



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO

MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

ANÁLISE MULTIVARIADA NA PREVISÃO DA RENDIBILIDADE DOS CAPITAIS PRÓPRIOS

PAULO MANUEL PINTO SIMÕES MARIANO

ORIENTAÇÃO:

PROF.^a DOUTORA ALEXANDRA BUGALHO DE MOURA

Documento especialmente elaborado para obtenção do grau de mestre

Documento Definitivo

DEZEMBRO – 2022

*“A distinção entre passado,
presente e futuro é apenas
uma ilusão teimosamente
persistente.” —*

Albert Einstein

LISTA DE SIGLAS E ACRÓNIMOS

ACP	Análise de Componentes Principais
AD	Análise Discriminante
ANOVA	<i>Analysis of Variance</i> (Análise de Variância)
CAPM	<i>Capital Asset Pricing Model</i> (Modelo de Equilíbrio de Ativos Financeiros)
EBIT	<i>Earnings Before Interest and Taxes</i> (Resultado Operacional)
EBITDA	<i>Earnings Before Interest, Taxes, Depreciation and Amortization</i> (Resultados antes de Receitas e Custos de Juros, Fiscais, Amortizações e Provisões)
EMH	<i>Efficient Market Hypothesis</i> (Hipótese da Eficiência do Mercado de Capitais)
FMI	Fundo Monetário Internacional
GLM	<i>Generalized Linear Model</i> (Modelo Linear Generalizado)
LDA	<i>Linear Discriminant Analysis</i> (Análise Discriminante Linear)
NACE	<i>Statistical Classification of Economic Activities in the European Community</i> (Classificação Estatística das Atividades Económicas da Comunidade Europeia)
PC	<i>Principal Component(s)</i> (Componente(s) Principal(ais))
PEAD	<i>Post-Earnings Announcement Drift</i> (Desvio Após o Anúncio de Resultados)
P/L	Profit/Loss (Lucros ou Prejuízos)
R&D	<i>Research and Development</i> (Pesquisa e Desenvolvimento)
RL	Regressão Logística
ROA	<i>Return on Assets</i> (Rendibilidade do Ativo)
ROCE	<i>Return on Capital Employed</i> (Rendibilidade do Capital Empregado)
ROE	<i>Return on Equity</i> (Rendibilidade dos Capitais Próprios)
STOXX 600	Índice Financeiro STOXX Europe 600

RESUMO

O objetivo principal por detrás do presente estudo é aplicar a Análise de Componentes Principais, a Análise Discriminante e a Regressão Logística para prever a rentabilidade dos capitais próprios a um ano à frente, enquanto indicador de sucesso ou insucesso das empresas. Para esse efeito, são utilizados dados de painel anuais de empresas que compunham o índice STOXX Europe 600 no período 2013-2021. Como variáveis independentes são utilizados dados extraídos das demonstrações financeiras das empresas, assim como diversos rácios baseados nessas informações e nas informações de mercado. Como variável dependente é criada uma variável binária derivada da rentabilidade dos capitais próprios, constituindo um indicador da medida de sucesso de cada empresa (empresas “bem-sucedidas” ou empresas “malsucedidas”). A população em estudo compreende apenas empresas ativas do tipo *corporate* listadas publicamente que compunham parte do índice STOXX Europe 600. A amostra é selecionada com base nos dados disponíveis para todas as variáveis de interesse. Na conceção dos modelos preditivos a partir da Análise Discriminante e Regressão Logística são utilizadas como variáveis independentes os componentes principais resultantes da Análise de Componentes Principais, permitindo uma redução da dimensionalidade dos dados e ultrapassar os problemas de multicolinearidade. Os resultados apontam para uma precisão fora da amostra de 79,82% com a Análise Discriminante e 80,73% com a Regressão Logística. Este estudo evidencia que com as metodologias utilizadas é possível encontrar uma boa taxa de acerto quanto ao nível da rentabilidade de capitais próprios. Além disso, mostra que as mesmas metodologias podem ser usadas por investidores, particulares, assim como, instituições financeiras ou administradores de fundos com vista a melhorar a sua capacidade de prever o sucesso ou insucesso das empresas.

PALAVRAS-CHAVE: Análise Fundamental; Previsão; Análise de Componentes Principais; Análise Discriminante; Regressão Logística.

CÓDIGOS DO JEL: C38; G11; G12; G14; G17.

ABSTRACT

The main objective behind the present study is to apply Principal Component Analysis, Discriminant Analysis and Logistic Regression to predict the return on equity one year ahead, as an indicator of success or failure of companies. For this purpose, annual panel data from companies that made up the STOXX Europe 600 index in the period 2013-2021 are used. As independent variables, data extracted from the companies' financial statements are used, as well as several ratios based on this information and on market information. As dependent variable, a binary variable derived from the return on equity is created, constituting an indicator of the measure of success of each company (successful companies or unsuccessful companies). The study population comprises only publicly listed active corporate companies that were part of the STOXX Europe 600 index. The sample is selected based on the data available for all variables of interest. In designing predictive models from Discriminant Analysis and Logistic Regression, the principal components resulting from the Principal Component Analysis are used as independent variables, allowing a reduction in the dimensionality of the data, and overcoming multicollinearity problems. The results point out to an out-of-sample precision of 79.82% with the discriminant analysis and 80.73% with the logistic regression. This study shows that, with the methodologies used, it is possible to find a good rate of success regarding the level of return on equity. Furthermore, it shows that the same methodologies can be used by investors, individuals, as well as financial institutions or fund managers in order to improve their ability to predict the success or failure of companies.

KEYWORDS: Fundamental Analysis; Forecasting; Principal Component Analysis; Discriminant Analysis; Logistic Regression.

JEL CODES: C38; G11; G12; G14; G17.

ÍNDICE

Lista de Siglas e Acrónimos.....	i
Resumo	ii
Abstract.....	iii
Índice	iv
Índice de Figuras	v
Índice de Tabelas	vi
Agradecimentos.....	vii
1. Introdução.....	1
2. Revisão da Literatura	4
3. Metodologia de Investigação	12
3.1. População e Dados	12
3.1.1. Base de dados	12
3.1.2. Variável dependente.....	15
3.1.3. Variáveis independentes	15
3.1.4. Transformações das variáveis.....	17
3.2. Métodos estatísticos aplicados – breves considerações teóricas	17
3.2.1. ACP – Análise de Componentes Principais.....	17
3.2.2. AD – Análise Discriminante.....	18
3.2.3. RL – Regressão Logística	19
4. Estudo sobre a Previsão do ROE - Resultados da Abordagem Empírica	21
4.1. Análise de Componentes Principais	21
4.2. Análise Discriminante	26
4.3. Regressão Logística.....	36
5. Conclusões	43
Referências Bibliográficas.....	46
Anexos.....	54

ÍNDICE DE FIGURAS

FIGURA 1: Valores próprios com os dados estandardizados.....	22
FIGURA 2: Análise Paralela de Horn	22
FIGURA 3: Representação das observações no sistema de eixos dado pelos PC para as 1115 empresas, de acordo com a medida de sucesso, após a aplicação da ACP, representando, os pontos a azul as empresas “malsucedidas” e os pontos a cor-de-rosa as empresas “bem-sucedidas” – (a): PC1 e PC2 em 2D (captam 31% da variabilidade dos dados); (b) e (c): PC1, PC2 e PC3 a 3D (captam 41% da variabilidade dos dados); (d): PC3 a 2D (capta 10% da variabilidade dos dados).	24
FIGURA 4: Teste de Wilks de adequação conjunta das variáveis	30
FIGURA 5: <i>Output</i> do R para a estimação da função discriminante	31
FIGURA 6: <i>Scores</i> para a amostra de treino após aplicação da função discriminante estimada	34
FIGURA 7: <i>Scores</i> para a amostra de teste após aplicação da função discriminante estimada	34
FIGURA 8: Histograma empilhado dos valores da função discriminante para as amostras de teste dos dois grupos	35
FIGURA 9: Sucesso/insucesso, dado o <i>score</i> do PC1.....	37
FIGURA 10: Sucesso/insucesso, dado o <i>score</i> do PC3.....	37
FIGURA 11: Coeficientes estimados do modelo de RL	38
FIGURA 12: Coeficientes estimados do modelo de RL restrito e teste à significância global do modelo.....	38
FIGURA 13: Probabilidade de sucesso de uma empresa e sucesso/insucesso efetivo utilizando a amostra de treino.....	40
FIGURA 14: Probabilidade de sucesso de uma empresa e sucesso/insucesso efetivo utilizando a amostra de teste.....	40

ÍNDICE DE TABELAS

TABELA I: Importância dos PC.....	23
TABELA II: Resultados do teste- <i>t</i> para a diferença de médias entre duas amostras ..	29
TABELA III: AD – Matriz de Classificação – <i>in-sample</i>	34
TABELA IV: AD – Matriz de Classificação – <i>out-of-sample</i>	34
TABELA V: Intervalos de confiança para os parâmetros a 95%	39
TABELA VI: Intervalos de confiança para os parâmetros a 90%	39
TABELA VII: Chance de sucesso de uma empresa e respectivos intervalos de confiança	40
TABELA VIII: RL – Matriz de Classificação – <i>in-sample</i>	41
TABELA IX: RL – Matriz de Classificação – <i>out-of-sample</i>	41
TABELA X: Principais Estatísticas Descritivas	58
TABELA XI: Variância das variáveis originais após transformações logarítmicas pertinentes.....	59
TABELA XII: Componentes Principais	60
TABELA XIII: <i>Loadings</i>	61

AGRADECIMENTOS

Quase finalizado este projeto, culminando com a ambicionada satisfação, não poderia deixar de agradecer a todos que contribuíram de forma direta ou indireta para a sua concretização. Não posso mencionar a todos, mas a todos expresso a minha gratidão.

À Doutora Alexandra Bugalho de Moura um merecido reconhecimento pelos seus conselhos, sugestões e disponibilidade; pela confiança e pelo estímulo.

Também, a todos os professores do mestrado de Métodos Quantitativos para a Decisão Económica e Empresarial, pelos seus ensinamentos, decisivos para que pudesse concluir este empreendimento.

Um particular agradecimento aos dirigentes da Autoridade Tributária, especialmente, a Sra. Diretora-Geral Dra. Helena Borges, o Sr. Subdiretor da área de Recursos Humanos Doutor Damasceno Dias e o Sr. Diretor de Finanças Dr. José Gante, que me permitiram conciliar a vida profissional com a vida académica.

Por último, desejo agradecer à minha família, em especial à minha esposa e aos meus filhos, pela paciência e sacrifícios, mas também, pelo encorajamento que sempre me souberam dar.

1. INTRODUÇÃO

A tomada de decisão tendente à escolha do momento certo para comprar ou vender títulos é geralmente determinada em função do resultado de uma análise técnica ou fundamental. Ambas estão focadas na previsão de tendências futuras do preço dos títulos. A análise técnica debruça-se, especialmente, sobre dados temporais, nomeadamente, o preço dos ativos formados em bolsa. Em contraposição, a análise fundamental parte de dados seccionais ou em painel, no sentido de se perceber a subavaliação ou sobreavaliação de títulos, através da determinação do seu valor intrínseco. Tais dados são extraídos das demonstrações financeiras das empresas, regularmente publicadas. O presente estudo procura identificar empresas que possam vir a apresentar rendibilidades dos capitais próprios acima da mediana, em última instância, dotar os investidores de informação que lhes permita, por um lado, selecionar ações para compor as suas carteiras e que lhes possam, de forma consistente, gerar rendibilidades superiores no futuro; e, por outro lado, a identificação de ações que possam vir a comprometer tais rendibilidades.

Será que a aplicação de técnicas estatísticas multivariadas aos dados financeiros e de mercado das empresas permite prever o seu sucesso ou insucesso, avaliados segundo a rendibilidade dos capitais próprios, na sua designação anglosaxónica, *Return on Equity* (ROE)? A previsão do desempenho de uma empresa pode dotar os investidores particulares, instituições financeiras, administradores de fundos ou outros atores de mercado, de informação crítica no que concerne à tomada de decisão tendente à seleção e composição das suas carteiras de investimento. Com o presente estudo pretende-se a obtenção de uma medida única de sucesso, reduzindo a extensa panóplia de indicadores usados pelos atores de mercado, ao mesmo tempo, preservando a informação contida em cada um deles. Para o efeito, são aplicados métodos estatísticos multivariados na seleção e avaliação de ações pertencentes ao índice financeiro STOXX Europe 600 (STOXX 600)¹, numa tentativa de se conseguir fazer uma discriminação baseada no seu potencial de valor para os acionistas.

¹ STOXX Europe 600 ou STOXX 600 é um índice financeiro composto por 600 pequenas, médias e, sobretudo, grandes empresas europeias, localizadas na Alemanha, Áustria, Bélgica, Dinamarca, Espanha, Finlândia, França, Holanda, Irlanda, Itália, Luxemburgo, Noruega, Polónia, Portugal, Reino Unido, Suécia e Suíça.

São utilizados dados de painel anuais das empresas que compunham o STOXX 600 no período 2013-2021. Como variáveis independentes são utilizados dados extraídos das demonstrações financeiras das empresas, assim como diversos rácios baseados nessas informações e nas informações de mercado. Como variável dependente é criada uma variável binária derivada da rentabilidade dos capitais próprios, constituindo um indicador da medida de sucesso de cada empresa (empresas “bem-sucedidas” ou empresas “malsucedidas”). A população em estudo compreende apenas empresas ativas do tipo *corporate* listadas publicamente que compunham parte do STOXX 600. A amostra é selecionada com base nos dados disponíveis para todas as variáveis de interesse.

A partir da base de dados estruturada efetua-se primeiro uma Análise de Componentes Principais (ACP), extraindo as principais conclusões que a mesma permite, bem como as vantagens da sua aplicação. Posteriormente, é conduzido um estudo sobre previsão do sucesso das empresas a um ano à frente, baseada no valor do ROE. Para o efeito são utilizadas duas técnicas estatísticas, nomeadamente, Análise Discriminante (AD) e Regressão Logística (RL), explorando os seus objetivos e discutindo os seus resultados. Finalmente, procura-se explorar a técnica de análise de dados mais adequada para o objetivo da previsão e a sua aplicação e discussão. Para as análises e cálculos efetuados utilizou-se o *software R*² (Versão 1.1.456 do *RStudio*). Na conceção dos modelos preditivos a partir da AD e RL são utilizadas como variáveis independentes os componentes principais (PC) resultantes da ACP, permitindo quer uma redução da dimensionalidade dos dados, quer ultrapassar os problemas de multicolinearidade.

Os resultados apontam para uma precisão fora da amostra de 79,82% com a AD e 80,73% com a RL. Este estudo evidencia que com as metodologias utilizadas é possível encontrar uma boa taxa de acerto quanto ao nível da rentabilidade de capitais próprios.

No Capítulo 2 é apresentada a revisão da literatura de estudos empíricos. A seguir, no Capítulo 3, é elaborada a descrição da base de dados e tecidas algumas considerações

² *R* é um *software* livre que pode ser utilizado para a análise de dados. Trata-se de uma linguagem de programação orientada a objetos, compreendendo a manipulação, análise e visualização de dados.

teóricas sobre a metodologia utilizada. No Capítulo 4, procede-se à análise empírica. Termina-se com o Capítulo 5, onde são apresentados os principais resultados, limitações e contribuições deste estudo.

2. REVISÃO DA LITERATURA

A análise quantitativa e previsão do comportamento dos valores dos títulos das empresas é atualmente um instrumento fundamental no suporte à gestão e tomada de decisão. Essa análise permite dotar os investidores de informação essencial, como por exemplo, a identificação de ações que possam comprometer a rentabilidade das suas carteiras.

Conforme [1], a análise fundamental postula que no curto prazo os mercados financeiros precificam incorretamente o preço de um ativo. Pelo contrário, no longo prazo o “preço correto” surgirá, resultando um equilíbrio de longo prazo. Isto acontece porque os investidores procuram a cada momento obter ganhos através da negociação de títulos incorretamente precificados. Ou seja, a análise fundamental permite aos investidores fazerem as suas próprias avaliações do valor das empresas e ignorarem o mercado no curto prazo. Por outro lado, na análise técnica é assumido que o preço de mercado de um ativo já incorpora todas as informações de mercado, sendo esse preço e o seu histórico a única informação relevante para a avaliação de um investimento. Neste caso, assume a existência de *trends*, cujo comportamento pode ser explicado na sequência das alterações de sentimento dos investidores. De acordo com [2], a especulação baseia-se no *momentum*, enquanto a análise fundamental baseia-se na ideia de que os preços tendem a gravitar para os fundamentos e o investidor, a eles ancorado, tem uma melhor perspectiva de longo prazo. Embora baseadas em diferentes paradigmas, ambas as técnicas são muitas vezes usadas em conjunto pelos investidores. Deste modo, os objetivos da análise fundamental consistem no desenvolvimento de modelos apoiados nos dados económicos e financeiros das empresas, com capacidade preditiva sobre o nível dos lucros futuros das empresas, encerrando em si a possibilidade de identificar títulos com potencial de valorização acrescido.

No contexto proposto, [3] refere-se à teoria do valor (*valuation theory*), segundo a qual o valor de uma empresa é o valor presente dos *cash-flows* que a empresa espera gerar no futuro. No sentido de estimar esses *cash-flows*, é necessário estimar os resultados futuros. Alinhado com esta ideia, é referido que ao longo do tempo os resultados são convertidos em *cash-flows* para investidores, credores e empresa. Tais *cash-flows* são fundamentais na estimação do valor intrínseco de uma empresa, refletido no preço das

ações. Segundo [2], a análise fundamental (ou análise de valor) consiste num conjunto de métodos que permite determinar o valor de um investimento. Nesse enquadramento, o valor intrínseco é o valor de um investimento que se justifica pelas informações sobre o seu retorno. Sendo que a análise fundamental permite a determinação do valor intrínseco de um ativo, a comparação deste com o seu valor de mercado irá identificar oportunidades de compra ou venda, tendo em vista a maximização da rendibilidade de uma carteira para níveis idênticos de risco.

No âmbito da problemática da subavaliação ou sobreavaliação dos títulos, têm vindo a ser desenvolvidos diversos trabalhos de natureza empírica. Uma das dimensões mais aprofundadas sobre este fenómeno é descrito em [4], onde é utilizado o conceito *post - earnings announcement drift* (PEAD). O PEAD refere-se a um atraso na resposta ao anúncio de resultados, ou seja, à tendência de serem observados retornos anormais de uma ação na direção de uma “surpresa de lucros” por várias semanas, após o seu anúncio. Em [5] o PEAD é referido como uma forma de *underreaction* às informações contabilísticas, onde se incluem outras formas, por exemplo, o *momentum* ou o anúncio do lançamento de um novo produto. Estes autores referem que a principal característica do PEAD é que os investidores parecem reagir de forma insuficiente às notícias de lucros e o preço das ações de uma empresa desvia-se na direção das notícias dos resultados após o seu anúncio. Por outro lado, [6] dá conta de diversos estudos sugerindo, igualmente, uma *overreaction* a determinadas informações contabilísticas. Ou seja, o problema de desvio (*drift*) pode ser identificado tanto como uma reação insuficiente, como uma reação exagerada a eventos relevantes. Em [7] refere-se que o fenómeno de PEAD é tanto menos intenso quanto a magnitude das participações institucionais que de alguma forma indicam o grau de sofisticação dos investidores. Em dissonância com a argumentação exposta, [8] e [9] sugerem que uma componente importante do PEAD pode ser explicada pelos custos de transação. Em [10] os autores fornecem evidências de que investidores transitórios (ou seja, aqueles que negociam ativamente para maximizar lucros de curto prazo) negociam para explorar o PEAD, obtendo com a arbitragem retornos extraordinários relevantes após custos de transação. No entanto, tais investidores negociam menos agressivamente para explorar o PEAD em empresas com altos custos de transação.

Contudo, a ideia expressa no parágrafo anterior desafia a hipótese de mercado eficiente ou *efficient market hypothesis* (EMH). De acordo com a EMH, o preço de uma

ação já incorpora toda a informação relevante e a única razão para os preços se alterarem entre o momento t e o momento $t + 1$ é a chegada de notícias novas ou eventos imprevistos [11]. Ou seja, esta teoria postula que não existem títulos subvalorizados ou sobrevalorizados. Os fundamentos teóricos da EMH encontram-se alinhados com o modelo de *martingale*, ou de forma mais restrita, com o modelo de passeio aleatório (*random walk*), segundo o qual um investidor não consegue superar o mercado sem assunção de maiores níveis de risco. Conforme observa [12], as melhores previsões para o dia de amanhã correspondem aos preços observados hoje. De acordo com [3], a EMH sugere que os mercados de capitais desenvolvidos incorporam no preço das ações todas as informações públicas e privadas disponíveis sobre o desempenho operacional presente e passado da empresa. No entanto, esses autores apontam para um importante corpo de pesquisas recentes em mercados emergentes que sugerem a violação da EMH. Já em [13] é defendido que os investidores não incorporam completamente as informações relacionadas com as medidas fundamentais. Em [14] foi testada a hipótese de *random walk* para retornos semanais do mercado de ações, concluindo pela sua forte rejeição. Tal situação tem vindo a ser documentada em numerosos estudos como “anomalias de mercado” relacionadas com ineficiências de informação no mercado de capitais, constituindo o PEAD um caso particular. Em [15] identificam-se anomalias relacionadas com oito atributos das empresas – dimensão, liquidez, resultados (*post-earnings announcement drift*), variabilidade do *cash-flow*, crescimento dos ativos, *price momentum*, *price-to-book ratio* e capacidade de gerar lucros. Os autores acrescentam que tais anomalias relacionadas com ineficiências de mercado são mais expressivas em mercados de capitais menos desenvolvidos. Estas anomalias concretizam-se na incorreta valorização de mercado dos títulos. Outra literatura dá conta de anomalias de calendário como são exemplo os trabalhos de [16] e [17]. Um exemplo de tais anomalias são as apelidadas *January Effect* e *Calendar Effect*.

Em [18] é descrita a relação entre o momento de publicação dos elementos contabilísticos e o comportamento das ações em torno dessa divulgação. Alerta-se, no entanto, que muitas informações contabilísticas são refletidas nos títulos muito antes da publicação desses elementos. Contudo, de acordo com os autores, os relatórios contabilísticos contêm informações que não são fornecidas por outras fontes. Já em [19] refere-se que as demonstrações financeiras anuais não são excessivamente avaliadas

como um meio oportuno, já que a maior parte do seu conteúdo (cerca de 85 a 90 %) é capturado por meios de comunicação que talvez incluam relatórios provisórios.

Em [20] sustenta-se que a tarefa central da análise fundamental deve concentrar-se na previsão dos resultados, ao invés de se centrar no retorno dos títulos. Trata-se de uma ideia recuperada de [21]. Este autor reforça o papel da contabilidade e a sua capacidade de distinguir riqueza (capacidade de gerar dividendos) e a sua criação (rendimentos e ganhos) da distribuição discricionária de riqueza na forma de dividendos. Refere que a análise fundamental envolve a descoberta do preço sem referência ao preço (entenda-se, este último, formado no mercado de capitais), sendo que na tarefa de determinação do verdadeiro valor de um título, a pesquisa do mercado de capitais é mal direcionada. Também nós direcionamos a nossa pesquisa na previsão dos resultados.

No trabalho apresentado em [22] usa-se a RL para estimar a probabilidade de incremento dos resultados no ano seguinte. Os autores referem que embora alguns trabalhos empíricos seminais indiquem que a informação contida nos resultados contabilísticos e nos seus componentes capturam informação que já está contida nos preços das ações, adotam uma perspectiva distinta alicerçada no que denominam como “análise fundamental tradicional”, sobre a qual os valores das empresas (“fundamentais”) são indicados por informações nas demonstrações financeiras. Os preços das ações tendem a desviar-se desses valores e gravitam lentamente em direção aos valores fundamentais. Deste modo, a análise das demonstrações financeiras publicadas pode permitir a descoberta de informação que não está refletida no preço das ações. Em vez de tomar os preços como referências de valor, os “valores intrínsecos” evidenciados nas demonstrações financeiras servem como referências com as quais os preços são comparados para identificar ações sobrevalorizadas e subvalorizadas. Usando como técnica de estimação um modelo LOGIT, constroem uma medida resumida de um conjunto alargado de rácios referentes à informação contida nas demonstrações financeiras que lhes permite indicar a direção das mudanças dos lucros um ano à frente. Consideram o posicionamento em ações com base nesta medida e observam os retornos associados. Os resultados alcançados indicam que a medida sumária prevê de forma robusta os retornos futuros das ações. As estratégias de negociação baseadas em previsões de lucros futuros a partir de informações de demonstrações financeiras “disponíveis ao público” capturam uma parcela significativa dos retornos com base no perfeito conhecimento

prévio desses lucros futuros. Além disso, consideram que os retornos dessas estratégias não são explicados por aspetos das empresas que foram nomeados como atributos de risco. A nossa abordagem recupera a ideia expressa com os autores deste trabalho na medida em que se foca na previsão dos resultados um ano à frente. Assim como estes autores, também se procede à construção de uma medida sumária a partir da estimação dum modelo LOGIT. No entanto, nesta tese, consideramos como variáveis independentes os primeiros PC, no sentido da análise paralela de Horn, resultantes do conjunto de variáveis de indicadores financeiros e de mercado. Em [23] foram também encontrados resultados empíricos robustos que indicam a valorização positiva pelos investidores dos resultados contabilísticos.

Ainda em [22], os autores referem que, dada a relevância dos resultados contabilísticos futuros, é desejável identificar descritores financeiros baseados na capacidade de os prever a longo prazo. E, neste aspeto, não partem de qualquer ideia pré-concebida sobre a capacidade preditiva das variáveis independentes, sendo antes conduzidos pelo mote: “*let the data speak*”³. No entanto, limitam a sua investigação a um ano à frente, apontando que a desconsideração de informações com mais de um ano de antecedência produz um viés conservador nos resultados, fazendo com que se aproximem mais da hipótese de eficiência do mercado. A variável de ganhos do ano seguinte é especificada como um resultado binário, ou seja, um aumento ou uma diminuição de ganhos. Neste aspeto também nós seguimos de perto a ideia expressa por estes autores. Não partimos de qualquer ideia pré-concebida sobre a capacidade preditiva das variáveis independentes, antes partilhando a ideia de “deixar os dados falar”. Embora com prejuízo para o número de observações disponíveis, como mais à frente veremos no ponto 3.1.1., opta-se por utilizar um grande número de variáveis (86 variáveis contínuas e 38 variáveis *dummy*), cujo número (apenas as variáveis contínuas) é posteriormente reduzido, pela aplicação da ACP, para utilização na AD e na RL⁴.

³ Em [20], embora partilhando a abordagem de [21] e outros autores de que a tarefa central da análise fundamental deve focar-se na previsão dos resultados contabilísticos, em vez de explicar os retornos dos títulos, baseiam-se apenas em preditores propostos pelos analistas.

⁴ As variáveis *dummy* apenas serão utilizadas na RL.

O presente trabalho recupera a linha de pesquisa de [24], onde os autores referem que a experimentação empírica sugere que os modelos de classificação superam os modelos de estimativa de nível, em termos de prever a direção do movimento do mercado de ações e maximizar os retornos da negociação de investimentos. Os modelos de classificação testados, que preveem a direção com base na probabilidade, incluem LOGIT, PROBIT e rede neural probabilística. Referem que apesar da abundância de pesquisas que se concentram em estimar o nível de retorno do índice do mercado de ações, há uma carência de estudos que examinem a previsibilidade da direção/sinal do movimento do índice de ações.

A gênese da análise multivariada tem sido atribuída a Francis Galton and Karl Pearson que, no final do século XIX, procederam à quantificação da relação entre a descendência e características parentais, e ao desenvolvimento do coeficiente de correlação [25]. Na primeira metade do séc. XX a sua difusão aplicacional estendeu-se a diversas áreas científicas, para a qual foram relevantes os trabalhos de [26]–[29]. O seu emprego na área económica surgiu mais tarde com propósitos direcionados na previsão de falência das empresas. Para isso foi significativo o trabalho desenvolvido em [30] que concebeu um *ranking* baseado na combinação ponderada de 5 rácios com recurso à AD, tendo surgido, entretanto, outros trabalhos igualmente baseados na AD. (*e.g.*, [31]). No domínio das finanças comportamentais e baseados na AD, em [32] é explorada a influência de fatores demográficos na tomada de decisão dos investidores. Em [33] são avaliados 24 potenciais fatores relacionados com o *momentum*, usando vários testes de ANOVA e AD múltipla.

Reportando-nos a [22], outros autores conduziram trabalhos com recurso à RL. Mesmo antes de [22], em [34] os autores levaram a cabo um trabalho empírico baseado nas metodologias proporcionadas pela RL, no sentido de testar a hipótese de *random walk*. Os últimos argumentam que esta hipótese apenas é verificável num sentido limitado. Um modesto alargamento da informação preditiva permite rejeitá-la. Mais tarde, em [35], concretiza-se um estudo inspirado em [22], com vista à previsão dos resultados, mas também do *cash flow* a um ano à frente. Os autores concluem que os seus achados reforçam evidências anteriores, mas também fornecem mais informações sobre a utilidade das variáveis contabilísticas na previsão dos lucros e alterações do *cash flow*, e a utilidade dos lucros futuros na previsão da rendibilidade das ações. Em [36] utiliza-se a RL e vários rácios financeiros como variáveis independentes para investigar indicadores

que afetam significativamente o desempenho das ações negociadas ativamente no mercado indiano. Integrando o conceito de investimento em valor na sua pesquisa, os autores de [37] analisam fundos brasileiros que adotam essa filosofia, no sentido de identificar os fatores que afetam as decisões de manter uma ação numa carteira. Fazem notar que esses fatores são uma maior estabilidade nos ganhos por ação, alto ROA, alto *Gross Margin*, dimensão da empresa e liquidez das ações. Para o efeito utilizaram um modelo de RL. Igualmente, em [38] é conduzida uma análise empírica utilizando as técnicas da RL no sentido de estimar a probabilidade sobre o desempenho de uma empresa no mercado de capitais com base nos seus rávios. Da mesma forma, utilizando a RL, [39] identifica retornos anormais em janeiro relacionados com eventos de *stock split*. Em [40] constrói-se um modelo de RL baseado nos rávios contabilísticos para prever o desempenho das ações no mercado de capitais paquistanês. Recorrendo à mesma técnica, em [41] efetua-se uma análise idêntica com empresas do mercado de capitais nigeriano.

Os autores de [42] propuseram um método baseado na AD para ajudar os investidores e analistas na tomada de decisão na seleção de ações. Reconhecendo a simplicidade e facilidade de aplicação do CAPM (*Capital Asset Pricing Model*) formulado independentemente em [43]–[45], asseveram basear-se em pressupostos muito restritivos sobre o funcionamento do mercado. Sustentam que nos anos mais recentes, as pesquisas têm mostrado que certas variáveis fundamentais complementam ou são ainda mais importantes na explicação dos retornos das ações. Em [46] desenha-se um modelo apoiado na AD, no sentido de auxiliar os investidores na tomada de decisão tendente à seleção de ações do mercado de capitais da Bósnia e Herzegovina. Em [47] efetua-se um estudo idêntico para o mercado de capitais nigeriano.

Reconhecendo o valor das informações proporcionadas pela informação contabilística na estimação do desempenho financeiro das empresas, em [48] é conduzido um trabalho empírico baseado em técnicas estatísticas multivariadas, nomeadamente a análise fatorial, AD e RL.

Mais recentemente outros trabalhos têm apelado à utilização de redes neurais (*e.g.*, [49]), e outros, a técnicas de *machine learning*. Em [50] são utilizados conjuntos de classificadores para a previsão da rendibilidade das ações. Também aqui a RL,

conjuntamente com outras técnicas, parece não ter perdido a sua importância no auxílio à construção deste tipo de modelos.

Alicerçados nos achados empíricos atrás referidos, reconhecemos o valor proporcionado pelas informações contabilísticas como reflexo do desempenho financeiro das empresas. Por isso, os modelos aqui estimados com base na ACP, na AD e na RL são direcionados para a previsão do ROE em detrimento da eventual previsão do preço dos títulos ou da sua rentabilidade, muitas vezes influenciados pelo *momentum*. Sustentado na análise fundamental, este estudo pretende constituir uma ferramenta auxiliar aos investidores e outros atores de mercado nas decisões de investimento de longo prazo, identificando empresas de valor, ao invés de explorar os movimentos de preços de curto prazo, mais direcionados à especulação. Apoiamo-nos na ideia atrás referida de que a construção de uma medida sumária prevê de forma robusta os retornos futuros das ações [22].

3. METODOLOGIA DE INVESTIGAÇÃO

Neste capítulo fazemos primeiramente uma descrição da extração e tratamento dos dados utilizados e posteriormente introduzimos as metodologias aplicadas.

3.1. População e Dados

3.1.1. Base de dados

Os dados foram obtidos a partir da base de dados Orbis Europe, com incidência nos anos de 2013 a 2021 (último ano para o qual existem dados publicados)⁵ [51]. A população em estudo compreende apenas empresas ativas do tipo *corporate* listadas publicamente que compõem parte do STOXX 600. A pesquisa dos dados foi conduzida tendo em vista a recolha de informação referente a 48 variáveis contínuas e respetiva variação percentual anual quando disponível (38 variáveis contínuas). Adicionalmente, foram criadas 20 variáveis *dummy* para o país e 18 variáveis *dummy* para o setor de atividade, segundo a nomenclatura estatística das atividades económicas NACE Rev. 2 [52].

Posteriormente, procedeu-se ao empilhamento dos dados agrupando todos os dados anuais às variáveis correspondentes, porém foi ignorada a estrutura de dados em painel. Assim, todas as observações foram tratadas como não correlacionadas, no sentido em que se ignorou a dimensão temporal dos dados em favor da sua dimensão seccional. O objetivo da recolha de dados de múltiplos anos foi permitir a maximização do número de observações. Por outro lado, com referência a certas variáveis e devido à inexistência de dados anuais para algumas das unidades de corte transversal, verificou-se estarmos perante um painel desequilibrado. O mesmo é dizer que ocorreu a ausência de dados em pelo menos um ano para pelo menos uma empresa. Tal circunstância é mais marcante por cada variável adicional considerada, já que a inexistência de dados para uma empresa para essa variável leva à desconsideração dessa observação, ainda que existam dados para as restantes variáveis. Ainda, no que concerne à utilização de dados em painel, convém referir que se a utilização dos modelos de efeitos fixos podem potenciar a capacidade de ajuste, tendo em conta a diferenciação dos termos independentes para cada empresa,

⁵ A Orbis compreende informações de mais de 400 milhões de empresas e entidades de todo o mundo – 41 milhões com informação financeira detalhada. Assume-se como um poderoso recurso com dados comparáveis de empresas privadas.

isolando as suas características, em contrapartida, podem limitar a sua capacidade preditiva na presença de unidades de corte transversal *out-of-sample* distintas das que foram utilizadas para o ajuste do modelo.

Considerando que o STOXX 600 comporta empresas de distintas geografias europeias, nem todas as empresas são cotadas em euro. As taxas de câmbio utilizadas na Orbis são recolhidas do site do FMI. No caso em estudo os dados foram recolhidos tendo em conta a aplicação das taxas de câmbio referentes à última data de encerramento das demonstrações financeiras. A divisa de referência selecionada foi o euro.

Quanto aos dados de mercado anuais, são apresentados com referência às demonstrações financeiras de cada ano. No entanto, tendo em conta que as demonstrações financeiras são publicadas com atraso, substancialmente, até 31 de março do ano seguinte a que respeitam, todas as variáveis relacionadas com dados de mercado foram ajustadas aos valores de fecho de mercado de março do ano seguinte a que respeitam. Se por um lado optámos por uma abordagem mais atualizada tendo em conta a data de disponibilização das demonstrações financeiras, por outro lado, não deixa de ser uma abordagem de algum modo conservadora, na medida que a 31 de março os dados de mercado de algumas empresas já terão captado parte significativa da informação relativa às demonstrações financeiras do ano anterior. Por isso, atendendo a que as datas de publicação das demonstrações financeiras, embora concentradas no primeiro trimestre do ano seguinte a que respeitam, não coincidem, a 31 de março estaremos já na posse de todos os dados das empresas relativos ao ano anterior. Deste modo, posicionámo-nos no momento em que estaríamos em condições de ajustar os modelos, ou seja, a 31 de março do ano seguinte, considerando todas as observações disponíveis.

Ainda no que concerne aos dados de mercado convém referir que estes se encontram ajustados pelos rácios de *stock split*⁶. Os rácios de *stock split* podem ser definidos como fatores de ajustamento que são aplicados quando ocorre o desdobramento de ações, isto é, quando uma empresa aumenta o número das suas ações em circulação. Estas operações têm normalmente o objetivo de aumentar a liquidez dos títulos, mas não alteram o valor da empresa, porquanto, o valor total de todas as ações permanece fundamentalmente o mesmo⁷.

⁶ A operação contrária designa-se por *reverse stock split*.

⁷ Existem estudos que parecem apontar para a evidência de retornos anormais dos títulos relacionados com os anúncios de *stock split* (e.g. [61]–[64]).

Após a recolha dos dados segundo os critérios definidos, verificou-se a existência de *missing values* com a indicação de NA (*Not Available*) ou NS (*Not Significant*). No primeiro caso, por não se encontrarem disponíveis ou, quando os valores resultam da aplicação de uma fórmula para a obtenção de um rácio, por este não poder ser calculado na falta de algum ou alguns dos seus elementos. No segundo caso, embora os valores que concernem a determinado rácio possam ser calculados, os mesmos encontram-se fora de um intervalo específico ou pode não fazer sentido o seu cálculo. A metodologia adotada conduziu à eliminação de todas as observações com *missing values*.

Tendo em conta as limitações apontadas quanto aos dados disponíveis, podemos concluir que existe um *trade-off* entre a quantidade de variáveis utilizadas e o número de observações disponíveis sem os *missing values*. Isto é, tendo em conta os dados disponíveis, a consideração de uma variável adicional implica quase sempre uma perda marginal de observações.

Por outro lado, a implicação da existência de *outliers* sobre os resultados dos modelos utilizados poderá condicionar os resultados preditivos. A eventual remoção de *outliers* será mais uma fonte de perda de observações.

A amostra resultante, antes da remoção dos *missing values*, conta com 4347 observações (483 observações para cada um dos anos). Após a remoção dos *missing values* resultaram 1115 observações (129 observações para o ano de 2013, 123 observações para o ano de 2014, 152 observações para o ano de 2015, 138 observações para o ano de 2016, 146 observações para o ano de 2017, 158 observações para o ano de 2018, 139 observações para o ano de 2019 e 130 observações para o ano de 2020). Em cada um dos anos, cada observação concerne a uma empresa que poderá estar ou não representada nos restantes anos.

A amostra utilizada conta com 125 variáveis: 1 variável binária dependente, 48 variáveis contínuas independentes com dados extraídos das demonstrações financeiras das empresas e dados de mercado, 38 variáveis contínuas independentes referentes à variação percentual anual de algumas das variáveis anteriores, 18 variáveis *dummy* independentes referentes às atividades desenvolvidas pelas empresas e 20 variáveis *dummy* independentes referentes ao país onde se encontram sediadas.

3.1.2. Variável dependente

A variável dependente (Y_{11}) é uma variável binária, constituindo um indicador da medida de sucesso de cada empresa e sendo derivada do valor do ROE [*using net income* (%) (ano $n + 1$)]. Foi-lhe atribuído o valor 0 para empresas com ROE no ano $n + 1$ menor ou igual à mediana ($Me = 13,885$) e valor 1 para as restantes empresas⁸. Daqui em diante designaremos como “bem-sucedidas” as empresas em que a variável dependente assume o valor 1 ($Y_{11} = 1$); e, “malsucedidas”, caso contrário (ou seja, $Y_{11} = 0$). O ROE é um acrónimo da designação anglosaxónica *Return on Equity*, ou seja, rendibilidade dos capitais próprios, e resulta da divisão do resultado líquido pelos capitais próprios. Embora apresentando algumas limitações, por exemplo, não tem em conta o financiamento, sendo que um ROE mais elevado pode ser conseguido à custa de maior endividamento, é esperado que se encontre em linha com a valorização das ações e dividendos distribuídos. Deste modo, é expectável que empresas que possam vir a apresentar um ROE superior podem assistir consistentemente a uma maior valorização de mercado, assim como, uma maior taxa de rendibilidade dos dividendos. Uma vez que a variável dependente é binária, os métodos utilizados são a AD e a RL.

3.1.3. Variáveis independentes

Como variáveis independentes são utilizados dados extraídos das demonstrações financeiras das empresas, assim como diversos rácios baseados nessas informações e nas informações de mercado. A utilização de rácios subsume-se na capacidade de aporte de uma maior quantidade de informação, embora, como referido, limite o número de observações aos que podem ser calculados. No presente estudo foram utilizados alguns rácios de rendibilidade, operacionais, estruturais e de mercado. Adicionalmente, e dada a pertinência, foi considerada a variação anual de algumas dessas variáveis. Por fim, foram consideradas diversas variáveis *dummy* no sentido da diferenciação do país a que pertencem as empresas e das atividades económicas por estas desenvolvidas, segundo a classificação estatística das atividades económicas da Comunidade Europeia (NACE Rev. 2 main section). No Quadro I encontram-se sumariadas as variáveis independentes utilizadas no presente estudo. O desenvolvimento do Quadro I pode ser consultado no Quadro II no anexo 1.

⁸ A mediana foi calculada com as 1115 observações após a remoção dos *missing values*.

QUADRO I: Sumário das Variáveis Independentes

Variável	Tipo	Variável	Tipo
X011 Total assets	Balance Sheet	X321 Solvency ratio (Liability based)	Structure Ratio
X012 % Δ Total assets		X322 % Δ Solvency ratio (Liability based)	
X021 Operating revenue	Profit & Loss Account	X331 Gearing	Structure Ratio
X022 % Δ Operating revenue		X332 % Δ Gearing	
X031 P/L before tax	Profit & Loss Account	X341 Market Capitalisation	Annual Stock Data
X032 % Δ P/L before tax		X342 % Δ Market Capitalisation	
X041 P/L for period [=Net income]	Profit & Loss Account	X351 Long Term Liabilities per share	Annual Stock Data
X042 % Δ P/L for period [=Net income]		X361 Price / earnings ratio	Annual Stock Valuation
X051 Cash flow	Profit & Loss Account	X371 Price / book value ratio	Annual Stock Valuation
X052 % Δ Cash flow		X372 % Δ Price / book value ratio	
X061 Non-current liabilities	Balance Sheet	X381 Market capitalisation / Total assets (Tobin's Q)	Annual Stock Valuation
X062 % Δ Non-current liabilities		X391 Earnings yield	Annual Stock Valuation
X071 Sales	Profit & Loss Account	X401 Market Capitalisation / shareholders funds	Annual Stock Valuation
X072 % Δ Sales		X411 Closing price / Cash Flow per Share ratio	Annual Stock Valuation
X081 Operating P/L [=EBIT]	Profit & Loss Account	X421 Earnings per share	Annual Stock Data
X082 % Δ Operating P/L [=EBIT]		X431 Cash flow per share	Annual Stock Data
X091 Financial P/L	Profit & Loss Account	X441 Dividends per share	Annual Stock Data
X092 % Δ Financial P/L		X451 Closing price	Annual Stock Data
X101 EBITDA	Profit & Loss Account	X452 % Δ Closing price	
X102 % Δ EBITDA		X461 Outstanding shares	Annual Stock Data
X111 ROE using P/L before tax	Profitability Ratio	X462 % Δ Outstanding shares	
X112 % Δ ROE using P/L before tax		X471 Net Cash from Operating Activities	Cash Flow Statement
X121 ROCE using P/L before tax	Profitability Ratio	X472 % Δ Net Cash from Operating Activities	
X122 % Δ ROCE using P/L before tax		X481 Shareholders funds	Balance Sheet
X131 ROA using P/L before tax	Profitability Ratio	X482 % Δ Shareholders funds	
X132 % Δ ROA using P/L before tax		W3 C - Manufacturing	Dummy Variable
X141 ROE using Net income	Profitability Ratio	W4 D - Electricity, gas, steam and air conditioning supply	Dummy Variable
X142 % Δ ROE using Net income		W5 E - Water supply; sewerage, waste management and remediation activities	Dummy Variable
X151 ROCE using Net income	Profitability Ratio	W6 F - Construction	Dummy Variable
X152 % Δ ROCE using Net income		W7 G - Wholesale and retail trade; repair of motor vehicles and motorcycles	Dummy Variable
X161 ROA using Net income	Profitability Ratio	W8 H - Transportation and storage	Dummy Variable
X162 % Δ ROA using Net income		W9 I - Accommodation and food service activities	Dummy Variable
X171 EBITDA margin	Profitability Ratio	W10 J - Information and communication	Dummy Variable
X172 % Δ EBITDA margin		W11 K - Financial and insurance activities	Dummy Variable
X181 EBIT margin	Profitability Ratio	W12 L - Real estate activities	Dummy Variable
X182 % Δ EBIT margin		W13 M - Professional, scientific and technical activities	Dummy Variable
X191 Profit margin	Profitability Ratio	W14 N - Administrative and support service activities	Dummy Variable
X192 % Δ Profit margin		W15 O - Public administration and defence; compulsory social security	Dummy Variable
X201 Gross margin	Profitability Ratio	W16 Q - Human health and social work activities	Dummy Variable
X202 % Δ Gross margin		W17 R - Arts, entertainment and recreation	Dummy Variable
X211 Cash flow / Operating revenue	Profitability Ratio	W18 S - Other service activities	Dummy Variable
X212 % Δ Cash flow / Operating revenue		Z1 Austria	Dummy Variable
X221 Market cap / Net Cash from Operating Activities	Profitability Ratio	Z2 Belgium	Dummy Variable
X231 Net assets turnover	Operational Ratio	Z3 Denmark	Dummy Variable
X232 % Δ Net assets turnover		Z4 Finland	Dummy Variable
X241 Interest cover	Operational Ratio	Z5 France	Dummy Variable
X242 % Δ Interest cover		Z6 Germany	Dummy Variable
X251 Collection period	Operational Ratio	Z7 Ireland	Dummy Variable
X252 % Δ Collection period		Z8 Isle Of Man (United Kingdom)	Dummy Variable
X261 Credit period	Operational Ratio	Z9 Italy	Dummy Variable
X262 % Δ Credit period		Z10 Jersey (United Kingdom)	Dummy Variable
X271 R&D expenses / Operating revenue	Operational Ratio	Z11 Luxembourg	Dummy Variable
X272 % Δ R&D expenses / Operating revenue		Z12 Malta	Dummy Variable
X281 Current ratio	Structure Ratio	Z13 Netherlands	Dummy Variable
X282 % Δ Current ratio		Z14 Norway	Dummy Variable
X291 Liquidity ratio	Structure Ratio	Z15 Poland	Dummy Variable
X292 % Δ Liquidity ratio		Z16 Portugal	Dummy Variable
X301 Shareholders liquidity ratio	Structure Ratio	Z17 Spain	Dummy Variable
X302 % Δ Shareholders liquidity ratio		Z18 Sweden	Dummy Variable
X311 Solvency ratio (Asset based)	Structure Ratio	Z19 Switzerland	Dummy Variable
X312 % Δ Solvency ratio (Asset based)		Z20 United Kingdom	Dummy Variable

3.1.4. Transformações das variáveis

Procedemos à logaritmização de todas variáveis que apresentavam valores suscetíveis de serem sujeitos a este tipo de transformação não linear, ou seja, que não apresentavam valores nulos ou negativos. Por um lado, esta transformação reduz em geral o valor da variância, por outro é por vezes usada na tentativa de aproximar os dados a uma distribuição gaussiana. Além disso, tendo em vista um dos principais objetivos da ACP na identificação de PC não correlacionados, a partir de variáveis correlacionadas, a independência entre os componentes só pode ser garantida quando se verifica normalidade multivariada das variáveis originais [53], embora não seja requisito. Considerando os dados univariados, apenas as variáveis $\ln(X041)$, $\ln(X061)$, $\ln(X171)$, $\ln(X211)$ e $\ln(X471)$, exibem um comportamento normal para um nível de $p < 0,05$ sob as hipóteses do teste de normalidade de Jarque-Bera, conforme Tabela X no anexo 2, onde se encontram resumidas as principais estatísticas descritivas das variáveis contínuas. Na mesma tabela é possível identificar as variáveis que foram objeto de transformação logarítmica.

3.2. Métodos estatísticos aplicados – breves considerações teóricas

3.2.1. ACP – Análise de Componentes Principais

A ACP é um método de redução do número de variáveis que permite a construção de novas variáveis não correlacionadas entre si (ortogonais) através de uma combinação linear das variáveis originais. No sentido proposto, a ACP baseia-se num problema de mudança de base, ela própria, uma combinação linear da base euclidiana. As novas variáveis são construídas de tal maneira que poucas delas explicam a maior variabilidade dos dados, sendo que consideramos, no máximo, o mesmo número de variáveis que o das variáveis originais. Desta forma, a primeira variável criada permite explicar o máximo possível da variabilidade dos dados; a segunda variável permite explicar o máximo possível da variabilidade dos dados não explicado pela primeira variável; e assim sucessivamente. Sob outra perspetiva, a capacidade explicativa das variáveis criadas reduz-se à medida que cada uma das novas variáveis é criada, sendo que, supondo que queremos explicar uma variabilidade total dos dados suficientemente grande, mas inferior ao total da sua variabilidade, podemos negligenciar as últimas variáveis obtidas.

Nesse sentido, se os dados forem descritos por p variáveis, podemos selecionar um número $m < p$ de PC que são responsáveis pela maior parte da variância (informação) contida nos dados. Então a questão que se coloca é saber qual o número de PC a reter. A representação dos dados num espaço de menor dimensão m , resulta numa perda de informação, a qual corresponde à soma das variâncias dos PC que não são usados na análise. Essa perda de informação poderá estar diretamente relacionada com os objetivos concretos do estudo que se pretende conduzir, isto é, da quantidade de informação que estaremos dispostos a perder. Neste contexto, foram sendo propostas algumas regras práticas com o objetivo de auxiliar nessa decisão. As principais referem-se ao critério de Kaiser, à inspeção do *scree-plot* e à análise paralela de Horn. A primeira compreende a retenção dos componentes cuja variância é superior a 1, no caso de dados estandardizados; a segunda concerne à inspeção gráfica das variâncias dos componentes principais e a última baseia-se em simulação. Adotamos no nosso trabalho a análise paralela de Horn.

A ACP não apresenta necessariamente a melhor solução para efeitos de discriminação, uma vez que as direções de maior variabilidade não coincidem propriamente com as direções de melhor discriminação.

Com referência à definição e aplicação da presente metodologia, seguimos de perto o autor de [54].

3.2.2. AD – Análise Discriminante

Como refere [55], a AD é uma técnica de classificação que permite encontrar as características a partir das quais podemos distinguir os membros de dois ou mais grupos, de modo que conhecendo as características de um novo objeto se possa prever a que grupo esse objeto pertence. Nesse contexto, pode ser utilizada para avaliar diferenças entre grupos, determinar formas para distinguir grupos e classificar novos grupos. Tem três grandes objetivos: o primeiro, identificar as variáveis que melhor discriminam dois ou mais grupos, porque nem todas as variáveis vão exibir poder de discriminação; o segundo, usar as variáveis identificadas para estimar equações lineares dessas variáveis (índices) que traduzam as diferenças entre os grupos; e, o terceiro, usar as variáveis identificadas e o índice para construir uma regra que classifique futuras observações num dos grupos.

Esta técnica de previsão assume que os dados (variáveis independentes) seguem uma distribuição normal multivariada. Sendo computacionalmente mais eficiente, a AD deve ser sempre usada quando a hipótese de normalidade dos dados não é violada. No entanto, a hipótese da normalidade da distribuição dos dados é violada quando as variáveis independentes incluem variáveis categóricas. A violação da hipótese de normalidade afeta os testes de significância e os métodos de classificação. Além da hipótese de normalidade dos dados, trata-se de uma técnica que exige um conjunto de pressupostos adicionais a serem observados. Nomeadamente, o número de variáveis independentes deve ser muito inferior ao número de casos, a variabilidade dentro dos grupos deve ser a mesma e porque a AD é bastante sensível a *outliers*, os mesmos devem ser removidos. A ACP não apresenta necessariamente a melhor solução para efeitos de discriminação, uma vez que as direções de maior variabilidade não coincidem propriamente com as direções de melhor discriminação, mas a sua utilização permitiu remover a multicolinearidade entre as variáveis sem subjetividade e perda de informação que a remoção empírica de variáveis poderia acarretar. Na AD coloca-se explicitamente o objetivo de separar subgrupos de indivíduos, subgrupos esses que são previamente conhecidos nos dados observados [55].

Para decidir sobre o poder discriminante individual de cada variável, faz-se um teste-*t* para a diferença de médias. Para avaliar sobre o poder discriminante conjunto de todas as variáveis realiza-se o teste de Wilks.

O trabalho conduzido com a AD baseou-se em [54] e [56].

3.2.3. RL – Regressão Logística

A RL surge como uma técnica de classificação alternativa que é normalmente recomendada quando as variáveis independentes não satisfazem a hipótese de normalidade multivariada, uma vez que não assume hipóteses sobre a distribuição dos dados. Por isso, as variáveis independentes podem ser uma mistura de variáveis contínuas e variáveis categóricas, nesse caso, sendo de aplicar a RL. Permite avaliar diferenças entre grupos e proceder à classificação dos dados. No problema em estudo, o objetivo é: dados os valores das variáveis independentes utilizadas na RL de uma observação, estimar a probabilidade de determinada empresa ser ou não “bem-sucedida”.

Na aplicação desta técnica inspirámo-nos essencialmente nos trabalhos [56] e [57].

A RL é um modelo de regressão linear generalizado em que a variável dependente é na verdade uma função que não é uma função linear – chamada de função *logit*. Se tivermos p variáveis independentes, o modelo de RL é dado por:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p. \quad (1)$$

A relação da probabilidade, P , de dado acontecimento (variável de interesse) com as variáveis independentes designa-se por *link function* e é neste caso dada pela função logística:

$$f(x) = \frac{1}{1+e^{-\beta x}} \quad (2)$$

Para validar o modelo, tal como na regressão linear múltipla, fazem-se testes- z à significância individual de cada variável explicativa. Para a significância global das variáveis independentes, realizam-se testes de rácio de verosimilhança (*likelihood-ratio*).

De outro modo, a adequação global do modelo pode ser avaliada através do Pseudo- R^2 de McFadden. Ou seja,

$$R^2 = 1 - \frac{\ln L}{\ln L_0} \quad (3)$$

Quanto maior for o valor do Pseudo- R^2 de McFadden, melhor o ajuste do modelo, embora seja de difícil interpretação em valor absoluto. R^2 corresponde à percentagem da variação total que é explicada pelos regressores; L , é a verosimilhança do modelo com as variáveis explicativas; e, L_0 , é a verosimilhança do modelo sem variáveis independentes.

4. ESTUDO SOBRE A PREVISÃO DO ROE - RESULTADOS DA ABORDAGEM EMPÍRICA

4.1. Análise de Componentes Principais

Tomando a base de dados recolhida a partir dos dados financeiros e de mercado das empresas procedemos primeiramente a uma ACP. Em geral, a ACP é realizada com dados corrigidos pela média e importa decidir a utilização de dados estandardizados ou não, considerando que os dados estandardizados dão soluções diferentes dos dados não estandardizados. Tal decisão é tomada tendo em conta a variância relativa das variáveis originais. Ou seja, quando existem discrepâncias acentuadas das variâncias das variáveis originais convém proceder à estandardização dos dados⁹. Doravante, a designação de variáveis originais, no caso das variáveis contínuas, inclui as variáveis que foram objeto de transformação logarítmica. No caso em análise, procedemos ao cálculo da matriz de variâncias e covariâncias das variáveis originais. Observando a sua diagonal verifica-se que as discrepâncias entre as variâncias das variáveis originais são bastante elevadas, o que sugere, desde logo, a estandardização dos dados (*vide* Tabela XI: Variância das variáveis originais após transformações logarítmicas pertinentes – no anexo 2).

Os PC vão ser obtidos na sequência da resolução do problema de valores e vetores próprios associado à matriz de variâncias e covariâncias dos dados originais ou, no caso em análise, uma vez que se procedeu à estandardização dos dados, à matriz de correlações. Os valores próprios obtidos através do *software R* encontram-se sintetizados na Tabela XII no anexo 2.

Encontrados os PC dados pelos vetores próprios importa escolher o número de componentes a reter. Nesse sentido e como referido, segundo o critério de Kaisser, para dados estandardizados deverão ser retidos os componentes com valores próprios $\lambda_i > 1, i = 1, \dots, p$.

⁹ Na aplicação da ACP não foram consideradas as variáveis *dummy*.

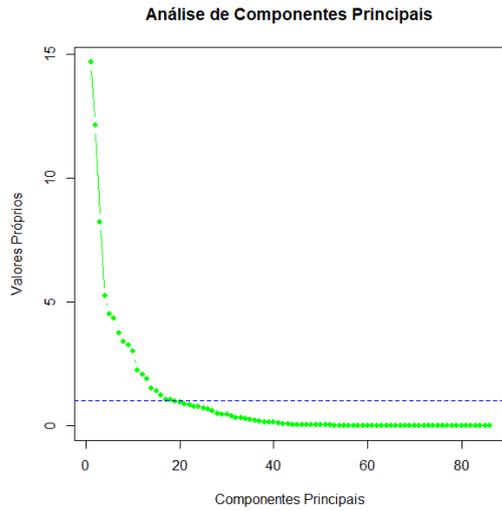


FIGURA 1: Valores próprios com os dados estandardizados

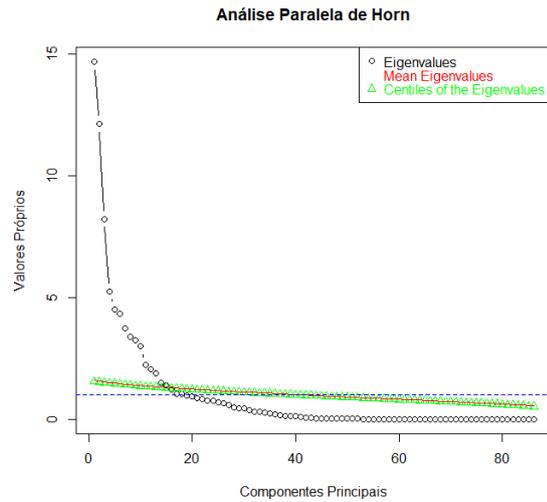


FIGURA 2: Análise Paralela de Horn

De acordo com a Figura 1 e conforme o Critério de Kaiser deverão ser retidos os primeiros 18 PC, ou seja, aqueles com variância superior a 1.

Procedendo à Análise Paralela de Horn vão ser retidos os PC cujos valores próprios calculados a partir dos dados superam a média dos valores próprios simulados. Na Figura 2, podemos observar as variâncias dos PC, *i.e.*, os seus valores próprios, e as que foram obtidas após a Análise Paralela de Horn. Uma análise mais detalhada destes valores (*vide* Tabela XII) irá dar suporte à retenção dos primeiros 15 PC, já que o valor próprio do componente principal 16 é menor que o valor médio simulado resultante da Análise Paralela de Horn ($PC16 = 1,22 < V16 = 1,29$).

Constituindo a Análise Paralela de Horn uma técnica mais consensual entre os diversos autores (*vide e.g.* [58]), optou-se então pela retenção de quinze PC, cujas características se encontram descritas na Tabela I abaixo, obtida através do \mathbf{R} .

TABELA I: Importância dos PC

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	3,8314	3,4824	2,8676	2,2890	2,1248	2,0846	1,9297	1,8356
Proportion of Variance	0,1707	0,1410	0,0956	0,0609	0,0525	0,0505	0,0433	0,0392
Cumulative Proportion	0,1707	0,3117	0,4073	0,4682	0,5207	0,5713	0,6146	0,6538
	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	
Standard deviation	1,7974	1,7278	1,4954	1,4283	1,3700	1,2303	1,1802	
Proportion of Variance	0,0376	0,0347	0,0260	0,0237	0,0218	0,0176	0,0162	
Cumulative Proportion	0,6913	0,7260	0,7520	0,7758	0,7976	0,8152	0,8314	

Conforme a Tabela I e sendo os valores próprios iguais à variância dos PC obtidos, podemos retirar, desde logo, algumas conclusões. Assim, verifica-se que os quinze PC retidos explicam aproximadamente 83% da variabilidade dos dados, sendo que individualmente o primeiro componente principal (PC1) explica aprox. 17% da variabilidade dos dados e o segundo (PC2) aprox. 14%. (*vide* Tabela XII para os restantes PC).

Os *scores* são as coordenadas dos dados na nova base, ou seja, são as novas variáveis. Com o auxílio do R foram calculados os *scores* (PC-*scores*). Para exemplificar, as ilustrações seguintes (Figura 3) representam as observações nos três primeiros PC para as 1115 empresas de acordo com a medida de sucesso obtida, representadas no novo sistema de eixos. Tanto as representações a duas, como a três dimensões, permitem uma discriminação visual entre as empresas “bem-sucedidas” e “malsucedidas”. Em particular, a Figura 3 (d) permite ilustrar que a introdução do terceiro componente ainda permite um incremento de informação relevante.

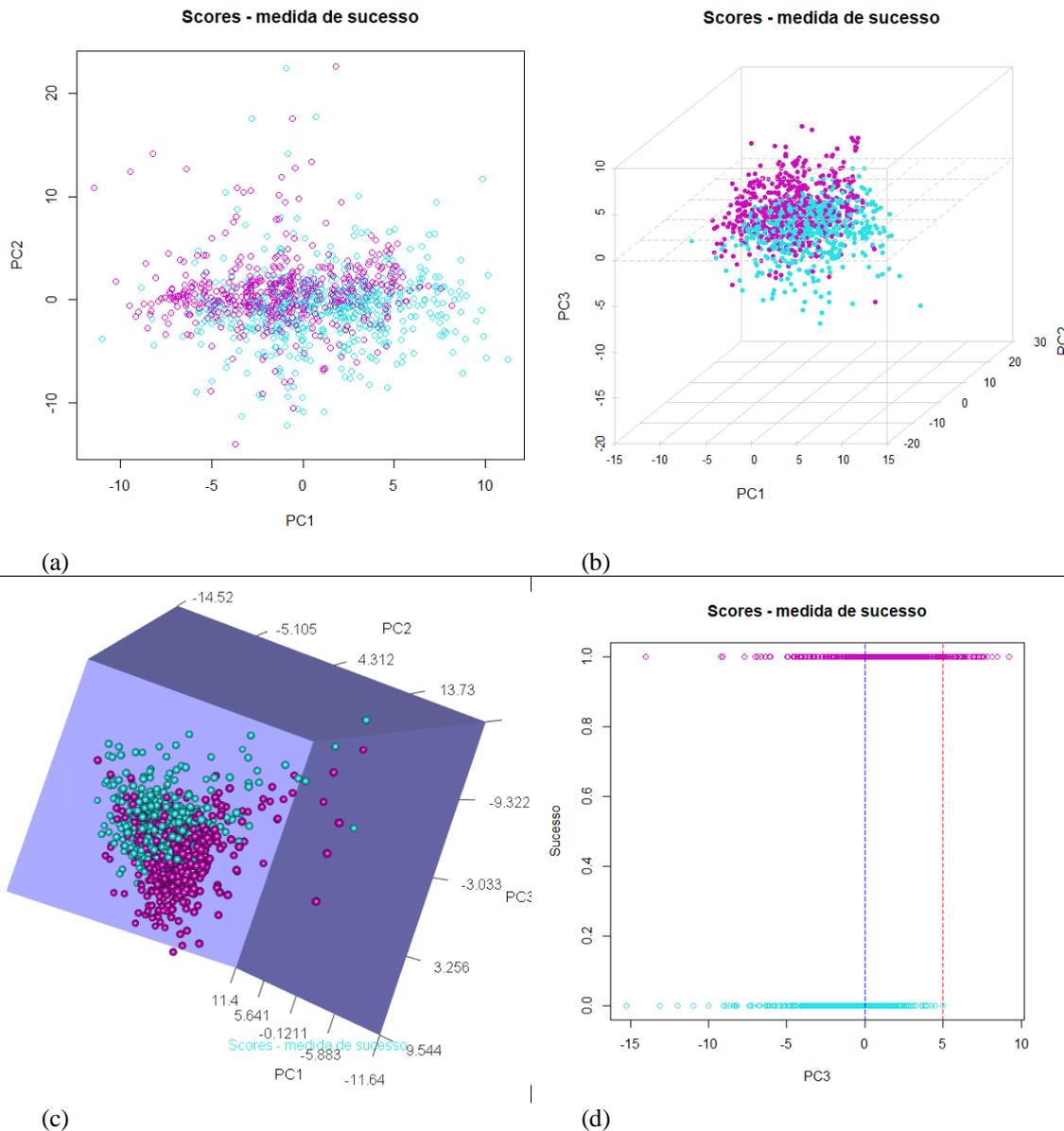


FIGURA 3: Representação das observações no sistema de eixos dado pelos PC para as 1115 empresas, de acordo com a medida de sucesso, após a aplicação da ACP, representando, os pontos a azul as empresas “malsucedidas” e os pontos a cor-de-rosa as empresas “bem-sucedidas” – (a): PC1 e PC2 em 2D (captam 31% da variabilidade dos dados); (b) e (c): PC1, PC2 e PC3 a 3D (captam 41% da variabilidade dos dados); (d): PC3 a 2D (capta 10% da variabilidade dos dados).

Ora, da análise da Figura 3 (c) podemos identificar duas regiões, cada uma delas registrando uma maior concentração de empresas “bem-sucedidas” ou “malsucedidas”. No entanto, a visualização das figuras anteriores não parecem indicar uma maior concentração ou dispersão de cada um dos grupos. Resulta ainda das figuras anteriores que o PC3 parece apresentar um maior poder de discriminação, tornando possível a

diferenciação dos dois grupos. Quanto aos PC1 e PC2, dadas as sobreposições observadas dentro dos grupos, o seu poder de discriminação parece inferior, especialmente o PC1. É assim aceitável supormos que a informação recuperada através do PC3, em comparação com o PC1 e PC2, é relevante na distinção entre os dois tipos de empresas.

Como é sabido (*vide e.g.* [54]) na ACP os PC são combinações lineares das variáveis originais – no nosso caso temos 86. É útil interpretar os PC retidos, atribuir-lhes um significado prático, usando-se para tal as covariâncias entre os PC-scores e os dados originais. Os *loadings* dão a correlação entre o PC e as variáveis originais. Quanto maior for o *loading* (correlação) entre uma variável e um PC, maior a influência dessa variável nesse PC. O *loading* permite-nos ver a influência de cada variável original em cada componente principal. Na Tabela XIII, no anexo 2, podemos observar os *loadings* dos primeiros 15 PC. A seguir, tentaremos esboçar uma possível interpretação para os primeiros 6 PC. Esta interpretação envolverá sempre alguma carga de subjetividade, sobretudo para os componentes λ de maior índice i ($\lambda_i, i = 1, \dots, p$).

Assim sendo, para uma possível interpretação, e considerando os *loadings*, resulta da ACP que o PC 1 é influenciado de forma mais significativa por 11 variáveis. Estas variáveis parecem associadas, por um lado, ao nível da dimensão das empresas [$\ln(X011)$ e $\ln(X481)$] ou ao nível de responsabilidades de longo prazo [$\ln(X061)$] – estas três variáveis pertencem ao balanço – e por outro lado, ao nível de resultados, ou a outras variáveis com influência importante na sua formação [$\ln(X021)$, $\ln(X031)$, $\ln(X041)$, $\ln(X051)$, $\ln(X071)$, $\ln(X081)$, $\ln(X101)$ e $\ln(X471)$] – neste caso, temos sete variáveis pertencentes à demonstração de resultados e, a última, à demonstração de fluxos de caixa¹⁰. O PC2 encontra-se igualmente relacionado significativamente com 11 variáveis, todas elas associadas à variação anual das variáveis subjacentes – sete variáveis associadas à variação anual de rácios de rentabilidade [$X112$, $X122$, $X132$, $X142$, $X152$, $X162$, $X182$ e $X192$] e, as restantes três, à variação anual de elementos da demonstração de resultados [$X032$, $X042$ e $X082$]. O PC3 encontra-se relacionado com o nível de seis rácios de rentabilidade [$\ln(X111)$, $\ln(X121)$, $\ln(X131)$, $\ln(X141)$, $\ln(X161)$ e $\ln(X191)$], e o nível de dois rácios que traduzem a valorização anual do mercado

¹⁰ Para uma descrição das variáveis pode ser consultado o Quadro I, onde se encontram sumariadas as variáveis independentes, ou o Quadro II, onde se encontram desenvolvidas as fórmulas de cálculo.

$[\ln(X371), \ln(X381)]$. Já o PC4 relaciona-se com o nível de quatro rácios de rendibilidade $[\ln(X171), \ln(X181), \ln(X201), \ln(X211)]$, o nível de um rácio operacional $[\ln(X231)]$ e o nível de dois rácios estruturais $[\ln(X301), \ln(X331)]$. O PC5 encontra-se relacionado com as variáveis $\ln(X171), \ln(X211), X302, X312, X322$ e $\ln(X391), \ln(X231), \ln(X361)$ e $X372$. O PC6 encontra-se associado às variáveis $\ln(X351), \ln(X421), \ln(X431), \ln(X441), \ln(X451)$ e $\ln(X461)$, todas elas conexas com dados do mercado.

Assim, pensamos constituir uma interpretação legítima associar o PC1 a um índice de resultados, o PC2 a um índice de taxas de crescimento da rendibilidade e o PC3 a um índice de rendibilidade. A interpretação do PC4, assim como, do PC5, revela-se difícil, porquanto estarem relacionados com uma combinação de diferentes tipos de variáveis. Já o PC6 parece constituir-se como um índice de mercado.

Certamente, avançar com uma interpretação mais técnica do significado dos PC envolveria um aprofundamento significativo na descrição das variáveis. Arriscamo-nos a avançar que os dados analisados, consumados na construção dos 15 PC, poderiam auxiliar na construção de índices de sucesso das empresas. De outro modo, considerando, por exemplo, os *scores* das empresas representados no PC3, quando maiores que cinco, permitem aferir com grande rigor o sucesso das empresas a um ano à frente, conforme facilmente se verifica na Figura 3 (d), representados à direita da linha tracejada vermelha.

4.2. Análise Discriminante

Com o propósito de conseguirmos prever o sucesso ou insucesso das empresas a um ano à frente, trataremos a seguir da aplicação da AD aos dados e a discussão dos resultados preditivos alcançados, comparando-os com os resultados da RL no capítulo 5.

Para a conceção do modelo de AD, optou-se por utilizar os primeiros 15 PC resultantes da ACP. Uma das principais vantagens da utilização dos PC prende-se com a eliminação da correlação entre as variáveis, já que as novas variáveis são todas ortogonais entre si. O objetivo é classificar as empresas como “bem-sucedidas” ou “malsucedidas”, informação essa que está contida na variável categórica dependente $Y11$.

Como amostra de treino serão utilizados os dados das variáveis de 2013 a 2019. Como amostra de teste serão utilizados os dados das variáveis de 2020 (*out-of-sample*). A variável Y_{11} , com a informação observada da classificação das empresas, reporta-se ao nível do ROE a um ano à frente.

Na condução do teste de Henze-Zirkler para a normalidade multivariada encontramos um valor crítico de $HZ = 2,538871$ com um valor- $p = 0$, rejeitando a hipótese nula de normalidade multivariada. No que concerne à normalidade univariada apenas o PC6 apresenta uma distribuição normal sob as hipóteses de teste de Anderson-Darling, com uma estatística = 0,6186 e um valor- $p = 0,1072$. Como referido no ponto 3.2.2., na AD assume-se que os dados (variáveis independentes) seguem uma distribuição normal multivariada. Verificando-se a violação da hipótese de normalidade dos dados, é recomendável a utilização da RL em detrimento da AD. No entanto, optámos por proceder à AD para posterior comparação dos resultados com a RL.

Na medida em que a existência de *outliers* poderá ter uma influência significativa na capacidade preditiva do modelo, procede-se à análise após a sua remoção. Pertinentemente, com o objetivo de percebermos a sua interferência, serão apresentados alguns resultados com a sua presença. Para a remoção de *outliers* foi utilizada a função *aq.plot* do pacote *mvoutlier* do **R** conforme [59], baseado no cálculo das distâncias de Mahalanobis robustas, considerando o quantil de ordem 0,975 ($Q_{0,975}$) da distribuição qui-quadrado (χ^2). Da aplicação do método resultou a identificação e remoção de 135 *outliers*.

No caso em estudo, temos 1115 empresas – ou 980 empresas, após a remoção de *outliers*. A amostra de treino é constituída pelos dados de 2013 a 2019 e a amostra de teste pelos dados de 2020. São considerados 2 grupos: empresas “bem-sucedidas” ($Y_{11} = 1$) e empresas “malsucedidas” ($Y_{11} = 0$).

Amostra de treino:

- 481 empresas “bem-sucedidas” – 424 após a remoção de *outliers*: $Y_{11} = 1$ (grupo 1);
- 504 empresas “malsucedidas” – 447 após a remoção de *outliers*: $Y_{11} = 0$ (grupo 0).

Amostra de teste:

- 76 empresas “bem-sucedidas” – 64 após a remoção de *outliers*: $Y_{11} = 1$ (grupo 1);
- 54 empresas “malsucedidas” – 45 após a remoção de *outliers*: $Y_{11} = 0$ (grupo 0).

Constituindo a AD uma forma de proceder à discriminação entre grupos, no contexto da presente análise, vamos ter dois grupos. E, para cada observação, temos informação a que grupo a empresa pertence. Deste modo:

- empresas “bem-sucedidas” – grupo 1;
- empresas “malsucedidas” – grupo 0.

Para aplicar a AD, pretende-se identificar primeiramente as variáveis discriminantes, ou seja, as variáveis que melhor discriminam entre as empresas “bem-sucedidas” e “malsucedidas”. Ora, para a identificação dessas variáveis procede-se à sua análise univariada e multivariada. Nesse âmbito, importa determinar o seu poder discriminante aferido através do teste de hipóteses.

Numa perspetiva univariada, é normalmente efetuado um teste- t para a diferença de médias, relativamente a cada um dos grupos. A condução deste teste parte dos pressupostos de normalidade de distribuição das variáveis em cada um dos grupos, assim como igualdade de variâncias. Ainda assim, efetuámos os testes para poder prosseguir com a AD. Na Tabela II seguinte encontram-se reproduzidos os resultados obtidos no **R** para esta estatística.

TABELA II: Resultados do teste-*t* para a diferença de médias entre duas amostras

	<i>t</i>	df	<i>p</i> -value	mean in group 0	mean in group 1
Comp.1	10,4180	869	0,0000	1,3102	-1,2341
Comp.2	-7,0170	869	0,0000	-0,7860	0,3655
Comp.3	-12,4370	869	0,0000	-0,6154	1,2399
Comp.4	0,7969	869	0,4257	0,3893	0,2759
Comp.5	-4,5797	869	0,0000	-0,4812	0,0752
Comp.6	7,0390	869	0,0000	0,2790	-0,6138
Comp.7	-1,0695	869	0,2851	-0,2088	-0,0900
Comp.8	1,7713	869	0,0769	-0,0467	-0,2059
Comp.9	5,7137	869	0,0000	0,3797	-0,2144
Comp.10	-1,3823	869	0,1672	0,0276	0,1455
Comp.11	-0,5040	869	0,6144	-0,0595	-0,0218
Comp.12	0,0092	869	0,9926	0,0534	0,0528
Comp.13	0,3231	869	0,7467	0,0896	0,0687
Comp.14	1,2370	869	0,2164	-0,0104	-0,0897
Comp.15	-3,3877	869	0,0007	-0,1640	0,0187

Têm poder discriminativo as variáveis Comp.1, Comp.2, Comp.3, Comp.5, Comp.6, Comp.9 e Comp.15, dando suporte à decisão de rejeitar a hipótese nula para a igualdade de médias com um nível de significância inferior a 0,1 %.

Sem a remoção de *outliers* foram encontrados resultados idênticos para todos os componentes, exceto para o Comp.15, cujo nível de significância encontrado para a rejeição da hipótese nula foi de 4,16%.

Numa perspectiva multivariada, a adequação das variáveis foi avaliada tendo em conta os resultados do teste de Wilks, permitindo testar conjuntamente a igualdade de médias das variáveis previamente selecionadas em resultado dos testes anteriores. Ou seja, importa compreender se as variáveis, conjuntamente, têm poder discriminante.

De acordo com os resultados do *R* reproduzidos na Figura 4 obtivemos um *p-value* igual a zero ($p - value < 2,2e - 16$), significando a adequação conjunta das variáveis. Ou seja, este resultado permite-nos rejeitar a hipótese nula do teste de Wilks. Foi ainda obtido o valor do Lambda de Wilks ($Wilks' Lambda = 0,6055$), assim como o valor da estatística de teste ($Chi2 - Value = 434,23$).

```
> Wilks.test(Y11 ~ Comp.1+Comp.2+Comp.3+Comp.5+Comp.6+Comp.9+Comp.15
+           ,data = my_data_3[1:871,])
```

One-way MANOVA (Bartlett Chi2)

data: x

Wilks' Lambda = 0,6055; Chi2-Value = 434,23; DF = 7,00; p-value < 2,2e-16

sample estimates:

	Comp.1	Comp.2	Comp.3	Comp.5	Comp.6	Comp.9	Comp.15
0	1,3102	-0,7860	-0,6154	-0,4812	0,2790	0,3797	-0,1640
1	-1,2341	0,3655	1,2399	0,0752	-0,6138	-0,2144	0,0187

FIGURA 4: *Teste de Wilks de adequação conjunta das variáveis*

Os resultados da figura anterior reproduzem ainda as médias de cada um dos grupos para cada uma das variáveis identificadas na análise univariada com poder discriminativo.

Identificadas as variáveis que melhor discriminam os dois grupos, as mesmas serão usadas para estimar uma equação (índice) que represente as diferenças entre as empresas “bem-sucedidas” e as empresas “malsucedidas”. Trata-se de determinar a função discriminante, a qual consiste numa combinação linear das variáveis discriminantes, onde a separação entre os dois grupos é máxima, e que assume a seguinte forma:

$$Z = W_1X_1 + W_2X_2 + \dots + W_nX_n. \quad (4)$$

Onde Z é a função discriminante e W_1 a W_n representam o conjunto de coeficientes (pesos) para as variáveis independentes X_1 a X_n , respetivamente.

A discriminação é estabelecida tendo em conta os pesos de cada variável independente de tal forma que seja maximizada a distinção entre grupos. Dito de outro modo, a função discriminante serve para separar da melhor forma as características de cada um dos grupos.

Ora, para que a nova variável Z resulte numa separação máxima (melhor discriminação) entre os 2 grupos importa maximizar o seguinte valor:

$$\lambda = \frac{SS_b}{SS_w}. \quad (5)$$

Ou seja, as médias dos dois grupos em Z têm de estar à maior distância possível (maximização da soma de quadrados entre grupos SS_b ¹¹) e as variâncias dos dois grupos em Z devem ser as menores possíveis (minimização da soma de quadrados dentro de grupos SS_W ¹²).

A estimação do modelo de AD foi levada a cabo no **R** recorrendo ao comando “lda” (*linear discriminant analysis*), e utilizando as sete variáveis com poder discriminante acima identificadas, assim como a amostra de treino, tendo-se obtido a estimativa para cada um dos coeficientes, conforme os resultados do **R**, reproduzidos na Figura 5.

```
> my_lda <- lda(Y11 ~ Comp.1+Comp.2+Comp.3+Comp.5+Comp.6+Comp.9+Comp.15
+             ,data = my_data_3[1:871,], CV = FALSE)
> my_lda
Call:
lda(Y11 ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.6 + Comp.9 +
    Comp.15, data = my_data_3[1:871, ], CV = FALSE)
```

Prior probabilities of groups:

	0	1
	0,5132	0,4868

Group means:

	Comp.1	Comp.2	Comp.3	Comp.5	Comp.6	Comp.9	Comp.15
0	1,3102	-0,7860	-0,6154	-0,4812	0,2790	0,3797	-0,1640
1	-1,2341	0,3655	1,2399	0,0752	-0,6138	-0,2144	0,0187

Coefficients of linear discriminants:

	LD1
Comp.1	-0,1734
Comp.2	0,1179
Comp.3	0,3024
Comp.5	0,1225
Comp.6	-0,2565
Comp.9	-0,2762
Comp.15	0,0760

FIGURA 5: Output do **R** para a estimação da função discriminante

A partir desses coeficientes é possível então obter a função discriminante, seguida do cálculo dos *scores*, com base nos quais são obtidas as pertinentes classificações. A função discriminante encontrada para os anos em análise pode ser observada na equação (6):

¹¹ Medida da diferença das médias entre grupos.

¹² Medida da homogeneidade dentro do grupo.

$$Z = -0,173Comp.1 + 0,118Comp.2 + 0,302Comp.3 + 0,123Comp.5 - \\ -0,256Comp.6 - 0,276Comp.9 + 0,076Comp.15. \quad (6)$$

O *output* obtido no R , além do valor da estimativa dos W_n , dá-nos a média dos grupos para cada uma das variáveis, e também, as probabilidades *a priori* de cada um dos grupos, no caso em estudo, 0,513 para o grupo 0 e 0,487 para o grupo 1 que correspondem às proporções de cada um dos grupos. Estas probabilidades *a priori* têm a ver com a quantidade de empresas “bem-sucedidas” e “malsucedidas” pertencentes a cada um dos grupos na amostra de treino, e surgem associadas ao conceito de valor de corte ou, do anglo-saxónico, *cut-off point*. As probabilidades *a priori* visam melhorar a taxa de previsões corretas e podem ser definidas de acordo com a estatística bayesiana. Este valor é geralmente resultado da média aritmética da média dos *scores* de cada grupo, e dado por:

$$cutoff = \frac{n_1\bar{z}_1+n_0\bar{z}_0}{n_1+n_0} \quad (7)$$

Onde, \bar{z}_j é o valor médio dos *scores* para o grupo j e n_j é o número de observações no grupo j , com $j = 1, 0$.

E, considerando as probabilidades *a priori*, vem:

$$cutoff = \frac{n_1\bar{z}_1+n_0\bar{z}_0}{n_1+n_0} + \ln\left(\frac{p_0}{p_1}\right) \quad (8)$$

Onde p_j é a probabilidade *a priori* de uma observação pertencer ao grupo j , com $j = 1, 0$.

O valor de corte permite dividir o espaço de discriminação em duas regiões mutuamente exclusivas e exaustivas. Isto é, a região dos valores de *score* que pertencem a um grupo e a região dos valores de *score* que pertencem a outro grupo.

Estabelecido o valor de corte com base na equação (8), torna-se possível classificar observações futuras num dos grupos. Então, dado o *score* de uma nova empresa

(observação) da amostra de teste, calculado a partir da função discriminante, podemos proceder à sua classificação.

Supondo R_1 – a região 1 das empresas “bem-sucedidas” – e R_0 – a região 0 das empresas “malsucedidas” –, a classificação de uma nova empresa vai depender do seu *score*. Podemos definir analiticamente as duas regiões da seguinte forma:

$$R_1: Z \geq \frac{n_1\bar{z}_1+n_0\bar{z}_0}{n_1+n_0} + \ln\left(\frac{p_0}{p_1}\right) \quad (9)$$

$$R_0: Z < \frac{n_1\bar{z}_1+n_0\bar{z}_0}{n_1+n_0} + \ln\left(\frac{p_0}{p_1}\right). \quad (10)$$

Eventualmente, poderiam ser definidos outros valores de corte, *e.g.*, um valor de corte incluindo os custos associados a uma decisão errada.

Estimado o modelo e obtido o valor de corte estamos finalmente em condições de saber como a função discriminante classifica as empresas, tanto as observações da amostra de treino, como as observações da amostra de teste, a partir das variáveis identificadas num dos grupos.

Deste modo, aplicando a expressão da função discriminante (6) aos dados da amostra de treino, obtemos os *scores*¹³ para cada uma das empresas, classificando-a como “bem-sucedida” ou “malsucedida” de acordo com o *cut-off point*. Identicamente, aplicando a mesma expressão aos dados da amostra de teste, obtemos os *scores* para cada uma das empresas, classificando-a de acordo com o *cut-off point*.

As figuras seguintes ilustram uma representação gráfica desses *scores* (os pontos a preto representam os *scores* das empresas efetivamente “malsucedidas” e os pontos a vermelho representam os *scores* das empresas efetivamente “bem-sucedidas”), assim como o valor de *cutoff* = 0,053 – reta tracejada a negro.

¹³ Os *scores* obtidos encontram-se corrigidos pela média.

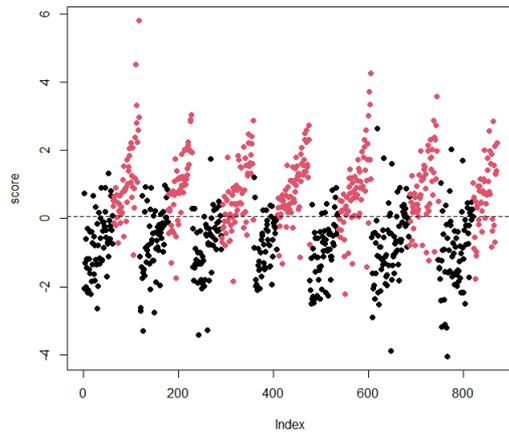


FIGURA 6: Scores para a amostra de treino após aplicação da função discriminante estimada

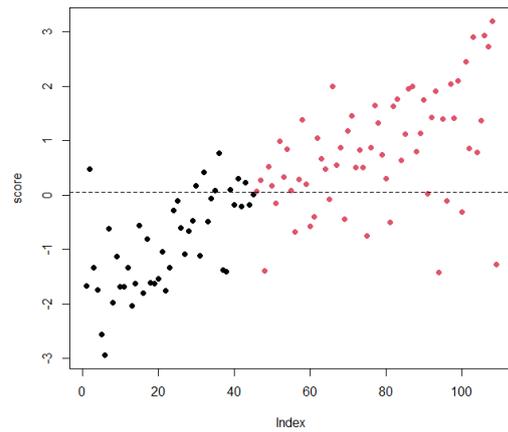


FIGURA 7: Scores para a amostra de teste após aplicação da função discriminante estimada

Se definirmos como “bem-sucedida” uma empresa com um *score* maior ou igual ao *cut-off point*; e, caso contrário, “malsucedida” – conforme as expressões (9) e (10), respetivamente –, sabemos como ficam classificadas as empresas nas amostras de teste. Equivalentemente, na Figura 6 e na Figura 7 são classificadas como “bem-sucedidas” as empresas acima da linha de *cut-off* e “malsucedidas” as empresas abaixo dessa linha.

Os resultados da classificação podem ser sumariados nas matrizes de classificação (*in-sample* e *out-of-sample*) reproduzidas nas tabelas III e IV.

TABELA III: AD – Matriz de Classificação – *in-sample*

Grupo	"Malsucedidas"	"Bem-sucedidas"	Total
"Malsucedidas"	364	83	447
"Bem-sucedidas"	85	339	424
Total	449	422	871

TABELA IV: AD – Matriz de Classificação – *out-of-sample*

Grupo	"Malsucedidas"	"Bem-sucedidas"	Total
"Malsucedidas"	37	8	45
"Bem-sucedidas"	14	50	64
Total	51	58	109

Verificamos na Tabela IV que oito empresas são classificadas como “bem-sucedidas”, embora a respetiva prestação efetiva corresponde à das empresas “malsucedidas”; por outro lado, foram classificadas como “malsucedidas” 14 empresas com uma prestação efetiva positiva. 87 empresas foram corretamente classificadas – 37 como “malsucedidas” e 50 como “bem-sucedidas”. Trata-se de uma classificação *out-of-sample*, ou seja, usou-se a função discriminante estimada a partir dos dados de treino para classificar as empresas fora da amostra.

A precisão alcançada *out-of-sample* com a AD foi assim de 79,82% que compara com uma precisão *in-sample* de 80,71%. Sem a remoção de *outliers* as capacidades preditivas alcançadas foram inferiores – 76,15% *out-of-sample* e 79,59% *in-sample*.

Na figura seguinte são apresentados resultados da Análise Discriminante Linear (LDA), por meio de um histograma empilhado dos valores da função discriminante, para as amostras de teste dos dois grupos.

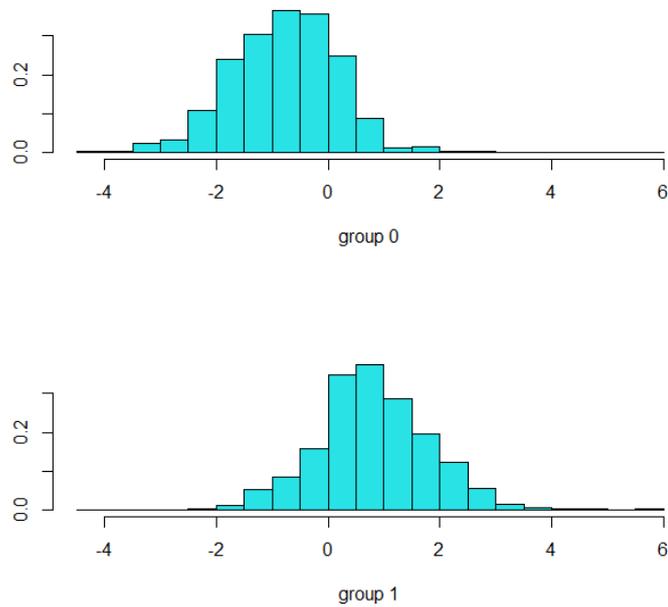


FIGURA 8: Histograma empilhado dos valores da função discriminante para as amostras de teste dos dois grupos

Ora, a validação externa da função estimada que consiste em calcular a precisão da função discriminante em novas observações (precisão fora da amostra – *out-of-sample*), cujos resultados acabámos de apresentar, foi feita dividindo a amostra total em duas subamostras mutuamente exclusivas. Ou seja, foi feita a estimação do modelo através da amostra de treino e a classificação utilizando os dados da amostra de teste recorrendo a esse modelo. Este é também conhecido como o método *holdout*. Alternativamente, para a validação externa da função estimada (*out-of-sample*) foi testada a validação cruzada (*leave-one-out cross validation*) apenas para os dados de 2013 a 2019, tendo-se obtido

uma precisão de 80,25%¹⁴. Através deste método, a função discriminante é estimada deixando uma observação de fora, a qual é depois classificada *out-of-sample*; e, isto é feito para todas as observações. Podemos concluir que as precisões *out-of-sample* encontradas foram ligeiramente inferiores à precisão *in-sample* (< 1%), aliás, como seria de esperar (devido a um sobre-ajustamento da função a esses dados), o que suporta a ideia de uma boa capacidade de generalização do modelo estimado tendo em vista a sua capacidade preditiva.

Tal como na ACP, com a AD procura-se encontrar um novo eixo a partir do qual é formada uma nova variável através da combinação linear das variáveis originais. A grande diferença entre estas técnicas reside no facto de na ACP se pretender encontrar o eixo que capta a maior percentagem de variância dos dados, enquanto na AD pretende-se maximizar a razão entre a soma dos quadrados entre grupos e a soma de quadrados dentro de grupos, no sentido da melhor discriminação.

4.3. Regressão Logística

A RL surge como uma técnica de classificação que é normalmente recomendada quando as variáveis independentes não satisfazem a hipótese de normalidade multivariada. Como alternativa à AD, a RL não faz pressupostos sobre a distribuição das variáveis independentes. No caso em estudo, apenas os PC são variáveis contínuas, mas verifica-se que a normalidade dos dados é violada. No presente trabalho, a RL é empregue no sentido de auxiliar na previsão de sucesso ou insucesso das empresas a um ano à frente.

Retomando o conteúdo do ponto anterior, tal como na AD, para a construção do modelo serão utilizados os primeiros 15 PC resultantes da ACP. Além dos referidos PC, foram também utilizadas as variáveis *dummy* conforme descritas no ponto 3.1.3. e a variável categórica dependente Y11, relativa ao ROE, como medida de sucesso das empresas. Tal como na AD iremos conduzir o estudo com a remoção de *outliers*, e quando pertinente, proceder à comparação de resultados com ou sem a sua presença.

As amostras de treino e de teste, como no ponto 4.2., serão constituídas pelas mesmas observações (com e sem a remoção de *outliers*) com vista à avaliação da diferença entre

¹⁴ Sem a remoção de *outliers* esta precisão cifrou-se em 79,19%.

o grupo das empresas “bem-sucedidas” (grupo 1) e o grupo das empresas “malsucedidas” (grupo 0).

Conforme Figura 9, dado o PC1 podemos ver a que grupo pertencem as empresas. Do mesmo modo, na Figura 10, dado o PC3 podemos ver o grupo a que pertencem essas empresas. Ora, poderíamos efetuar um exercício idêntico para todas as variáveis, em que comparássemos os seus valores com o sucesso ou insucesso. Tomando as variáveis isoladamente, é possível observar alguma capacidade de diferenciação dos grupos através dos *scores* obtidos com o PC1 e PC3.

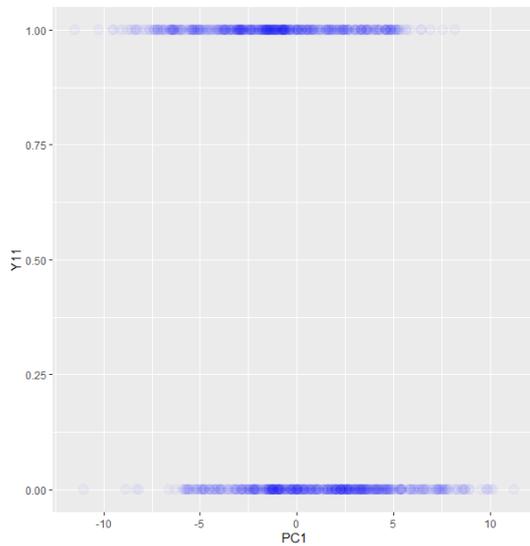


FIGURA 9: Sucesso/insucesso, dado o score do PC1

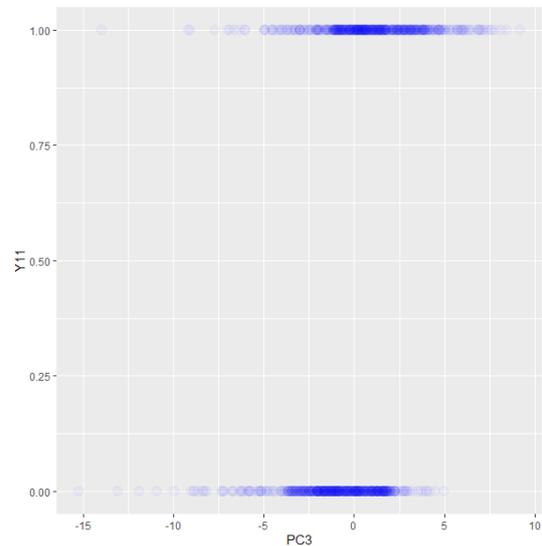


FIGURA 10: Sucesso/insucesso, dado o score do PC3

A estimação do modelo de RL foi levada a cabo no \mathbf{R} pelo método da máxima verosimilhança, utilizando o comando “glm” (*generalized linear model*), com a família de distribuições binomial, utilizando todas as variáveis (15 PC e 38 variáveis *dummy*), tendo-se obtido a estimativa para cada um dos coeficientes ($\beta_0, \beta_1, \dots, \beta_p$), com p igual ao número de variáveis utilizadas. Foi utilizado o método de pesquisa passo a passo (*stepwise*) “retroceder” (*backward*) fixando um nível de significância individual inferior a 5%.

Call:
 glm(formula = Y11 ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.6 +
 Comp.9 + Comp.11 + W3 + W12 + Z13, family = binomial(link = logit),
 data = my_data_3[1:871,])

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3,3556	-0,6312	-0,1142	0,6624	2,7426

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0,4572	0,1626	-2,8130	0,0049 **
Comp.1	-0,2842	0,0287	-9,8920	0,0000 ***
Comp.2	0,1976	0,0432	4,5720	0,0000 ***
Comp.3	0,5834	0,0558	10,4550	0,0000 ***
Comp.5	0,2124	0,0542	3,9180	0,0001 ***
Comp.6	-0,4991	0,0576	-8,6620	0,0000 ***
Comp.9	-0,5454	0,0690	-7,9060	0,0000 ***
Comp.11	0,1934	0,0886	2,1820	0,0291 *
W3	0,6111	0,2058	2,9700	0,0030 **
W12	2,9531	1,3082	2,2570	0,0240 *
Z13	-1,4133	0,4420	-3,1970	0,0014 **

 Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
 Null deviance: 1206,9 on 870 degrees of freedom
 Residual deviance: 729,8 on 860 degrees of freedom
AIC: 751,8
 Number of Fisher Scoring iterations: 5

FIGURA 11: Coeficientes estimados do modelo de RL

Call:
 glm(formula = Y11 ~ 1, family = binomial(link = logit),
 data = my_data_3[1:871,])

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1,1550	-1,1550	-1,1550	1,2000	1,2000

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0,0528	0,0678	-0,7790	0,4360

(Dispersion parameter for binomial family taken to be 1)
 Null deviance: 1206,9 on 870 degrees of freedom
 Residual deviance: 1206,9 on 870 degrees of freedom
AIC: 1208,9
 Number of Fisher Scoring iterations: 3

```
> # usando likelihood ratio
> LR <- -2*logLik(my_logit_null) + 2*logLik(my_logit)
> LR
'log Lik.' 477.0511 (df=1)
> pchisq(LR, 2, lower.tail = FALSE)
'log Lik.' 2.568401e-104 (df=1)
```

FIGURA 12: Coeficientes estimados do modelo de RL restrito e teste à significância global do modelo

Na Figura 11 apresenta-se o *output* do R , incluindo as estimativas dos coeficientes de regressão β_j e os testes de significância individual.

Considerando a estimação do modelo sem a remoção de *outliers*, para um nível de significância de 5%, verificou-se que os regressores da variável $W3$ e $W12$ não apresentam significância estatística. Em sentido oposto, os regressores das variáveis $Comp.4$, $Comp.15$, $W2$, $W13$, $Z5$ e $Z10$ apresentam significância estatística ao nível considerado.

Conclui-se que existe significância global da regressão, conforme decorre do *output* do R apresentado na Figura 12, em que o valor- p do teste de rácio de verosimilhança é nulo.

Avaliando a adequação global do modelo através do R^2 de McFadden ou Pseudo- R^2 , obteve-se, no R , um valor de $R^2 = 0,40$ que parece indicar um bom ajuste. No entanto, relembramos que o R^2 apenas pode dar uma indicação da qualidade do ajuste do modelo, sendo difícil interpretar a adequação do modelo em termos absolutos. Sem remoção de *outliers*, o valor de R^2 alcançado foi também de 0,40.

Nas tabelas V e VI, observam-se os intervalos de confiança para os parâmetros.

TABELA V: Intervalos de confiança para os parâmetros a 95%

	2,50%	97,50%
(Intercept)	-0,7796	-0,1414
Comp.1	-0,3422	-0,2295
Comp.2	0,1143	0,2841
Comp.3	0,4776	0,6967
Comp.5	0,1077	0,3206
Comp.6	-0,6151	-0,3889
Comp.9	-0,6840	-0,4132
Comp.11	0,0208	0,3688
W3	0,2110	1,0187
W12	0,0855	5,5011
Z13	-2,3411	-0,5916

TABELA VI: Intervalos de confiança para os parâmetros a 90%

	5%	95%
(Intercept)	-0,7272	-0,1918
Comp.1	-0,3327	-0,2381
Comp.2	0,1275	0,2700
Comp.3	0,4941	0,6779
Comp.5	0,1243	0,3029
Comp.6	-0,5960	-0,4063
Comp.9	-0,6612	-0,4340
Comp.11	0,0484	0,3403
W3	0,2749	0,9525
W12	0,5934	5,0754
Z13	-2,1826	-0,7185

Estimado o modelo podemos obter a previsão da chance de sucesso de uma empresa usando a regressão estimada.

TABELA VII: Chance de sucesso de uma empresa e respetivos intervalos de confiança

	Chance	2,50%	97,50%
(Intercept)	0,633	0,459	0,868
Comp.1	0,753	0,710	0,795
Comp.2	1,219	1,121	1,329
Comp.3	1,792	1,612	2,007
Comp.5	1,237	1,114	1,378
Comp.6	0,607	0,541	0,678
Comp.9	0,580	0,505	0,662
Comp.11	1,213	1,021	1,446
W3	1,842	1,235	2,770
W12	19,166	1,089	244,970
Z13	0,243	0,096	0,553

Aplicando a expressão da função logística (2) à regressão obtida e usando os dados da amostra de treino das variáveis utilizadas nessa regressão, podemos obter a previsão da probabilidade (probabilidade estimada) para cada uma das empresas da amostra de treino ser “bem-sucedida”. A Figura 13 contém uma representação gráfica da probabilidade de uma empresa pertencente à amostra de treino ser “bem-sucedida” ou “malsucedida”. Por sua vez, a Figura 14 representa a mesma probabilidade para as empresas pertencentes à amostra de teste, utilizando o mesmo modelo para o efeito.

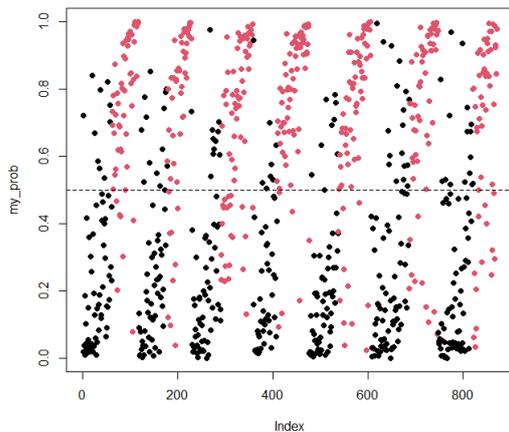


FIGURA 13: Probabilidade de sucesso de uma empresa e sucesso/insucesso efetivo utilizando a amostra de treino

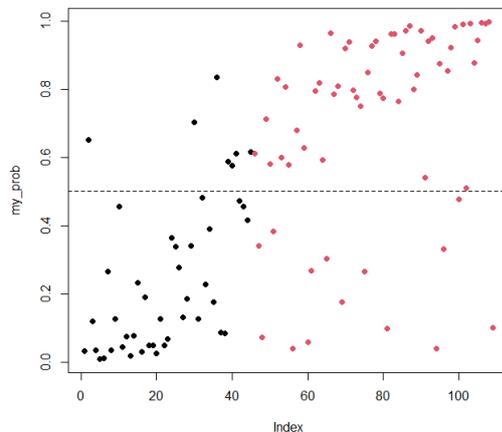


FIGURA 14: Probabilidade de sucesso de uma empresa e sucesso/insucesso efetivo utilizando a amostra de teste

O objetivo agora é classificar. Isto é, dados os valores de uma observação da amostra de treino para cada uma das variáveis utilizadas na RL, estimar a probabilidade de sucesso de uma empresa – para uma classificação *in-sample*; ou, *ceteris paribus*, dados os valores de uma observação da amostra de teste – para uma classificação *out-of-sample*. Se definirmos como “bem-sucedida” uma empresa com probabilidade de sucesso superior 0,5, e “malsucedida”, caso contrário, podemos observar como ficam classificadas as empresas nas nossas amostras. Ora, essa definição corresponde à linha de corte das figuras anteriores. Podemos então verificar nessas figuras “falsos positivos” (pontos negros acima da linha de corte) e “falsos negativos” (pontos vermelhos abaixo da linha de corte), ou seja, todos os valores cuja prestação efetiva da empresa difere da classificação dada pela probabilidade estimada para os intervalos definidos. De outro modo, a parte superior do gráfico corresponde às empresas que classificamos como sendo “bem-sucedidas”, e a parte inferior do gráfico corresponde às empresas que classificamos como sendo “malsucedidas”. No caso da Figura 13, trata-se de uma classificação *in-sample*, ou seja, usou-se a regressão estimada para classificar elementos da amostra de treino. Por sua vez, no caso da Figura 14, trata-se de uma classificação *out-of-sample*, ou seja, usou-se a regressão estimada para classificar elementos fora da amostra.

No sentido proposto, os resultados da classificação podem se sumariados nas seguintes matrizes de classificação (*in-sample* e *out-of-sample*) nas tabelas VIII e IX.

TABELA VIII: RL – Matriz de Classificação – *in-sample*

Grupo	"Malsucedidas"	"Bem-sucedidas"	Total
"Malsucedidas"	361	86	447
"Bem-sucedidas"	78	346	424
Total	439	432	871

TABELA IX: RL – Matriz de Classificação – *out-of-sample*

Grupo	"Malsucedidas"	"Bem-sucedidas"	Total
"Malsucedidas"	38	7	45
"Bem-sucedidas"	14	50	64
Total	52	57	109

Ainda relativamente aos valores das tabelas anteriores, a leitura em linha dá-nos a prestação efetiva das empresas, enquanto a leitura em coluna dá-nos as previsões dessa prestação.

A precisão alcançada *out-of-sample* com a RL foi assim de 80,73%. *In-sample*, essa precisão foi de 81,17%. Sem a remoção de *outliers* estas capacidades preditivas foram de 76,92 % e 81,73%, respetivamente.

À semelhança do que fizemos com a AD, a validação externa da função estimada foi levada a cabo dividindo a amostra total em duas subamostras (as mesmas que anteriormente) mutuamente exclusivas. Ou seja, foi feita a estimação do modelo através da amostra de treino e a classificação utilizando os dados da amostra de teste recorrendo a esse modelo.

Tal como se registou na condução da AD, as precisões *out-of-sample* encontradas foram ligeiramente inferiores à precisão *in-sample* (< 1%), mais uma vez, em linha com o expectável, igualmente suportando a ideia de uma boa capacidade de generalização do modelo estimado visando a sua capacidade preditiva.

5. CONCLUSÕES

Em estudos em que existam muitas variáveis, a ACP pode constituir-se como uma técnica de construção de novas variáveis, assim como, redução do seu número. Além de permitir a definição de novas variáveis não correlacionadas, permite captar os padrões dominantes da variação dos dados num número reduzido de variáveis. Embora, a ACP constitua um excelente método de redução da dimensionalidade das variáveis, os resultados não são muitas vezes direta ou facilmente interpretáveis. Nesse contexto, muito mais que a interpretação dos PC obtidos, centrámo-nos no objetivo último do nosso estudo: prever o nível de sucesso de uma empresa de acordo com o seu ROE. No sentido proposto, a ACP serviu para a criação de 15 novas variáveis (PC), facilitando a aplicação da AD e da RL, eliminando-se a correlação entre as variáveis originais e minimizando-se a perda de informação.

Em consequência, na senda das possibilidades de utilização dos dados na previsão do sucesso ou insucesso das empresas pertencentes ao STOXX 600, avaliadas segundo a rendibilidade dos capitais próprios, discutiu-se, alternativamente, a AD e a RL, duas técnicas com idênticos objetivos, mas com diferentes pressupostos. Ambas as técnicas apresentam uma efetiva facilidade interpretativa e, em última instância, inequívoca aplicabilidade prática.

No presente estudo, verificou-se a violação da hipótese de normalidade multivariada das variáveis independentes, de algum modo comprometendo a aplicabilidade da AD. Em todo o caso, a AD foi concretizada possibilitando a comparação dos resultados com a RL. E, para os dados analisados, foi recomendada a RL, uma vez que permite proceder à relaxação de alguns dos pressupostos da AD, não assumindo quaisquer hipóteses distribucionais sobre os dados. Tal permitiu ainda a consideração de 38 variáveis *dummy* adicionais não utilizadas na ACP e na AD.

Os resultados alcançados confirmaram ligeira vantagem na utilização da RL, tanto *in-sample* como *out-of-sample*, com ou sem a remoção de *outliers*. As precisões *in-sample* com a AD e a RL foram de 80,71% e 81,17%, respetivamente – sem a remoção de *outliers*, 79,59% e 81,73%, respetivamente. Alternativamente, as precisões alcançadas *out-of-sample* com a AD e a RL foram de 79,82% e 80,73%, respetivamente – sem a remoção de *outliers*, 76,15% e 76,92%, respetivamente. Verifica-se assim que,

out-of-sample, a RL superou a AD em apenas 0,91%. Mais expressivo foi o acréscimo verificado com a remoção de *outliers* que permitiu à AD um acréscimo de precisão de 3,67% e à RL de 3,81%.

Este estudo evidencia que com as metodologias utilizadas é possível encontrar uma boa taxa de acerto quanto ao nível da rendibilidade de capitais próprios. Além disso, mostra que as mesmas metodologias podem ser usadas por investidores, particulares, assim como instituições financeiras ou administradores de fundos com vista a melhorar a sua capacidade de prever o sucesso ou insucesso das empresas.

Por um lado, em face dos resultados do presente estudo, embora possamos apresentar uma avaliação relativa das empresas em função da previsão do ROE a um ano à frente, o mesmo não dá resposta à subavaliação ou sobreavaliação dos títulos no mercado de capitais no presente. Nesse sentido, a utilização da metodologia apresentada, visando a compra ou venda de ações que compõem o STOXX 600, baseada em análise fundamental, poderá ser sempre completada com a análise de outros indicadores que possam traduzir uma ligação com o mercado, ajudando a encontrar os melhores pontos de entrada ou saída. Nesse aspeto, conciliar a análise fundamental com a análise técnica poderá desempenhar um papel importante no que toca ao estabelecimento de estratégias lucrativas na composição de uma carteira de ações. Em particular, futuros trabalhos poderão direcionar-se precisamente na simulação de carteiras com base nos resultados do presente estudo. De qualquer forma, o fundamento do presente trabalho encontra-se reforçado na ideia de que o ROE tem um efeito positivo sobre o preço das ações. Esta ideia é corroborada em [60], onde é referido que quanto maior for a rendibilidade de uma empresa, maior será o interesse dos investidores e isso tem um impacto positivo no preço das ações.

Por outro lado, uma vez que optámos por empresas ativas do tipo *corporate* listadas publicamente que compunham parte do STOXX 600 nos anos em análise, trabalhos semelhantes poderão ser desenvolvidos sobre outros índices ou conjuntos de empresas, com referência a outros períodos, com recursos às técnicas que aqui utilizamos. O facto de não termos avançado para a determinação da sobreavaliação ou subavaliação de títulos dota este estudo com um carácter mais generalista, permitindo a utilização das técnicas apresentadas em empresas não cotadas.

Adicionalmente, a existência de manipulação contabilística pode comprometer a capacidade preditiva dos modelos desenvolvidos, na medida em que a informação contida nos valores das variáveis contabilísticas, de forma verdadeira e apropriada, não traduza a posição financeira da empresa numa certa data e o desempenho para o período abrangido. Nesse contexto, a aplicação destes métodos não perde importância, antes pode vir reforçada no sentido de auxiliar propósitos direcionados para a identificação de empresas a serem alvo de auditorias, com vista a explorar eventuais irregularidades, dando suporte aos desvios encontrados.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] R. D. Navas, “Accounting Fundamentals and Volatility in the Euronext 100 index,” Ph.D. dissertation, Universidade da Beira Interior, 2017. [Online]. Available: <http://hdl.handle.net/10400.6/4459> (Accessed Jan. 07, 2022)
- [2] S. H. Penman, *Financial Statement Analysis and Security Valuation*, 4th Ed. New York: McGraw-Hill Irwin, 2010.
- [3] C. A. D. Dosamantes, “The Relevance of Using Accounting Fundamentals in the Mexican Stock Market,” *Journal of Economics, Finance and Administrative Science*, vol. 18, no. S, pp. 2–10, 2013, doi: 10.1016/S2077-1886(13)70024-6.
- [4] V. Bernard and J. Thomas, “Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?,” *Journal of Accounting Research*, vol. 27, pp. 1–36, 1989.
- [5] S. Richardson, I. Tuna and P. Wysocki, “Accounting anomalies and fundamental analysis: A review of recent research advances,” *Journal of Accounting and Economics*, vol. 50, no. 2–3, pp. 410–454, 2010, doi: 10.1016/j.jacceco.2010.09.008.
- [6] S. P. Kothari, “Capital markets research in accounting,” *Journal of Accounting and Economics*, vol. 31, no. 1–3, pp. 105–231, 2001, doi: 10.1016/S0165-4101(01)00030-1.
- [7] E. Bartov, S. Radhakrishnan and I. Krinsky, “Investor sophistication and patterns in stock returns after earnings announcements,” *Accounting Review*, vol. 75, no. 1, pp. 43–63, 2000, doi: 10.2308/accr.2000.75.1.43.
- [8] R. R. Mendenhall, “Arbitrage Risk and Post-Earnings - Announcement Drift,” *Journal of Business*, vol. 77, no. 4, pp. 875–894, 2004, [Online]. Available: <https://www.jstor.org/stable/10.1086/422627>
- [9] J. Ng, T. O. Rusticus and R. S. Verdi, “Implications of Transaction Costs for the Post-Earnings Announcement Drift,” *Journal of Accounting*

- Research*, vol. 46, no. 3, pp. 661–696, 2008, doi: 10.1111/j.1475-679X.2008.00290.x.
- [10] B. Ke and S. Ramalingegowda, “Do institutional investors exploit the post-earnings announcement drift?,” *Journal of Accounting and Economics*, vol. 39, no. 1, pp. 25–53, 2005, doi: 10.1016/j.jacceco.2004.02.002.
- [11] K. Cuthbertson, *Quantitative Financial Economics: Stocks, Bonds and Foreign Exchange*. Chichester: John Wiley & Sons Ltd, 1996.
- [12] J. Caiado, *Métodos de Previsão em Gestão - Com Aplicações em Excel*, 2.^a Edição. Lisboa: Edições Sílabo, Lda, 2016.
- [13] J. S. Abarbanell and B. J. Bushee, “Abnormal returns to a fundamental analysis strategy,” *Accounting Review*, vol. 73, no. 1, pp. 19–45, 1998, doi: 10.2139/ssrn.40740.
- [14] A. W. Lo and A. C. MacKinlay, “Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test,” *Review of Financial Studies*, vol. 1, pp. 41–66, 1988, doi: 10.1515/9781400829095.17.
- [15] T. J. Agrawal, S. Sehgal and V. Vasisht, “Firm Attributes, Corporate Fundamentals and Investment Strategies: An Empirical Study for Indian Stock Market,” *Management and Labour Studies*, vol. 45, no. 3, pp. 366–387, 2020, doi: 10.1177/0258042X20927995.
- [16] M. S. Rozeff and W. R. Kinney, “Capital market seasonality: The case of stock returns,” *Journal of Financial Economics*, vol. 3, no. 4, pp. 379–402, Oct. 1976, doi: 10.1016/0304-405X(76)90028-3.
- [17] M. R. Reinganum, “The anomalous stock market behavior of small firms in January. Empirical tests for tax-loss selling effects,” *Journal of Financial Economics*, vol. 12, no. 1, pp. 89–104, 1983, doi: 10.1016/0304-405X(83)90029-6.

- [18] A. E. Chambers and S. H. Penman, “Timeliness of Reporting and the Stock Price Reaction to Earnings Announcements,” *Journal of Accounting Research*, vol. 22, no. 1, pp. 21–47, 1984, doi: 10.2307/2490700.
- [19] R. Ball and P. Brown, “An Empirical Evaluation of Accounting Income Numbers,” *Journal of Accounting Research*, vol. 6, no. 2, pp. 159–178, 1968.
- [20] J. S. Abarbanell and B. J. Bushee, “Fundamental Analysis, Future Earnings and Stock Prices,” *Journal of Accounting Research*, vol. 35, no. 1, pp. 1–24, 1997, doi: 10.2307/2491464.
- [21] S. H. Penman, “Return to Fundamentals,” *Journal of Accounting, Auditing & Finance*, vol. 7, no. 4, pp. 465–483, 1992, doi: 10.1177/0148558X9200700403.
- [22] J. A. Ou and S. H. Penman, “Financial statement analysis and the prediction of stock returns,” *Journal of Accounting and Economics*, vol. 11, no. 4, pp. 295–329, Nov. 1989, doi: 10.1016/0165-4101(89)90017-7.
- [23] R. Ball and P. Brown, “An Empirical Evaluation of Accounting Income Numbers,” *Journal of Accounting Research*, vol. 6, no. 2, pp. 159–178, 1968, [Online]. Available: <https://doi.org/10.2307/2490232>
- [24] M. T. Leung, H. Daouk and A. S. Chen, “Forecasting stock indices: A comparison of classification and level estimation models,” *International Journal of Forecasting*, vol. 16, no. 2, pp. 173–190, 2000, doi: 10.1016/S0169-2070(99)00048-5.
- [25] F. R. S. Galton, “Co-relations and Their Measurement, Chiefly from Anthropometric Data,” *Studies in individual differences: The search for intelligence*, vol. 45, pp. 135–145, 1888, doi: 10.1037/11491-002.
- [26] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Human Genetics*, vol. 7, pp. 179–188, 1936.
- [27] H. Hotelling, “The Economics of Exhaustible Resources,” *Journal of Political Economy*, vol. 39, no. 2, pp. 137–175, 1931, doi: 10.1007/978-3-030-76753-2_2.

- [28] M. S. Bartlett, “Multivariate Analysis,” *Journal of the Royal Statistical Society*, vol. 9, no. 2, pp. 176–197, 1947, [Online]. Available: <https://www.jstor.org/stable/2984113>
- [29] S. S. Wilks, “Certain Generalizations in the Analysis of Variance,” *Biometrika*, vol. 24, no. 3/4, pp. 471–494, 1932, [Online]. Available: <https://www.jstor.org/stable/2331979>
- [30] E. I. Altman, “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968, doi: 10.1111/j.1540-6261.1968.tb00843.x.
- [31] E. B. Deakin, “A Discriminant Analysis of Predictors of Business Failure,” *Journal of Accounting Research*, vol. 10, no. 1, pp. 167–179, 1972, doi: 10.2307/2490225.
- [32] J. Meliza, I. Sadalia and K. A. Fachrudin, “The Financial Behavior of Investors Using Discriminant Analysis Approach,” *Journal of Management and Business*, vol. 12, no. 2, pp. 216–226, 2015, doi: 10.24123/jmb.v12i2.20.
- [33] J. D. Henning, “Multiple discriminant analysis of the price momentum anomaly and reversal event signals,” Ph.D. dissertation, Northcentral University, Prescott Valley, Arizona, 2016.
- [34] R. N. Freeman, J. A. Ohlson and S. H. Penman, “Book Rate-of-Return and Prediction of Earnings Changes: An Empirical Investigation,” *Journal of Accounting Research*, vol. 20, no. 2, Part II, pp. 639–653, 1982, doi: 10.2307/2490890.
- [35] A. Charitou and G. Panagiotides, “Financial analysis, future earnings and cash flows and the prediction of stock returns: Evidence for the UK,” *Accounting and Business Research*, vol. 29, no. 4, pp. 281–298, 1999, doi: 10.1080/00014788.1999.9729589.
- [36] A. Dutta, G. Bandopadhyay and S. Sengupta, “Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression,”

- International Journal of Business and Information*, vol. 7, no. 1, pp. 105–136, 2012.
- [37] P. Holloway, R. Rochman and M. Laes, “Factors Influencing Brazilian Value Investing Portfolios,” *Journal of Economics, Finance and Administrative Science*, vol. 18, no. S, pp. 18–22, 2013, doi: 10.1016/S2077-1886(13)70026-X.
- [38] M. Mironiuc and M.-A. Robu, “Obtaining a Practical Model for Estimating Stock Performance on an Emerging Market Using Logistic Regression Analysis,” *Procedia - Social and Behavioral Sciences*, vol. 81, pp. 422–427, 2013, doi: 10.1016/j.sbspro.2013.06.454.
- [39] H. Beladi, C. C. Chao and M. Hu, “Another January effect - Evidence from stock split announcements,” *International Review of Financial Analysis*, vol. 44, pp. 123–138, 2016, doi: 10.1016/j.irfa.2016.01.007.
- [40] S. S. Ali, M. Mubeen, I. Lal and A. Hussain, “Prediction of stock performance by using logistic regression model: evidence from Pakistan Stock Exchange (PSX),” *Asian Journal of Empirical Research*, vol. 8, no. 7, pp. 247–258, 2018, doi: 10.18488/journal.1007/2018.8.7/1007.7.247.258.
- [41] C. Ajekwe and A. Ibiameke, “Financial Statement Analysis to Predict Stock Returns of Listed Consumer Goods Firms in Nigeria,” *SSRN Electronic Journal*, vol. 6, no. 1, pp. 27–33, 2018, doi: 10.2139/ssrn.3247805.
- [42] E. Siqueira, T. Otuki and N. da Costa Jr., “Stock return and fundamental variables: A discriminant analysis approach,” *Applied Mathematical Sciences*, vol. 6, no. 113–116, pp. 5719–5733, 2012.
- [43] W. F. Sharpe, “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *The Journal of Finance*, vol. 19, no. 3, pp. 425–442, 1964, doi: 10.1111/j.1540-6261.1964.tb02865.x.
- [44] J. Lintner, “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,” *The Review of*

- Economics and Statistics*, vol. 47, no. 1, pp. 13–37, 1965, [Online].
Available: <https://www.jstor.org/stable/1924119>
- [45] J. Mossin, “Equilibrium in a Capital Asset Market,” *Econometrica*, vol. 34, no. 4, pp. 768–783, 1966, doi: 10.2307/1910098.
- [46] J. Okičić, S. Remetić-Horvath and B. Büyükdemir, “Stock Selection Based on Discriminant Analysis: Case of Capital Market of Bosnia and Herzegovina,” *Journal of Economic and Social Studies*, vol. 4, no. 2, pp. 5–30, 2014, doi: 10.14706/JECOSS11421.
- [47] Maureen. T. Nwakuya and A. Ofokansi, “A discriminant function analysis approach to stock selection in the capital market of Nigeria,” *European Journal of Business, Economics and Accountancy*, vol. 4, no. 7, pp. 79–92, 2016.
- [48] A. Fernando, S. Ufes, L. Claudio and L. Cepead, “Aplicação de Técnicas Estatísticas Multivariadas a Indicadores Contábeis: Uma Contribuição de Evidenciação das Métricas Contábeis,” *ConTexto*, vol. 15, no. 31, 2015.
- [49] N. Baba and M. Kozak, “An intelligent forecasting system of stock price using neural networks,” in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, 1992, pp. 371–377. doi: 10.1109/IJCNN.1992.287183.
- [50] C.-F. Tsai, Y.-C. Lin, D. C. Yen and Y.-M. Chen, “Predicting stock returns by classifier ensembles,” *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 2452–2459, 2011, doi: 10.1016/j.asoc.2010.10.001.
- [51] “ORBIS Europe (Bureau Van Dijk).”
<https://orbiseurope.bvdinfo.com/version-2022714/Orbis4Europe/1/Companies/Search> (accessed Jun. 28, 2022).
- [52] “Regulamento (CE) n.º 1893/2006, do Parlamento Europeu e do Conselho, de 20 de dezembro de 2006,” *Jornal Oficial da União Europeia*, vol. L, no. 393.
- [53] D. Kim and S. K. Kim, “Comparing patterns of component loadings: Principal Component Analysis (PCA) versus Independent Component

- Analysis (ICA) in analyzing multivariate non-normal data,” *Behavior Research Methods*, vol. 44, no. 4, pp. 1239–1243, 2012, doi: 10.3758/s13428-012-0193-1.
- [54] J. D. Jobson, *Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods*. New York, NY: Springer, 1992. doi: 10.1007/978-1-4612-0921-8.
- [55] T. F. G. G. Cova, “Aplicações em quimiometria: do diagnóstico médico à cientometria,” M.S. thesis, FCTUC, Universidade de Coimbra, 2011. [Online]. Available: <http://hdl.handle.net/10316/16193> (Accessed Jan. 07, 2022)
- [56] P. K. Dunn and G. K. Smyth, *Generalized Linear Models With Examples in R*. New York, NY: Springer Nature, 2018. doi: 10.1007/978-1-4419-0118-7.
- [57] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, 2nd ed. New York, NY: Springer Nature, 2021.
- [58] S. Sharma, *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc., 1996.
- [59] P. Filzmoser, K. Hron and C. Reimann, “Interpretation of multivariate outliers for compositional data,” *Computers and Geosciences*, vol. 39, pp. 77–85, 2012, doi: 10.1016/j.cageo.2011.06.014.
- [60] R. Bintara, S. M. Wahyudi and Molina, “Analysis of Fundamental Factors on Stock Price,” *International Journal of Academic Research in Accounting, Finance and Management Sciences*, vol. 9, no. 4, pp. 169–181, 2020, doi: 10.6007/ijarafms/v9-i4/6813.
- [61] R. M. Conroy and R. S. Harris, “Stock Splits and Information: The Role of Share Price,” *Financial Management*, vol. 28, no. 3, pp. 28–40, 1999, doi: 10.2307/3666181.

- [62] S. Menéndez and S. Gómez-Ansón, “Stock splits: motivations and valuation effects in the Spanish market,” *Investigaciones Económicas*, vol. 27, no. 3, pp. 459–490, 2003.
- [63] C. G. Lamoureux and P. Poon, “The Market Reaction to Stock Splits,” *The Journal of Finance*, vol. 42, no. 5, pp. 1347–1370, 1987, [Online]. Available: <https://www.jstor.org/stable/2328531>
- [64] K. Lyroudi, A. Dasilas and A. Varnas, “The valuation effects of stock splits in NASDAQ,” *Managerial Finance*, vol. 32, no. 5, pp. 401–414, 2006, doi: 10.1108/03074350610657427.

ANEXOS

ANEXO 1 - QUADROS

QUADRO II (continua): Desenvolvimento do Quadro I

Variável	Descrição / Fórmula	Tipo
Y11	Medida de Sucesso ROE (ano n+1)	Binary Profitability Ratio
Y01	ROE using Net income (%) (ano n+1)	Profitability Ratio
X011	Total assets th EUR (ano n)	Balance Sheet
X012	Total assets EUR (%) (ano n)/(ano n-1)	
X021	Operating revenue (Turnover) th EUR (ano n)	Profit & Loss Account
X022	Operating revenue (Turnover) EUR (%) (ano n)/(ano n-1)	
X031	P/L before tax th EUR (ano n)	Profit & Loss Account
X032	P/L before tax EUR (%) (ano n)/(ano n-1)	
X041	P/L for period [=Net income] th EUR (ano n)	Profit & Loss Account
X042	P/L for period [=Net income] EUR (%) (ano n)/(ano n-1)	
X051	Cash flow th EUR (ano n)	Profit & Loss Account
X052	Cash flow EUR (%) (ano n)/(ano n-1)	
X061	Non-current liabilities th EUR (ano n)	Balance Sheet
X062	Non-current liabilities EUR (%) (ano n)/(ano n-1)	
X071	Sales th EUR (ano n)	Profit & Loss Account
X072	Sales EUR (%) (ano n)/(ano n-1)	
X081	Operating P/L [=EBIT] th EUR (ano n)	Profit & Loss Account
X082	Operating P/L [=EBIT] EUR (%) (ano n)/(ano n-1)	
X091	Financial P/L th EUR (ano n)	Profit & Loss Account
X092	Financial P/L EUR (%) (ano n)/(ano n-1)	
X101	EBITDA th EUR (ano n)	Profit & Loss Account
X102	EBITDA EUR (%) (ano n)/(ano n-1)	
X111	ROE using P/L before tax (%) (ano n)	Profitability Ratio
X112	ROE using P/L before tax (%) (ano n)/(ano n-1)	
X121	ROCE using P/L before tax (%) (ano n)	Profitability Ratio
X122	ROCE using P/L before tax (%) (ano n)/(ano n-1)	
X131	ROA using P/L before tax (%) (ano n)	Profitability Ratio
X132	ROA using P/L before tax (%) (ano n)/(ano n-1)	
X141	ROE using Net income (%) (ano n)	Profitability Ratio
X142	ROE using Net income (%) (ano n)/(ano n-1)	
X151	ROCE using Net income (%) (ano n)	Profitability Ratio
X152	ROCE using Net income (%) (ano n)/(ano n-1)	
X161	ROA using Net income (%) (ano n)	Profitability Ratio
X162	ROA using Net income (%) (ano n)/(ano n-1)	
X171	EBITDA margin (%) (ano n)	Profitability Ratio
X172	EBITDA margin (%) (ano n)/(ano n-1)	
X181	EBIT margin (%) (ano n)	Profitability Ratio
X182	EBIT margin (%) (ano n)/(ano n-1)	
X191	Profit margin (%) (ano n)	Profitability Ratio
X192	Profit margin (%) (ano n)/(ano n-1)	
X201	Gross margin (%) (ano n)	Profitability Ratio
X202	Gross margin (%) (ano n)/(ano n-1)	
	SE [ROE using Net income (%) (ano n+1) > Mc: 1; 0]	Binary Profitability Ratio
	(Profit before tax / Shareholders funds) * 100	Profitability Ratio
	Total assets (Fixed assets + Current assets)	Balance Sheet
	Total operating revenues (Net sales + Other operating revenues+ Stock variations)	Profit & Loss Account
	Operating profit + financial profit	Profit & Loss Account
	Profit after taxation + Extraordinary and other profit	Profit & Loss Account
	Profit for period + Depreciation	Profit & Loss Account
	Long term liabilities of the company (Long term financial debts + other long term liabilities and provisions)	Balance Sheet
	Net sales	Profit & Loss Account
	Operating revenue - Cost of goods sold - Other Operating Expenses	Profit & Loss Account
	Financial revenue - Financial expenses	Profit & Loss Account
	Operating profit + Depreciation	Profit & Loss Account
	(Profit before tax / Shareholders funds) * 100	Profitability Ratio
	(Profit before tax + Interest paid) / (Shareholders funds + Non-current liabilities) * 100	Profitability Ratio
	(Profit before tax / Total assets) * 100	Profitability Ratio
	(Net income / Shareholder funds) * 100	Profitability Ratio
	(Net income + Interest paid) / (Shareholder funds + Non-current liabilities) * 100	Profitability Ratio
	(Net income / Total Assets) * 100	Profitability Ratio
	(EBITDA / Operating revenue) * 100	Profitability Ratio
	(EBIT / Operating revenue) * 100	Profitability Ratio
	(Profit before tax / Operating revenue) * 100	Profitability Ratio
	(Gross profit / Operating revenue) * 100	Profitability Ratio

QUADRO II (continuação): Desenvolvimento do Quadro I

Variável	Descrição / Fórmula	Tipo
X211 Cash flow / Operating revenue (%) (ano n)	(Cash flow / Operating revenue) * 100	Profitability Ratio
X212 Cash flow / Operating revenue (%) (ano n)/(ano n-1)		Profitability Ratio
X221 Market cap March (ano n+1)/Net Cash from Operating Activities (ano n)	(X)Annual Market Cap (Mar ano n+1) / Net Cash from Operating Activities	Operational Ratio
X231 Net assets turnover (ano n)	Operating revenue / (Shareholders funds + Non current liabilities)	Operational Ratio
X232 Net assets turnover (%) (ano n)/(ano n-1)		Operational Ratio
X241 Interest cover (ano n)	Operating profit / Interest paid	Operational Ratio
X242 Interest cover (%) (ano n)/(ano n-1)		Operational Ratio
X251 Collection period (days) (ano n)	(Debtors / Operating revenue) * 360	Operational Ratio
X252 Collection period (days) (%) (ano n)/(ano n-1)		Operational Ratio
X261 Credit period (days) (ano n)	(Creditors / Operating revenue) * 360	Operational Ratio
X262 Credit period (days) (%) (ano n)/(ano n-1)		Operational Ratio
X271 R&D expenses / Operating revenue (%) (ano n)	(Research & development / Operating revenue) * 100	Operational Ratio
X272 R&D expenses / Operating revenue (%) (ano n)/(ano n-1)		Operational Ratio
X281 Current ratio (ano n)	Current assets / Current liabilities	Structure Ratio
X282 Current ratio (%) (ano n)/(ano n-1)		Structure Ratio
X291 Liquidity ratio (ano n)	(Current assets - Stocks) / Current liabilities	Structure Ratio
X292 Liquidity ratio (%) (ano n)/(ano n-1)		Structure Ratio
X301 Shareholders liquidity ratio (ano n)	Shareholders funds / Non current liabilities	Structure Ratio
X302 Shareholders liquidity ratio (%) (ano n)/(ano n-1)		Structure Ratio
X311 Solvency ratio (Asset based) (%) (ano n)	(Shareholders funds / Total assets) * 100	Structure Ratio
X312 Solvency ratio (Asset based) (%) (ano n)/(ano n-1)		Structure Ratio
X321 Solvency ratio (Liability based) (%) (ano n)	(Shareholders funds / (Non current liabilities + Current liabilities)) * 100	Structure Ratio
X322 Solvency ratio (Liability based) (%) (ano n)/(ano n-1)		Structure Ratio
X331 Gearing (%) (ano n)	((Non current liabilities + Loans) / Shareholders funds) * 100	Structure Ratio
X332 Gearing (%) (ano n)/(ano n-1)		Structure Ratio
X341 Monthly - Market Capitalisation - March th EUR (ano n+1)	Number of outstanding shares * share price of the main security	Annual Stock Data
X342 Monthly - Market Capitalisation - March th % (ano n+1)/(ano n)		Annual Stock Data
X351 Long Term Liabilities per share EUR (ano n+1)	Non-current liabilities/outstanding shares (Mar ano n+1)	Annual Stock Valuation
X361 Price / earnings ratio - close March (ano n+1)	Market price/earnings per share	Annual Stock Valuation
X371 Price / book value ratio - close March (ano n+1)	Market price/book value per share	Annual Stock Valuation
X372 Price / book value ratio - close March % (ano n+1)/(ano n)		Annual Stock Valuation
X381 Market capitalisation March (ano n+1) / Total assets (ano n) (Tobin's Q)	Market capitalization/total assets	Annual Stock Valuation
X391 Earnings yield (%) - close (ano n+1)	(Earnings per share/market price)*100	Annual Stock Valuation
X401 Monthly - Market Capitalisation - March (th) (ano n+1) / shareholders funds (ano n)	Market capitalization/shareholder funds	Annual Stock Valuation
X411 Monthly - Closing price - March EUR (ano n+1)/Cash Flow per Share ratio	Market price/cash flow per share	Annual Stock Valuation
X421 Earnings per share EUR March (ano n+1)	Net income/outstanding shares	Annual Stock Data
X431 Cash flow per share EUR (ano n+1)	Cash flow/outstanding shares	Annual Stock Data
X441 Dividends per share EUR March (ano n+1)	Dividends/outstanding shares	Annual Stock Data
X451 Monthly - Closing price - March EUR (ano n+1)	Market price - march	Annual Stock Data
X452 Monthly - Closing price - March % EUR (ano n+1)/(ano n)		Annual Stock Data
X461 Monthly - Outstanding shares - March th (ano n+1)	Shares outstanding	Cash Flow Statement
X462 Monthly - Outstanding shares - March % (ano n+1)/(ano n)		Cash Flow Statement
X471 Net Cash from Operating Activities th EUR (ano n)	Net Cash from Operating Activities	Balance Sheet
X472 Net Cash from Operating Activities th EUR (ano n)/(ano n-1)		Balance Sheet
X481 Shareholders funds th EUR (ano n)	Total equity (Capital + Other shareholders funds)	Balance Sheet
X482 Shareholders funds % EUR (ano n)/(ano n-1)		Balance Sheet

Quadro II (conclusão): Desenvolvimento do Quadro I

Variável	Descrição / Fórmula	Tipo
W1 A - Agriculture, forestry and fishing		Dummy Variable
W2 B - Mining and quarrying		Dummy Variable
W3 C - Manufacturing		Dummy Variable
W4 D - Electricity, gas, steam and air conditioning supply		Dummy Variable
W5 E - Water supply; sewerage, waste management and remediation activities		Dummy Variable
W6 F - Construction		Dummy Variable
W7 G - Wholesale and retail trade; repair of motor vehicles and motorcycles		Dummy Variable
W8 H - Transportation and storage		Dummy Variable
W9 I - Accommodation and food service activities		Dummy Variable
W10 J - Information and communication		Dummy Variable
W11 K - Financial and insurance activities		Dummy Variable
W12 L - Real estate activities		Dummy Variable
W13 M - Professional, scientific and technical activities		Dummy Variable
W14 N - Administrative and support service activities		Dummy Variable
W15 O - Public administration and defence; compulsory social security		Dummy Variable
W16 Q - Human health and social work activities		Dummy Variable
W17 R - Arts, entertainment and recreation		Dummy Variable
W18 S - Other service activities		Dummy Variable
Z1 Austria		Dummy Variable
Z2 Belgium		Dummy Variable
Z3 Denmark		Dummy Variable
Z4 Finland		Dummy Variable
Z5 France		Dummy Variable
Z6 Germany		Dummy Variable
Z7 Ireland		Dummy Variable
Z8 Isle Of Man (United Kingdom)		Dummy Variable
Z9 Italy		Dummy Variable
Z10 Jersey (United Kingdom)		Dummy Variable
Z11 Luxembourg		Dummy Variable
Z12 Malta		Dummy Variable
Z13 Netherlands		Dummy Variable
Z14 Norway		Dummy Variable
Z15 Poland		Dummy Variable
Z16 Portugal		Dummy Variable
Z17 Spain		Dummy Variable
Z18 Sweden		Dummy Variable
Z19 Switzerland		Dummy Variable
Z20 United Kingdom		Dummy Variable
	NACE Rev. 2 main section	
	Country	

ANEXO 2 - TABELAS

TABELA X: Principais Estatísticas Descritivas

	X012	LN_X02_L	X022	LN_X03_L	X032	LN_X04_L	X042	LN_X05_L	X052	LN_X06_L	X062	LN_X07_L	X072	LN_X08_L	X082	X091	X092	LN_X10_L	X102	LN_X11_L	X112	LN_X12_L	
Mean	16,18517	16,09100	16,27612	16,13664	16,27612	16,09100	16,27612	16,13664	16,27612	16,09100	16,27612	16,13664	16,27612	16,09100	16,27612	16,13664	16,27612	16,09100	16,27612	16,13664	16,27612	16,09100	
Median	16,09100	16,18517	16,27612	16,13664	16,09100	16,18517	16,27612	16,13664	16,09100	16,18517	16,27612	16,13664	16,09100	16,18517	16,27612	16,13664	16,09100	16,18517	16,27612	16,13664	16,09100	16,18517	
Maximum	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108	19,50108
Minimum	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080	12,47080
Std. Dev.	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856	1,257856
Skewness	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264	0,118264
Kurtosis	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016	2,837016
Janque-Bera	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329	8,700329
Probability	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905	0,012905
Sum	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46	18046,46
Sum Sq. Dev.	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963	2053,963
Observations	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115	1115

TABELA XI: Variância das variáveis originais após transformações logarítmicas pertinentes

Variância das Variáveis Originais			
ln.X011.	1,84	ln.X231.	0,53
X012	231,54	X232	181,15
ln.X021.	1,66	ln.X241.	0,74
X022	153,87	X242	3510,30
ln.X031.	1,72	X251	652,97
X032	4491,82	X252	610,84
ln.X041.	1,74	X261	705,91
X042	8682,04	X262	808,75
ln.X051.	1,74	X271	11,81
X052	1134,84	X272	1158,92
ln.X061.	2,38	ln.X281.	0,16
X062	1646,55	X282	413,50
ln.X071.	1,66	ln.X291.	0,15
X072	161,47	X292	583,24
ln.X081.	1,63	ln.X301.	0,40
X082	1368,70	X302	691,88
X091	729501700000,00	ln.X311.	0,10
X092	11067,15	X312	156,07
ln.X101.	1,72	ln.X321.	0,21
X102	547,84	X322	327,07
ln.X111.	0,43	ln.X331.	0,35
X112	3976,65	X332	789,74
ln.X121.	0,27	ln.X341.	1,27
X122	1448,25	X342	1073,46
ln.X131.	0,44	ln.X351.	1,98
X132	4186,88	ln.X361.	0,38
ln.X141.	0,55	ln.X371.	0,51
X142	7562,93	X372	0,10
ln.X151.	0,29	ln.X381.	0,54
X152	2210,74	ln.X391.	0,38
ln.X161.	0,57	ln.X401.	0,51
X162	8348,69	ln.X411.	0,36
ln.X171.	0,37	ln.X421.	1,81
X172	425,04	ln.X431.	1,57
ln.X181.	0,47	ln.X441.	1,62
X182	1187,13	ln.X451.	1,59
ln.X191.	0,56	X452	1027,82
X192	4246,60	ln.X461.	1,84
ln.X201.	0,27	X462	30,38
X202	187,12	ln.X471.	1,79
ln.X211.	0,39	X472	3348,44
X212	1092,16	ln.X481.	1,73
ln.X221.	0,42	X482	382,04

TABELA XII: Componentes Principais

Componente Principal	Valor Próprio: Variância	Variância Acumulada	Proporção da Variância	Proporção Acumulada	Componente Principal	Valores próprios:	
						Componente Principal	Análise Paralela de Horn
1	14,68	14,68	17%	17%	V1		1,60
2	12,13	26,81	14%	31%	V2		1,56
3	8,22	35,03	10%	41%	V3		1,53
4	5,24	40,27	6%	47%	V4		1,50
5	4,51	44,78	5%	52%	V5		1,48
6	4,35	49,13	5%	57%	V6		1,46
7	3,72	52,85	4%	61%	V7		1,44
8	3,37	56,22	4%	65%	V8		1,42
9	3,23	59,45	4%	69%	V9		1,40
10	2,99	62,44	3%	73%	V10		1,38
11	2,24	64,68	3%	75%	V11		1,37
12	2,04	66,72	2%	78%	V12		1,35
13	1,88	68,59	2%	80%	V13		1,34
14	1,51	70,11	2%	82%	V14		1,32
15	1,39	71,50	2%	83%	V15		1,31
16	1,22	72,72	1%	85%	V16		1,29
17	1,06	73,78	1%	86%	V17		1,28
18	1,03	74,80	1%	87%	V18		1,26
19	0,96	75,77	1%	88%	V19		1,25
20	0,92	76,69	1%	89%	V20		1,24
21	0,86	77,56	1%	90%	V21		1,23
22	0,85	78,40	1%	91%	V22		1,21
23	0,76	79,16	1%	92%	V23		1,20
24	0,75	79,91	1%	93%	V24		1,19
25	0,70	80,62	1%	94%	V25		1,18
26	0,67	81,29	1%	95%	V26		1,16
27	0,57	81,86	1%	95%	V27		1,15
28	0,48	82,34	1%	96%	V28		1,14
29	0,46	82,80	1%	96%	V29		1,13
30	0,43	83,24	1%	97%	V30		1,12
31	0,37	83,61	0%	97%	V31		1,11
32	0,31	83,92	0%	98%	V32		1,09
33	0,30	84,22	0%	98%	V33		1,08
34	0,27	84,49	0%	98%	V34		1,07
35	0,23	84,72	0%	99%	V35		1,06
36	0,20	84,92	0%	99%	V36		1,05
37	0,16	85,08	0%	99%	V37		1,04
38	0,14	85,22	0%	99%	V38		1,03
39	0,13	85,35	0%	99%	V39		1,02
40	0,12	85,47	0%	99%	V40		1,01
41	0,09	85,56	0%	99%	V41		1,00
42	0,08	85,63	0%	100%	V42		0,99
43	0,07	85,70	0%	100%	V43		0,98
44	0,04	85,74	0%	100%	V44		0,97
45	0,04	85,79	0%	100%	V45		0,96
46	0,04	85,82	0%	100%	V46		0,95
47	0,04	85,86	0%	100%	V47		0,94
48	0,03	85,89	0%	100%	V48		0,93
49	0,03	85,91	0%	100%	V49		0,92
50	0,02	85,93	0%	100%	V50		0,91
51	0,01	85,94	0%	100%	V51		0,90
52	0,01	85,95	0%	100%	V52		0,89
53	0,01	85,96	0%	100%	V53		0,88
54	0,01	85,97	0%	100%	V54		0,87
55	0,01	85,98	0%	100%	V55		0,86
56	0,01	85,98	0%	100%	V56		0,85
57	0,00	85,99	0%	100%	V57		0,84
58	0,00	85,99	0%	100%	V58		0,84
59	0,00	85,99	0%	100%	V59		0,83
60	0,00	86,00	0%	100%	V60		0,82
61	0,00	86,00	0%	100%	V61		0,81
62	0,00	86,00	0%	100%	V62		0,80
63	0,00	86,00	0%	100%	V63		0,79
64	0,00	86,00	0%	100%	V64		0,78
65	0,00	86,00	0%	100%	V65		0,77
66	0,00	86,00	0%	100%	V66		0,76
67	0,00	86,00	0%	100%	V67		0,75
68	0,00	86,00	0%	100%	V68		0,74
69	0,00	86,00	0%	100%	V69		0,73
70	0,00	86,00	0%	100%	V70		0,72
71	0,00	86,00	0%	100%	V71		0,71
72	0,00	86,00	0%	100%	V72		0,70
73	0,00	86,00	0%	100%	V73		0,69
74	0,00	86,00	0%	100%	V74		0,68
75	0,00	86,00	0%	100%	V75		0,67
76	0,00	86,00	0%	100%	V76		0,66
77	0,00	86,00	0%	100%	V77		0,65
78	0,00	86,00	0%	100%	V78		0,64
79	0,00	86,00	0%	100%	V79		0,63
80	0,00	86,00	0%	100%	V80		0,62
81	0,00	86,00	0%	100%	V81		0,61
82	0,00	86,00	0%	100%	V82		0,60
83	0,00	86,00	0%	100%	V83		0,59
84	0,00	86,00	0%	100%	V84		0,57
85	0,00	86,00	0%	100%	V85		0,56
86	0,00	86,00	0%	100%	V86		0,54

ANÁLISE MULTIVARIADA NA PREVISÃO DA RENDIBILIDADE DOS CAPITAIS PRÓPRIOS

TABELA XIII: *Loadings*

Variável	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
In.X011.	0,253	0,024	0,010	0,008	0,049	0,003	0,058	0,052	0,036	0,039	0,005	0,023	0,004	0,012	0,009
X012	-0,047	-0,010	0,016	-0,122	0,012	-0,014	-0,216	0,361	0,020	0,136	0,076	0,137	-0,143	0,064	0,044
In.X021.	0,221	0,022	0,010	0,151	0,144	-0,063	0,048	0,079	-0,008	0,056	-0,002	0,019	0,030	0,012	0,015
X022	-0,061	0,053	-0,016	-0,049	-0,033	-0,064	-0,156	0,117	0,176	0,415	-0,094	-0,128	0,062	0,064	-0,017
In.X031.	0,214	0,100	0,125	0,049	0,035	-0,041	0,050	0,065	0,054	0,029	-0,015	0,015	-0,022	-0,008	-0,019
X032	-0,023	0,212	-0,155	-0,051	0,078	0,019	-0,016	-0,032	0,027	0,043	-0,117	0,151	-0,118	-0,071	-0,055
In.X041.	0,204	0,119	0,129	-0,059	0,011	-0,045	0,043	0,066	0,035	0,003	-0,012	-0,005	-0,025	0,043	0,030
X042	-0,016	0,214	-0,156	-0,041	0,034	0,019	-0,007	-0,006	0,012	-0,031	-0,043	0,039	-0,045	0,158	0,272
In.X051.	0,239	0,065	0,069	0,028	0,035	-0,033	0,065	0,061	0,029	0,006	0,012	-0,010	0,029	0,013	0,037
X052	-0,036	0,187	-0,093	-0,039	-0,029	0,008	-0,086	0,122	0,004	-0,019	0,176	-0,141	0,245	0,161	0,176
In.X061.	0,250	0,016	0,007	-0,076	0,025	0,009	0,049	0,042	0,000	0,034	-0,027	0,008	0,024	0,007	-0,003
X062	-0,024	-0,024	0,004	-0,129	0,107	-0,066	-0,222	0,312	0,032	-0,040	-0,040	0,068	-0,120	-0,022	0,152
In.X071.	0,220	0,021	0,011	-0,153	0,145	-0,064	0,048	0,080	-0,007	0,058	-0,005	0,019	0,030	0,012	0,016
X072	-0,061	0,042	-0,016	-0,055	-0,033	-0,015	-0,151	0,105	0,177	0,421	-0,107	-0,128	0,043	0,060	-0,005
In.X081.	0,227	0,083	0,100	0,023	0,039	-0,038	0,070	0,058	0,047	0,038	-0,016	0,020	0,014	-0,013	-0,023
X082	-0,035	0,215	-0,109	-0,029	0,067	0,009	-0,028	0,086	0,049	0,115	-0,014	0,005	-0,017	-0,125	-0,220
X091	-0,088	0,009	0,027	0,082	0,014	0,034	-0,080	-0,003	-0,003	-0,051	-0,021	0,014	-0,101	0,019	0,025
X092	0,004	-0,006	0,009	-0,028	0,029	-0,015	-0,052	0,088	0,097	0,061	0,076	-0,072	0,063	-0,106	-0,151
In.X101.	0,242	0,059	0,067	0,016	0,044	-0,031	0,071	0,056	0,038	0,026	0,000	0,003	0,033	-0,011	0,003
X102	-0,049	0,172	-0,078	-0,049	0,015	0,001	-0,106	0,188	0,057	0,113	0,165	-0,138	0,245	-0,045	-0,186
In.X111.	-0,060	0,145	0,221	0,000	0,030	-0,182	-0,052	-0,036	-0,133	0,045	-0,039	-0,020	-0,031	-0,030	-0,068
X112	-0,010	0,202	-0,152	-0,057	0,120	-0,014	-0,016	-0,116	0,065	-0,011	-0,135	0,120	-0,090	-0,078	-0,065
In.X121.	-0,102	0,139	0,210	0,111	0,017	-0,146	-0,015	-0,017	-0,034	0,024	-0,003	-0,012	-0,023	-0,026	-0,039
X122	-0,025	0,225	-0,123	0,000	0,056	0,004	0,033	-0,124	0,071	0,057	-0,038	0,017	0,004	-0,072	-0,116
In.X131.	-0,095	0,149	0,229	0,081	-0,030	-0,088	-0,020	0,022	0,033	-0,024	-0,040	-0,019	-0,053	-0,042	-0,055
X132	-0,013	0,213	-0,158	-0,029	0,075	0,018	0,021	-0,100	0,019	0,013	-0,134	0,130	-0,093	-0,083	-0,055
In.X141.	-0,069	0,163	0,202	0,018	-0,016	-0,168	-0,058	-0,030	-0,152	-0,007	-0,028	-0,053	-0,033	0,064	0,025
X142	-0,006	0,209	-0,154	-0,044	0,069	-0,006	-0,010	-0,072	0,044	-0,070	-0,057	0,011	-0,021	0,157	0,265
In.X151.	-0,107	0,151	0,186	0,109	-0,017	-0,139	-0,015	-0,022	-0,075	-0,016	0,007	-0,042	-0,004	0,056	0,055
X152	-0,018	0,224	-0,120	0,008	0,005	0,004	0,019	-0,077	0,024	-0,035	0,008	-0,089	0,071	0,135	0,162
In.X161.	-0,099	0,165	0,206	0,088	-0,067	-0,084	-0,029	0,021	-0,004	-0,066	-0,029	-0,051	-0,052	0,053	0,036
X162	-0,008	0,213	-0,158	-0,022	0,031	0,018	0,019	-0,058	0,006	-0,058	-0,058	0,019	-0,022	0,154	0,276
In.X171.	0,052	0,080	0,123	-0,285	-0,210	0,066	0,051	-0,048	0,098	-0,064	0,006	-0,035	0,007	-0,049	-0,025
X172	-0,018	0,163	-0,082	-0,027	0,031	0,011	-0,029	0,145	-0,056	-0,157	0,273	-0,063	0,255	-0,098	-0,218
In.X181.	0,007	0,114	0,166	-0,240	-0,198	0,047	0,040	-0,042	0,103	-0,035	-0,025	0,001	-0,031	-0,047	-0,071
X182	-0,014	0,209	-0,111	-0,014	0,082	0,012	0,021	0,053	-0,024	-0,054	0,042	0,078	-0,037	-0,167	-0,246
In.X191.	-0,006	0,137	0,202	-0,174	-0,185	0,035	0,006	-0,022	0,108	-0,046	-0,023	-0,007	-0,090	-0,035	-0,058
X192	-0,010	0,206	-0,155	-0,041	0,086	0,017	0,010	-0,054	-0,014	-0,056	-0,086	0,196	-0,130	-0,095	-0,063
In.X201.	0,032	0,045	0,106	-0,246	-0,131	0,071	0,040	-0,054	0,072	-0,033	0,013	0,052	0,031	0,081	-0,018
X202	0,002	0,044	-0,030	-0,007	0,018	0,008	0,009	0,111	-0,075	-0,171	0,274	0,005	0,215	-0,074	-0,163
In.X211.	0,050	0,093	0,125	-0,252	-0,222	0,060	0,038	-0,034	0,078	-0,102	0,029	-0,060	-0,001	0,002	0,049
X212	-0,013	0,170	-0,093	-0,022	-0,021	0,015	-0,031	0,083	-0,075	-0,211	0,238	-0,088	0,243	0,125	0,167
In.X221.	-0,153	0,009	0,104	-0,038	0,127	0,094	0,092	0,051	0,074	0,042	0,049	0,169	0,121	0,021	0,083
In.X231.	-0,091	-0,006	-0,005	0,302	0,206	-0,152	-0,038	0,028	-0,110	0,053	0,019	0,001	0,034	0,012	0,016
X232	-0,014	0,053	-0,038	0,072	-0,048	-0,014	0,053	-0,283	0,184	0,291	-0,055	-0,182	0,179	0,061	-0,087
In.X241.	-0,109	0,101	0,172	0,150	0,034	-0,059	-0,013	0,052	0,098	-0,035	0,006	0,015	-0,067	-0,075	-0,038
X242	-0,026	0,179	-0,071	0,024	0,042	0,031	0,072	-0,005	-0,017	0,003	0,015	0,098	-0,056	-0,080	-0,126
X251	-0,009	-0,013	0,033	-0,041	-0,002	0,045	0,001	0,000	-0,036	-0,028	-0,056	0,176	0,023	0,532	-0,253
X252	-0,001	-0,006	0,016	-0,062	-0,001	-0,001	-0,027	0,076	-0,050	-0,099	0,072	0,135	-0,129	-0,011	-0,189
X261	0,074	-0,010	-0,033	-0,011	-0,010	-0,016	-0,005	-0,020	-0,040	0,013	0,121	0,075	-0,139	0,288	-0,112
X262	0,008	-0,008	0,032	-0,038	0,037	-0,043	-0,019	0,041	-0,047	-0,086	0,127	0,152	-0,228	0,101	-0,097
X271	0,016	0,030	0,078	-0,033	0,051	0,058	0,085	0,006	0,061	-0,015	-0,062	0,154	0,134	0,141	0,034
X272	-0,007	-0,004	0,004	0,031	0,010	0,005	0,003	0,010	-0,046	-0,122	0,082	0,101	-0,007	0,005	-0,037
In.X281.	-0,074	-0,014	0,015	0,074	0,061	0,112	0,038	0,144	-0,005	-0,122	-0,342	-0,019	0,164	0,251	-0,236
X282	-0,015	0,018	-0,001	-0,047	0,014	0,094	0,059	0,209	-0,167	-0,129	-0,355	-0,184	0,067	-0,218	0,104
In.X291.	-0,040	-0,012	0,021	0,021	0,036	0,118	0,032	0,143	-0,068	-0,115	-0,357	-0,018	0,176	0,298	-0,263
X292	-0,014	0,014	-0,004	-0,049	0,025	0,098	0,055	0,218	-0,147	-0,134	-0,345	-0,191	0,067	-0,224	0,091
In.X301.	-0,105	0,017	0,014	0,287	-0,019	0,080	0,037	0,070	0,249	-0,069	0,075	0,032	-0,072	-0,002	0,038
X302	-0,026	0,078	-0,021	0,141	-0,201	0,117	0,201	-0,064	-0,138	0,186	0,115	0,047	0,034	0,039	-0,050
In.X311.	-0,074	0,013	0,020	0,166	-0,122	0,188	0,064	0,117	0,337	-0,140	-0,004	0,002	-0,046	-0,023	0,026
X312	-0,017	0,074	-0,019	0,115	-0,222	0,144	0,197	0,073	-0,264	0,144	0,002	0,070	0,003	-0,036	0,057
In.X321.	-0,075	0,015	0,021	0,163	-0,125	0,187	0,068	0,115	0,335	-0,142	-0,006	0,003	-0,042	-0,025	0,024
X322	-0,019	0,079	-0,018	0,116	-0,223	0,150	0,198	0,074	0,257	0,137	0,002	0,066	0,004	-0,036	0,064
In.X331.	0,106	-0,019	-0,020	-0,282	0,013	-0,091	-0,049	-0,089	-0,254	0,085	-0,045	-0,040	0,048	0,010	-0,044
X332	0,007	-0,066	0,034	-0,155	0,196	-0,152	-0,224	0,068	0,153	-0,140	-0,009	-0,041	-0,018	0,027	0,048
In.X341.	0,192	0,065	0,144	-0,008	0,125	0,010	0,147	0,075	0,080	0,051	0,012	0,082	0,060	0,005	0,066
X342	-0,076	0,021	0,001	-0,087	0,172	0,124	0,238	0,110	-0,053	0,057	0,127	-0,294	-0,278	0,098	-0,039
In.X351.	0,153	0,006	0,040	-0,050	0,079	0,294	-0,196	-0,086	-0,094	0,039	0,004	-0,028	-0,003	-0,009	-0,021
In.X361.	-0,084	-0,134	-0,011	-0,141	0,203	0,115	0,175	-0,004	0,072	0,087	0,046	0,159	0,163	-0,083	0,057
In.X371.	-0,143	0,053	0,200	-0,104	0,160	-0,074	0,093	-0,034	-0,094	0,068	0,010	0,082	0,106	-0,005	0,077
X372	-0,047	-0,007	0,004	-0,096	0,253	0,069	0,226	-0,039	0,024	-0,047	0,094	-0,344	-0,223	0,093	-0,051
In.X381.	-0,172	0,057	0,203	-0,028	0,102	0,010	0,119	0,018	0,056	0,005	0,009	0,082	0,084	-0,015	0,085
In.X391.	0,084	0,134	0,011	0,141	-0,203	-0,115	-0,175	0,004	-0,072	-0,087	-0,046	-0,159	-0,163	0,083	-0,057
In.X401.	-0,144	0,053	0,200	-0,103	0,160	-0,074	0,093	-0,034	-0,095	0,069	0,011	0,084	0,107		