodologias (Bellotti et al., 2021).

Ainda numa perspetiva de tentar prever componentes bancárias, (Bastos, 2010) , avaliou a capacidade preditiva de uma regressão de resposta fracionária paramétrica e de um modelo de árvore de regressão não paramétrica para prever *Loss Given Default* de empréstimos concedidos pelo Banco Comercial Português (Bastos, 2010). Comparando indicadores de performance como a raiz do erro quadrático médio e erro relativo absoluto obtidos nos dados de teste do modelo, foi possível confirmar que o modelo de árvore de regressão obteve previsões com mais exatidão, quando comparado com os outros modelos testados, tornando este modelo, o melhor para conseguir prever a componente de *Loss Given Default*.

Tendo em conta a pesada carga regulatória por parte dos bancos centrais, no controlo das *Loss Given Defaults*, as instituições financeiras acabam por juntar grandes esforços numa tentativa de tentar prever esta componente. Num outro estudo, desta vez levado a cabo por 4 investigadores da Universidade de Antuérpia (Tobback et al, 2014), tentaram prever as LGD's recorrendo a técnicas não lineares como o *Support Vector Regressions*, árvores de regressão e outros modelos que são transformações lineares do *Support Vector Regressions* (Tobback et al., 2014). Depois de estudados os modelos propostos, foi concluído que uma variante do modelo SVP combinado com uma regressão linear, apresentou melhores resultados quando avaliadas as previsões *out-of-time*. Por outro lado, o modelo de árvore de regressões não paramétricas foi aquele que mostrou melhores resultados quando avaliadas as previsões *out-of-sample*.

Outra componente que as autoridades bancárias têm especial atenção, é o balanço do banco. O controlo deste conjunto de componentes pode-se tornar muito importante no que toca a garantir a rentabilidade de um banco e por isso, deve-se tentar prever as principais fontes de rendimento para as instituições financeiras. Uma das principais fontes de lucro dos bancos são os depósitos e numa tentativa de prever a taxa de depósitos bancários, um grupo de investigadores de várias universidades do Irão comparou a capacidade preditiva de modelos como o ARIMA e de *Artificial Neural Networks* (Cheshti et al., 2014). Mais uma vez, foi concluído que apesar da boa capacidade preditiva do ARIMA, métodos mais sofisticados como *Artificial Neural Networks* acabam por ser mais precisos na sua previsão. Comparando os indicadores de performance, o modelo ARIMA obteve sempre piores resultados quer em relação ao REQM quer ao R^2 .

Por fim, num estudo também direcionado para a previsão da taxa de levantamento de dinheiro de um banco polaco, foram comparadas as capacidades preditivas do modelo ARMAX e MPL, uma variante de *Artificial Neural Networks* (Bielak et al., 2015). Ao comparar o erro percentual absoluto médio de ambos os modelos, o modelo ARMAX foi o que obteve melhores resultados para todos os grupos de testes criados pelos autores. Estes resultados vêm contrariar a tendência de que os melhores modelos são aqueles apoiados em técnicas de *machine learning*, levando assim a concluir que o tipo de série a modelar, acaba por ser um fator fundamental para a previsão da mesma.

3 | Análise Descritiva

Apesar de numa forma global, a Caixa Geral de Depósitos ter tido um crescimento do valor das comissões recebidas, quando analisadas as comissões em estudo o mesmo já não acontece.

Analisando a evolução mensal das comissões de crédito habitação (Anexo 2), foram recebidos em média entre 2015 e 2023 cerca de 3,4 milhões de euros, atingindo um valor máximo de 4,35 milhões de euros no início de 2021 e um valor mínimo de 2,3 milhões de euros em meados de 2015. Ao longo do período estudado estas comissões revelaram uma tendência bastante constante em torno da média, apesar do ligeiro decréscimo verificado no período final da amostra. Quanto às comissões de crédito pessoal, apesar de durante grande parte do período em estudo o valor rondar os 427 000€,

em meados de 2021 este valor caiu abruptamente registando um mínimo histórico de 200 000€ de comissões recebidas. Para as comissões de crédito automóvel e descobertos e limites, apesar de um aumento substancial do valor recebido de comissões de crédito automóvel, ambas as comissões registaram uma tendência decrescente ao longo do período em análise e onde se verificou a maior amplitude de valores. Para as comissões automóvel a amplitude entre o valor máximo e o valor mínimo da série ultrapassou os 150 000€ e para as comissões de descobertos e limites este indicador ultrapassou os 700 000€. Por fim, as comissões de cartões de crédito foram as que apresentaram os valores com uma tendência mais constante rondando sempre os 2.5 milhões de euros de comissões recebidas.

Quando realizado o estudo de sazonalidade nas séries estudadas, não foram encontradas evidências de sazonalidade em nenhuma das comissões de crédito concedidas a particulares.

Comissões	Média	Desvio Padrão	Mínimo	Máximo	Sazonalidade
Crédito Habitação	3 391 291,53 €	473962,51	2 316 760,19 €	4 356 095,76 €	Não
Crédito Pessoal	427 132,91 €	114382,28	201 947,86 €	748 516,52 €	Não
Crédito Automóvel	69 999,09 €	43547,65	13 346,04 €	182 867,08 €	Não
Cartões de Crédito	2 609 868,47 €	408438,42	1 775 507,11 €	3 558 964,21 €	Não
Descobertos e Limites	426 061,22 €	183470,16	176 589,24 €	900 978,48 €	Não

Tabela 1 – Análise descritiva das comissões escolhidas

4 | Metodologia

4.1. Modelos Univariados

Existem diferentes modelos estatísticos que conseguem tentar prever uma determinada variável tendo em conta os seus valores passados. Os modelos univariados são o principal exemplo onde tendo em conta os valores passados e transformações desse histórico, é possível tentar prever com alguma eficácia os valores futuros da variável.

4.1.1. Métodos de Alisamento Exponencial

O alisamento exponencial trata-se de um dos métodos de previsão com menos complexidade metodológica de se implementar numa série temporal, uma vez que é um método que utiliza médias ponderadas de observações históricas, mas vai dando progressivamente menos peso a observações mais antigas e mais peso a observações mais recentes. As versões mais usadas em aplicações práticas são o método de Alisamento Exponencial Duplo, o método de *Holt* e o método de *Holt Winters* em que cada um deles é adequado para determinadas situações, conforme se explica de seguida.

O método de Alisamento Exponencial Duplo é apropriado para séries temporais que apresentam uma tendência linear e consiste na aplicação do método de alisamento exponencial simples duas vezes (Caiado, 2022). O alisamento duplo é efetuado seguindo as duas expressões seguintes:

- $M_t = \alpha Y_t + (1 - \alpha)M_{t-1}$
- $D_t = \alpha M_t + (1 - \alpha)D_{t-1}$,

onde M_t corresponde à série de alisamento exponencial simples, D_t corresponde à série de alisamento exponencial duplo e α é a constante de alisamento e varia entre 0 e 1.

Quanto ao método de *Holt*, este alisamento é adequado a séries com tendência linear e sem movimentos de carácter sazonal, sendo assim um modelo alternativo ao método de alisamento exponencial duplo (Caiado, 2022). O método de *Holt* é efetuado seguindo as seguintes equações de atualização:

- $a(t) = \alpha Y_t + (1 - \alpha)[a(t - 1) + b(t - 1)]$, $0 < \alpha < 1$
- $b(t) = \beta[a(t) - a(t - 1)] + (1 - \beta)b(t - 1)$, $0 < \beta < 1$,

onde $a(t)$ representa a equação de atualização do nível e $b(t)$ representa a equação de atualização da tendência ou declive da série, e α e β correspondem às respetivas constantes de alisamento.

As previsões a h passos à frente com base no método de *Holt* são obtidas com base na expressão:

$$P_{t+h} = a(t) + b(t) \times h, h = 1, 2, 3, \dots,$$

Esta função de previsão é também usada pelo método de Alisamento Exponencial Duplo, mas aqui as grandezas do nível e do declive de tendência são atualizadas com base nas expressões $a(t) = 2M_t - D_t$ e $b(t) = (M_t - D_t)\left[\frac{\alpha}{(1-\alpha)}\right]$, respetivamente.

Por fim, o método de *Holt-Winters* é apropriado para séries que apresentam tendência linear e movimentos sazonais (Caiado, 2022), acabando por ser um modelo mais completo quando comparado com os modelos anteriormente referidos. Existem duas versões deste método, aditivo e multiplicativo. O primeiro é mais ajustado para séries com tendência linear e sazonalidade com variações aditivas. Por outro lado, o método multiplicativo é mais adequado para séries com tendência multiplicativa e sazonalidade com variações multiplicativas.

Na forma aditiva do modelo, o método de *Holt-Winters* é definido pelas seguintes equações de atualização:

- $a(t) = \alpha(Y_t - S_{t-s}) + (1 - \alpha)[a(t - 1) + b(t - 1)]$, $0 < \alpha < 1$
- $b(t) = \beta[a(t) - a(t - 1)] + (1 - \beta)b(t - 1)$, $0 < \beta < 1$
- $S_t = \gamma[Y_t - a(t)] + (1 - \gamma)S_{t-s}$, $0 < \gamma < 1$

As previsões h passos à frente são obtidas com o auxílio da seguinte expressão:

$$P_{T+h} = a(T) + b(T) \times h + S_{T+h-s}, h = 1,2,3, \dots$$

Na forma multiplicativa do modelo, o método de *Holt-Winters* é definido pelas seguintes equações:

- $a(t) = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)[a(t - 1) + b(t - 1)]$, $0 < \alpha < 1$
- $b(t) = \beta[a(t) - a(t - 1)] + (1 - \beta)b(t - 1)$, $0 < \beta < 1$
- $S_t = \gamma \frac{Y_t}{a(t)} + (1 - \gamma)S_{t-s}$, $0 < \gamma < 1$,

onde $a(t)$ é o nível, $b(t)$ o declive e S_t a componente sazonal, e α , β e γ são as constantes de alisamento.

As previsões h passos à frente são obtidas da forma abaixo:

$$P_{t+h} = [a(t) + b(t) \times h] \times S_{t+h-s}, h = 1,2,3, \dots$$

4.1.2. Autoregressive Integrated Moving Average (ARIMA)

O modelo ARIMA (*Autoregressive Integrated Moving Average*) é um modelo estatístico utilizado para a previsão de séries temporais. Este modelo univariado é baseado num modelo menos sofisticado, modelos ARMA, que apenas consegue prever os valores futuros de uma determinada série caso a mesma seja estacionária. Como uma grande maioria das séries não são estacionárias, a incorporação da diferenciação no modelo acaba por tornar o mesmo mais robusto e permitir a sua aplicação numa maior variedade de séries temporais. Dito isto, os modelos ARIMA seguem a seguinte expressão genérica:

$$\phi_p(B) (1 - B)^d y_t = \theta_q(B) \varepsilon_t,$$

onde $(1 - B)^d y_t$ corresponde ao processo de tornar a série estacionária através da diferenciação da função d vezes; $\phi_p(B)$ corresponde ao polinómio autorregressivo de ordem p (estrutura autorregressiva não sazonal do modelo); $\theta_q(B)$ correspondem ao polinómio médias móveis de ordem q (estrutura médias móveis não sazonal do modelo); e ε_t corresponde a um ruído branco com média zero.

Estes modelos são geralmente expressos na forma ARIMA (p, d, q) , com os inteiros p, d e q a assumirem a interpretação já mencionada.

Para a eleição dos parâmetros que melhor se ajustam à série, costumam ser utilizados as funções de autocorrelação (FAC) e as funções de autocorrelação parcial (FACP). A primeira mede a correlação entre o valor atual da série e os seus desfasamentos e segue a seguinte expressão:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad k=1, 2, \dots, n-1, \text{ onde } n \text{ corresponde ao número de}$$

observações da amostra e k a ordem do desfasamento.

Já a função de autocorrelação parcial analisa a correlação entre as observações atuais de uma série com as observações de outro determinado momento removendo a influência de observações que se encontrem entre as duas observações em análise (Caiado, 2022). Esta função segue através de um método recursivo a seguinte expressão:

$$p_{kk} = \frac{r_k - \sum_{j=1}^{k-1} p_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} p_{k-1,j} r_j},$$

onde $p_{kk} = r_1$ (inicialização) e $p_{kj} = p_{k-1} - p_{kk}p_{k-1,k-j}, j = 1, 2, \dots, k - 1$

É através da análise conjunta destas duas funções que se irão encontrar os parâmetros ideais para os modelos autorregressivos e de médias móveis, que compõem os modelos ARIMA.

Quando a FAC denota uma queda acentuada num determinado *lag* e ao mesmo tempo, a FACP denota uma queda lenta, então o *lag* em questão na FAC constituirá a ordem do modelo MA (médias móveis). Por outro lado, no modelo AR (autorregressivo) a FAC apresenta uma queda lenta e a FACP denota uma queda mais acentuada no *lag* que será a ordem do modelo.

4.2. Modelos Multivariados

Outra forma de tentar prever o valor futuro de uma variável é através de modelos multivariados. Estes modelos são em grande parte semelhantes aos modelos univariados explicados anteriormente, no entanto, contêm uma parte em que o comportamento de uma variável exógena tenta explicar e ajudar a prever o comportamento da variável endógena.

4.2.1. *Autoregressive Integrated Moving Average Exogenous* (ARIMAX)

O modelo ARIMAX trata-se de um modelo multivariado baseado no modelo univariado ARIMA, que inclui outros preditores como variáveis exógenas. Ou seja, os valores futuros da variável dependente não só irão depender dos seus valores passados, como também dos valores presentes e passados de outras variáveis explicativas ou exógenas. Por norma, são escolhidas uma ou mais variáveis exógenas que tenham correlação com a variável dependente e que tenham algum sentido económico.

Os modelos ARIMAX com uma variável exógena seguem a seguinte expressão genérica:

$$\phi_p(B)(1 - B)^d Y_t = \varphi_r(B) X_t + \theta_q(B) \varepsilon_t,$$

onde X_t representa a variável exógena no momento t e r corresponde à ordem do polinómio que representa a variável exógena.

4.2.2. Modelo de Vetor Autorregressivo (VAR)

Os modelos de Vetor Autorregressivo são um dos principais modelos estatísticos para a análise e previsão de séries temporais multivariadas. Estes modelos são baseados no conceito de autocorrelação, que mede a relação da variável *target* com ela mesma em diferentes períodos, mas também na relação entre duas ou mais variáveis e de que forma cada uma consegue explicar e prever a outra.

Assim, este tipo de modelos são especialmente úteis para capturar relações de interdependência entre duas ou mais variáveis permitindo entender as relações dinâmicas entre múltiplas variáveis ao longo do tempo. Devido a esta capacidade de interdependência entre variáveis, os modelos VAR, ao contrário dos modelos anteriormente estudados, têm como principal particularidade a utilização de mais do que uma variável endógena o que permite explorar os efeitos mútuos e as interdependências que possam existir entre as variáveis endógenas. Existem algumas condições para este tipo de modelos, sendo a principal delas a obrigatoriedade das variáveis que integram o modelo serem estacionárias.

Os modelos VAR bivariados seguem a seguinte expressão genérica (Clarke & Granato, 2005):

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t$$

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + \dots + \gamma_p X_{t-p} + \lambda_1 Y_{t-1} + \dots + \lambda_t Y_{t-p} + \xi_t$$

4.2.3. Vector Error Correction

O modelo *Vector Error Correction* trata-se de uma especificação do modelo VAR anteriormente exposto, com a particularidade de ser um modelo para variáveis não estacionárias e cointegradas. Segundo Winarno (2021), o modelo VEC consiste num modelo VAR de ordem $p-1$ diferenciando as variáveis que integram o modelo e um termo de correção do erro derivado da já conhecida relação de cointegração das variáveis. (Winarno et al., 2021)

Dito isto, os modelos VEC seguem a seguinte expressão genérica (Winarno et al., 2021):

$$\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t,$$

onde:

- Δ : Operador de diferenciação onde $\Delta y_t = y_t - y_{t-1}$;
- y_{t-1} : Vetor da variável endógena com *lag* de 1 período;
- ε_t : Vetor dos resíduos;
- Γ_i : Matriz do coeficiente da variável endógena da i -ésima variável;
- α : Vetor de ajustamento, matriz de ordem $(k \times r)$;
- β : Vetor de cointegração (longo prazo) matriz $(k \times r)$.

4.2.4. Variáveis explicativas

Para variáveis exógenas utilizadas nos modelos multivariados, foram selecionadas um conjunto de variáveis com periodicidade mensal, que tendo em conta a tipologia dos créditos estudados, faziam sentido serem utilizadas como variáveis com um elevado poder explicativo económico. Algumas das variáveis selecionadas apesar de parecerem um pouco fora do âmbito, como a taxa de juro das obrigações do tesouro alemãs ou o valor de fecho do PSI (Bolsa de valores portuguesa), foram selecionadas uma vez que estas variáveis tinham apresentado bons resultados de previsão, em exercícios semelhantes realizados na Caixa Geral de Depósitos.

As variáveis utilizadas foram as seguintes:

- Euribor_3M – Taxa Euribor a 3 meses;
- Euribor_6M – Taxa Euribor a 6 meses;
- Euribor_12M – Taxa Euribor a 12 meses;
- Emp_Part_CH – Percentagem de empréstimos a particulares para habitação;
- MEPHab – Montante de empréstimos cedidos a particulares para habitação;
- OT_GER_10Y – Taxa de juro de obrigações do tesouro alemãs a 10 anos;
- OT_PT_10Y – Taxa de juro de obrigações do tesouro portuguesas a 10 anos;
- PSI20 – Valor de fecho do índice PSI20;
- Rac_Cred_Ven – Rácio de crédito em incumprimento e crédito total concedido;
- Tx_Juro_Emp_PT – Taxa de juro em empréstimos concedidos em Portugal;
- Tx_Juro_Emp_EURO – Taxa de juro em empréstimos concedidos na Zona Euro;
- Tx_Desemprego – Taxa de Desemprego;
- Tx_Inflacao_Homo – Taxa de Inflação Homologa.

Com o intuito de obter os melhores resultados, para estas variáveis foram calculadas as suas variações, primeiras e segundas diferenças ordinais, bem como *lags* de 6, 12 e 24 meses, de forma a conseguir retirar toda a sua capacidade explicativa.

5 | Testes a realizar

Os modelos de previsão anteriormente expostos são uma ferramenta poderosa para uma das principais formas de tentar prever o valor de certas variáveis com exatidão. No entanto, é importante avaliar a qualidade desses modelos e determinar se os mesmos são adequados tendo em conta as variáveis que estão a ser alvo de modelação. Uma maneira de fazer isso é através de testes estatísticos. Neste capítulo, irão ser expostos os principais testes estatísticos a ser aplicados aos modelos univariados e multivariados para assim garantir os melhores resultados dos mesmos.

5.1. Testes a modelos univariados

5.1.1. Teste *Dickey-Fuller* Aumentado (ADF)

O teste de *Dickey-Fuller* Aumentado (ADF) é um dos diversos testes existentes para raízes unitárias em séries temporais. Sendo uma versão aumentada do teste original *Dickey-Fuller*, o mesmo foi concebido com o objetivo de ser aplicado a modelos mais complexos de séries temporais. Especificamente, este teste é realizado de modo a verificar a estacionariedade das séries, isto é, analisa se uma série contém ou não raiz unitária.

Desta forma, a seguinte regressão é estudada:

$$y_t = \alpha y_{t-1} + \epsilon_t$$

Onde α corresponde ao coeficiente autorregressivo da série temporal.

Este teste considera as seguintes hipóteses estatísticas:

- $H_0 : \alpha = 1$ indicando que a série é não estacionária;
- $H_1 : |\alpha| < 1$ a série é estacionária.

A estatística de teste é a seguinte:

$$t_{obs} = \frac{\hat{\alpha}}{\hat{\sigma}_{\alpha}}$$

Se $t_{obs} < t_{crit}$, rejeita-se a hipótese H_0 e conclui-se que a série não contém uma raiz unitária. Ao passo que, se $t_{obs} > t_{crit}$, não se rejeita a hipótese nula (H_0), e, portanto, é possível concluir que a série original é não estacionária.

5.1.2. Significância dos parâmetros

Ao estimar os parâmetros dos modelos apresentados, podem ser retiradas conclusões sobre a ajustabilidade do valor das comissões no modelo estimado. Para confirmar se efetivamente os valores das comissões se ajustam bem ao modelo, pode ser estudada a significância dos parâmetros do modelo estimado em que caso os parâmetros não sejam significativos, o modelo não tem significância estatística.

Desta forma são efetuados testes de significância estatística com base na seguinte estatística de teste:

$$t = \frac{\widehat{\beta}_i}{\sigma_{\widehat{\beta}_i}} \cap t_{(n-m)},$$

onde:

- $\widehat{\beta}_i$ representa o valor do parâmetro estimado pelo modelo;
- $\sigma_{\widehat{\beta}_i}$ representa o desvio padrão do parâmetro estimado pelo modelo.

A regra de decisão do teste é a seguinte:

- Se $|t| > t_{(n-m)}$ a hipótese nula (H_0) é rejeitada e é possível concluir que o parâmetro do modelo é estatisticamente significativo.

5.1.3. Teste de *Ljung-Box*

O teste de *Ljung-Box* consiste em testar a nulidade conjunta das autocorrelações dos resíduos estimados do modelo, até um determinado desfasamento. Este teste é utilizado para testar a qualidade de ajustamento do modelo à série temporal em estudo.

As seguintes hipóteses são consideradas:

- H_0 : o modelo ajusta-se bem aos dados (os resíduos são próximos de um ruído branco);

- H_1 : o modelo não se ajusta bem aos dados (os resíduos não são próximos de um ruído branco).

A estatística de teste é a representada de seguida:

$$Q(m) = n(n + 2) \sum_{j=1}^m \frac{r_j^2}{n-j},$$

onde:

- R_j é a autocorrelação estimada da série no *lag* j ;
- m é o número de desfasamentos (*lags*) que está a ser testado.

5.1.4. Teste Normalidade dos Resíduos *Shapiro-Wilk Test*

Os testes de normalidade são utilizados para averiguar se a distribuição de probabilidade associada a um conjunto de dados pode ser aproximada pela distribuição Normal. O teste de *Shapiro-Wilk* é um dos testes de normalidade existentes que se destina a verificar a aderência de uma amostra à função de distribuição acumulada. Este teste é utilizado quando o número de observações é inferior a 2000. Consequentemente, este é o teste indicado para verificar a distribuição dos resíduos dos modelos estimados neste projeto.

O teste considera as seguintes hipóteses:

- H_0 : Os dados seguem uma distribuição Normal;
- H_1 : Os dados não seguem uma distribuição Normal.

A estatística de teste é definida por:

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}$$

Onde $x_{(i)}$ são os valores da amostra ordenados ($x_{(1)}$ é o menor), e a constante b é determinada da seguinte forma:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} \times (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é par} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} \times (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é ímpar} \end{cases}$$

em que a_{n-1+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição normal, e a_i são coeficientes presentes na tabela *Shapiro-Wilk*.

5.2. Testes a modelos multivariados

Para os modelos multivariados, uma vez que têm um nível de complexidade superior aos modelos univariados, para além dos testes apresentados anteriormente, também é necessário garantir que a informação adicional a entrar no modelo, não entre em conflito com a outra respeitando assim os princípios de ausência de multicolineariedade.

5.2.1. Teste de multicolineariedade (VIF)

A multicolineariedade ocorre quando numa regressão duas ou mais variáveis independentes estão correlacionadas entre si. Em particular, é um problema no ajuste do modelo que pode causar impactos na estimativa dos parâmetros uma vez que não é possível distinguir entre os efeitos individuais das variáveis independentes sobre a variável dependente.

Um dos testes que pode detetar a existência de multicolineariedade é o *Variance Inflation Factor* (VIF).

Os elementos da diagonal principal são $(X'X)^{-1}$ também úteis para detetar multicolinearidade. O j -ésimo elemento da diagonal principal $(X'X)^{-1}$, C_{jj} pode ser escrito como:

$$C_{jj} = (1 - R_j^2)^{-1}, j = 1, \dots, p$$

Onde:

- R_j^2 é o coeficiente de determinação da regressão de X_j sobre as outras variáveis explicativas;
- C_{jj} é o fator de inflação da variância e outra notação usada é VIF_j . Assim, o VIF_j é dado por:

$$VIF_j = \frac{1}{1 - R_j^2}$$

O VIF_j mede o quanto a variância do coeficiente $\hat{\beta}_j$ é inflacionada pela sua colinearidade.

Se R_j^2 for próximo de 1, significa que existe uma alta correlação entre a variável X_j e as restantes, então $1 - R_j^2$ estará próximo de 0 (zero), e conseqüentemente, o VIF assume um elevado valor. Assim, um valor superior a 10 indica que a multicolineariedade pode estar a influenciar as estimativas de mínimos quadrados.

5.2.2. Teste de Cointegração

O teste de cointegração de *Johansen* é um teste que permite verificar se duas variáveis têm uma relação de longo prazo forte, ou seja, que apesar de por vezes se afastarem do equilíbrio, existe um conjunto de forças que funcionam de forma a manter este equilíbrio a longo prazo. Este teste em conjunto com o teste ADF irá ser utilizado de forma a perceber qual os modelos a utilizar para as comissões. No caso de existir cointegração entre as variáveis, irá ser estimado um modelo VEC. Caso não exista cointegração, deverá ser estimado um modelo VAR em níveis ou em diferenças, dependendo dos resultados obtidos nos testes de estacionariedade. Para mais detalhes, veja-se, por exemplo, Walter Enders (2014, *Applied Econometric Time Series*, 4th Edition, Wiley)

5.2.3. Teste de Causalidade de *Granger*

O teste de Causalidade de *Granger* foi introduzido em 1969 por Clive Granger (Granger, 1969) onde foi verificado que a relação entre duas variáveis, pode ser estudada através da decomposição dos mecanismos causais existentes nas variáveis.

Desta forma, ajustando-se um modelo VAR à variável de interesse Y , considerando uma variável explicativa, X , se X causa Y então os valores desfasados de X devem ser significativos para explicar (ou causar á Granger) Y , para além dos valores desfasados de Y . Considerando o seguinte modelo:

$$Y_t(t) = \sum_{i=1}^m \beta_i Y_{t-i} + \sum_{i=1}^m \gamma_i X_{t-i} + \varepsilon_t$$

São testadas as seguintes hipóteses:

- $H_0: \gamma_1 = \gamma_2 = \dots = \gamma_m = 0$;
- $H_1: \exists \gamma_i \neq 0$.

A regra de decisão é a seguinte:

- A hipótese nula (H_0) é rejeitada quando o *p-value* for inferior a 5%, portanto, conclui-se que a variável X causa à Granger Y .

5.3. Indicadores de Seleção de Modelos

Depois de escolhidos os melhores modelos através da sua prestação nos testes apresentados no capítulo anterior, também é necessário verificar como é que os dados se ajustam aos modelos testados e quais os modelos que têm o melhor desempenho para prever os valores futuros das comissões de crédito. Para tal, vamos ter em conta diferentes indicadores de seleção de modelos (AIC) e de avaliação da qualidade preditiva (Erro Quadrático Médio, Erro Absoluto Médio e o Erro Percentual Absoluto Médio).

5.3.1. Critério de *Akaike* (AIC)

O critério de informação *Akaike* (AIC) consiste em minimizar a fórmula que se segue:

$$AIC = -2 \ln(\hat{\theta}) + 2k$$

Onde:

- $\ln(\hat{\theta})$ corresponde ao logaritmo da função de máxima verossimilhança dos vetores dos parâmetros estimados;
- k é o número de parâmetros independentes.

Deste modo, é possível obter o número máximo ótimo de desfasamentos a incluir nos testes previamente referidos. O modelo com a estatística AIC menor será o que se ajusta melhor aos dados.

5.3.2. Erro Quadrático Médio

O Erro Quadrático Médio (EQM) consiste em calcular a média dos erros quadráticos, ou seja, a média quadrática da diferença entre os valores reais e os valores previstos da série temporal em estudo.

A fórmula a ser aplicada é a seguinte:

$$EQM = \frac{\sum_{t=1}^n e_t^2}{n}$$

Onde:

- e_t representa a diferença entre o valor observado e a sua previsão.
- n representa o número de observações.

Devido à grandeza dos valores das comissões a ser estudadas, muitas vezes o valor do EQM acaba por tomar valores tão grandes acabando por dificultar a sua interpretação. Para colmatar esta limitação podem ser usadas as medidas do Erro Absoluto Médio (EAM) e do Erro Percentual Absoluto Médio (EPAM) a seguir apresentadas.

5.3.3. Erro Absoluto Médio

O Erro Absoluto Médio (EAM) consiste em calcular a média das diferenças em valor absoluto entre os valores observados e as previsões.

A fórmula a ser aplicada é a seguinte:

$$EAM = \frac{1}{n} \sum_{t=1}^n |e_t|$$

Onde:

- e_t representa a diferença entre o valor observado e a previsão;
- n representa o número de observações.

5.3.4. Erro Percentual Absoluto Médio

O Erro Percentual Absoluto Médio (EPAM) consiste em calcular a média da percentagem de erro absoluto entre os valores observados e as suas previsões.

A fórmula a ser aplicada é a seguinte:

$$EPAM = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{Y_t} \times 100$$

Onde:

- e_t representa a diferença entre o valor observado e a previsão;
- Y_t representa o valor observado;
- n representa o número de observações.

6 | Resultados

6.1. Estacionariedade

Retirando apenas o modelo de alisamento exponencial, todos os outros modelos estudados teriam de garantir que a variável *target* necessita de ser estacionária, ou seja, não pode conter uma raiz unitária. Foi utilizado o teste de Dickey-Fuller Aumentado para testar a estacionariedade. Ao analisar a decomposição sazonal de cada uma das comissões (Anexo 1), foi possível constatar que todas elas não apresentam uma componente sazonal significativa. Através do teste ADF, conseguimos identificar os inteiros da diferenciação simples (ou não sazonal) para os modelos ARIMA e ARIMAX.

A tabela seguinte apresenta os resultados dos testes ADF de estacionarização das comissões de crédito:

Variável	Em Nível	Primeira Diferença Ordinária
Com_CH	Não Estacionária	Estacionária
Com_CP	Não Estacionária	Estacionária
Com_Auto	Não Estacionária	Estacionária
Com_CC	Não Estacionária	Estacionária
Com_DL	Não Estacionária	Estacionária
Com_CH_Var	Estacionária	-
Com_CP_Var	Estacionária	-
Com_Auto_Var	Estacionária	-
Com_CC_Var	Estacionária	-
Com_DL_Var	Não Estacionária	Estacionária

Tabela 2 - Resumo da Estacionariedade das comissões de crédito

Ao analisar a tabela, podemos verificar que, em níveis, todas as comissões contêm uma raiz unitária. Assim, procedeu-se à diferenciação ordinal das mesmas, levando a que todas as comissões em primeiras diferenças já se encontram estacionárias. Para as comissões transformadas pela taxa de variação mensal, apenas a variável das Comissões de Descobertos e Limites é não estacionária.

6.2. Modelos Univariados

6.2.1. Alisamento Exponencial

Para os métodos de alisamento exponencial, foram utilizadas as versões dos métodos de *Holt*, *Holt-Winters* Aditivo e *Holt-Winters* Multiplicativo para prever cada uma das séries de comissões quer em níveis quer em variações mensais.

Tendo em conta os indicadores de performance e os testes a realizar apresentados anteriormente, foram escolhidos para cada uma das comissões, os modelos que minimizaram as funções habituais dos erros de previsão. Assim, foram obtidos os seguintes resultados:

Variável	Modelo	Método	EAM	EPAM	EQM	REQM
Com_CH	Holt	Aditivo	303871,148	8,465	151113865421,487	388733,669
Com_CP	Holt	Aditivo	247686,895	98,811	93367330001,086	305560,681
Com_Auto	Holt	Aditivo	38316,788	235,780	1603084381,720	40038,536
Com_CC	Holt	Aditivo	212640,177	9,581	68585994278,183	261889,279
Com_DL	Holt	Multiplicativo	60777,140	22,581	7653351832,578	87483,437
Com_CH_Var	Holt	Aditivo	0,091	99,362	0,012	0,112
Com_CP_Var	Holt	Aditivo	0,122	114,402	0,023	0,151
Com_Auto_Var	Holt	Aditivo	0,104	117,535	0,019	0,139
Com_CC_Var	Holt	Aditivo	0,078	100,417	0,010	0,101
Com_DL_Var	Holt	Aditivo	0,304	240,119	0,181	0,425

Tabela 3 – Melhores métodos de alisamento exponencial escolhidos para cada comissão

Analisando os resultados apresentados na tabela 2, é possível verificar que considerando as comissões em nível, o modelo de *Holt* aditivo é o que consegue obter melhor qualidade preditiva para as séries de comissões de crédito habitação, crédito pessoal, crédito automóvel e cartões de crédito. Para as comissões de descobertos e limites, o *Holt-Winters* multiplicativo é o que apresenta melhores resultados.

Analisando desta vez as comissões em variação, o modelo *Holt* Aditivo foi novamente o que revelou maior exatidão na previsão de todas as comissões.

6.2.2. ARIMA

Para os modelos ARIMA, foram estimados vários modelos com diferentes especificações de parâmetros, tanto para as comissões em nível como para as comissões em taxa de variação. Para a escolha dos melhores modelos, foram escolhidos aqueles que apresentavam melhores resultados nos habituais testes de avaliação do diagnóstico e que revelaram melhor qualidade preditiva. Foram assim obtidos os seguintes modelos:

Variável Target	Ordem ARIMA	EAM	EPAM	EQM	REQM	AIC
Com_CH	(1, 1, 0)	292651,746	8,156	140699349272,214	375099,119	2008,458
Com_CP	(1, 1, 1)	56667,239	15,810	7183847085,229	84757,578	1774,693
Com_Auto	(2, 1, 3)	4679,172	25,233	46589476,117	6825,648	1606,757
Com_CC	(0, 1, 1)	145551,759	6,703	31847295574,692	178458,106	2032,516
Com_DL	(1, 1, 0)	72439,851	27,538	11876034513,735	108977,220	1809,720
Com_CH_Var	(1, 0, 0)	0,083	131,490	0,011	0,105	-134,253
Com_CP_Var	(3, 0, 2)	0,150	159,447	0,045	0,212	-83,745
Com_Auto_Var	(0, 0, 4)	0,112	112,162	0,032	0,179	-10,043
Com_CC_Var	(0, 0, 1)	0,062	107,738	0,006	0,078	-72,578
Com_DL_Var	(1, 1, 1)	0,232	245,324	0,130	0,360	-18,456

Tabela 4 – Melhores modelos ARIMA escolhidos para cada comissão

Ao analisar a tabela 3, é possível encontrar os melhores modelos para cada uma das variáveis *target* sendo possível analisar não só os erros de previsão, como também a performance do modelo. Comparando os erros de previsão dos modelos ARIMA com os métodos de alisamento exponencial analisados no capítulo anterior, é possível verificar uma melhoria substancial de todos estes indicadores. Tomando as comissões de crédito automóvel como exemplo o erro percentual absoluto diminuiu de 235,78% para 25,23% refletindo assim um melhor ajuste do modelo à variável.

Variável Target	Ordem ARIMA	Teste <i>Ljung-Box</i>	Teste <i>Shapiro-Wilk</i> (Resíduos)	Significancia dos Parâmetros
Com_CH	(1, 1, 0)	Passa	Não Passa	Passa
Com_CP	(1, 1, 1)	Passa	Não Passa	Passa
Com_Auto	(2, 1, 3)	Passa	Não Passa	Passa
Com_CC	(0, 1, 1)	Passa	Não Passa	Passa
Com_DL	(1, 1, 0)	Passa	Não Passa	Passa
Com_CH_Var	(1, 0, 0)	Passa	Passa	Passa
Com_CP_Var	(3, 0, 2)	Passa	Passa	Não Passa
Com_Auto_Var	(0, 0, 4)	Passa	Passa	Não Passa
Com_CC_Var	(0, 0, 1)	Passa	Não Passa	Passa
Com_DL_Var	(1, 1, 1)	Passa	Não Passa	Passa

Tabela 5 – Resultados dos testes realizados nos modelos ARIMA

Analisando os resultados obtidos dos testes realizados a cada um dos modelos selecionados apresentados na tabela 4, é possível verificar que em todos os modelos, a ausência de autocorrelação foi garantida, bem como a significância dos parâmetros onde apenas os modelos selecionados para as comissões de crédito pessoal e automóvel em variação não passaram neste teste. Por fim, no teste *Shapiro-Wilk* que garante a independência dos erros, não foram obtidos na generalidade dos modelos bons resultados neste teste, onde apenas as comissões de crédito habitação, pessoal e automóvel passaram.

6.3. Modelos Multivariados

6.3.1. ARIMAX

Para a estimação dos modelos ARIMAX, foram estimados modelos com diferentes combinações de variáveis exógenas que apresentavam valores de correlação elevados com a variável *target*. Depois de estimados todos esses modelos com os diferentes parâmetros que compõem os modelos ARIMAX, foram escolhidos os modelos que apresentam melhores indicadores de performance e que passam no maior número de testes. Foram obtidos os seguintes resultados:

Desta vez, utilizando variáveis exógenas nos modelos, foram obtidos melhores resultados quando comparados com o modelo univariado ARIMA. Para as comissões em nível de crédito habitação, crédito pessoal e descobertos e limites em nível e variação, o modelo que melhor se ajustou nos três, foi um modelo com duas variáveis exógenas. Para as restantes comissões, integrar no modelo apenas uma variável exógena foi a forma de obter os melhores resultados, quer analisando o erro da previsão quer analisando o AIC.

Variável Target	Ordem ARIMA	Variável Exógena	EAM	EPAM	EQM	REQM	AIC
Com_CH	(3, 1, 2)	MEPHab_Var, Tx_Inflacao_Homo_Lag_6	292283,869	8,118	119379180644,695	345512,924	1840,488
Com_CP	(2, 1, 3)	MEPHab_Var, 2dif_Tx_Inflacao_Homo	62196,910	16,429	7716449211,229	87843,322	1716,660
Com_Auto	(1, 1, 0)	2dif_MEPHab	6328,310	35,801	67598294,564	8221,818	1559,776
Com_CC	(2, 1, 4)	Tx_Inflacao_Homo_Lag_6	174955,477	8,035	54770310292,283	234030,576	1871,945
Com_DL	(0, 1, 1)	MEPHab_Var, 2dif_OT_GER_10Y	81716,673	32,033	11504162023,053	107257,457	1748,366
Com_CH_Var	(0, 0, 1)	MEPHab_Var	0,062	84,978	0,007	0,086	-149,776
Com_CP_Var	(1, 0, 3)	MEPHab_Var	0,127	128,903	0,031	0,176	-92,602
Com_Auto_Var	(0, 0, 4)	2dif_Tx_Inflacao_Homo	0,114	111,041	0,026	0,160	-6,958
Com_CC_Var	(2, 0, 0)	2dif_Euribor_3M	0,111	180,514	0,024	0,156	-84,675
Com_DL_Var	(0, 1, 2)	2dif_OT_GER_10Y, 2dif_Tx_Desemprego	0,258	199,228	0,165	0,406	-15,847

Tabela 6 – Melhores modelos ARIMAX escolhidos para cada comissão

Analisando as variáveis exógenas escolhidas, é possível verificar que para as comissões de crédito habitação, os montantes de empréstimos concedidos a particulares e a taxa de inflação têm uma grande capacidade explicativa da variável. De realçar também variáveis como a taxa de desemprego, Euribor a 3 meses e a taxa de juro das obrigações do tesouro alemãs que revelaram um poder muito explicativo das comissões estudadas.

Variável Target	Ordem ARIMA	Variável Exógena	Teste <i>Ljung-Box</i>	Teste <i>Shapiro-Wilk</i> (Resíduos)	Teste VIF	Significância dos Parâmetros
Com_CH	(3, 1, 2)	MEPHab_Var, Tx_Inflacao_Homo_Lag_6	Passa	Não Passa	Passa	Passa
Com_CP	(2, 1, 3)	MEPHab_Var, 2dif_Tx_Inflacao_Homo	Passa	Não Passa	Passa	Passa
Com_Auto	(1, 1, 0)	2dif_MEPHab	Passa	Não Passa	N/A	Passa
Com_CC	(2, 1, 4)	Tx_Inflacao_Homo_Lag_6	Passa	Não Passa	N/A	Passa
Com_DL	(0, 1, 1)	MEPHab_Var, 2dif_OT_GER_10Y	Passa	Não Passa	Passa	Passa
Com_CH_Var	(0, 0, 1)	MEPHab_Var	Passa	Passa	N/A	Passa
Com_CP_Var	(1, 0, 3)	MEPHab_Var	Passa	Passa	N/A	Passa
Com_Auto_Var	(0, 0, 4)	2dif_Tx_Inflacao_Homo	Passa	Passa	N/A	Não Passa
Com_CC_Var	(2, 0, 0)	2dif_Euribor_3M	Passa	Passa	N/A	Passa
Com_DL_Var	(0, 1, 2)	2dif_OT_GER_10Y, 2dif_Tx_Desemprego	Passa	Passa	Passa	Não Passa

Tabela 7 - Resultados dos testes realizados nos modelos ARIMAX

Quanto aos resultados dos testes realizados nos modelos, é possível verificar que todos os modelos passam no teste de ausência de autocorrelação. Esta situação também acontece para os modelos com mais do que uma variável exógena, sendo que os quatro modelos selecionados passam no teste de ausência de multicolineariedade.

Para o teste de *Shapiro-Wilk* que mede a independência dos resíduos de cada modelo, verificou-se que apenas os modelos selecionados para as comissões em variação passam neste teste. Ou seja, todos os modelos selecionados, assim como todos os modelos testados, não passam neste teste utilizando como variável *target* as comissões em nível, o que nos permite concluir que a utilização da variação consegue de certa forma eliminar problemas de independência os resíduos.

Por fim, quanto à significância dos parâmetros do modelo, apenas dois modelos selecionados não passam nos testes de significância. São estes os modelos que utilizam como variável *target* as comissões de crédito automóvel e descobertos e limites em taxa de variação.

6.3.2. Modelo de Vetor Autorregressivo

Para os modelos VAR, à semelhança dos outros modelos anteriormente estimados, foram estudadas várias combinações de variáveis endógenas a integrar o modelo com as comissões já estacionárias, uma vez que se trata de uma das especificações

do modelo. Apenas foram testadas variáveis endógenas estacionárias, que não fossem cointegradas com a variável *target* e que passassem no teste de causalidade de *Granger*. Todas as comissões em taxa de variação, com exceção das comissões de descobertos e limites, eram cointegradas com todas as variáveis exógenas testadas e por isso não foram testadas para este modelo.

Para cada uma destas combinações, foram estimados modelos com 1 a 10 *lags*. Por entre todos estes modelos, foram escolhidos os que apresentam melhores resultados, tanto a nível dos resultados dos testes, como a nível dos indicadores de performance. Foram obtidos os seguintes resultados:

Variável Target	Lag	Variável Endógena	EAM	EPAM	EQM	REQM	AIC
1dif_Com_CH	1	2dif_Tx_Juro_Emp_EURO	333580,551	106,522	157910880732,885	397380,021	9,296
1dif_Com_CP	10	2dif_Euribor_3M	147002,568	881,990	30016261168,483	173252,016	5,343971629
1dif_Com_Auto	1	2dif_Tx_Juro_Emp_EURO	2190,577	106,462	12630587,822	3553,954	3,839311671
1dif_Com_CC	3	2dif_Euribor_6M	172415,905	101,544	45917800212,176	214284,391	9,321971053
1dif_Com_DL	2	2dif_Euribor_3M	81197,218	120,091	14913357182,784	122120,257	6,33484197
1dif_Com_DL_Var	3	Emp_Part_CH_Var	0,616856336	108,4689082	0,612793377	0,783	-5,008441803

Tabela 8 - Melhores modelos VAR escolhidos para cada comissão

Analisando os resultados é possível verificar que na generalidade dos modelos testados, os modelos de ordens 1, 2 e 3 foram os que melhor se ajustaram. A única exceção foram as comissões de crédito pessoal em que o modelo que apresentou melhores resultados foi um modelo com desfasamentos de 10 períodos. Quanto às variáveis endógenas escolhidas, apesar de terem sido seleccionadas através de um critério diferente das variáveis exógenas dos modelos ARIMAX, variáveis como taxa Euribor a 3 e 6 meses, taxa de juro de empréstimos na zona euro e a percentagem de empréstimos a particulares para habitação continuaram a ser as que melhor serviam de impulso para as comissões testadas.

Variável Target	Lag	Variável Endógena	Teste <i>Shapiro-Wilk</i> (Resíduos)	Teste Cointegração
1dif_Com_CH	1	2dif_Tx_Juro_Emp_EURO	Não Passa	Passa
1dif_Com_CP	10	2dif_Euribor_3M	Não Passa	Passa
1dif_Com_Auto	1	2dif_Tx_Juro_Emp_EURO	Não Passa	Passa
1dif_Com_CC	3	2dif_Euribor_6M	Não Passa	Passa
1dif_Com_DL	2	2dif_Euribor_3M	Não Passa	Passa
1dif_Com_DL_Var	3	Emp_Part_CH_Var	Não Passa	Passa

Tabela 9 - Resultados dos testes realizados nos modelos VAR

Quanto aos testes realizados, foi testada a independência dos resíduos em que novamente, as variáveis em níveis não passaram neste teste. Ao contrário do que acontecia no modelo ARIMAX, o modelo VAR utilizando como variável *target* as comissões de descobertos e limites, não passou no teste de independência dos resíduos.

6.3.3. Vector Error Correction

Por fim, para os modelos VEC todo o processo de seleção do melhor modelo e de testes realizados foram similares aos realizados nos modelos VAR. Desta vez, conhecendo as especificações do modelo apresentadas anteriormente, foram utilizadas variáveis endógenas não estacionárias, cointegradas com as comissões e que revelam existência de relações de causalidade à *Granger*. Foram estimados modelos de ordens 1 a 10 *lags* e também para os *ranks* de cointegração entre as variáveis endógenas significativos a 1%. Por entre todos estes modelos, foram escolhidos os que apresentam melhores resultados, tanto a nível dos resultados dos testes, como ao nível dos indicadores de performance. Foram obtidos os seguintes resultados:

Variável Target	Lag Order	Rank Cointegração	Variável Endógena	EAM	EPAM	EQM	AIC
Com_CH	9	1	Euribor_3M_Lag_24	321866,050	8,598	175350247599,988	425,3889148
Com_CP	10	1	Euribor_6M	211336,325	75,297	59901920854,073	620,1315822
Com_Auto	9	1	OT_PT_10Y_Lag_12	11221,348	67,984	143550449,257	606,6594611
Com_CC	9	1	Tx_Desemprego_Lag_24	228791,237	10,325	90738616621,295	594,7059454
Com_DL	10	1	Tx_Juro_Emp_PT_Lag_24	75587,684	28,056	11510505891,566	429,0615962
Com_DL_Var	1	1	Euribor_3M	0,290746884	235,34859	0,169176258	-968,3825373

Tabela 10 - Melhores modelos VEC escolhidos para cada comissão

Analisando os resultados obtidos é possível verificar que o *lag* que melhor se ajustou às variáveis foram os lags 9 e 10 com exceção do modelo para as comissões de descobertos e limites em variação em que apenas 1 desfasamento foi o que apresentou melhores resultados. Quanto às variáveis endógenas escolhidas, as taxas da Euribor foram as variáveis que apresentaram melhores resultados para metade das comissões testadas. Para as restantes, variáveis como taxa de juro de empréstimos concedidos em Portugal, taxa de desemprego e taxa de obrigações do tesouro portuguesas foram as melhores variáveis.

Variável Target	Lag	Rank Cointegração	Variável Endógena	Teste <i>Shapiro-Wilk</i> (Resíduos)
Com_CH	9	1	Euribor_3M_Lag_24	Não Passa
Com_CP	10	1	Euribor_6M	Não Passa
Com_Auto	9	1	OT_PT_10Y_Lag_12	Não Passa
Com_CC	9	1	Tx_Desemprego_Lag_24	Não Passa
Com_DL	10	1	Tx_Juro_Emp_PT_Lag_24	Não Passa
Com_DL_Var	1	1	Euribor_3M	Não Passa

Tabela 11 - Resultados dos testes realizados nos modelos VEC

Quanto ao teste realizado, a tendência verificada nos modelos estudados anteriormente de maior parte dos modelos não passar no teste de ausência de autocorrelação dos resíduos mantém-se, uma vez que todos os modelos VEC testados não passam no teste de *Shapiro-Wilk*.

6.4. Comparação de resultados

Analisando os resultados obtidos da previsão dos modelos estudados, foram escolhidos entre eles os modelos que apresentavam melhores resultados quer a nível do erro e dentro destes, escolher o que passava na generalidade dos testes realizados para o modelo. Assim, das dez comissões estudadas, seis delas apresentaram os melhores resultados nos modelos ARIMA, sendo elas as comissões de crédito pessoal, cartão de crédito em nível e variação. Também foi o que melhor se ajustou para as comissões de crédito automóvel e descobertos e limites ambas em variação.

De seguida, o modelo que se ajustou melhor para mais comissões foi o modelo ARIMAX, que apresentou melhores resultados para as variáveis de comissões de crédito habitação em nível e em variação e crédito pessoal em variação. Curiosamente, a variável de montante de empréstimos cedidos a particulares para habitação em variação foi a que apresentou melhores resultados nos modelos com as variáveis em variação.

O modelo de alisamento exponencial, apesar de menos complexo do que todos os outros modelos estudados, apresentou os melhores resultados para as comissões de descobertos e limites onde não só passava em todos os testes como também apresenta os melhores resultados de erro de previsão.

O modelo que melhor se ajustou para as comissões de crédito automóvel foi o modelo VAR, utilizando como variável endógena a segunda diferença da taxa de juro de empréstimos concedidos na zona euro. Este modelo não só apresentava os melhores resultados ao nível de minimização do erro, como também passou na generalidade dos testes realizados. Por fim, nenhuma das comissões testadas nos modelos VEC, apresentou melhores resultados do que em outros modelos testados anteriormente. A tabela apresentada abaixo mostra em resumo os modelos escolhidos para cada uma das comissões:

Modelo	Variável Target	Ordem ARIMA	Lag	Variável Endógena	Variável Exógena
ARIMAX	Com_CH	(3, 1, 2)	N/A	N/A	MEPHab_Var, Tx_Inflacao_Homo_Lag_6
ARIMA	Com_CP	(1, 1, 1)	N/A	N/A	N/A
VAR	1dif_Com_Auto	N/A	1	2dif_Tx_Juro_Emp_E URO	
ARIMA	Com_CC	(0, 1, 1)	N/A	N/A	N/A
Alisamento Exponencial (Holt Multiplicativo)	Com_DL	N/A	N/A	N/A	N/A
ARIMAX	Com_CH_Var	(0, 0, 1)	N/A	N/A	MEPHab_Var
ARIMAX	Com_CP_Var	(1, 0, 3)	N/A	N/A	MEPHab_Var
ARIMA	Com_Auto_Var	(0, 0, 4)	N/A	N/A	N/A
ARIMA	Com_CC_Var	(0, 0, 1)	N/A	N/A	N/A
ARIMA	Com_DL_Var	(1, 1, 1)	N/A	N/A	N/A

Tabela 12 - Modelos finais

6.5. Previsão

Para a realização da previsão para os dois anos seguintes de cada uma das previsões, uma vez que existem dois modelos multivariados entre os melhores modelos encontrados para cada uma das previsões, é necessário estimar os seguintes vinte e quatro meses para estas variáveis. Dados os bons resultados obtidos para as comissões, foi estimado para as variáveis exógenas do modelo, os modelos ARIMA que melhor se ajustavam a estas variáveis. Para a variável “MEPHab_Var” foi escolhido o modelo ARIMA (4,1,5) para utilizar nos dois modelos ARIMAX tanto para as comissões de crédito habitação como crédito pessoal em variação.

Analisando a distribuição das previsões dos modelos ARIMA todos eles apresentaram resultados que acabaram por se tornar constantes ao longo dos 24 períodos estudados. Esta situação é justificada pela pouca complexidade do modelo, onde em maior parte deles existe apenas um coeficiente autorregressivo ou de médias móveis.

Verificando-se esta situação e utilizando já valores estimados para estimar valores seguintes, acabam por tornar os valores em torno do último valor observado. Assim, o valor das comissões de crédito habitação vai rondar os 285 535€ ao longo dos dois anos em análise. Para as comissões de crédito pessoal o valor já vai rondar os 209 161€ e para cartões de crédito o valor das comissões vai rondar os 230 000€. Para as comissões em variação dos modelos ARIMA, tanto as comissões de crédito automóvel e cartões de crédito estimam se manter bastante constantes ao longo do período em análise. O mesmo não acontece com as comissões de descobertos e limites em variação que ao longo do período em análise, estima-se que apesar de um ligeiro decréscimo no início de 2023, até ao final de 2024 o valor destas comissões irá aumentar em 5% mensalmente.

Quanto às comissões de crédito automóvel em nível, o modelo que apresentou melhores resultados foi o VAR e este estima que ao longo dos dois anos em análise, vai existir uma redução mensal a rondar os 680€. Ainda assim, no início deste período estima-se que existirá um período de grande oscilação deste valor, sendo que em fevereiro de 2023 estima-se uma redução de quase 7000€ em relação ao mês anterior, apesar de ser compensado no mês seguinte com um aumento de cerca de 2500€.

Quanto ao modelo de alisamento exponencial este revela uma tendência decrescente para as comissões de descobertos e limites em nível, sendo que no final do período em análise, o modelo estima que o valor destas comissões será de cerca de 160 000€. Uma redução de 75 000€ em relação ao início do período em análise.

Por fim, para as comissões em que modelos ARIMAX foram os que apresentaram melhores resultado, comissões de crédito habitação e crédito pessoal em variação, verifica-se uma oscilação bastante mais errática quando comparado com comissões em variação como crédito automóvel e cartões de crédito. Para o caso das comissões de crédito habitação, é possível constatar uma tendência de crescimento da taxa de variação mensal ao longo do período em análise onde se verificou uma média de crescimento mensal de 1,9% destas comissões. O mesmo se verificou nas comissões de crédito pessoal em variação que ao longo do período registou o mesmo movimento ascendente da taxa de variação mensal, registando uma média de crescimento mensal de 1,12% durante o período em análise.

7 | Conclusão

Tal como referido na introdução do trabalho, a previsão de componentes como comissões de crédito podem ser um fator muito importante na gestão de uma instituição financeira. Os resultados provenientes destas previsões podem influenciar estratégias de investimento da instituição financeira, onde caso se verifique que um certo tipo de crédito concedido não só traz grandes retornos de juros, mas também a nível de comissões, incentivando o banco a potenciar as vendas deste tipo de crédito.

Tendo em conta a estrutura da carteira de crédito da Caixa Geral de Depósitos, onde segundo o relatório de contas do primeiro semestre de 2023, a Caixa Geral de Depósitos detinha 19% de todos os créditos concedidos em Portugal a particulares, tendo uma posição de liderança no crédito habitação onde 23% dos créditos habitação concedidos em Portugal são concedidos pela Caixa Geral de Depósitos. Devido a estes números, o tratamento e análise do valor das comissões de créditos concedidos a particulares acaba por ser um trabalho de enorme relevância para o sistema financeiro português, dada a grande representatividade da amostra em questão.

Numa primeira instância deste trabalho, foram realizados testes estacionariedade e de cointegração, uma vez que se tratavam de requisitos fundamentais para a aplicação dos modelos propostos, sejam eles univariados ou multivariados.

Comparando os resultados obtidos, principalmente os indicadores de performance como o erro quadrático médio ou o erro percentual absoluto médio, foi possível verificar que em mais de metade das comissões os modelos ARIMA e ARIMAX foram os que melhores previsões produziram. O modelo de alisamento exponencial, apesar de ser o menos complexo de todos os modelos estudados, foi o que apresentou melhores resultados numa das comissões estudadas. Comparando resultados de comissões de crédito automóvel, enquanto o melhor modelo de alisamento exponencial apresentava um erro percentual absoluto médio de 235,78%, no melhor modelo ARIMA a mesma variável apresentou para o mesmo indicador 25,23% realçando assim o maior ajuste do modelo à variável. O modelo VAR, apesar da sua grande complexidade, apenas apresentou os melhores resultados numa das variáveis estudadas.

Não obstante dos resultados obtidos, a facilidade de implementação de modelos como alisamento exponencial, ARIMA e ARIMAX, tornam estes modelos opções mais apetecíveis para a previsão de qualquer tipo de componentes. A elevada complexidade dos modelos VAR e VEC, principalmente o último devido ao número muito elevado de condições, testes e parâmetros do modelo como *lag* e *rank* de cointegração, tornam-se pontos negativos quando comparados com modelos com menor complexidade e capacidade de previsão semelhante.

Apesar da complexidade dos modelos multivariados, estes também oferecem pontos positivos que os modelos com melhor complexidade não oferecem. A relação de variáveis macroeconómicas com as comissões estudadas e a medição da causalidade entre elas, é uma das principais vantagens dos modelos VAR comparativamente com modelos que também integram variáveis exógenas. O teste de causalidade de *Granger* acaba por tornar a análise da interação entre duas variáveis muito mais sofisticada e com melhores resultados do que a análise simples das correlações.

A capacidade de poder integrar no modelo variáveis que têm uma relação de correlação de longo prazo sem que isso tenha efeitos nocivos nos resultados do modelo são uma das principais vantagens que os modelos VEC têm quando comparados com os restantes modelos.

Como sugestão para futuros trabalhos nesta área, podem ser explorados algoritmos de *machine learning*, como redes neurais, *Random Forests* ou *Long Short Term Memory* (LSTM), a fim de aprimorar ainda mais a capacidade de previsão em séries temporais. Esses métodos podem ter uma capacidade de previsão mais precisa, uma vez que oferecem uma abordagem flexível e poderosa, capaz de capturar padrões complexos e não lineares presentes nos dados. Esta abordagem inovadora promete ser uma extensão de grande valor para as técnicas tradicionais de modelagem de séries temporais, enriquecendo a análise e abrindo novas perspectivas para a pesquisa futura.

8 | Referências Bibliográficas

- Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, 2014, 1–7. <https://doi.org/10.1155/2014/614342>
- Anggraeni, W., Vinarti, R. A., & Kurniawati, Y. D. (2015). Performance Comparisons between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study. *Procedia Computer Science*, 72, 630–637. <https://doi.org/10.1016/j.procs.2015.12.172>
- Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking & Finance*, 34(10), 2510–2517. <https://doi.org/10.1016/j.jbankfin.2010.04.011>
- Bellotti, A., Brigo, D., Gambetti, P., & Vrins, F. (2021). Forecasting recovery rates on non-performing loans with machine learning. *International Journal of Forecasting*, 37(1), 428–444. <https://doi.org/10.1016/j.ijforecast.2020.06.009>
- Bielak, J., Burda, A., Kowerski, M., & Pancierz, K. (2015). Modelling and Forecasting Cash Withdrawals in the Bank. *Barometr Regionalny. Analizy i Prognozy*, 13(4), 165–177. <https://doi.org/10.56583/br.719>
- Chatterjee, A., Bhowmick, H., & Sen, J. (2021). Stock Price Prediction Using Time Series, Econometric, Machine Learning, and Deep Learning Models. *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, 289–296. <https://doi.org/10.1109/MysuruCon52639.2021.9641610>
- Cheshti, M., Taher Ahmadi Shadmehri, M., & Safaye Nikoo, H. (2014). Forecasting Bank Deposits Rate: Application of ARIMA and Artificial Neural Networks. *Research Journal of Applied Sciences, Engineering and Technology*, 7(3), 527–532. <https://doi.org/10.19026/rjaset.7.286>
- Clarke, H. D., & Granato, J. (2005). Time Series Analysis in Political Science. Em *Encyclopedia of Social Measurement* (pp. 829–837). Elsevier. <https://doi.org/10.1016/B0-12-369398-5/00321-2>
- Enders, W. (2015). Applied econometric time series (Fourth edition). Wiley.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424. <https://doi.org/10.2307/1912791>

Kumar, R., Kumar, P., & Kumar, Y. (2022). Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms. *International Journal of Information Technology*, 14(1), 359–373. <https://doi.org/10.1007/s41870-021-00741-8>

Meyler, A., Kenny, G., & Quinn, T. (1998). *FORECASTING IRISH INFLATION USING ARIMA MODELS*, 1-48.

Tobback, E., Martens, D., Van Gestel, T., & Baesens, B. (2014). Forecasting Loss Given Default models: Impact of account characteristics and the macroeconomic state. *Journal of the Operational Research Society*, 65(3), 376–392. <https://doi.org/10.1057/jors.2013.158>

Winarno, S., Usman, M., Warsono, Kurniasari, D., & Widiarti. (2021). Application of Vector Error Correction Model (VECM) and Impulse Response Function for Daily Stock Prices. *Journal of Physics: Conference Series*, 1751(1), 012016. <https://doi.org/10.1088/1742-6596/1751/1/012016>

9 | Anexo

9.1. Evolução das comissões na Caixa Geral de Depósitos

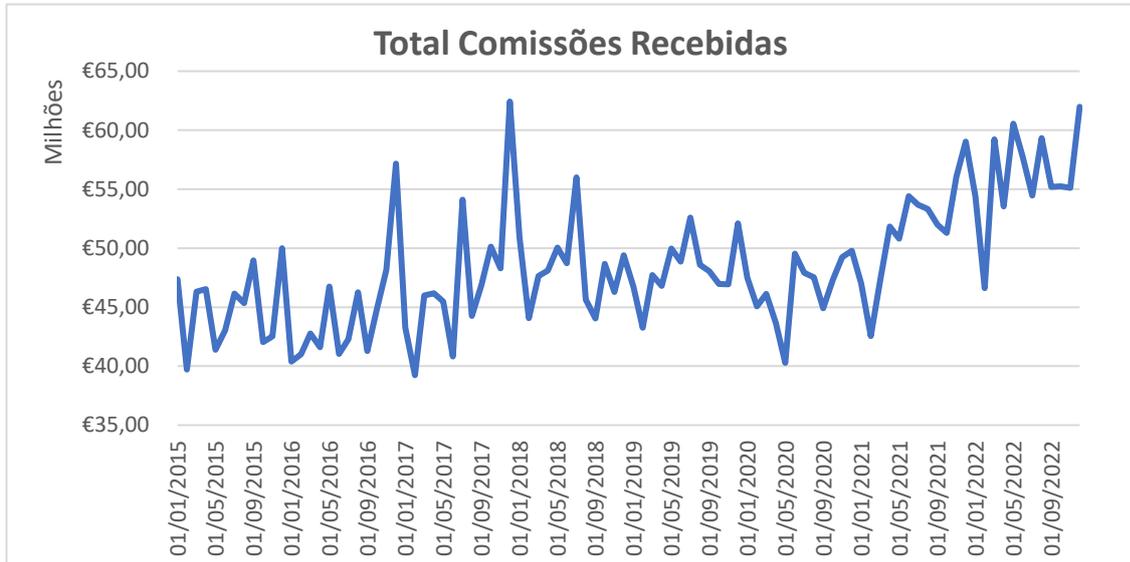


Gráfico 1 - Evolução das Comissões Recebidas entre 2015 e 2022 – CGD

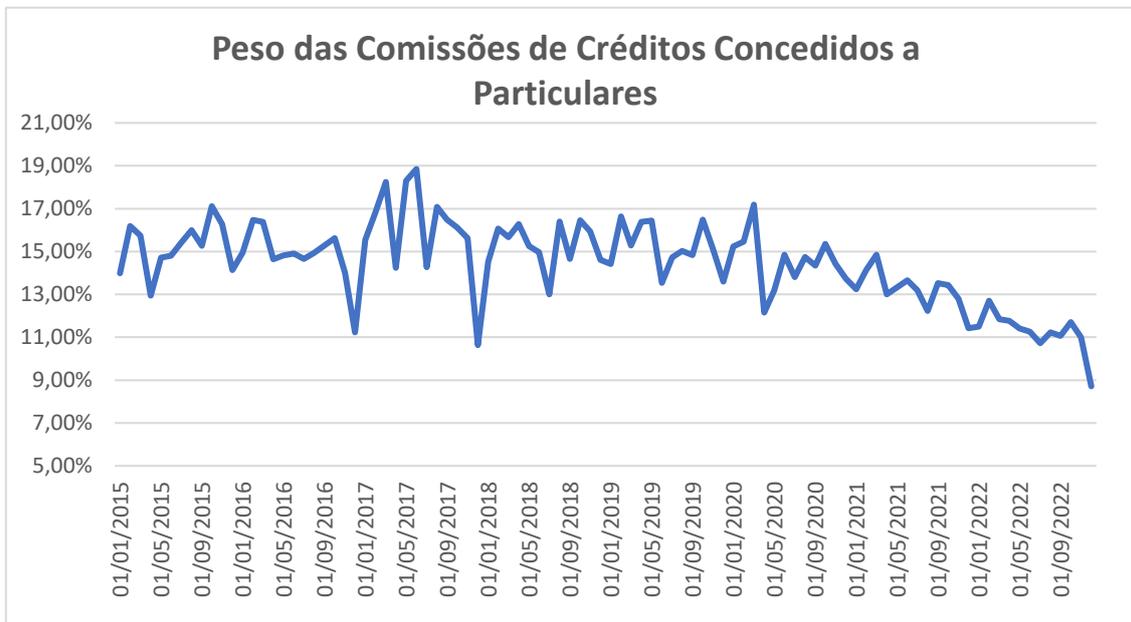


Gráfico 64 - Peso das Comissões de Crédito Concedidos a Particulares sobre o total de Comissões Recebidas

9.2. Evolução das comissões de crédito para particulares

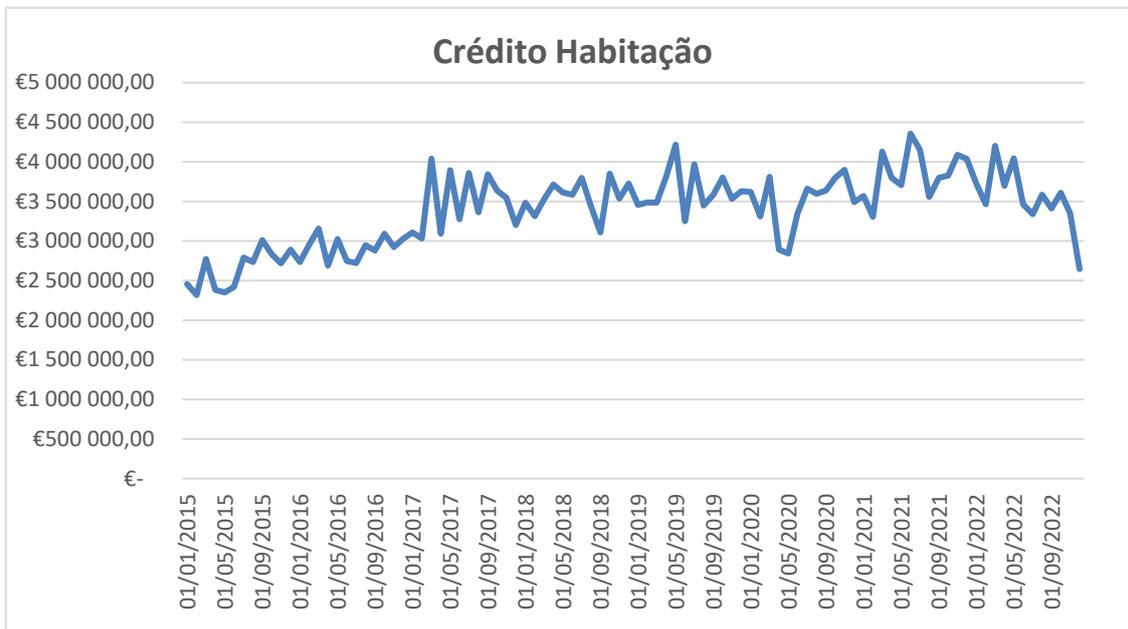


Gráfico 3 – Evolução das comissões de Crédito Habitação

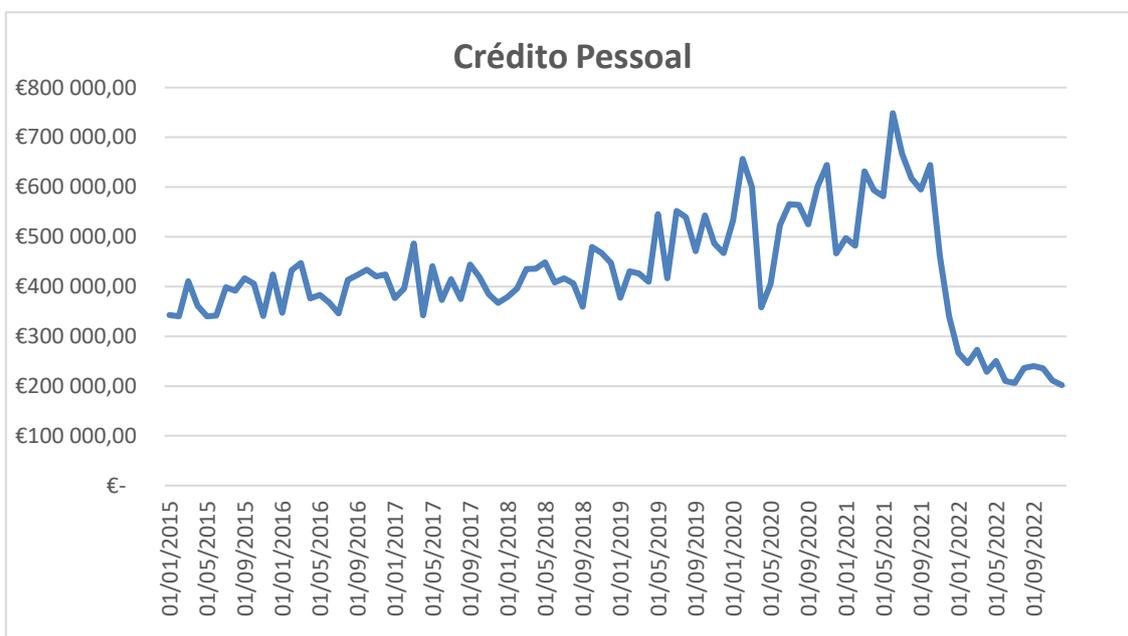


Gráfico 4 – Evolução das comissões de Crédito Pessoal

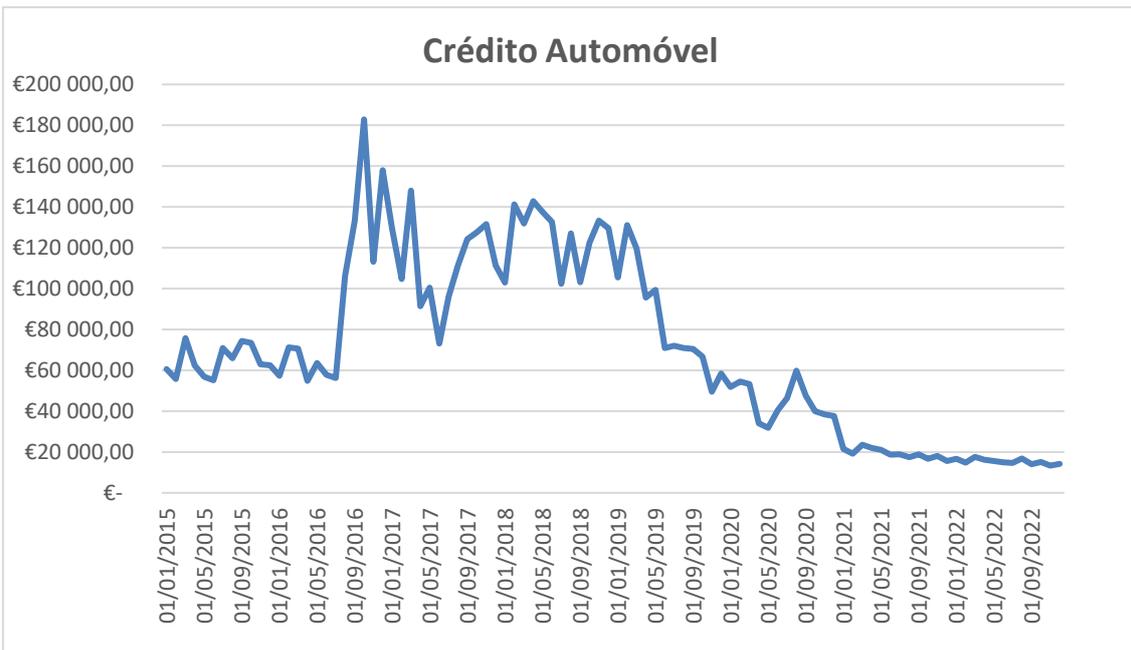


Gráfico 5 – Evolução das comissões de Crédito Automóvel

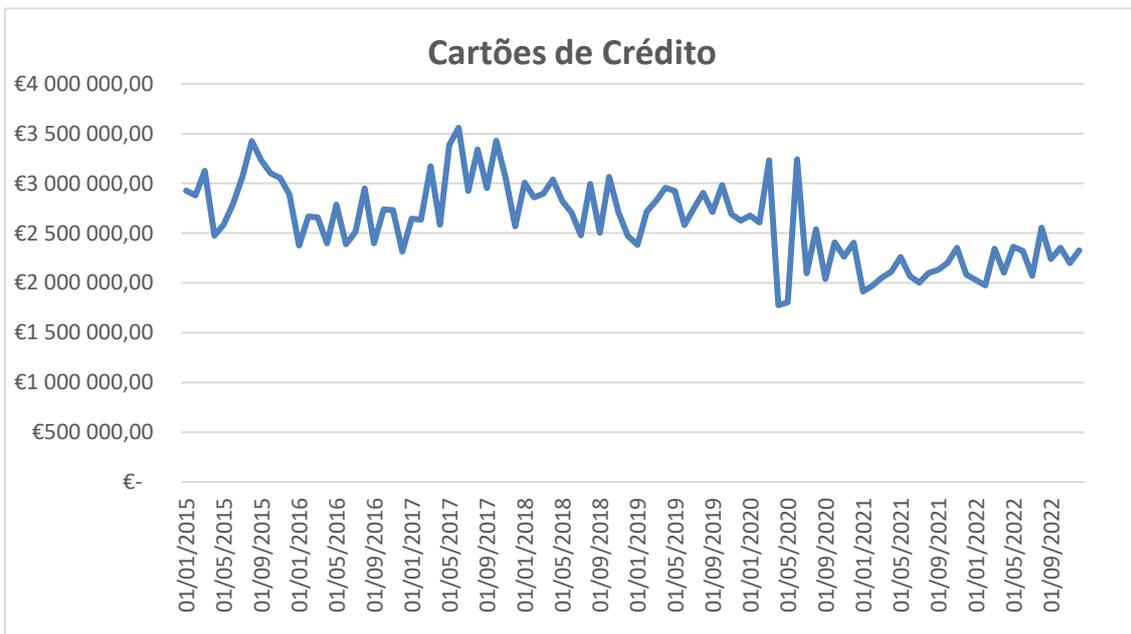


Gráfico 6 – Evolução das comissões de Cartões de Crédito



Gráfico 7 – Evolução das comissões de Descobertos e Limites