

INSTITUTO SUPERIOR DE CIÊNCIAS DO TRABALHO E DA EMPRESA
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DE LISBOA

DEPARTAMENTO DE FINANÇAS
DEPARTAMENTO DE MATEMÁTICA



Ciências
ULisboa

ISCTE  Business School
Lisbon University Institute

Modelos de Notação de Risco de Crédito – *Rating* de Empresas

Mestrado em Matemática Financeira

Raquel Ramos Martins

Dissertação supervisionada por: Professora Doutora Diana Mendes

2018

Agradecimentos

Em primeiro lugar queria agradecer à professora Diana Mendes por ter aceite ser minha orientadora e pela disponibilidade, apoio e motivação dados para a conclusão desta tese.

Ao GCA, um agradecimento também especial por ter permitido a disponibilização dos dados e, em particular a toda a equipa da Direcção de Risco Global pelo apoio e ajuda que me deram.

Queria também agradecer com carinho a toda minha família e amigos, mas especialmente à minha mãe, ao meu pai, ao meu irmão, ao meu avô e ao meu namorado por estarem presentes em todos os momentos da minha vida e terem tornado o meu percurso académico muito mais fácil com o vosso apoio, paciência e suporte.

Abstract

In the sequence of the main economic crises in the financial sector over the past few years, there has been a bigger intervention by supervisory authorities aimed at risk prevention and mitigation, as well as credit risk.

Therefore, with the emergence of increasingly demanding regulations, namely the Basel agreements that condition the quality and amount of credit risk that Credit Institutions can be exposed through the establishment of standard limits, arose the need to build and implement more rigorous/effective internal models.

In this sense, both from a credit granting perspective and from a portfolio monitoring perspective, the existence of credit rating models is essential in order to assist the Financial Institution's good risk management.

This thesis aims to approach the concepts of credit risk and credit rating models, as well as to develop a rating model, using a logistic regression model, in order to measure the companies ability to accomplish with their responsibilities.

Keywords: Financial Mathematics, Econometrics, Credit Risk, Rating of Companies, Logistic Regression.

Resumo

Na sequência das principais crises económicas sentidas no sector financeiro tem-se vindo a observar ao longo dos últimos anos uma maior intervenção das autoridades de supervisão que visa a prevenção e mitigação do risco, entre os quais o risco de crédito.

Assim, com o surgimento de regulamentações cada vez mais exigentes nomeadamente os acordos de Basileia que condicionam a qualidade e a quantidade de risco de crédito a que as Instituições de Crédito podem estar expostas através do estabelecimento de limites *standard*, criou-se a necessidade de construir e implementar modelos internos mais rigorosos/ eficazes.

Neste sentido, quer numa óptica de concessão de crédito quer numa óptica de acompanhamento da carteira é fundamental a existência de modelos de notação de risco de forma a auxiliar a boa gestão de risco da Instituição Financeira.

Com a presente tese pretende-se abordar os conceitos de risco de crédito e de modelos de notação de risco, bem como desenvolver um modelo de *rating*, recorrendo a um modelo de regressão logística, de forma a medir a capacidade das empresas cumprirem com as suas responsabilidades.

Palavras-chave: Matemática Financeira, Econometria, Risco de Crédito, *Rating* de Empresas, Regressão Logística.

Abreviaturas

AUROC - *Area Under Receiving Operating Characteristic*;

BCBS – Comité de Basileia de Supervisão Bancária (*Basel Committee on Banking Supervision*);

BCE - Banco Central Europeu;

BdP - Banco de Portugal;

CAE - Código de Actividade Económica;

CRR - Regulamentação de Requisitos de Capital (*Capital Requirements Regulation*);

CVA - Ajuste de Avaliação de Crédito (*Credit Valuation Adjustment*);

DA - Análise Discriminante (*Discriminant Analysis*);

EAD - Exposição ao Incumprimento (*Exposure at Default*);

EL - Perda Esperada (*Expected Losses*);

GCA - Grupo Crédito Agrícola;

GPME - Grandes e Pequenas e Médias Empresas;

IF - Instituição Financeira;

IRB - Internal Rating Based;

KS - Kolmogorov-Smirnov;

LGD - Perda dado o Incumprimento (*Loss Given Default*);

MAR - Modelo de Avaliação de Riscos;

MDA - Análise Discriminante Múltipla (*Multiple Discriminant Analysis*);

OLS - Mínimos Quadrados Ordinários (*Ordinary Least Squares*);

PD - Probabilidade de Incumprimento (*Probability of Default*);

PIT - *Point in Time*;

PME - Pequenas e Médias Empresas;

ROC - *Receiving Operating Characteristic*;

RWA - Activos Ponderados pelo Risco (*Risk Weighing Assets*);

TTC - *Through the cycle*;

VN - Volume de Negócios.

Conteúdo

1. Introdução	1
1.1. Enquadramento	2
2. Risco de Crédito	9
2.1. Modelos de <i>Rating</i>	10
2.1.1. Metodologias de <i>Rating</i>	11
2.1.2. Modelos Estatísticos	12
2.1.2.1. Modelo de Regressão Linear	12
2.1.2.2. Modelo de Análise Discriminante	13
2.1.2.3. Modelo <i>Hazard</i>	13
2.1.2.4. Modelo de Regressão Logística	14
2.1.2.5. Comparação entre os Modelos Estatísticos	16
2.1.3. Modelos Heurísticos	17
2.1.4. Modelos Híbridos	18
3. Desenvolvimento dos Modelos de <i>Rating</i> - Parte Teórica	19
3.1. Selecção do Modelo	19
3.2. Metodologia PIT vs Metodologia TTC	19
3.3. Testes Estatísticos	20
3.3.1. Teste de Kolmogorov-Smirnov	20
3.3.2. Teste de Kruskal-Wallis	21
3.3.3. Teste de contingência	22

Conteúdo

3.3.4. Coeficiente de Kendall Tau-b	23
3.3.5. Análise de Componentes Principais	24
3.4. Critérios de Selecção do Modelo	25
3.4.1. Curva de ROC	25
3.4.2. Kolmogorov-Smirnov	26
3.4.3. Critério de Informação de Akaike	27
3.4.4. Critério de Informação de Schwarz	27
3.5. Testes de Significância	28
3.5.1. Teste de Wald	28
3.5.2. Teste de Score	28
4. Desenvolvimento dos Modelos de <i>Rating</i> - Parte Prática	31
4.1. Análise e construção de variáveis	31
4.1.1. Análise da qualidade das variáveis	31
4.2. Selecção da Amostra	32
4.3. Análise Univariada	33
4.4. Análise Multivariada	35
4.5. Categorização das variáveis	36
4.6. Resumo da Selecção de Variáveis	36
4.7. Aplicação da Regressão Logística	37
4.8. <i>Performance</i> do modelo	38
4.9. Integração da Informação Qualitativa e Quantitativa	39
4.10. Níveis de Risco	39
4.11. Interpretação dos Resultados	40
5. Conclusão	45
Bibliografia	46
A. Anexo A	49

Lista de Figuras

1.1. Os 3 pilares de Basileia II.	5
1.2. Sistematização de Basileia III.	6
1.3. Sequência cronológica da evolução dos acordos de Basileia.	7
2.1. <i>Rating</i> Final - Combinação do <i>Rating</i> Quantitativo com o <i>Rating</i> Qualitativo	18
3.1. Gráfico ilustrativo com curvas ROC.	26
4.1. Divisão da Amostra	33
4.2. Testes de capacidade discriminante	34
4.3. Testes de correlação e tendência	34
4.4. Testes de análise multivariada	35
4.5. Resumo do número de variáveis.	37
4.6. Variáveis selecionadas para cada modelo	37
4.7. Medidas de <i>performance</i> para os modelos de <i>rating</i>	39
4.8. Níveis de risco pelos intervalos de PD.	41
4.9. Distribuição da carteira por níveis de risco.	41
4.10. Distribuição da carteira por modelo e por níveis de risco agregado.	42

Lista de Tabelas

A.1. Descrição das variáveis	55
--	----

1. Introdução

Nos dias de hoje, dada a complexidade dos mercados financeiros e as exigências regulamentares cada vez mais rígidas e abrangentes devem ser implementados mecanismos que auxiliem a boa gestão do ciclo de crédito, desde a concessão ao acompanhamento até à maturidade, por forma, a prevenir o risco a que as Instituições Financeiras se encontram expostas.

Das várias medidas aconselhadas/ impostas pelo regulador, no âmbito da presente tese, salienta-se a existência de modelos de notação de risco que permitem a monitorização do perfil de risco da carteira de crédito.

Os modelos de notação de risco têm como principal objectivo a identificação de variáveis/ características relacionadas com a previsibilidade de um mutuário não ter capacidade de assumir os seus compromissos financeiros, ou seja, corresponde a uma estimativa da sua probabilidade de incumprimento. Portanto, faz sentido primeiramente segmentar a carteira em grupos homogéneos (i.e. por tipo de cliente, tipologia de crédito, etc.) e posteriormente construir modelos adequados/ ajustados a cada segmento.

O objectivo desta tese é construir um modelo que permita classificar as empresas em níveis de risco de acordo com a probabilidade de incumprimento (PD) associada.

Desde cedo que têm vindo a ser estudados modelos de previsão de falência de acordo com os indicadores financeiros das empresas. O primeiro trabalho nesta área foi introduzido por Fitzpatrick (1932) onde através de métodos de observação foi estabelecido um padrão entre os indicadores e o desempenho da empresa. No

entanto, foi na década de 60 que foram publicados dois artigos de Beaver (1966) e de Altman (1968) em que se passou a considerar técnicas estatísticas para estimar a previsão de falência das empresas (ver [16]).

Estes artigos impulsionaram o interesse de outros autores, tendo-se iniciado uma crescente pesquisa referente a este tema, não só a nível teórico como também a nível empírico (ver [16]).

Na literatura científica as revisões mais relevantes neste âmbito podem ser encontradas em Zavgren (1985), Altman (1983), Jones (1987), Altman e Narayanan (1997), Altman e Saunders (1998) e Balcaena e Oogh (2006) (ver [16]) onde as principais metodologias enunciadas para obter uma classificação de crédito que quantifica o risco associado a um cliente são: modelos de regressão linear, modelos de análise discriminantes e modelos *Logit/ Probit*.

A estrutura desta tese é a seguinte: A secção 1.1, apresenta uma visão geral sobre a regulamentação, acordos de Basileia, que foi a base para a introdução de modelos de notações de risco no sector financeiro. O Capítulo 2 apresenta uma breve introdução ao risco de crédito enunciando os principais conceitos e inicia o tema em foco nesta tese, os modelos de *rating*. No que diz respeito aos modelos de *rating* são descritas as principais metodologias e modelos utilizados para a quantificação do nível de risco atribuído às empresas. O Capítulo 3, dá início à parte empírica deste trabalho, onde são apresentadas as noções teóricas relevantes para o desenvolvimento de um modelo de notação de risco. No Capítulo 4, que corresponde à parte prática do desenvolvimento dos modelos de *rating* será estruturado o processo de aplicação destes modelos e enunciados alguns resultados. Por fim, no Capítulo 5 são apresentadas as conclusões deste estudo.

1.1. Enquadramento

As instituições financeiras, em particular as instituições de crédito “cuja atividade consiste em receber do público depósitos ou outros fundos reembolsáveis e em conce-

der crédito por conta própria” (Decreto-Lei n.o 23-A/2015 de 26 de Março de 2015) estão expostas a um conjunto significativo de riscos que podem afetar diretamente a sua capacidade de solvabilidade e liquidez. Neste sentido, e também devido às crises económicas dos últimos anos têm vindo a ser implementadas medidas de supervisão cada vez mais complexas e exigentes cujo objectivo consiste em uniformizar as normas a nível Europeu de modo a auxiliar na mensuração e monitorização do perfil de risco das Instituições.

O Comité de Supervisão Bancária de Basileia (BCBS) tem sido a principal referência na definição da regulamentação prudencial, destacando-se os acordos de Basileia I, Basileia II e Basileia III.

Para garantir a estabilidade financeira é necessário implementar medidas que minimizem a probabilidade de falência do sistema bancário, sendo a definição rigorosa de requisitos de capital importante nesta matéria, dado que garantem uma melhor absorção de perdas inesperadas e impulsionam a uma melhor gestão do risco a que as Instituições estão expostas.

Neste contexto, de forma a uniformizar a concorrência entre as instituições financeiras dos diversos países surgiu, em Julho de 1988, o acordo de Basileia I onde foram estabelecidos níveis mínimos de capital assegurando que os bancos não concediam crédito em proporções superiores às que tinham capacidade de suportar em caso de incumprimento das contrapartes.

Neste acordo foram introduzidos dois grandes conceitos, o de capital e o de activos ponderados com base no risco de incumprimento associado (RWA). Sendo que o Capital surge dividido em duas partes:

- Capital Principal (*Tier I*): Corresponde ao capital com maior capacidade efectiva de absorção de perdas, preservando a continuidade da IF.

- Capital Suplementar (*Tier II*): Corresponde ao capital com capacidade de absorção de perdas em caso de se constatar inviabilidade do funcionamento da IF.

Assim o capital elegível deve ser igual ou superior a 8% dos activos ponderados

pelo risco, ou seja, foi estabelecido o Rácio de Solvabilidade da seguinte forma:

$$\frac{\text{Total de Fundos Próprios}}{\text{Risco de Crédito}} \geq 8\% \quad (1.1)$$

As medidas impostas pelo acordo de Basileia I verificaram-se insuficientes na prevenção do risco, não tendo garantido a estabilidade financeira pretendida no sector bancário, uma vez que não evitou as várias crises financeiras que ocorreram no passado.

Desta forma, de modo a contornar as limitações existentes no primeiro acordo de Basileia, em particular a consideração exclusiva do risco de crédito para o cálculo dos requisitos mínimos de capital, foi revista a regulamentação e, em Junho de 2004, surge um novo acordo designado por Basileia II com entrada em vigor apenas em 2007.

O acordo de Basileia II teve como principal objectivo adequar os requisitos de capital de forma a incorporar os riscos a que as Instituições Financeiras se encontram expostas, nomeadamente o risco de crédito, o risco de mercado e o risco operacional e introduzir o conceito de transparência e divulgação de informação através da Disciplina de Mercado para que exista uma maior comparabilidade entre as Instituições no que diz respeito à gestão entre o risco e a rentabilidade.

Relativamente ao Rácio de Solvabilidade apesar de se manter a definição presente no acordo anterior, isto é, o capital elegível deve ser igual ou superior a 8% dos activos ponderados pelo risco a expressão passou a ter a seguinte notação:

$$\frac{\text{Total de Fundos Próprios}}{\text{Risco de Crédito} + \text{Risco de Mercado} + \text{Risco Operacional}} \geq 8\% \quad (1.2)$$

O cálculo dos RWA passou a ter em consideração não só o risco de crédito como também o risco de mercado e o risco operacional.

Basileia II veio estruturar a regulamentação bancária em três pilares que, de

forma geral, dizem respeito à mensuração, monitorização e prevenção do risco e, posteriormente, à divulgação da informação com base na transparência de resultados.

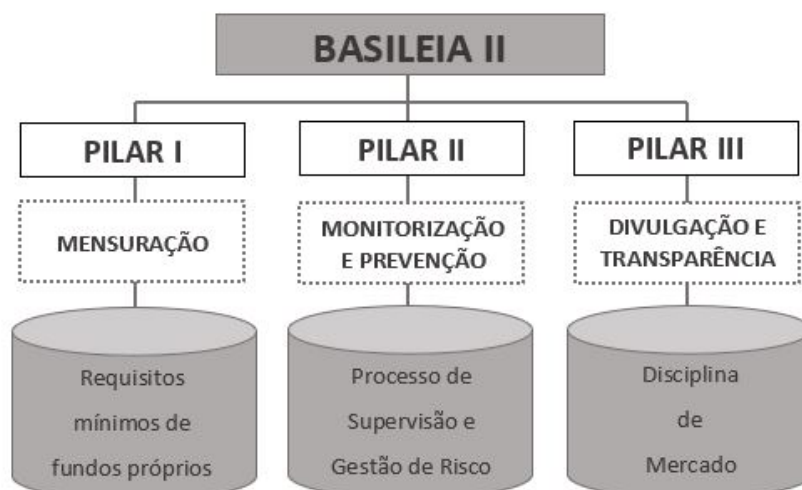


Figura 1.1. Os 3 pilares de Basileia II.

O primeiro Pilar surgiu com o intuito de tornar o regime prudencial mais sensível ao risco, alterando as regras de cálculo de acordo com a equação 1.2, ou seja, prevê a determinação de requisitos de fundos próprios para risco de crédito, risco de mercado e risco operacional. O segundo Pilar teve como objectivo introduzir o conceito de processo de supervisão, ou seja, fortalecer a interação entre as instituições supervisionadas e os respectivos supervisores. Assim, as instituições devem seguir as recomendações do supervisor no que diz respeito a estratégias, processos e mecanismos de controlo, enquanto que as autoridades de supervisão (BdP e BCE) são responsáveis por produzir as referidas recomendações. O terceiro Pilar foi criado por forma a tornar transparente a divulgação de informação ao público relativamente à solvabilidade e ao perfil de risco da Instituição, (Banco de Portugal, [7]).

Para a determinação dos requisitos de capital foram introduzidos dois novos riscos, o risco de mercado e o risco operacional, no entanto foi sobre o risco de crédito, já existente no antigo acordo, que incidiram as maiores alterações relativamente às metodologias utilizadas. Assim sendo, passaram a existir três possíveis metodologias

(metodologia *standard*, metodologia IRB *foundation* e metodologia IRB *advanced*) que serão explicadas em detalhe mais à frente no Capítulo 2.

Embora tenham sido efectuadas melhorias ao acordo de Basileia I com o intuito de melhorar o comportamento das Instituições perante cenários adversos do mercado, os requisitos definidos no acordo de Basileia II não foram suficientemente conservadores para assegurar a estabilidade do sector bancário pelo que em 2008 assistimos a uma grande crise financeira, pelo que houve novamente uma revisão da regulamentação, dando origem, em Setembro de 2010 ao acordo de Basileia III.

O acordo de Basileia III não teve como objectivo alterar a metodologia de cálculo dos requisitos de capital, no entanto introduziu uma definição mais estrita de capital regulamentar e definiu, pela primeira vez, requisitos de liquidez e de alavancagem. Foram também adicionados requisitos de constituição de reservas de fundos próprios, quer de natureza estrutural, quer de natureza contracíclica, por forma a fortalecer a estabilidade do sistema financeiro.

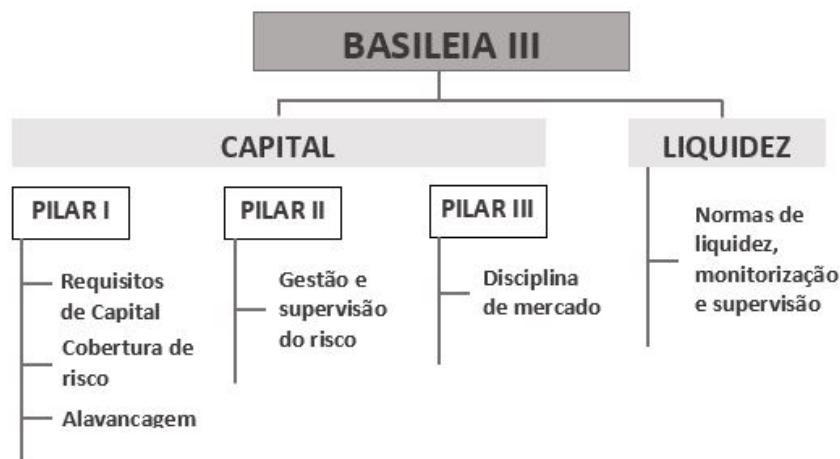


Figura 1.2. Sistematização de Basileia III.

Em conclusão, ao longo dos anos tem vindo a ser necessário reformular os acordos de Basileia quer por serem consideradas metodologias desactualizadas face à constante evolução dos sistemas/ mercados financeiros quer devido ao estabelecimento de padrões que na prática se revelaram insuficientes para manter o equilíbrio no

sector financeiro.

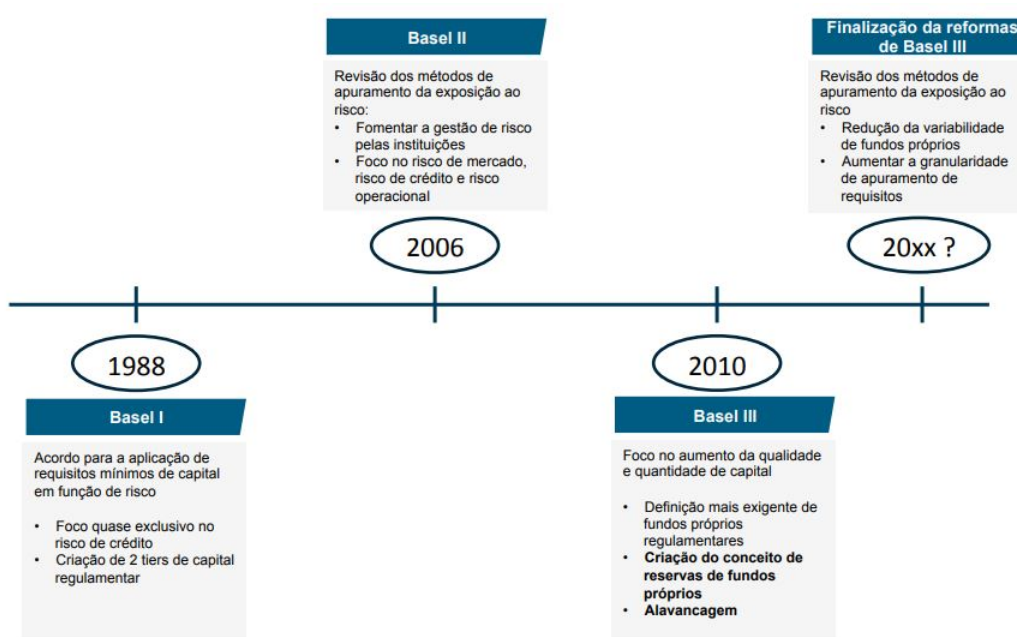


Figura 1.3. Sequência cronológica da evolução dos acordos de Basileia.¹

Tendo em consideração o já mencionado relativamente à regulamentação em vigor bem como a extrema necessidade das IF quantificarem/ avaliarem o risco, em particular o risco de crédito, a que estão expostas torna-se claro a implementação de modelos internos de classificação de risco que serão objecto de estudo nos capítulos seguintes.

¹Fonte: https://www.bportugal.pt/sites/default/files/anexos/documentos-relacionados/intervpub20170508_1.pdf

2. Risco de Crédito

O Risco de Crédito é definido no MAR como sendo a “probabilidade de ocorrência de impactos negativos nos resultados ou no capital, devido à incapacidade de uma contraparte cumprir os seus compromissos financeiros perante a instituição, incluindo possíveis restrições à transferência de pagamentos do exterior. O risco de crédito existe, principalmente, nas exposições em crédito (incluindo o titulado), linhas de crédito, garantias e derivados.”.

De acordo com [12], para Bessis (2010) o risco de crédito é o risco mais importante no sector bancário e pode-se dividir em várias componentes, entre as quais:

- a) O risco de incumprimento, que corresponde ao risco de um mutuário não cumprir com as suas responsabilidades, resultando num evento de incumprimento;
- b) O risco de concentração, que é o risco de ocorrerem perdas derivadas da concentração de créditos elevados a um pequeno número de clientes, ou seja, diz respeito a uma diversificação inadequada da carteira;
- c) O risco de degradação da garantia, que é o risco associado à desvalorização do colateral no mercado.

Desta forma, uma vez que o risco de incumprimento diz respeito à probabilidade de um mutuário não ter capacidade de assumir os seus compromissos financeiros, para avaliar o risco associado a cada cliente com base na sua perda esperada (EL) tem de se ter em consideração os seguintes parâmetros de risco:

- i) A Probabilidade de incumprimento (PD) que diz respeito à estimativa da probabilidade de ocorrência de um incumprimento num determinado horizonte temporal.

Normalmente são considerados dois horizontes temporais, um a 12 meses e outro correspondente à vida útil da operação (*lifetime*);

ii) A Exposição em incumprimento (EAD) que corresponde ao montante total a que a instituição está exposta no momento em que se observa o incumprimento por parte do mutuário;

iii) A Perda dado o incumprimento (LGD) que representa a estimativa da perda resultante de um incumprimento.

Neste sentido, através dos parâmetros de risco anteriormente mencionados é possível calcular a perda esperada de um cliente utilizando a seguinte expressão:

$$EL = EAD \times PD \times LGD. \quad (2.1)$$

Assim, para se determinar o risco de crédito de cada cliente da carteira é necessário estimar os factores de risco associados, em particular a PD, que deve variar de acordo com a distribuição e evolução das notações de *rating/ scoring* atribuídas preferencialmente por modelos internos (IRB). No âmbito desta tese será estudado um modelo de notação de risco para clientes do tipo empresa, ou seja, serão utilizados modelos de *Rating*.

2.1. Modelos de *Rating*

Neste capítulo pretende-se aprofundar o conceito de *Rating* bem como apresentar os diversos modelos estatísticos existentes para a construção e estimação de modelos de *rating*.

A notação de *rating* corresponde a uma classificação que visa caracterizar o nível de risco de uma determinada contraparte de acordo com a sua probabilidade de incumprimento.

A monitorização do risco representado por cada cliente da carteira de crédito é fundamental pois permite às Instituições Financeiras terem uma avaliação da qua-

lidade dos seus clientes tanto na concessão de crédito como no acompanhamento da carteira e auxilia ainda noutros processos como por exemplo, no cálculo do capital regulamentar, no modelo de capital económico e no cálculo de imparidade.

Os modelos estatísticos usam métodos econométricos para classificar os mutuários de acordo com o seu risco, através da selecção das variáveis explicativas que devem fornecer uma boa estimativa relativamente à qualidade financeira do cliente. Pelo contrário, os modelos estruturais baseiam-se em modelos económicos, portanto consideram conexões casuais entre as variáveis em vez de estudar a sua correlação.

2.1.1. Metodologias de *Rating*

Por forma a prevenir o risco de incumprimento, conforme já mencionado foram definidas em Basileia II e mantidas em Basileia III, três possíveis metodologias:

i) A metodologia *standard*, em que os factores de risco são atribuídos de acordo com o rating externo fornecido pelas agências de notação.

ii) A metodologia IRB *foundation*, em que a instituição pode determinar a PD através de modelos internos, no entanto os restantes parâmetros de risco são fornecidos pela entidade reguladora.

iii) A metodologia IRB *advanced*, em que todos os parâmetros de risco podem ser calculados pela Instituição, através de modelos internos.

Assim, a abordagem IRB permitiu que os bancos utilizassem os seus próprios modelos de *rating* na estimação da probabilidade de incumprimento (PD), desde que fossem cumpridos os requisitos mínimos estabelecidos.

Apesar da existência destas três abordagens, a diversificação dos mercados e a evolução do conhecimento neste ramo condicionam cada vez mais as IF a utilizarem metodologias IRB *advanced* onde os parâmetros de risco são estimados em função de cada carteira e da sua respectiva segmentação em grupos homogéneos.

2.1.2. Modelos Estatísticos

A presente tese incide sobre o desenvolvimento de modelos de notação de risco para empresas e, para tal, foram definidos conceitos estatísticos e econométricos já presentes na literatura (ver por exemplo [2], [16] e [10]).

A nível académico já foram abordados temas semelhantes ao em análise, embora na generalidade direcionados para outras vertentes (ver por exemplo [1], [9] e [3]).

2.1.2.1. Modelo de Regressão Linear

O modelo de regressão ([2], Engelman e Rauhmeier, 2006) estabelece uma relação linear entre as características do mutuário e a variável de incumprimento:

$$y_i = \beta' \cdot x_i + u_i, \quad (2.2)$$

onde y_i é uma variável binária que indica se o mutuário está em incumprimento ($y = 1$) ou não ($y = 0$), x_i é um vector coluna com as características do cliente observadas antes do incumprimento, β é um vector coluna de parâmetros que captam o impacto de uma mudança das características, x_i , na variável de incumprimento e, por fim u_i corresponde à variável residual que contém a variação não captada pelas características x_i .

O procedimento usual é estimar (2.2) através do método dos mínimos quadrados (OLS) que produz uma estimativa b do vector β que pode ser utilizada para calcular a classificação (*score*) do cliente, da seguinte forma:

$$S_i = E(y_i|x_i) = b' \cdot x_i \quad (2.3)$$

ou seja, a classificação representa o valor esperado da variável incumprimento sabendo as características do cliente.

A variável S_i é contínua, portanto, o resultado da equação (2.3) não pode ser analisado como uma probabilidade de incumprimento. Desta forma a classificação

deve ser interpretada considerando que valores mais altos de S_i implicam maior risco de incumprimento.

2.1.2.2. Modelo de Análise Discriminante

A análise discriminante ou análise discriminante múltipla ([16], Trueck e Rachev, 2008) é baseada na combinação linear de duas ou mais variáveis independentes com o intuito de separar grupos previamente definidos, neste caso empresas em incumprimento e empresas em cumprimento, de acordo com determinadas características.

Neste caso em que se tem apenas dois grupos a análise da função discriminante também pode ser considerada como uma regressão múltipla, resultando assim de combinações lineares através da seguinte expressão:

$$Z = w_0 + w_i \cdot X_i, \quad (2.4)$$

onde Z é a variável discriminante, que representa a classificação (*score*) do cliente, w_0 é uma constante, w_i (com $i = 1, 2, \dots, n$) são os coeficientes discriminantes e X_i (com $i = 1, 2, \dots, n$) as variáveis independentes.

2.1.2.3. Modelo Hazard

O modelo *Hazard* ([2], Engelmann e Rauhmeier, 2006) é o modelo de regressão mais geral porque não assenta sobre pressupostos à cerca da natureza ou forma da distribuição de sobrevivência.

Este modelo assume que a taxa *Hazard* corresponde a uma função das variáveis independentes. Desta forma, a função *Hazard* pode ser representada através da seguinte expressão:

$$h_i(t|x_i) = h_0(t) \cdot e^{\beta' \cdot x_i}, \quad (2.5)$$

onde $h_i(t|x_i)$ representa o resultado da função *Hazard*, tendo em consideração as

covariáveis dos mutuários e o seu respectivo tempo de sobrevivência e a variável $h_0(t)$ representa o risco associado ao cliente quando todos os valores das variáveis independentes são iguais a zero.

Não existem pressupostos sobre a função *Hazard*, porém a equação (2.5) especifica uma relação multiplicativa entre a função *Hazard* e a função log-linear das variáveis explicativas.

2.1.2.4. Modelo de Regressão Logística

O modelo de regressão logística ([10], Witzany, 2017) é utilizado para quantificar a probabilidade de ocorrência de uma variável binária, neste caso o incumprimento, de acordo com as variáveis explicativas seleccionadas, características dos mutuários.

Neste sentido a variável não observada, é calculada da seguinte forma:

$$y^* = \beta' \cdot x_i + u_i, \quad (2.6)$$

onde x_i é um vector coluna com as características do mutuário, β é um vector coluna de parâmetros que captam o impacto de uma mudança das características, x_i , na variável de incumprimento e, por fim u_i corresponde à variável residual.

A variável de incumprimento é obtida através da equação (2.6), isto é:

$$y = \begin{cases} 1 & \text{se } y_i^* \geq 0 \\ 0 & \text{caso contrário} \end{cases} \quad (2.7)$$

Assim, a probabilidade de incumprimento condicionada resulta da expressão seguinte:

$$p_i = P[y_i = 1|x_i] = P[\beta' \cdot x_i + u_i \leq 0] = F[-\beta' \cdot x_i], \quad (2.8)$$

sendo $F(\cdot)$ a função distribuição cumulativa associada à variável residual, u_i . Esta

função transforma a classificação:

$$S_i = \beta' \cdot x_i = \beta_0 + \sum_{i=1}^n \beta_i x_i, \quad (2.9)$$

que toma valores entre $] - \infty, +\infty[$ na respectiva probabilidade de incumprimento, ou seja pertence ao intervalo $[0,1]$.

Quando a variável residual, u_i segue uma distribuição normal, ou seja, $F(\cdot) = \Phi(\cdot)$ deve-se aplicar o modelo *probit*, no entanto, se for assumido que os resíduos têm uma distribuição logística tem-se:

$$F_i(x) = p_i = \Lambda(x) = \frac{e^x}{1 + e^x} = \frac{1}{e^{-x} + 1}. \quad (2.10)$$

A equação anterior pode ser reescrita como a probabilidade de acontecimentos positivos sobre a probabilidade de acontecimentos negativos (*odds*):

$$\begin{aligned} p_i = \Lambda(x) &\Leftrightarrow \\ \Leftrightarrow \frac{1 - p_i}{p_i} = e^{\beta' \cdot x_i} &\Leftrightarrow \\ \Leftrightarrow \ln\left(\frac{1 - p_i}{p_i}\right) = \beta' \cdot x_i & \end{aligned} \quad (2.11)$$

onde:

$$odds = \frac{\text{probabilidade de acontecimentos bons (não incumprimento)}}{\text{probabilidade de acontecimentos maus (incumprimento)}} = \frac{1 - p_i}{p_i} \quad (2.12)$$

Neste contexto não é possível quantificar especificamente a probabilidade de incumprimento de um determinado mutuário. É, no entanto, possível agrupar a amostra em grupos com características, x_i , semelhantes e posteriormente calcular a probabilidade de incumprimento de cada grupo através do método dos mínimos quadrados mas esta abordagem envolve alguns problemas, nomeadamente o reduzido número de observações.

Posto isto, a metodologia mais utilizada para estimar os parâmetros do modelo *Logit* é a estimativa de máxima verossimilhança.

2.1.2.4.1 Método de Máxima Verossimilhança

O método de máxima verossimilhança estima os coeficientes de regressão que maximizam a probabilidade de encontrar os acontecimentos observados da variável dependente, que é o incumprimento.

Assumindo que as observações são independentes e identicamente distribuídas, a função de verossimilhança que é maximizada é dada por:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n F(-b' \cdot x_i)^{y_i} (1 - F(-b' \cdot x_i))^{1-y_i} \\ &= \prod_{i=1}^n p_i^{y_i} \cdot (1 - p_i)^{1-y_i} \end{aligned} \quad (2.13)$$

Por simplicidade em vez de se usar a função de verossimilhança, enunciada anteriormente, utiliza-se o logaritmo da função verossimilhança, isto é:

$$\begin{aligned} \ln L(\beta) &= \sum_{i=1}^n [y_i \cdot \ln F(-b' \cdot x_i) + (1 - y_i) \cdot \ln (1 - F(-b' \cdot x_i))] \\ &= \sum_{i=1}^n [y_i \cdot \ln p_i + (1 - y_i) \cdot \ln (1 - p_i)] \end{aligned} \quad (2.14)$$

2.1.2.5. Comparação entre os Modelos Estatísticos

A técnica estatística implícita nos modelos de regressão linear é relativamente simples de aplicar, desde que estejam disponíveis os dados sobre os clientes (variáveis x_i), no entanto, o facto da probabilidade estimada de incumprimento não se restringir apenas ao intervalo $[0,1]$ faz com que o resultado obtido não tenha sentido quer a nível estatístico quer a nível económico.

Desta forma, o modelo de regressão linear não é frequentemente utilizado, uma vez que, existem outros modelos estatísticos capazes de combater a desvantagem

anteriormente mencionada, nomeadamente o modelo de regressão logística que limita a probabilidade de incumprimento, estimada pelo modelo de regressão linear, ao intervalo $[0,1]$.

Relativamente ao modelo MDA, uma vez que temos apenas dois grupos, empresas em incumprimento e empresas que não estão em incumprimento, esta metodologia é equivalente a aplicar-se um modelo de regressão linear, como tal pelo motivo enunciado anteriormente também não é um modelo muito utilizado.

Em comparação com o modelo de regressão logística, o modelo *Hazard*, apresenta a mesma função de probabilidade e a mesma matriz de variâncias/ covariâncias pelo que as duas estimativas são similares. A principal diferença é que o modelo *Hazard* observa as variações das covariáveis ao longo do tempo e considera observações anuais resultando em estimativas mais eficientes.

Neste sentido, de acordo com o estudo efectuado por Abdullah *et al.* (2008) verificou-se que o modelo *Logit* nem sempre apresenta um quadro real das características das empresas quando comparando com os resultados obtidos através do modelo *Hazard* que tem em consideração o espaço temporal. No entanto, em contrapartida, o modelo *Logit* é apresentado como o melhor na previsão da variável dependente, o incumprimento, amostra de validação. Em conclusão este estudo indica o modelo de regressão logística como o que permite obter resultados de maior confiança, no caso da Malásia, e o modelo *Hazard* no caso de países ocidentais.

Portanto, o melhor modelo a aplicar depende da amostra em análise, sendo que para Portugal de acordo com as práticas observadas no sector bancário (*benchmarking*), o modelo mais utilizado para a atribuição de notações de risco a empresas é o modelo de regressão logística.

2.1.3. Modelos Heurísticos

Os modelos heurísticos são modelos qualitativos que dependem da avaliação subjectiva de cada analista de crédito. Desta forma, é atribuído um *rating* qualitativo

com base no conhecimento e experiência passada do analista (*expert judgement*) sobre determinados factores, tais como o risco do sector, a competitividade e a qualidade de gestão.

A forma mais geral para a quantificação qualitativa do *rating* é através do preenchimento de um questionário sendo atribuída uma ponderação a cada resposta. E, de acordo com a ponderação final obtida no questionário resulta uma notação de risco.

2.1.4. Modelos Híbridos

Os modelos híbridos representam uma combinação entre os modelos estatísticos e os modelos heurísticos, abrangendo as duas vertentes, qualitativa e quantitativa, num só modelo.

Estes são possivelmente os modelos mais utilizados devido à sua capacidade de tratar dois tipos de informação de formas diferentes, ficando a componente qualitativa a cargo de analistas e a componente quantitativa a cargo de modelos estatísticos.

Neste sentido, o *rating* final corresponde à soma de uma determinada percentagem, p , do *rating* qualitativo e de uma percentagem, $1 - p$, do *rating* quantitativo.

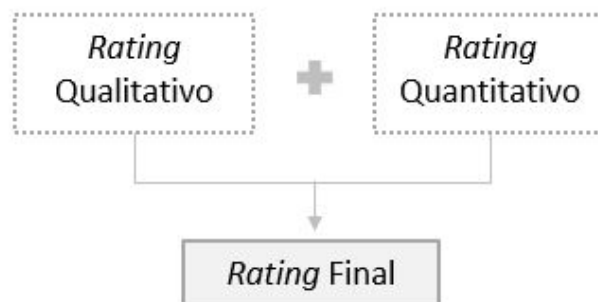


Figura 2.1. *Rating* Final - Combinação do *Rating* Quantitativo com o *Rating* Qualitativo

3. Desenvolvimento dos Modelos de *Rating* - Parte Teórica

3.1. Selecção do Modelo

Para o desenvolvimento dos modelos de *rating* foi escolhido um modelo híbrido que equilibra a informação qualitativa com a informação quantitativa num único *rating*. A informação qualitativa, que resulta num *rating* qualitativo será obtida através das bases de dados do GCA, enquanto que a informação quantitativa será quantificada, no próximo capítulo, com base no modelo de regressão logística.

3.2. Metodologia PIT vs Metodologia TTC

Existem duas possíveis abordagens para a quantificação do *rating*, sendo uma *point-in-time* (PIT) e outra *through-the-cycle* (TTC) (Witzany, 2017).

A metodologia PIT descreve a qualidade de crédito com base num determinado horizonte temporal pré-definido, onde normalmente as observações dos anos mais recentes são mais ponderadas do que as observações dos anos mais antigos.

A metodologia TTC descreve a qualidade de crédito ao longo da sua vida útil, ou seja, até à maturidade, tendo em consideração os ciclos económicos. Neste contexto, a estimativa TTC não será condicionada pela alteração da qualidade de crédito devido a variáveis macroeconómicas desde que estas espelhem o comportamento do mercado.

Neste caso, por forma a reduzir a volatilidade associada à metodologia PIT onde em períodos de recessão, a distribuição das classificações desloca-se para a direita (agrava os níveis de risco) e em períodos de maior liquidez a distribuição das classificações desloca-se para a esquerda (melhora os níveis de risco), o modelo desenvolvido teve por base uma abordagem TTC. Assim, é expectável que não existam oscilações nas notações de risco, a menos que efectivamente ocorram alterações nas características das empresas que conduzam a um aumento da probabilidade de incumprimento e conseqüentemente a um aumento da notação de risco.

3.3. Testes Estatísticos

3.3.1. Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov (KS) é utilizado para testar a igualdade entre duas distribuições, ou seja, é comparada a distribuição da amostra com a distribuição assumida.

As hipóteses podem ser definidas da seguinte forma:

- $H_0 : F = F_0$
- $H_1 : F \neq F_0$

Onde a hipótese nula corresponde à igualdade entre a distribuição empírica dos dados e a distribuição assumida para os dados e a hipótese alternativa assume a negação da igualdade das distribuições.

Assumindo que a amostra de dimensão n está ordenada por ordem crescente, a função distribuição empírica $F_n(x)$ corresponde à proporção de valores menores ou iguais a x e tem a seguinte expressão:

$$F_n(x) = \frac{\#A}{n} = \begin{cases} 0 & \text{se } x \leq x_1 \\ \frac{i}{n} & \text{se } x_i \leq x \leq x_{i+1} \\ 1 & \text{se } x \geq x_{i+1} \end{cases}, \text{ em que } A = \{x_i, i=1, \dots, n : x_i \leq x\} \quad (3.1)$$

sendo a estatística de KS, que mede a distância entre os gráficos das duas funções, dada por:

$$D_n = \text{Sup}_{x \in R} |F_n(x) - F_x| \quad (3.2)$$

Considera-se um nível de significância de 1% e portanto a hipótese nula, (H_0), do teste de KS vai ser rejeitada se *p-value* inferior a 1%.

3.3.2. Teste de Kruskal-Wallis

O teste de Kruskal-Wallis avalia se as funções de distribuição das amostras são iguais, através do seguinte teste de hipóteses:

- $H_0 : F_1 = F_2 = \dots = F_k$
- $H_1 : \exists F_i \neq F_j \text{ com } i, j = 1, 2, \dots, k$

Portanto, a hipótese nula significa que as k amostras têm a mesma distribuição, enquanto que a hipótese alternativa diz que existe pelo menos uma amostra com distribuição diferente das outras.

Para se aplicar este método é necessário ordenar de forma crescente as N observações das k amostras.

Neste caso o teste estatístico utilizado é dado por:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (3.3)$$

onde R_i corresponde à soma das posições para a amostra i (com $i = 1, \dots, k$), n_i é o número de observações da amostra i (com $i = 1, \dots, k$) e N é o número total de observações das k amostras.

De forma semelhante ao teste de KS, a hipótese nula, (H_0), vai ser rejeitada para um valor de *p-value* inferior a 1%.

3.3.3. Teste de contingência

O teste de contingência será utilizado para analisar a relação entre variáveis categóricas, desta forma, para uma tabela de dimensão $k \times g$, com N observações, deve-se primeiramente calcular o seguinte:

- A frequência relativa observada em cada linha L_i

$$L_i = \frac{\sum_{j=1}^g o_j}{N}, \text{ com } i = 1, \dots, k \quad (3.4)$$

onde $\sum_{j=1}^g o_j$ corresponde à soma dos valores observados da linha i da tabela em análise.

- A frequência relativa observada em cada coluna C_j

$$C_j = \frac{\sum_{i=1}^k o_i}{N}, \text{ com } j = 1, \dots, g \quad (3.5)$$

sendo $\sum_{i=1}^k o_i$ equivalente à soma dos valores observados da coluna j da tabela em análise.

- A frequência esperada de cada combinação (L_i, C_j)

$$E_{i,j} = \frac{L_i \times C_j}{N} \quad (3.6)$$

Resultando assim a estatística de teste da relação entre a frequência observada e a frequência esperada, de acordo com a expressão que se segue:

$$\chi^2 = \sum_{j=1}^g \sum_{i=1}^k \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \approx \chi_{(k-1)(g-1), \alpha}^2 \quad (3.7)$$

O coeficiente de Cramer, que é uma medida baseada na estatística chi-quadrado (χ^2), assume valores entre 0 e 1, que correspondem à força de associação entre as

variáveis. Este coeficiente pode ser calculado através da seguinte expressão:

$$V = \sqrt{\frac{\frac{\chi^2}{N}}{\min(k-1, g-1)}} \quad (3.8)$$

Desta forma, foi considerado que as variáveis discretas apresentam uma boa capacidade discriminante quando:

- 1- Nos testes de χ^2 obtém-se um *p-value* inferior a 1%;
- 2- O coeficiente de Cramer apresenta um valor superior a 0.1.

3.3.4. Coeficiente de Kendall Tau-b

O coeficiente de Kendall Tau-b (τ) é um teste de hipóteses não paramétrico, utilizado como medida de correlação.

Primeiramente é necessário criar um par ordenado, (X_i, Y_i) , com cada combinação possível das v.a. X e Y . E, posteriormente tem de se calcular as observações $(X_i - X_j, Y_i - Y_j)$ para cada par i e j .

Quando o par ordenado, $(X_i - X_j, Y_i - Y_j)$, tem o mesmo sinal na abcissa e na ordenada diz-se que este é concordante, caso contrário o mesmo diz-se discordante.

Desta forma, o coeficiente de Kendall é calculado através da seguinte expressão:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}, \quad (3.9)$$

sendo que as variáveis n_c e n_d dizem respeito, respectivamente, ao número de pares concordantes e discordantes. O denominador corresponde ao número total de combinações de pares, o que implica que este coeficiente está restrito ao intervalo $[-1,1]$.

Assim, quando a equação (3.9) é zero indica que existe ausência de relação, ou seja, as variáveis são independentes.

Neste teste são consideradas as seguintes hipóteses estatísticas:

- $H_0 : \tau = 0$

- $H_1 : \tau \neq 0$

No caso em que a análise é univariada a hipótese nula, (H_0), vai ser rejeitada para um valor de *p-value* inferior a 1%.

Para o caso em que a análise é multivariada, considera-se uma forte correlação entre as variáveis quando o coeficiente de Kendall é superior a 0,5.

3.3.5. Análise de Componentes Principais

A análise de componentes principais (ACP) tem como principal objectivo agrupar as variáveis que contêm as mesmas informações.

A análise de *clusters* foi efectuada através do seguinte algoritmo:

1. Para cada variável x_1, \dots, x_k são calculados os componentes principais (F_k) através da maximização das seguintes funções:

$$\begin{aligned}
 F_1 &= \text{Max}_{||w||=1} \text{Var}(w_1 x_1 + w_k x_k) \\
 &\dots \\
 F_k &= \text{Max}_{||w||=1} \text{Var}(F_{k-1} - (w_1 x_1 + w_k x_k))
 \end{aligned}
 \tag{3.10}$$

Os componentes principais são calculados, iterativamente, até F_k a partir da qual já não existe variabilidade a ser explicada.

2. Agrupamento das variáveis de forma iterativa. Primeiramente, aloca-se as k variáveis em L sub-grupos (G_1, \dots, G_L) mutuamente exclusivos. De seguida, para cada sub-grupo faz-se a ACP, acima descrita.

O algoritmo distribui as k variáveis de forma a que a variabilidade que o primeiro componente principal não conseguiu explicar seja reduzida.

3.4. Critérios de Selecção do Modelo

3.4.1. Curva de ROC

Para avaliar a *performance* do modelo de regressão logística pode-se recorrer à curva ROC.

A curva ROC é calculada com base em duas variáveis:

- A sensibilidade, que corresponde à probabilidade do modelo classificar correctamente os incumprimentos, isto é:

$$\text{Sensibilidade} = \frac{VP}{FN + VP}, \quad (3.11)$$

sendo que a variável VP diz respeito ao número de incumprimentos classificados correctamente ("Verdadeiros Positivos") e a variável FN corresponde ao número de "não incumprimentos" classificados incorrectamente ("Falsos Negativos").

- A especificidade, que representa a probabilidade do modelo classificar correctamente os "não incumprimentos", isto é:

$$\text{Especificidade} = \frac{VN}{FP + VN}, \quad (3.12)$$

sendo que a variável VN diz respeito ao número de "não incumprimentos" classificados correctamente ("Verdadeiros Negativos") e a variável FP corresponde ao número de incumprimentos classificados incorrectamente ("Falsos Positivos").

Tendo em consideração o ROC determinado é classificada a *performance* do modelo de regressão logística (Hosmer e Lemeshow, 2013) de acordo com os seguintes intervalos:

- Se, $ROC \leq 0,5$, sugere que o modelo não apresenta capacidade de discriminação;
- Se, $0,5 < ROC < 0,7$, sugere que o modelo apresenta pouca capacidade de discriminação;
- Se, $0,7 \leq ROC < 0,8$, sugere que o modelo apresenta uma capacidade de discriminação;

minação aceitável;

- Se, $0,8 \leq \text{ROC} < 0,9$, sugere que o modelo apresenta uma capacidade de discriminação excelente;

- Se, $\text{ROC} \geq 0,9$ sugere que o modelo apresenta uma capacidade de discriminação excepcional.

A área abaixo da curva ROC está associada ao poder discriminante, ou seja, quanto maior for a área melhor será a qualidade do modelo. De acordo com a figura (3.1) o modelo que apresenta uma maior capacidade de previsão é o que está associado à curva ROC3.

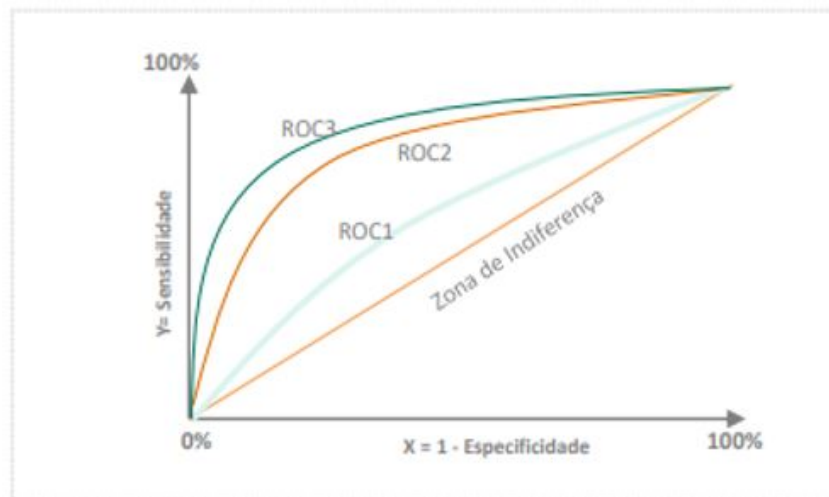


Figura 3.1. Gráfico ilustrativo com curvas ROC.

3.4.2. Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov, pode também ser utilizado para analisar a capacidade do modelo discriminar correctamente os possíveis valores da variável dependente (incumprimento e "não-incumprimento") testando as diferenças entre as duas distribuições acumuladas das probabilidades estimadas.

A distância de Kolmogorov-Smirnov é dada pela seguinte expressão:

$$KS_{nm} = \text{Sup}_{x \in R} |F_{1,n}(x) - F_{2,m}(x)|, \quad (3.13)$$

sendo $F_{1,n}(x)$ e $F_{2,m}(x)$ as funções de distribuição empíricas das amostras de incumprimentos e "não-incumprimentos" e as variáveis n e m correspondem ao número de observações de cada população.

A partir da equação anterior pode-se calcular a estatística normalizada pelas dimensões das populações:

$$KSa_{nm} = KS_{nm} \sqrt{\frac{nm}{n+m}}, \quad (3.14)$$

O melhor modelo, segundo este critério, é aquele que maximiza a distância entre as duas distribuições acumuladas de probabilidade.

3.4.3. Critério de Informação de Akaike

O critério de informação de Akaike corresponde a um estimador que indica a qualidade do modelo estatístico e é calculado através da seguinte expressão:

$$AIC = -2L + 2k, \quad (3.15)$$

sendo que L representa o logaritmo da função de verossimilhança de acordo com a equação (2.14) e a variável k é o número de parâmetros utilizados no modelo estatístico.

Neste sentido, a comparação entre modelos, com base no critério Akaike é feita assumindo que quanto menor o valor da equação (3.15) melhor é a qualidade do modelo.

3.4.4. Critério de Informação de Schwarz

O critério de informação de Schwarz corresponde a um estimador que indica a qualidade do modelo estatístico e é calculado através da seguinte expressão:

$$SC = -2L + k \times \ln(n), \quad (3.16)$$

sendo que L representa o logaritmo da função de verossimilhança de acordo com a equação (2.14), k é o número de parâmetros utilizados no modelo estatístico e n é o número de observações da amostra.

Desta forma, com base neste critério de selecção, o modelo que apresenta melhor qualidade é o que tem o valor do critério de informação de Schwarz mais baixo.

3.5. Testes de Significância

3.5.1. Teste de Wald

O teste de Wald é um teste paramétrico que tem como objectivo aferir o grau de significância de cada coeficiente da equação logística, isto é, verifica se as estimativas são significativamente diferentes de zero.

Temos então a seguinte hipótese nula (e alternativa) para o teste de Wald:

- $H_0 : \widehat{\beta}_k^2 = 0$
- $H_1 : \widehat{\beta}_k^2 \neq 0$

Sendo a estatística de teste utilizada a seguinte:

$$W = \frac{\widehat{\beta}_k^2}{\widehat{Var}(\beta_k)} \quad (3.17)$$

onde $\widehat{Var}(\beta_k)$ corresponde à variância estimada do parâmetro calculado pelo modelo.

Considera-se, um nível de significância de 1% assim a hipótese nula, (H_0), do teste de Wald vai ser rejeitada para um valor de *p-value* inferior a 1%.

3.5.2. Teste de Score

O teste de Score, à semelhança do teste de Wald, é um teste paramétrico que mede o grau de significância de cada coeficiente da equação logística, isto é, verifica se os coeficientes estimados são significativamente diferentes de zero.

Temos assim o seguinte teste de hipóteses para o teste de Score:

- $H_0 : \widehat{\beta}_k^2 = 0$
- $H_1 : \widehat{\beta}_k^2 \neq 0$

Sendo a estatística de teste utilizada calculada da seguinte forma:

$$S = \frac{\left(\frac{\partial \ln L(\beta)}{\partial \beta_k}\right)^2}{-E\left(\frac{\partial^2}{\partial \beta_k^2} \ln L(\beta)\right)} \quad (3.18)$$

onde $E\left(\frac{\partial^2}{\partial \beta_k^2} \ln L(\beta)\right)$ corresponde ao valor esperado da segunda derivada do logaritmo da máxima verossimilhança em relação a β_k .

É considerado, uma vez mais, um nível de significância de 1% portanto a hipótese nula, (H_0), do teste de Wald vai ser rejeitada para um valor de *p-value* inferior a 1%.

Observação sobre os testes de Wald e de Score

Apesar de os testes enunciados nesta secção testarem a mesma hipótese nula, a análise é feita através de abordagem diferentes. O teste de Wald testa se remover uma variável não diminui a capacidade preditiva do modelo, enquanto que o teste de Score testa se incluir uma variável no modelo melhora significativamente a sua previsão.

4. Desenvolvimento dos Modelos de *Rating* - Parte Prática

4.1. Análise e construção de variáveis

Para o desenvolvimento dos modelos de *rating* foram consideradas variáveis da base de dados do Crédito Agrícola que dizem respeito, na generalidade, a indicadores financeiros das empresas e indicadores que ilustram a relação do cliente com a Instituição/ Banca.

As variáveis em questão encontram-se identificadas na tabela A.1 do Anexo A, com a respectiva descrição.

4.1.1. Análise da qualidade das variáveis

Primeiramente, foi feita uma análise às variáveis disponíveis para tratar campos com uma elevada percentagem de informação vazia (*missings*) e valores muito díspares da restante amostra (*outliers*) de forma a não se verificar um enviesamento dos resultados por falta de qualidade de algumas variáveis.

Assim, de forma a minimizar as inconsistências existentes na base de dados, começa-se por assumir o seguinte:

- Remoção das variáveis que apresentem mais de 30% de *missings*;
- Identificação e truncagem de *outliers*.

Foram considerados *outliers* quaisquer valores abaixo do primeiro quartil (Q_1) e acima do terceiro quartil (Q_3) por uma distância superior a 4 vezes a distância

interquartil. Desta forma, os limites (superior e inferior) são calculados de acordo com as seguintes expressões:

$$\begin{aligned}\text{Limite Superior} &= Q_1 - 4(Q_2 - Q_1) \\ \text{Limite Inferior} &= Q_3 + 4(Q_2 - Q_1)\end{aligned}\tag{4.1}$$

Desta forma, os valores atípicos inferiores foram limitados ao ponto de corte inferior e os valores atípicos superiores ao ponto de corte superior. No entanto, para evitar uma elevada percentagem de valores truncados foi considerado como limite os percentis de 1% e 99% para o limite inferior e superior, respectivamente.

4.2. Selecção da Amostra

Foi considerado um histórico de quatro anos e meio, desde Dezembro de 2012 a Junho de 2017, com informação sobre as empresas, resultando numa amostra com 59 434 observações. No entanto, no mercado existem vários tipos de empresas portanto é necessário segmentá-las em grupos homogéneos de acordo com determinadas características (por exemplo, volume de negócios, sector CAE, etc).

Desta forma, a amostra foi dividida em três segmentos:

- Construção e Imobiliário (RE);
- Grandes, Pequenas e Médias Empresas (GPME);
- Negócios (NEG).

Neste sentido, foram construídos três modelos independentes para cada segmento.

Antes do desenvolvimento dos modelos de *rating* procedeu-se à divisão das observações em amostras de treino e de validação, em que:

- A amostra de treino serve de base para a modelização/ construção do modelo e, portanto, deve ser estatisticamente significativa, representado 70% da amostra inicial.

- A amostra de validação serve de base à avaliação da aplicabilidade do modelo. Desta forma, aplicando o modelo a uma amostra distinta da que serviu para modelização é possível avaliar a sua assertividade e poder preditivo. A amostra de validação corresponde a 30% da amostra inicial.

A figura (4.1) ilustra o número de observações considerado para cada um dos modelos de acordo com a respectiva amostra, teste e validação.

	Segmento			
	RE	GPME	NEG	
Amostra de teste	3 611	11 013	26 978	70%
Amostra de validação	1 548	4 721	11 563	30%
	5 159	15 734	38 541	

Figura 4.1. Divisão da Amostra

As amostras foram selecionadas de forma aleatória, garantindo, no entanto, a mesma percentagem de empresas que não estão em incumprimento em ambas.

Nas secções que se seguem as análises produzidas serão efectuadas individualmente para cada segmento/ amostra.

4.3. Análise Univariada

Com o objectivo de seleccionar as variáveis que, individualmente, têm maior poder discriminante face ao incumprimento, foram realizadas várias análises, aplicando diferentes testes consoante a tipologia de cada variável, conforme descrito na figura (4.2).

De seguida, é verificada a correlação e a coerência da tendência de cada variável com o incumprimento, uma vez mais, de acordo com a tipologia das variáveis foram

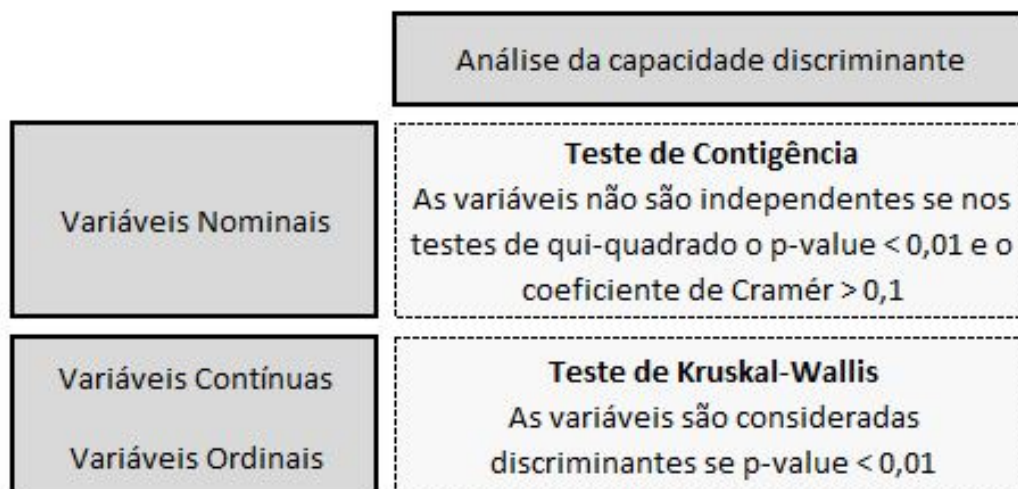


Figura 4.2. Testes de capacidade discriminante

aplicados testes distintos. As metodologias utilizadas por tipo de variável encontram-se sintetizadas na figura (4.3).

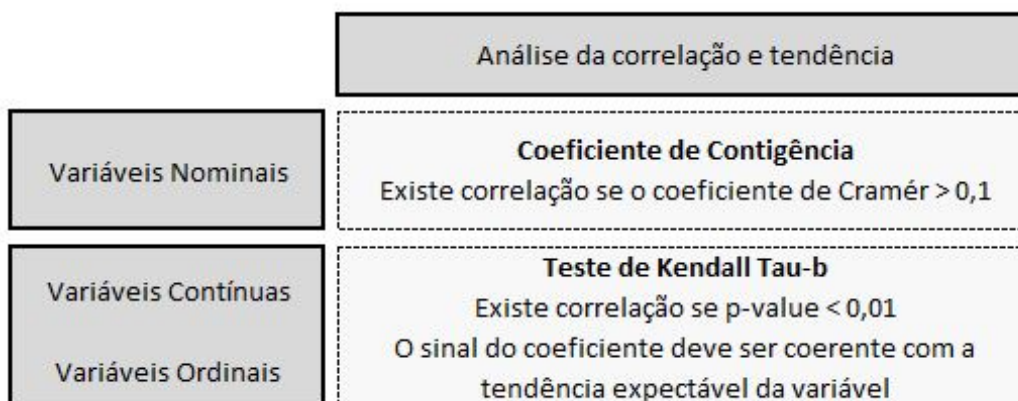


Figura 4.3. Testes de correlação e tendência

Na sequência dos testes efectuadas, uma vez que, o intuito da presente análise foi seleccionar as variáveis estatisticamente mais explicativas, foram consideradas quaisquer variáveis financeiras e de qualidade creditícia da empresa, por forma a quantificar numericamente a sua relação com o incumprimento. No entanto, conforme se pode visualizar na figura (4.5), nesta primeira fase (análise univariada) foram excluídas uma quantidade significativa de variáveis, representando uma percentagem de

cerca de 84%, 47% e 42%, respectivamente para os segmentos RE, GPME e NEG.

Deste modo, após exclusão das variáveis que não passaram nos testes enunciados, deve-se prosseguir o estudo recorrendo a análises multivariadas, para aferir o efeito conjuntos das variáveis perante o incumprimento.

4.4. Análise Multivariada

Após efectuada a análise univariada das variáveis em estudo, efectuou-se uma análise multivariada com o intuito de identificar/ excluir variáveis que apresentem informações semelhantes devido a serem redundantes para a estimação da probabilidade de incumprimento. Para a referida análise foram utilizados os testes enunciados na figura (4.4).

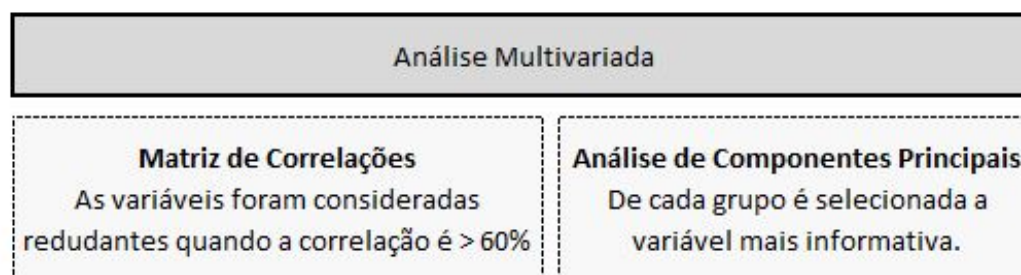


Figura 4.4. Testes de análise multivariada

Neste sentido foi calculada a correlação para cada par de variáveis e identificadas as que possuem uma correlação superior a 60%. Em seguida, para estes casos, foi selecionada a variável com maior poder discriminante face ao incumprimento, no entanto, existem situações em que as variáveis apresentam uma correlação com o incumprimento semelhante. Neste contexto, de forma a auxiliar na selecção das variáveis utilizou-se a análise de componentes principais.

Nesta última análise, após o agrupamento das variáveis com informação idêntica, foi selecionada a variável mais informativa de cada grupo.

Com o prosseguimento dos testes observa-se uma diminuição do número de va-

riáveis em estudo, o que coincide com o objectivo dado que pretende-se seleccionar um número reduzido de variáveis com uma grande capacidade explicativa para a ocorrência de um incumprimento. De acordo com a figura (4.5) a percentagem de exclusão de variáveis corresponde a 47%, 54% e 69%, para os modelos RE, GPME e NEG, respectivamente.

4.5. Categorização das variáveis

As variáveis seleccionadas, devem ser categorizadas para garantir que a variável apresenta uma relação linear com o incumprimento, isto é, os intervalos de cada categoria são seleccionados de forma a assegurar que a probabilidade de incumprimento é sempre crescente ou decrescente.

Para se obter a referida categorização recorreu-se ao método de árvore de decisão (ver Engelmann e Rauhmeier, 2006). Neste método é utilizado o teste de qui-quadrado (χ^2) para determinar a melhor categorização e respectivos pontos de corte.

Posteriormente deve-se efectuar novamente a análise univariada e a análise multivariada sobre as variáveis categorizadas.

O intuito destas análises, uma vez mais, passa por garantir que existe uma boa capacidade de discriminação em relação ao incumprimento e que as variáveis não são suficientemente correlacionadas entre si.

4.6. Resumo da Selecção de Variáveis

Resumidamente, os testes estatísticos efectuados tiveram por base a eliminação de variáveis, que por não serem suficientemente explicativas ou por serem redundantes (devido à elevada correlação com outras variáveis), não são seleccionadas para a regressão logística.

Em conclusão, na figura (4.5), encontra-se esquematizado o número de variáveis eliminadas em cada análise.

	Antes da Categorização				Depois da Categorização	
	Variáveis disponíveis	Após análise de qualidade	Após análise univariada	Após análise multivariada	Após análise univariada	Após análise multivariada
RE	191	187	30	16	16	9
GPME	191	187	99	46	41	16
NEG	191	187	108	33	29	17

Figura 4.5. Resumo do número de variáveis.

4.7. Aplicação da Regressão Logística

Conforme anteriormente mencionado, foi aplicado o modelo de regressão logística às variáveis selecionadas de forma a estimar a probabilidade de incumprimento.

As variáveis finais, ver figura (4.6), são selecionadas de entre as variáveis da figura (4.5) para o modelo de acordo com um algoritmo iterativo de selecção de variáveis cujo nome é *Stepwise Selection*. Este algoritmo consiste na adição ou remoção de variáveis consoante estas adicionem ou retirem poder discriminante ao modelo. Para a tomada de decisão são utilizados os testes de Wald e de Score descritos no capítulo 3.

Variáveis	Modelos		
	RE	GPME	NEG
	<ul style="list-style-type: none"> - Rendibilidade dos Capitais Próprios - Ativo Corrente sobre o Ativo Total - EBITDA sobre os Juros e gastos similares suportados - Autonomia financeira - Responsabilidades no GCA sobre o Total do Rendimento - Crédito vencido no BdP sobre as Responsabilidades no BdP - Prazo médio de Recebimentos (dias) - Net Debt sobre o EBITDA - Meios Financeiros líquidos sobre o Passivo Corrente 	<ul style="list-style-type: none"> - Flag de descoberto não autorizado nos últimos 12 meses - Crédito vencido no BdP sobre as Responsabilidades no BdP - Capital Próprios sobre o Ativo não Corrente - Capital Próprios sobre o Passivo Corrente - Resultado Antes de Impostos face às responsabilidades na GCA - Autonomia Financeira - Prazo Médio de Pagamentos (dias) - Net Debt sobre o EBITDA - Juros e gastos similares sobre o Activo total - Responsabilidades no GCA sobre o Total do Rendimento 	<ul style="list-style-type: none"> - Indicador de crédito vencido nos últimos 12 meses no GCA - Rácio de Liquidez Geral - Resultado Antes de Impostos face às responsabilidades no BdP - (Caixa + Investimento financeiro + Clientes - Fornecedores) sobre o Ativo Total - Total de Financiamentos obtidos sobre o Total do Passivo - Autonomia financeira - Número (máximo) de dias em atraso nos últimos 9 meses - Juros e gastos similares suportados sobre o Resultado antes de Imposto - Passivo não Corrente sobre o Capital Próprio - (Ativo não Corrente - Capitais Próprios) sobre o EBITDA - Prazo Médio de Pagamentos (dias) - Responsabilidades no BdP sobre o Volume de Negócios

Figura 4.6. Variáveis selecionadas para cada modelo

De seguida, após notar a amostra de treino com a probabilidade de incumprimento associada aos clientes, de acordo com os modelos contruídos para cada segmento,

é necessário aplicar os modelos à amostra de validação de forma a avaliar a sua qualidade e capacidade de previsão.

Esta metodologia permitiu obter uma probabilidade de incumprimento para cada cliente, no entanto, o objectivo deste estudo corresponde a classificar cada empresa num nível de risco. Portanto, neste sentido, é necessário converter a PD estatística numa pontuação (*score*) estatística, isto é:

$$PD_{estatística} = \frac{1}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i X_i}} \quad (4.2)$$

$$Score_{estatístico} = \ln\left(\frac{1 - PD_{estatística}}{PD_{estatística}}\right)$$

4.8. Performance do modelo

De forma a avaliar a qualidade do modelo, deve-se ter em consideração várias medidas estatísticas, nomeadamente a curva ROC e o coeficiente de Kolmogorov-Smirnov.

Os três modelos apresentaram uma boa qualidade quer para a amostra de treino como para a amostra de validação o que significa que apresentam um bom nível de assertividade na previsão de incumprimentos. A figura (4.7) ilustra os resultados obtidos para as medidas ROC e para o coeficiente de KS. Relativamente à medida ROC temos valores percentuais na casa dos 70% o que indica, de acordo com os intervalos definidos na secção (3.4), que sugere que os modelos apresentam uma capacidade de discriminação excelente. O coeficiente de KS mede a distância entre a distribuição da amostra de incumprimentos e a distribuição da amostra de não-incumprimentos, portanto quanto maior for esta distância, em valor absoluto, maior é a capacidade discriminante do modelo.

	Amostra de Treino		Amostra de Validação	
	ROC	KS	ROC	KS
RE	80%	0,51	81%	0,58
GPME	82%	0,46	83%	0,59
NEG	85%	0,55	82%	0,50

Figura 4.7. Medidas de *performance* para os modelos de *rating*

4.9. Integração da Informação Qualitativa e Quantitativa

Posteriormente ao desenvolvimento dos modelos de *rating* estatísticos e respectiva notação da carteira, deve-se obter o *rating* final ponderando a componente quantitativa com a componente qualitativa (calculada com base na análise do analista).

A componente qualitativa é importante pois requer uma análise individual para cada empresa de acordo com o conhecimento que os analistas têm do cliente.

$$Score_{final} = p \times Score_{estatístico} + (1 - p) \times Score_{qualitativo}, \quad (4.3)$$

sendo que o $score_{qualitativo}$ foi obtido directamente das bases de dados do GCA. A probabilidade, p , é igual a 70%, isto é, para o apuramento final do *rating* foi considerado que a componente qualitativa vale 30% e a componente quantitativa vale 70%.

4.10. Níveis de Risco

Nesta fase, em que cada cliente tem atribuída uma pontuação, $Score_{final}$, deve-se segmentar a carteira em classes de *score* com distribuição uniforme em termos de percentis. E, de seguida calcula-se a taxa de incumprimento de cada classe de *score*.

Seguidamente, para cada classe de *score* são calculados as *log-odds* associadas à

taxa de incumprimento, ou seja:

$$\text{logg-odds} = \ln\left(\frac{\text{Taxa Incumprimento}_n}{1 - \text{Taxa Incumprimento}_n}\right) \quad (4.4)$$

A relação entre o *score* e os *logg-odds* deve ser linear permitindo aplicar uma regressão linear, nomeadamente:

$$\text{logg-odds} = -\alpha - \beta \text{score}_n. \quad (4.5)$$

Desta forma, é possível obter a PD_{final} para cada observação da amostra através da seguinte equação:

$$PD(\text{Score}_{final}) = \frac{1}{1 + e^{-\alpha - \beta \text{score}_n}} \quad (4.6)$$

4.11. Interpretação dos Resultados

A escala de risco foi definida por forma a que a PD tenha uma evolução exponencial com o nível de risco e garantindo que não foram definidos níveis de risco sem população.

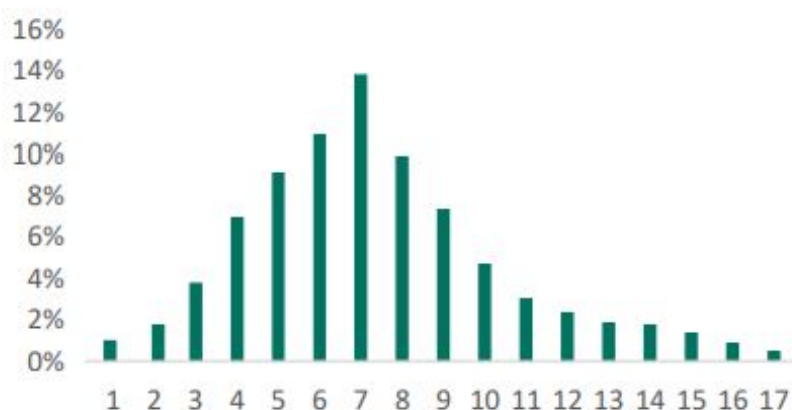
Por fim, estando definidos os intervalos mínimos e máximos da probabilidade de incumprimento para cada nível de risco, o mesmo é mapeado para a carteira de acordo com o intervalo em que a probabilidade de incumprimento, $PD(\text{Score}_{final})$, de cada cliente se insere.

Na tabela da figura (4.8) são ilustrados os intervalos de PD para cada nível de risco e, conforme se pode constatar existe uma relação linear entre a notação/ nível de risco e a PD, isto é, quanto maior é a probabilidade de incumprimento de um cliente maior será o seu nível de risco. Os cinco primeiros níveis de risco representam um baixo risco, ou seja, a empresa apresenta uma situação financeira sólida. Os seguintes seis níveis simbolizam um risco médio, isto é, as condições financeiras da

Descrição	Nível de Risco	Notação de Risco	Ponderação Final	
			MIN	MÁX
Extremamente sólido Muito sólido Sólido	1	1	0,00%	0,30%
	2	1,5	0,30%	0,42%
	3	2	0,42%	0,60%
	4	2,5	0,60%	0,85%
	5	3	0,85%	1,20%
Razoável	6	3,5	1,20%	1,70%
	7	4	1,70%	2,40%
	8	4,5	2,40%	3,40%
	9	5	3,40%	4,85%
Moderadamente Vulnerável	10	5,5	4,85%	6,82%
	11	6	6,82%	9,64%
	12	6,5	9,64%	13,60%
Vulnerável	13	7	13,60%	19,32%
	14	7,5	19,32%	27,41%
Muito Vulnerável	15	8	27,41%	39,80%
	16	8,5	39,80%	55,00%
Excessivamente Vulnerável	17	9	55,00%	100,00%

Figura 4.8. Níveis de risco pelos intervalos de PD.

empresa são razoáveis, mas já apresentam indícios de vulnerabilidade. Por fim, os últimos níveis de risco classificam as empresas como vulneráveis, na medida em que o risco de não assumirem os seus compromissos financeiros é elevado.



Por motivos de confidencialidade os valores apresentados são apenas ilustrativos, não correspondendo à distribuição oficial da carteira por nível de risco.

Figura 4.9. Distribuição da carteira por níveis de risco.

Desta forma, a definição da carteira em níveis de risco permite que exista uma melhor monitorização do perfil de risco de cada cliente, neste caso de clientes empresas, e facilita ainda uma maior comparabilidade entre os perfis de risco, uma vez que a escala é uniforme para cada modelo.

Neste momento torna-se possível fazer várias análises de risco de crédito, como por exemplo obter a distribuição da carteira por nível de risco dando directamente a visibilidade da qualidade da mesma. Se a distribuição tiver demasiado concentrada na cauda esquerda significa que a qualidade da carteira é ótima (composta por clientes de baixo perfil de risco) se, por outro lado, a distribuição tiver demasiado concentrada na cauda direita significa que a qualidade da carteira é má (composta por clientes de elevado perfil de risco). No entanto, o expectável é que exista uma distribuição aproximadamente normal com maior concentração nos níveis intermédios e caudas pouco densas.

Por observação da figura (4.10) é possível constatar que tal como mencionado anteriormente existe, em todos os modelos, uma maior concentração da carteira nos níveis intermédios (risco médio) quando comparado com os restantes níveis (risco reduzido e risco elevado).

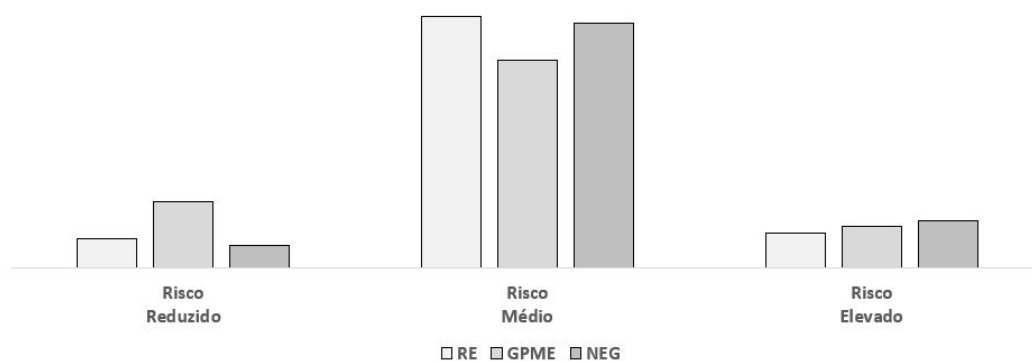


Figura 4.10. Distribuição da carteira por modelo e por níveis de risco agregado.

Embora para a previsão de incumprimento as variáveis explicativas variem de modelo para modelo (ver figura (4.6)) de acordo com as características da amostra

em análise, observa-se que os resultados finais mostram uma distribuição por nível de risco uniforme entre os três modelos.

Detalhadamente, conforme figura (4.10), é possível verificar que existe uma maior percentagem de empresas classificadas como GPME nos níveis de risco reduzido o que está alinhado com a segmentação efectuada, uma vez que, o segmento GPME inclui as empresas com volume de negócios mais elevado, portanto, na maioria, é expectável que se verifique uma maior estabilidade financeira face ao segmento NEG, que inclui empresas com um volume de negócios mais reduzido. Relativamente ao segmento RE a segmentação não se encontra feita por volume de negócio mas sim por sector de actividade (construção e imobiliário), pelo que não existe um termo de comparação directo com os restantes segmentos, no entanto, a percentagem de empresas no nível de risco reduzido encontra-se alinhado com o dos restantes modelos, sendo mais semelhante à do modelo NEG.

Nos níveis de risco médio temos, uma vez mais, uma percentagem muito semelhante de empresas do segmento NEG e do segmento RE. Em contrapartida, o segmento GPME tem uma menor percentagem de empresas nestes níveis devido a ter uma distribuição ligeiramente mais à esquerda que os restantes.

Por fim, nos níveis de risco elevado observa-se uma distribuição muito semelhante nos três modelos, correspondendo a percentagem mais reduzida ao segmento RE e a mais elevada ao segmento NEG. Assim, o segmento NEG apresenta a cauda direita ligeiramente mais densa do que a dos restantes modelos, o que está alinhado com a visão do mercado, dado que é nos pequenos negócios que se verifica uma maior volatilidade quanto à sua estabilidade financeira.

Em conclusão, a distribuição da carteira pelos níveis de risco para os modelos NEG e RE é bastante semelhante e, dado que a amostra de clientes do segmento RE é reduzida quando comparada com a dos restantes modelos, podia-se testar os resultados obtidos através da segmentação da amostra em apenas dois segmentos, o que implicaria a construção de apenas dois modelos. Relativamente à distribuição

das empresas do segmento GPME pelos níveis de risco aparenta estar de acordo com o expectável.

5. Conclusão

Na presente tese, procedeu-se à estimação de três modelos, híbridos, de notação de risco onde foi combinada uma componente qualitativa com uma componente quantitativa para se obter um nível de risco final associado a cada cliente da amostra.

Para se obter a componente quantitativa, recorreu-se a vários testes estatísticos com o intuito de seleccionar as variáveis mais explicativas possíveis. E, em seguida aplicando a regressão logística às variáveis seleccionadas foi possível notar cada observação da amostra com a sua respectiva probabilidade de incumprimento (estatística). Por fim, converteu-se a probabilidade de incumprimento numa classificação (*score*).

Relativamente à componente qualitativa não foram desenvolvidos, quaisquer modelos, tendo sido apenas um *input* das bases de dados.

Os modelos de notação de risco são essenciais para uma boa gestão e monitorização da carteira de crédito permitindo aos analistas fazerem análises rápidas e detalhadas da carteira com um elevado grau de confiança tendo em consideração a boa capacidade discriminante do modelo.

No entanto, os modelos de *rating* requerem uma revisão frequente dado que a sua *performance* pode-se degradar com o passar do tempo devido a vários factores (por exemplo, mudança de políticas de crédito, mudanças no comportamento dos clientes, mudanças macroeconómicas, etc). Neste sentido de forma a dar continuidade a este trabalho era interessante considerar uma nova amostra (por exemplo de Junho 2017 a Junho 2018) e validar o modelo através da metodologia de *Backtesting*. Está

metodologia permitiria aferir se o modelo passado um ano continuaria com uma boa capacidade discriminante em relação ao incumprimento.

Bibliografia

- [1] Mendes A. (2013). *A transição de Basileia II para Basileia III: qual o enfoque que é dado aos riscos nos Acordos de Basileia?* - Instituto Superior de Economia e Gestão.
- [2] Engelman B. e Rauhmeier, R. (2006). *The Basel II Risk Parameters - Estimations, Valuation, Stress Testing - with Applications to Loan Risk Management*. Springer.
- [3] Silva D. (2011). *Modelos econométricos de classificação de rating de instituições financeiras*. - ISCTE Business School Instituto Universitário de Lisboa.
- [4] Banco de Portugal (2007). Mar - modelo de avaliação de riscos - available at: https://www.bportugal.pt/sites/default/files/anexos/documentos-relacionados/consulta_bp_2_07_mar.pdf.
- [5] Banco de Portugal (2008). Liquidez, solvabilidade e risco nas instituições de crédito:
Rácios Financeiros. Available at:
https://www.bportugal.pt/sites/default/files/anexos/documentos-relacionados/intervpub20170508_1.pdf.
- [6] Banco de Portugal (2015). Lei n.º 23-a/2015, de 26 de março - available at: <https://www.bportugal.pt/legislacao/lei-no-23-a2015-de-26-de-marco>.

- [7] Banco de Portugal (2018). Regras prudenciais - available at: <https://www.bportugal.pt/page/micro-regras-prudenciais>.
- [8] Ferreira E. (2010). *Ratings - 50 Perguntas e Respostas*. Centro Atlântico.
- [9] Santos J. (2011). *Metodologia de implementação do Método das Notações Internas*. - Faculdade de Engenharia da Universidade do Porto.
- [10] Witzany J. (2017). *Credit Risk Management - Pricing, Measurement, and Modeling*. Springer.
- [11] Hosmer J. e Lemeshow S. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.
- [12] Amaral M. (2015). Tipos de riscos na actividade bancária. available at: <http://www.oroc.pt/fotos/editor2/revista/69/gestao.pdf>.
- [13] Abdullah N.; Halim A.; Hamilton A.; e Rus R. (2008). *Predicting corporate failure of Malaysia's listed companies: Comparing multiple discriminant analysis, logistic regression and the hazard model*. EuroJournals Publishing, Inc.
- [14] Antão P. e Lacerda A. (2008). Avaliação dos Requisitos de Capital sob Basileia II: O Caso Português. Available at: https://www.bportugal.pt/sites/default/files/anexos/papers/ar200803_p.pdf.
- [15] Czepiel S. (202). Maximum likelihood estimation of logistic regression models: theory and implementation. Available at czep.net/stat/mlelr.pdf.
- [16] Trueck S. e Rachev, S. (2008). *Rating Based Modeling of Credit Risk - Theory and Application of migration matrices*. Academic Press.
- [17] Greene W. (2002). *Econometric Analysis*. Prentice Hall.

A. Anexo A

Na tabela que se segue são apresentadas todas as variáveis utilizadas para modelização.

Variáveis	Descrição
badloan_iM	Nº de vezes em que a conta está vencida, abatida ou renegociada nos últimos i meses, $i = 3, 6, 9$ e 12 .
con_decbal_iM	Nº de vezes em que há descidas do saldo mensal das contas à ordem do cliente nos últimos i meses, $i = 3, 6, 9$ e 12 .
con_incbal_iM	Nº de vezes em que há subidas do saldo mensal das contas à ordem do cliente nos últimos i meses, $i = 3, 6, 9$ e 12 .
Max_incump_dda_iM	Máximo do estado de incumprimento da conta à ordem nos últimos i meses, $i = 3, 6, 9$ e 12 .
No_times_Overdraft_iM	Nº vezes em que se verificaram descobertos não autorizados nos últimos i meses, $i = 3, 6, 9$ e 12 .
Var_Saldo_DO_iM	Varição do saldo da conta à ordem nos últimos i meses, $i = 3, 6, 9$ e 12 .
X_UTR_CD_iM	Indicador de Utilizador de Risco no BdP

	nos últimos i meses, i = 3, 6, 9 e 12.
CHEQUES_DEV_iM	Indicador de cheques devolvidos nos últimos i meses, i = 0, 3, 6, 9 e 12.
OVERDRAFT_LIMIT_FLG_iM	Indicador de descobertos não autorizados nos últimos i meses, i = 0, 3, 6, 9 e 12.
PAST_DUE_CNT_iM	Máximo do número de dias em atraso da conta nos últimos i meses, i = 3, 6, 9 e 12.
Sector_CD	Sector CAE
I1	Volume de Negócios
I2	Cash-Flow
I3	Fundo de Maneio
I4	Produção
I5	Valor Acrescentado Bruto (VAB)
I6	Gastos Variáveis
I7	Gastos Fixos
I8	Margem Bruta de Exploração
I10	Margem Líquida de Exploração
I12	EBITDA
I13	Margem EBITDA
I14	Net Debt
I15	Net Debt to EBITDA
I16	Rácio de Liquidez Geral
I17	Rácio de Liquidez Reduzida
I18	Rácio de Liquidez Imediata
I19	Cobertura do Activo não corrente
I20	Cobertura de Encargos Financeiros
I21	Autonomia Financeira
I22	Autonomia Financeira Corrigida

I23	Solvabilidade
I24	Endividamento
I25	Estrutura do Endividamento
I26	Período de Recuperação de Empréstimos (anos)
I27	Cash-Flow to Liabilities
I28	Cash-Flow to Assets
I29	Cash-Flow to Equity
I30	Prazo Médio de Recebimentos (dias)
I31	Prazo Médio de Pagamentos (dias)
I32	Prazo Médio de Stocks de Matérias-primas (dias)
I33	Prazo Médio de Stocks de Produtos Acabados (dias)
I34	Rotação do Activo
I35	Rendibilidade do Activo
I36	Rendibilidade dos Capitais Próprios
I37	Rendibilidade do Negócio
I38	Peso do Fundo de Maneio no Volume de Negócios
I39	Peso dos Encargos com Pessoal no VAB
I40	Peso dos Encargos Financeiros no VAB
I41	Importância dos Credores Preferenciais (SPE)
B1	Activos fixos tangíveis
B2	Propriedades de investimento
B3	Goodwill
B4	Activos intangíveis
B5	Activos biológicos (Activo não corrente)
B6	Participações financeiras – método de equivalência patrimonial
B7	Participações financeiras - outros métodos
B8	ANC - Accionistas/sócios

B9	Outros activos financeiros
B10	Activos por impostos diferidos
B11	Investimentos financeiros (2)
B12	Inventários
B13	Activos biológicos (Activo corrente)
B14	Clientes
B15	Adiantamentos a fornecedores
B16	AC - Estado e outros entes públicos
B17	AC - Accionistas/sócios (Activo Corrente)
B18	Outras contas a receber
B19	Diferimentos
B20	Activos financeiros detidos para negociação
B21	Outros activos financeiros
B22	Activos não correntes detidos para venda
B23	Outros activos correntes
B24	Caixa e depósitos bancários
B25	Capital realizado
B26	Acções (quotas) próprias
B27	Outros instrumentos de capital próprio
B28	Prémios de emissão
B29	Reservas legais
B30	Outras reservas
B31	Resultados transitados
B32	Ajustamentos em activos financeiros
B33	Excedentes de revalorização
B34	Outras variações no capital próprio
B35	Resultado líquido do período
B36	Dividendos antecipados

B37	Interesses minoritários
B38	Provisões
B39	Financiamentos obtidos
B40	Responsabilidades por benefícios pós-emprego
B41	Passivos por impostos diferidos
B42	Outras contas a pagar
B43	Fornecedores
B44	Adiantamentos de clientes
B45	PC - Estado e outros entes públicos
B46	PC - Accionistas/sócios
B47	Financiamentos obtidos
B48	Outras contas a pagar
B50	Diferimentos
B51	Passivos financeiros detidos para negociação
B52	Outros passivos financeiros
B53	Passivos não correntes detidos para venda
B54	Outros passivos correntes
BD52	Matérias-primas, subsidiárias e de consumo
BD53	Mercadorias
BD54	Produtos acabados e intermédios
BSS1	Sub-total (ACTIVO NÃO CORRENTE)
BSS2	Sub-total (ACTIVO CORRENTE)
BSS3	TOTAL DO CAPITAL PRÓPRIO
BSS5	Sub-total (PASSIVO NÃO CORRENTE)
BSS6	Sub-total (PASSIVO CORRENTE)
D55	Vendas e serviços prestados (VSP)
D56	Subsídios à exploração
D57	Ganhos/Perdas imput. subs., assoc. e emp. conj.

D58	Varição nos inventários da produção
D60	Trabalhos para a própria entidade
D61	Custo mercadorias vendidas e matérias consumidas
D62	Fornecimentos e serviços externos
D63	Gastos com o pessoal
D64	Imparidade de inventários (perdas/reversões)
D65	Imparidade de dívidas a receber (perdas/reversões)
D66	Provisões (aumentos/reduções)
D67	Imparidade de inventários. não deprec./amortiz.
D69	Aumentos/reduções de justo valor
D70	Outros rendimentos e ganhos
D71	Outros gastos e perdas
D72	Gastos/reversões de depreciação e de amortização
D73	Imparidade de investimentos. deprec./amortiz.
D74	Juros e rendimentos similares obtidos
D75	Juros e gastos similares suportados
D76	Imposto sobre o rendimento do período
D77	Resultados activ. descont. (líquido impostos) no RL
D78	Detentores do Capital da Empresa-Mãe
D79	Interesses Minoritários
D80	Número de Colaboradores
D81	Compras
D82	Vendas e Serviços Prestados - mercado interno
D148	Resultado por acção básico
D158	Vendas e Serviços Prestados - mercado externo
D159	Outras imparidades/Imparidades (perd./revers.) (2)
DSS9	Resultado antes deprec., gast. financ., imp. (EBITDA)
DSS10	Resultado operacional (antes gast. financ. e impostos)

DSS11	Resultado antes de impostos
DSS12	Resultado líquido do período
DSS13	Resultado líquido do período atribuível a: (1)
DSS14	Informação complementar
RELATIONSHIP_TO_CREDIT_CD	Nível de responsabilidade
CREDIT_SIT_CD	Situação do crédito
CRC_SICAM_CREDIT_AMT	Montante de responsabilidades GCA
CRC_BANCA_CREDIT_AMT	Montante de responsabilidades total

Tabela A.1.: Descrição das variáveis

