



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO

ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

RELATÓRIO DE ESTÁGIO

**ESTUDO SOBRE EVOLUÇÃO DO BALANÇO DE UM BANCO EM
SITUAÇÃO DE STRESS ECONÓMICO - MODELO
MACROECONÓMICO DE PD**

BERNARDO VIEIRA GONÇALVES PEREIRA

10-2019



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO EM ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

RELATÓRIO DE ESTÁGIO

**ESTUDO SOBRE EVOLUÇÃO DO BALANÇO DE UM BANCO EM
SITUAÇÃO DE STRESS ECONÓMICO - MODELO
MACROECONÓMICO DE PD**

BERNARDO VIEIRA GONÇALVES PEREIRA

ORIENTAÇÃO:

DR. NUNO CANDEIAS SANTOS

PROF. DR. NUNO RICARDO MARTINS SOBREIRA

10-2019

Abstract

Este relatório de estágio enquadra-se no âmbito do Trabalho Final de Mestrado para o Mestrado de Econometria Aplicada e Previsão do Instituto Superior de Economia e Gestão (ISEG).

O estágio decorreu na Deloitte Consultores, S.A., num dos escritórios de Lisboa, com o intuito de me preparar para a realidade do mercado de trabalho através de um processo de integração na empresa durante a realização das atividades propostas.

Neste relatório pretende-se expor um retrato daquilo que foram os meus primeiros seis meses de trabalho na Deloitte, através da descrição aprofundada do tema de trabalho, procurando sempre relacioná-lo com os conceitos adquiridos no decorrer do mestrado.

A Deloitte faz parte das Big 4, grupo de empresas que domina o setor de serviços profissionais, sendo de realçar o seu grau de exigência e profissionalismo nos projetos a que se compromete. Na Deloitte, as minhas funções passaram pela análise das taxas de incumprimento e de migrações entre classe de risco de uma carteira de clientes empresa de um banco, tendo estabelecido um modelo econométrico que permite a previsão da evolução dessas taxas com base em variáveis macroeconómicas. Este tipo de modelos é aplicado na obtenção da perda esperada usada, por exemplo, no cálculo de provisões por imparidade e em testes de stress.

KEYWORDS: Risco de Crédito; Probabilidade de Default; Modelo de Desfasamentos Distribuídos; Perda de Crédito Esperada; Séries Temporais.

JEL CODES: C52; C53; C32; E17; G21

ÍNDICE

ABSTRACT.....	I
ÍNDICE.....	II
ÍNDICE DE FIGURAS.....	III
AGRADECIMENTOS.....	IV
1. ENQUADRAMENTO TEÓRICO.....	1
1.1. RISCO, INCERTEZA E PROBABILIDADE.....	1
1.2 RISCO DE CRÉDITO E PERDA ESPERADA.....	4
1.3 CLASSES DE RISCO E MODELO DE RATING.....	5
1.4 MÉTODOS MACROECONÓMICOS E PREVISÃO.....	8
1.5 MODELOS ADL.....	10
1.5.1 INTRODUÇÃO TEÓRICA.....	10
1.5.2 ESCOLHA DO DESFASAMENTO MÁXIMO.....	11
1.5.3 TESTES DE ESPECIFICAÇÃO.....	12
1.5.3.1 BREUSCH-GODFREY E DURBIN-WATSON.....	12
1.5.3.2 TESTE RESET.....	14
1.5.3.3 TESTES-T E -F.....	14
2. APLICAÇÃO PRÁTICA.....	16
2.1 APRESENTAÇÃO DOS DADOS E ANÁLISE DESCRITIVA.....	16
2.2 ESTIMAÇÃO DA SÉRIE Z.....	17
2.3 SELEÇÃO DO MODELO.....	21
2.3.1 ANÁLISE PRELIMINAR.....	21
2.3.2 ESTIMAÇÃO DO MODELO.....	22
3. DISCUSSÃO.....	28
3.1 ANÁLISE DOS RESULTADOS.....	28
3.2 APLICAÇÃO DO MODELO.....	29
3.3 PROBLEMAS E HIPÓTESES DE DESENVOLVIMENTO FUTURO.....	31
4. REFERÊNCIAS BIBLIOGRÁFICAS.....	33
5. ANEXO.....	35
5.1 ANEXO A.....	35
5.2 ANEXO B.....	40

Índice de Figuras

<i>TABELA I</i>	17
<i>TABELA II</i>	18
<i>TABELA III</i>	19
<i>TABELA IV</i>	20
<i>TABELA V</i>	22
<i>TABELA VI</i>	40
<i>GRÁFICO 1</i>	29
<i>GRÁFICO 3</i>	30
<i>GRÁFICO 2</i>	30
<i>FIGURA 4</i>	35
<i>FIGURA 5</i>	36
<i>FIGURA 6</i>	37
<i>FIGURA 7</i>	38
<i>GRÁFICO 8</i>	40
<i>GRÁFICO 9</i>	41
<i>GRÁFICO 10</i>	42
<i>GRÁFICO 11</i>	43
<i>GRÁFICO 12</i>	43
<i>GRÁFICO 13</i>	44
<i>GRÁFICO 14</i>	44
<i>GRÁFICO 15</i>	45

AGRADECIMENTOS

Gostaria de agradecer, em primeiro lugar, aos meus orientadores: do lado da empresa, um agradecimento ao Nuno Candeias, pelo apoio constante ao longo do projeto e pela disponibilização de parte do seu tempo quase todas as semanas durante o período de estágio; do lado do ISEG, agradecer ao Professor Nuno Sobreira por todos os conhecimentos transmitidos ao longo do semestre, que me ajudaram imenso neste trabalho, e pela prontidão de resposta aos mails enviados (que não foram muitos e deveriam ter sido mais).

Um agradecimento também a todos os outros professores deste mestrado maravilhoso, que desde o início nunca escondi ser o meu mestrado de sonho, e não tive um professor que me tivesse feito duvidar desse facto.

De seguida, agradecer aos meus pais, não só pelo apoio financeiro, mas também, e mais importante, pelo apoio psicológico e motivacional, por serem compreensivos e por me suportarem durante épocas mais stressantes desta fase da minha vida de estudante universitário, e por me aturarem desde nascença, tudo o que sou devo-lhes a eles, e é uma dívida que nunca conseguirei pagar. Agradecer também à minha irmã por me ensinar o que são rotinas de estudo, e por me lembrar constantemente que estudo pouco. Agradecer também todas as noites em que me impediu de terminar a tese porque tinha dúvidas de matemática e há que estabelecer prioridades, e a prioridade é ela, e no fundo essa é a verdade.

Um agradecimento e um grande abraço a todos os meus amigos que me convidaram a ir beber uns copos para desanuviar, em especial ao Hugo e ao César, por todas aquelas sextas no bar do Mickas a beber cerveja e a descascar amendoins e a conversar sobre a bola e o nosso Benfica, momentos que levo para sempre na minha bagagem e amizades que guardo comigo.

Por último, um agradecimento especial à pessoa que se eu tivesse acabado o relatório uns meses antes não teria sido incluída nos agradecimentos. Por todas as manhãs, tardes e noites que passámos juntos, por todo o apoio e carinho que todos os dias me faz questionar se o mereço, por me teres ensinado que de facto sim, eu mereço, e por tanta coisa mais que não tenho espaço para mencionar, mas acima de tudo por seres tudo o que eu sempre quis e pela honra de ser teu, deixo aqui o meu agradecimento à rapariga dos meus sonhos, a minha sorte grande.

I. ENQUADRAMENTO TEÓRICO

I.1. RISCO, INCERTEZA E PROBABILIDADE

Risco e incerteza são dois conceitos que têm vindo a ser extensivamente discutidos pela comunidade económica, e não só, sendo que a literatura existente sobre este tema é ainda incapaz de fornecer uma definição global para qualquer uma das noções e a própria relação entre elas, da mesma maneira que não existe ainda um consenso em relação ao método mais correto de quantificação/cálculo destes fatores, visto que depende forte e logicamente das circunstâncias e da área em que é aplicado. Neste trabalho, procurar-se-á desenvolver um dos muitos métodos existentes para o cálculo do risco de crédito.

Na procura pela correta definição destes conceitos (de risco e incerteza), numa perspetiva económica, Knight, F. (1921) opta por fazer uma distinção entre risco e incerteza, afirmando que a verdadeira incerteza surge quando o risco a si associado é na verdade desconhecido. Ou seja, de acordo com Knight, incerteza é risco incalculável, e sob o qual é impossível atuar de forma a mitigar perdas, através de seguros, ou outras medidas de prevenção. Knight adota esta definição para poder fundamentar a sua teoria de lucro, onde considera a incerteza, e a capacidade de “prever/adivinhar” o desconhecido, como um dos determinantes para a obtenção de lucro.

Seguindo esta lógica de ideias, Stigler, G. J. (1987) afirma que Knight procurava distinguir três tipos de acontecimentos futuros: o primeiro são acontecimentos cuja probabilidade matemática de acontecerem pode ser calculada direta e facilmente (por exemplo, o acontecimento de tirar uma carta de um baralho, ao acaso, e sair o naipe de espadas); o segundo são acontecimentos que se conseguem prever mais facilmente com o uso de informação histórica, apesar de serem impossíveis de prever *a priori* (por exemplo, a ocorrência de um desastre natural, na medida em que são imprevisíveis mas existem diversos indicadores e relatórios históricos que podem facilitar o cálculo da probabilidade de acontecerem); finalmente o terceiro, e assumidamente mais raro, são acontecimentos cuja probabilidade de se darem é impossível de ser determinada, de qualquer forma.

Na verdade, Knight gerou muita controvérsia na comunidade económica, fruto das diferentes interpretações do seu livro “*Risk, Uncertainty and Profit*” (1921), obrigando alguns economistas a publicarem as suas visões em relação à obra. Uma grande crítica a esta obra não advém

propriamente do seu conteúdo, mas sim da sua fraca aplicabilidade para circunstâncias reais, onde a aferição das probabilidades não é exatamente clara e objetiva. Holton, G. refere que

Frank Knight's famous definition of risk is problematic for being unrelated to common usage. It depends upon an objectivist interpretation of probability and is somewhat parochial, being more useful in certain fields than in others.

In: Holton, G. (2004)

sem descurar a importância da ligação entre risco e probabilidade, mas fortalecendo essa vertente subjetiva da probabilidade, tornando impossível a definição objetiva do primeiro, meramente em função do segundo. Aven, T. (2012) refere ainda que a definição teórica de risco poderá nem ser sequer necessária de entender para analisar e gerir risco, *“as long as we are precise on what we describe when performing a risk assessment, there is consistency in terminology, and we are aware of and able to point to the limitations of the perspective adopted.”*

No mesmo ano que Knight, Keynes, J. M. (1921) publicou um livro intitulado *“A Treatise on Probability”*, onde expôs as suas crenças sobre probabilidade e incerteza. Keynes começa por separar conhecimento direto do conhecimento por argumentação (indireto), afirmando que a noção de probabilidade só faz sentido ser empregue para o último tipo, uma vez que o conhecimento direto é 100% certo. Keynes associa depois probabilidade ao grau de crença numa qualquer proposição, crença essa que pode ser racional ou irracional. É aqui que Keynes define a linha que separa as vertentes objetiva e subjetiva da probabilidade. Ou seja, por um lado, a crença irracional (sem qualquer fundamento lógico), muito presente no comum dia-a-dia, apela à subjetividade do conceito de probabilidade. Por outro lado, com crenças racionais, comuns em estudos e matérias científicas, o conceito de probabilidade é obrigatoriamente objetivo.

Numa análise a esta obra, Sakai, Y. (2016) considera que a posição de Keynes é muito delicada:

“On the one hand, his concept of probability is more general and comprehensive than it is adopted in the mathematical theory of probability. On the other hand, it is more academically and scientifically used than in everyday life.”

In: Sakai, Y. (2016)

Em relação ao risco, Keynes, J.M. (1921) faz menção a uma distinção similar à de Knight: “(...) *underwriters themselves distinguish between risks which are properly insurable, either because their probability can be estimated (...), and other risks which cannot be dealt with in this way and which cannot form the basis of a regular business of insurance*”, focando-se principalmente no risco “calculável”, oferecendo uma explicação prática e simplista com um exemplo: Seja A um resultado positivo de um acontecimento, e p a probabilidade de verificarmos esse resultado, então o valor esperado, E , desse acontecimento pode ser dado por $E = pA$, que pode ser interpretado como o sacrifício necessário para obter A ; e, conseqüentemente, o risco, R , será dado por $R = p(A - E) = p(1 - p)A = pqA = qE$, onde $q = 1 - p$ pode ser interpretado como a probabilidade de esse sacrifício ser em vão, fazendo de R o valor a minimizar, o valor do seguro. Esta definição culmina com a demonstração de que este tipo de risco pode ser completamente “transferido”, com recurso a uma propriedade básica das sucessões geométricas: suponha-se que se pode aplicar sucessivos seguros a um acontecimento, de forma que $R_2 = qR_1 = q^2E$, temos então,

$$E + R_1 + R_2 + R_3 + \dots = E(1 + q + q^2 + q^3 + \dots) = \frac{E}{1-q} = \frac{E}{p} = A. \quad (1)$$

No entanto, esta definição de risco e respetiva demonstração baseiam-se em hipóteses que cada vez mais se consideram impossíveis dada a complexidade do mundo real (“*The old assumptions, that all quantity is numerical and that all quantitative characteristics are additive, can be no longer sustained*” - In: Keynes, J. M. (1921).

Finalmente, em relação ao risco incalculável, Keynes, J. M. remete para a tendência, de um agente racional, em “fugir” desse risco, referindo que “*in an uncertain and unknown world, economic agents prefer to retain money rather than make investment decisions*”.

Perante isto, é facilmente perceptível a grande hesitação por parte das instituições financeiras na tomada de decisão em relação à concessão de crédito para os empréstimos, colocando em risco o seu próprio capital. A própria variedade de tipos de risco (que pode ser dividida em três grandes grupos: financeiro, não financeiro, e outros) é evidência do alto nível de preocupação e cautela que se deve ter na tomada de tais decisões. Temos, entre outros, risco de reputação, risco de mercado,

risco operacional, e risco de crédito. Este último será o tópico central desta tese, por estar inserido no plano de funções do meu trabalho na Deloitte Consultores, S.A..

1.2 RISCO DE CRÉDITO E PERDA ESPERADA

O risco de crédito é a possibilidade de uma contraparte incumprir um pagamento, por exemplo de um empréstimo ou obrigação.

Uma das métricas relacionadas com risco de crédito é a Expected Credit Loss (ECL) (perda esperada), obtida pela fórmula (2), composta por Loss Given Default (LGD), Probability of Default (PD), e Exposure at Default (EAD):

$$ECL_t = LGD_t \times PD_t \times EAD_t. \quad (2)$$

Destas três variáveis, desenvolvem-se modelos mais complexos para cálculo de perfis de LGD e PD. Para este projeto dar-se-á foco ao método de cálculo da PD, pelo que é importante perceber primeiro o seu conceito.

A probabilidade de incumprimento consiste, tal como o nome indica, na probabilidade de uma contraparte entrar em incumprimento, ou seja, não ser capaz de cumprir com os pagamentos previamente acordados aquando da concretização do acordo de cedência de crédito. Por sua vez, a LGD reflete a percentagem de exposição que o banco espera perder no caso de a contraparte entrar em *default*, e por último, a EAD reflete uma estimativa do montante em dívida (quantias utilizadas e prováveis levantamentos futuros de linhas de crédito ainda não utilizadas) no caso de a contraparte entrar no dito *default*.

Os parâmetros acima descritos dependem não só de uma componente idiossincrática, exclusiva para cada contraparte, como de uma componente sistemática, comum a todas as contrapartes. Esta componente sistemática pode ser explicada por fatores externos, tais como condições macroeconómicas.

Uma das aplicações importantes do conceito de perda esperada, numa perspetiva de enquadramento macroeconómico, é no cálculo da imparidade. A imparidade da carteira de crédito de um banco equivale à diferença entre o valor dos créditos e o valor que se espera conseguir recuperar desses créditos. Neste contexto, a norma IFRS9 define uma nova metodologia de reconhecimento das perdas por imparidade de ativos financeiros com base em perdas esperadas -

expected loss model. A substituição do modelo de mensuração das perdas por imparidade de ativos financeiros com base em perdas incorridas - *incurred loss model* - por um modelo que se baseia em perdas esperadas, permitirá a detecção antecipada do aumento significativo do risco de crédito com a incorporação de elementos *forward looking* (com informação macroeconómica, decorrente de vários cenários: pessimista, base e otimista) na estimação da perda esperada.

Outra aplicação imediata é o da construção de cenários de *stress*. Estes são aplicados nos testes de *stress* que os bancos têm de realizar recorrentemente.

A PD é a variável em foco neste projeto, sendo que no que toca ao método de estimação, seguir-se-á o processo de Belkin, B. e Suchower, S. (1998), para a obtenção de uma série temporal que descreve a evolução sistemática das transições entre classes de risco (incluído transições para o default). Posteriormente, recorrer-se à uma estimação por desfasamentos distribuídos, com o objetivo de estabelecer relação com variáveis macroeconómicas e assim obter previsões para a PD em diferentes enquadramentos macroeconómicos. Estes métodos serão apresentados, em detalhe, em secções seguintes.

O trabalho foi desenvolvido recorrendo uma base de clientes empresa de um banco e respetiva classificação em classes de risco.

1.3 CLASSES DE RISCO E MODELO DE RATING

Para a classificação de empresas em classes de risco, um banco recorre habitualmente a modelos internos de *rating*, podendo, no entanto, para grandes empresas recorrer a ratings externos.

De acordo com Altman, E. e Saunders, A. (1998), existiu uma grande evolução na computação do risco de crédito num período de 20 anos, entre 1977 e 1997, período em que se passou de análises subjetivas por parte de especialistas, com base na regra dos 4 C's (Character (reputação), Capital (alavancagem), Capacity (volatilidade dos ganhos), e Colateral), para modelos de cariz mais objetivo, que produzem valores de classificação de crédito ou de probabilidade de incumprimento para serem comparados com os valores críticos previamente definidos.

Beaver, W. H. (1966) executou um estudo onde obteve vários tipos de rácios em formato "*cash-flow*", para empresas em falência e para empresas financeiramente sustentáveis, comparando-os de forma a conseguir estabelecerlos como indicadores da probabilidade de falência

para novas empresas (entenda-se por falência “*the inability of a firm to pay its financial obligations as they mature (...) when any of the following events have occurred: bankruptcy, bond default, an overdrawn bank account, or nonpayment of a preferred stock dividend*” - In Beaver, W.H. (1966), p.71).

Mais tarde, Altman, E. I. (1968) introduziu um método diferente de análise de risco, baseado igualmente em demonstrações financeiras e dados contabilísticos, mas mais focada para facilitar a tomada de decisão relativamente à concessão de crédito, por parte das instituições financeiras, a empresas públicas ou privadas. Através da criação de um Z-score, único a cada empresa, e baseado nas demonstrações financeiras da mesma, consegue-se obter uma medida de aprovação/desaprovação de crédito. Mais concretamente, a função para o Z-score de uma empresa de capital aberto é dada por:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + X_5, \quad (3)$$

onde $X_1 := \text{Capital de Giro} / \text{Ativo Circulante} / \text{Total de Ativos}$;

$X_2 := \text{Lucros Acumulados} / \text{Total de Ativos}$; $X_3 := \text{EBIT} / \text{Total de Ativos}$; $X_4 := \text{Capitalização de Mercado} / \text{Dívida Total}$; $X_5 := \text{Vendas} / \text{Dívida Total}$. Calculado o Z-score, compara-se o resultado obtido ao valor crítico de $VC = \{Z_{VC}: Z_{VC} = 1.81\}$, sendo a regra de decisão, para determinar se uma empresa de capital aberto seria “boa” para emprestar dinheiro ou não, a seguinte:

$$\begin{cases} Z > 1.81, & \text{"boa" empresa, conceder crédito} \\ Z < 1.81, & \text{"má" empresa, não conceder crédito} \end{cases}$$

Para empresas privadas, o cenário muda ligeiramente: função para o Z-score dada por

$$Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5, \quad (4)$$

onde $X_4 := \text{Valor Contabilístico do Capital Próprio/Dívida Total}$; e os restantes X_i ($i = 1, 2, 3, 5$) representam as mesmas variáveis que anteriormente, na equação (3); o critério de decisão passa a ser um pouco mais complexo, deixando de se basear num único valor, e com o aparecimento de uma “Zona Cinzenta”:

$$\begin{cases} Z' > 2.90, & \text{Zona Segura ("boa" empresa, conceder crédito)} \\ 1.23 \leq Z' \leq 2.90, & \text{Zona Cinzenta} \\ Z' < 1.23, & \text{Zona de Stress ("má" empresa, não conceder crédito)} \end{cases} .$$

Com o passar dos anos e o constante aparecimento de nova literatura, o próprio Altman viu-se obrigado a visitar o seu método, e em 1995 criou o final Z'' , um *score* geral, tanto para empresas públicas como privadas:

$$Z'' = 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4 \quad (5)$$

$$Z'_S = Z'' + 3.25 \quad (6)$$

onde Z'_S é o Z'' estandardizado. O critério de escolha passa também a ser:

$$\begin{cases} Z'_S > 5.85, & \text{Zona Segura ("boa" empresa, conceder crédito)} \\ 4.50 \leq Z'_S \leq 5.85, & \text{Zona Cinzenta} \\ Z'_S < 4.50, & \text{Zona de Stress ("má" empresa, não conceder crédito)} \end{cases} ,$$

aparecendo mais uma vez a “Zona Cinzenta”, onde é incerto qual a melhor decisão a tomar em relação à concessão de crédito.

Atualmente os bancos recorrem a modelos estatísticos para estimar diretamente a probabilidade de incumprimento e dessa forma classificarem as empresas em classes de risco. Um dos modelos mais usuais é o da regressão logística. Nestes modelos a probabilidade de incumprimento é dada por

$$p = \frac{e^F}{1 + e^F} \quad (7)$$

com $F = \beta_0 + \sum_i \beta_i X_i$ onde X_i são as variáveis explicativas e β_i parâmetros a serem estimados (recorrendo por exemplo ao método de máxima verosimilhança). Os modelos desenvolvidos são suportados principalmente pela situação financeira da empresa, por fatores comportamentais de relação com a instituição bancária e o sistema financeiro, e parâmetros qualitativos analisados pelas redes comerciais e direções de crédito do banco.

1.4 MÉTODOS MACROECONÓMICOS E PREVISÃO

Numa perspectiva de previsão e de desenvolvimento de cenários de *stress*, motivados por fatores externos, com vista a responder aos desafios expostos atrás (cálculo de imparidade, testes de *stress*), os bancos viram-se obrigados a desenvolver modelos em que a variável de incumprimento, pode ser potencialmente explicada por fatores macroeconómicos, exógenos à variável dependente. A formulação destes modelos econométricos tem o objetivo de tentar prever a probabilidade de incumprimento, mas também a das várias transições entre classes de risco, representadas tipicamente por uma matriz de migrações. Alguns autores responsáveis pela introdução destas ideias no mundo económico são Duffie e Singleton (1999), Jarrow e Turnbull (1995), e ainda Belkin e Suchower (1998).

O método abordado neste trabalho para avaliação do risco de crédito foi introduzido no mundo económico por Belkin, B. e Suchower, S. (1998) e enquadra-se no grupo dos modelos mais objetivos, fornecendo uma representação da probabilidade de *default* e das matrizes de migração através de um único parâmetro denominado de fator-z (não confundir com o Z-score de Altman). Na sua publicação “*A one-parameter representation of credit risk and transition matrices*”, Belkin e Suchower referem-se a esta variável Z como correspondendo a uma componente sistemática da variável que mede as transições de valor de crédito (*credit rating*), refletindo as probabilidades de *default* do fim do período, através da média histórica dessas mesmas transições. Desta forma, um valor positivo para Z é sinal de um “bom” ano em termos de menor número de *defaults* e maior número de subidas de classe de crédito do que descidas; já os valores negativos de Z deverão

refletir a situação oposta. O método de estimação para a componente Z, proposto por Belkin e Suchower, é apresentado de seguida.

O primeiro passo é considerar a variável “original”, designada no *paper* por X , ou seja, a variável que mede as transições de classe de risco, que se assume como tendo uma distribuição Normal ($X \sim N(0, 1)$) e definir intervalos, designados por *bins* ($x_g^G; x_{g+1}^G[$, de tal maneira que a probabilidade da variável X se encontrar nesse intervalo seja igual à média histórica dessa correspondente transição. Ou seja, a condição de definição dos *bins* pode ser escrita da seguinte forma:

$$P(G, g) = \Phi(x_{g+1}^G) - \Phi(x_g^G), \quad (8)$$

onde $P(G, g)$ representa a média histórica da probabilidade de transição da classe de risco G para g ; e Φ é naturalmente a função distribuição cumulativa da Normal.

Quer isto dizer que a cada *classe de risco* da matriz de migração corresponderá um *bin*, composto por um limite inferior (x_g^G) e um limite superior (x_{g+1}^G), sendo que para a classe de *default* o limite inferior será $-\infty$ e, na ponta oposta, para a classe de crédito máximo, o limite superior será $+\infty$, de forma a obter a lógica unidade para a soma de todas as probabilidades de transição para cada *rating* ($\sum_i P(G, i) = 1$, i corresponde a cada classe).

Depois de obtidos os *bins* para a matriz de migração global, que representa uma espécie de média das matrizes anuais (Point-in-Time), o método de estimação não passa de um simples método de mínimos quadrados, ou seja, um método de estimação com o objetivo de minimizar o quadrado das diferenças entre as probabilidades de transição do modelo estimado a partir dos *bins* da matriz global e as observadas para cada período.

Será então expectável que para anos piores, ou abaixo da média, com discrepâncias negativas, se obtenham valores menores (negativos) para o fator Z, e que a anos acima da média, em termos de transição de classe, correspondam valores mais elevados para esse fator (positivos).

1.5 MODELOS ADL

1.5.1 Introdução Teórica

Considere-se o seguinte modelo linear para séries temporais,

$$Y_t = \alpha + \phi Y_{t-1} + D(L)X_t + \epsilon_t, \quad (9)$$

onde $D(L) = \delta_0 + \delta_1 L + \dots + \delta_s L^s$, com L a representar o operador de defasamento, ou seja $LX_t = X_{t-1}$; $L^2 X_t = X_{t-2}$; ...; $L^s = X_{t-s}$; X_t representa o conjunto de variáveis explicativas do modelo, e ϵ_t representa o termo do erro, seguindo, por hipótese, um processo de ruído branco independente ($\epsilon_t \sim iid(0, \sigma^2)$). Esta é a versão mais simples e explícita de um modelo *ADL* (*s*) (*Autoregressive Distributed Lag*), para a qual se assume a satisfação da condição de estabilidade, $|\phi| < 1$, para não cairmos no problema de raízes unitárias. No entanto, na prática nem sempre se verifica essa condição.

Um dos pontos fortes deste modelo é a sua generalidade, ou seja, de uma perspectiva mais conservadora, é prática comum iniciar a análise de estimação com este modelo geral, deixando depois os dados adaptarem-se ao caso particular mais apropriado. Na verdade, a lista de casos particulares que advêm deste modelo é muito extensa, indo desde as próprias regressões estáticas, sem dinâmica ($\phi = 0$ e $\delta_1 = \delta_2 = \dots = \delta_s = 0$) até aos Modelos de Correção de Erros, que na verdade são obtidos por mera reparametrização, não se impondo qualquer restrição. Um potencial problema deste modelo é a presença de raízes unitárias na componente autorregressiva do modelo, que pode condicionar a consistência dos estimadores OLS. No entanto, em caso de existência de cointegração esse problema não só deixa de existir, como os estimadores passam até a ser super consistentes, ou seja, convergem para o valor real do parâmetro a uma taxa superior à usual. Em todo o caso, é prática comum testar a presença de raízes unitárias através dos testes de Dickey-Fuller (DF e ADF).

Um caso particular, mais simples, deste modelo consiste na eliminação da própria componente autorregressiva por evidência de não significância estatística ($\phi = 0$), e designa-se de modelo DL (Modelo de Desfasamentos Distribuídos Finitos), um método muito comum na modelização de

efeitos dinâmicos por considerar defasamentos das variáveis explicativas. Este modelo mais simples pode ser representado da seguinte forma:

$$y_t = \mu + D(L)x_t + u_t, t = 1, 2, \dots, T, \quad (10)$$

onde mais uma vez $D(L) = \delta_0 + \delta_1 L + \dots + \delta_s L^s$ é o polinómio de defasamento. Os problemas associados à estimação deste tipo de modelos serão expostos de seguida.

1.5.2 ESCOLHA DO DESFASAMENTO MÁXIMO

O primeiro problema relacionado com este tipo de modelos recai na escolha do nível de defasamento máximo s , de forma a não perder informação importante, mas procurando sempre seguir o princípio da parcimónia. Obviamente que este não é meramente um problema superficial de escolha, tendo sim algumas implicações maioritariamente ao nível da inferência. Mais concretamente, por um lado, escolher um s demasiado pequeno, significa perda de informação importante para explicar o modelo, o que pode levar à autocorrelação dos erros do modelo, invalidando toda a inferência dos testes estatísticos, uma vez que se viola uma das hipóteses fundamentais sob a qual estes se baseiam (hipótese da ausência de autocorrelação); por outro lado, escolher um s demasiado grande pode implicar uma perda significativa de observações, uma vez que teremos de considerar os valores da nossa amostra só a partir do ponto imediatamente a seguir ao valor máximo de defasamento. Isto implica, por sua vez, uma maior incerteza em relação à inferência estatística, visto que nos afastamos cada vez mais de uma perspetiva assintótica, e muitas hipóteses, como a da consistência dos estimadores, são válidas apenas assintoticamente.

Perante este problema, os econométristas recorrem a dois tipos de soluções mais comuns. São elas: o uso de critérios de informação, tais como o AIC ou o SBC, em que o método subjacente consiste em escolher o modelo que minimiza a estatística refletora do tal problema de escolha do defasamento máximo (a explicação detalhada destes critérios vai além do âmbito deste projeto, mas recomenda-se a leitura de Akaike (1998) para os mais interessados); e o uso de testes-F sequenciais sobre os últimos coeficientes de defasamento, determinando a significância estatística conjunta destes e a plausibilidade da sua inclusão, denominado de método GTS (*General to Specific*). Gilbert, C. S. (1989) afirma que um dos incentivos para o aparecimento deste método

foi a demonstração, de Ted Anderson, de que processos que se baseiam em testes sequenciais de hipóteses cada vez mais restritas são mais potentes. De notar ainda que ambos procedimentos necessitam de uma escolha prévia de um valor fixo de defasamento máximo (S_{max}), para o qual se tem sempre $s \leq S_{max}$. Este valor deve ser escolhido tendo em conta a dimensão da amostra e a periodicidade dos dados (por exemplo, para uma amostra de dimensão pequena com dados anuais, não faz sentido definir um $S_{max} > 4$).

1.5.3 TESTES DE ESPECIFICAÇÃO

O segundo problema reside na estimação dos parâmetros da regressão, ou seja, dos coeficientes das variáveis explicativas. Como é sabido, o estimador OLS é BLUE (Best Unbiased Linear Estimator) sob certas hipóteses e condições, mais concretamente se X_t for estritamente exógeno e os erros forem *iid*, por exemplo. No entanto, em séries macroeconómicas é muito comum obter valores de autocorrelação muito elevados entre o valor contemporâneo (atual) e os seus defasamentos e entre estes, originando problemas de colinearidade dos regressores, violando a primeira condição. Para perceber melhor este problema, pensemos que se uma série y_t segue um determinado padrão temporal, logicamente que se a autocorrelação de primeira ordem for muito elevada, como acontece em muitos casos, y_{t-1} seguirá um padrão semelhante, resultando na correlação das duas séries. Sob estas condições, estimar por OLS pode resultar em estimações imprecisas, com coeficientes estatisticamente insignificantes, ou até mesmo com sinais incorretos. Para garantir a satisfação desta e outras hipóteses necessárias para validar a inferência estatística, bem como para garantir a validade do modelo final escolhido, executam-se diversos testes, apresentados e explicados de seguida.

1.5.3.1 BREUSCH-GODFREYE DURBIN-WATSON

É bastante comum, em séries temporais, encontrar dados autocorrelacionados, provocando autocorrelação dos resíduos do modelo estimado. Consequentemente, esta autocorrelação põe em causa a validade e eficiência das previsões baseadas no modelo, previsões essas que constituem a base do projeto. Desta forma, far-se-á uso de dois testes muito comuns no meio econométrico para identificar a presença de autocorrelação.

Primeiramente, o teste de Durbin-Watson (DW) (1950), baseia-se na hipótese nula de autocorrelação dos resíduos:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

onde ρ é o coeficiente da regressão de teste

$$e_t = \rho e_{t-1} + v_t, \quad (11)$$

com e_t a representar as estimativas dos erros do modelo (resíduos do modelo). A estatística de teste é obtida pela seguinte fórmula:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \cong 2(1 - \hat{\rho}), \quad (12)$$

onde $\hat{\rho}$ representa a autocorrelação dos resíduos da amostra. Ou seja, como $-1 \leq \hat{\rho} \leq 1$, significa que a estatística de teste varia entre 0 e 4, sendo que os valores desejados para a não rejeição da hipótese nula de ausência de autocorrelação serão valores próximos de 2 ($\hat{\rho} = 0$).

O outro teste à presença de autocorrelação no modelo que será aplicado neste projeto é o de Breusch-Godfrey (1978). Neste caso, o teste torna-se mais geral, ou seja, a hipótese nula passa a considerar a ausência de autocorrelação até à ordem p :

$$\begin{cases} H_0: \rho_i = 0, i = 1, 2, \dots, p \\ H_1: \exists \rho_i \neq 0, i = 1, 2, \dots, p \end{cases}$$

onde os ρ 's provêm da regressão de teste

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t, \quad (13)$$

de onde são retirados os resíduos por OLS (e_t), e implementados na regressão auxiliar

$$e_t = \alpha_0 + \alpha_1 X_{t,1} + \alpha_2 X_{t,2} + \dots + \alpha_k X_{t,k} + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p}, \quad (14)$$

ou seja, a regressão de teste para o teste de DW é agora aumentada pelas variáveis explicativas do modelo original $X_{t,1}, X_{t,2}, \dots, X_{t,k}$. A estatística de teste é obtida a partir do R^2 da regressão auxiliar ($nR^2 \sim \chi_p^2$, onde $n = T - p$, T é o número de observações da série original).

1.5.3.2 TESTE RESET

O teste RESET (*Regression Equation Specification Error Test*) de Ramsey (1969), é um mero teste de especificação, ou seja, é um teste que pode detetar problemas resultantes de omissão de variáveis significativas ou de incorreta especificação da própria forma funcional do modelo. O teste baseia-se na hipótese de que se combinações não lineares das variáveis explicativas têm algum poder explicativo sobre a variável dependente, então o modelo estará mal especificado. Esta ideia pode ser traduzida pela seguinte regressão auxiliar:

$$y = \alpha x + \gamma_1 \hat{y}^2 + \dots + \gamma_{k-1} \hat{y}^k, \text{ onde } \hat{y} = E(y|x) = \beta x, \quad (15)$$

ou seja, aumenta-se a regressão original com potências da variável dependente. A hipótese nula passa assim a ser descrita da seguinte forma:

$$\begin{cases} H_0: \gamma_i = 0, i = 1, \dots, k - 1 \\ H_1: \exists \gamma_i \neq 0, i = 1, \dots, k - 1 \end{cases}$$

A estatística de teste é baseada num teste-F, introduzido de seguida.

1.5.3.3 TESTES-T E -F

A versão mais simples e comum de testes de significância são de facto os testes-t (para significância individual) e os testes-F (para significância global). Estas duas versões são muito semelhantes, no sentido em que o teste-F pode ser visto como uma adaptação do teste-t quando temos mais que uma restrição a testar.

Considere-se uma simples regressão linear:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k. \quad (16)$$

Neste caso, a hipótese nula para o teste-t seria $H_0: \beta_1 = 0$, por exemplo, enquanto que a hipótese nula do teste-F poderia ser generalizada da seguinte forma $H_0: \beta_i = 0, i = 1, \dots, k$. A estatística de teste para um teste de significância individual pode ser dada por:

$$t = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}/\sqrt{n}}, \quad (17)$$

sendo que $\hat{\beta}$ é o estimador OLS do modelo, $\hat{\sigma}_{\hat{\beta}}$ é a estimativa do desvio-padrão do estimador OLS, e n é o número de observações.

Para o caso do teste-F, é necessário estimar uma regressão auxiliar, impondo as restrições a testar, ou seja, tornando todos os coeficientes sob teste iguais a zero. A estatística de teste pode ser então dada por:

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(T - p)}, \quad (18)$$

sendo que R_{UR}^2 é a medida de R^2 do modelo não restrito, e R_R^2 do modelo com as restrições impostas, q representa o número de restrições a testar, e $T - p$ é o número de graus de liberdade do teste.

Pretende-se com este trabalho, fundamentalmente, relacionar o risco de crédito com indicadores macroeconómicos, por forma a tentar prever as possíveis perdas na carteira de um banco. Desta forma, na Secção 2 iniciarei com uma análise à base de dados sob estudo, passando à descrição dos métodos de estimação usados, procedendo também à apresentação e breves comentários dos resultados do estudo. Por fim, na Secção 3 irei expor as conclusões deste projeto, bem como algumas dificuldades e possíveis melhorias.

2. APLICAÇÃO PRÁTICA

2.1 APRESENTAÇÃO DOS DADOS E ANÁLISE DESCRITIVA

No âmbito do que é uma das responsabilidades e uma das funções da Deloitte enquanto empresa de serviços profissionais, foi meu encargo, durante este período de estágio, elaborar as matrizes de migração para clientes de um banco não especificado, por questões de confidencialidade. Para tal, procedeu-se primeiramente à preparação dos dados através do software estatístico SAS. Esta preparação dos dados engloba, numa fase inicial, uma análise da qualidade dos mesmos¹, bem como uma análise à estatística descritiva. A nossa amostra contém um total de 330,155 observações, divididas por 11,719 clientes, ao longo de 8 anos (09/2011-09/2018), e consiste numa listagem das datas de referência (D_REFERENCIA), no nº identificativo do cliente (ID_CLIENTE), no *rating* do cliente (RATINGCLIENTE), calculado pelo banco e numa variável binária indicadora do estado de incumprimento do cliente ($default = 1$ se cliente está em incumprimento)
($default = 0$ caso contrário).

A base de dados é composta por 330 155 registos, de 11 719 clientes diferentes. Verificou-se, numa primeira análise, que nos anos iniciais da amostra (entre 2011 e 2013) os dados têm uma periodicidade trimestral, enquanto que a partir de 2014 passam a ter periodicidade mensal. Desta forma, para poder garantir a consistência do estudo, optou-se por considerar uma periodicidade trimestral na elaboração das matrizes de migração. Ao retirar da amostra todos os meses que não Março, Junho, Setembro e Dezembro, sobraram ainda 148 395 registos. Destes, no entanto, verificou-se que existiam 3 624 registos sem valor de classificação de crédito, cerca de 2.44% da amostra total, tendo sido necessário desprezar estas entradas. A amostra final fica então com 144 771 registos, que serão considerados para a elaboração das matrizes de migração.

Considerando agora a elevada granularidade da variável de classificação de crédito, com algumas classificações a terem 3 casas decimais, houve uma necessidade de agrupar esta variável, em valores inteiros, e de forma a obter uma distribuição em forma de sino (*bell-shaped*), ou seja, onde os valores extremos são mais raros, e a maioria dos valores encontram-se no meio da distribuição. Para além disto, impõe-se também, por questões de lógica económica, que a PD seja

¹ Entenda-se por qualidade, a quantidade de valores em falta (missings) e de valores inconsistentes com a teoria económica

crecente, ou seja, que a valores menores de classificação correspondam PDs mais altas. Desta forma, o *trimming* encontrado foi o seguinte, apresentado na Tabela II:

TABELA I

ESTIMATIVA DOS CUTS PARA O *RATING*

	1	2	3	4	5	6	7
CUT	3.5	4.3	4.7	5	5.5	6	8.9
%ACUM	6.21%	21.68%	41.44%	63.41%	84.16%	94.10%	100%
%	6.21%	15.46%	19.77%	21.97%	20.75%	9.94%	5.90%

resultando em 7 valores de *rating* distintos, mais a classe de *default*, sendo 1 a pior classe e 7 a melhor. A última linha da tabela mostra-nos a tal desejada distribuição *bell-shaped* para os *ratings*, fornecida pelos dados.

2.2 ESTIMAÇÃO DA SÉRIE Z

Depois de preparados os dados, elaborou-se a matriz de migração através de SAS, tendo-se obtido matrizes de transição globais, ou seja, um registo de todas as transições ocorridas ao longo do período total da amostra, independente do trimestre em que ocorreram, sendo que a cada linha da matriz corresponde o *rating* inicial e a cada coluna o *rating* final (ou seja, cada valor introduzido na matriz é calculado através de uma simples divisão entre o número de transições do *rating* correspondente à linha da matriz para o *rating* correspondente à coluna desta, sobre todos os clientes com *rating* inicial correspondente à linha, independentemente do *rating* para onde transitaram); e cronológicas, com o registo apenas das transições ocorridas especificamente em cada trimestre.

Com as devidas matrizes obtidas, aplicou-se o método introduzido por Belkin e Suchower intitulado de “*one-parameter representation of credit risk and transition matrices*” para a estimação do fator Z. Este método assenta numa hipótese central de que as matrizes de migração são resultado de um processo de “intervalização” de uma variável aleatória X que segue uma distribuição Normal. Esta variável pode ser decomposta por uma componente idiossincrática Y e uma componente sistemática, o tal fator Z, da seguinte forma:

$$X = \sqrt{1 - \rho}Y + \sqrt{\rho}Z, \quad (19)$$

onde $\rho > 0$, representa o fator de correlação entre a componente sistemática e a variável refletora das matrizes de migração, ou seja, ρ explica parte da variância da variável X , a parte que pode ser estimada através do método aqui aplicado. Aplicou-se, então, este processo às matrizes trimestrais, tendo-se obtido, no entanto, apenas 29 matrizes trimestrais (uma global e 28 cronológicas, correspondente a cada trimestre a partir de setembro de 2011 até junho de 2018). Os *bins* (mais concretamente, os limites superiores dos *bins*) foram obtidos a partir da matriz de migração global (Tabela II), através da inversa da distribuição Normal,

TABELA II
MATRIZ DE MIGRAÇÃO GLOBAL (THROUGH-THE-CYCLE)

TTC	7	6	5	4	3	2	1	D
7	92,2%	4,9%	1,9%	0,8%	0,1%	0,1%	0,01%	0,01%
6	2,5%	90,1%	5,4%	1,3%	0,5%	0,1%	0,0%	0,02%
5	0,4%	2,7%	90,6%	4,5%	1,3%	0,4%	0,0%	0,1%
4	0,1%	0,5%	4,4%	88,7%	4,5%	1,5%	0,2%	0,1%
3	0,03%	0,2%	1,2%	4,6%	87,9%	5,0%	0,7%	0,4%
2	0,03%	0,04%	0,4%	1,2%	4,8%	88,2%	3,7%	1,6%
1	0,01%	0,01%	0,2%	0,3%	1,0%	5,5%	84,1%	8,9%
D	-	0,01%	0,0%	0,1%	0,2%	0,7%	2,2%	96,8%

$$\Phi^{-1}[P(G, g) + \Phi(x_g^G)], \quad (20)$$

resultado derivado da equação (8), onde $P(G, g)$ representa a função probabilidade de transição do *rating* inicial G para o *rating* final g , explícita nessa mesma matriz, e $\Phi(x_g^G)$ representa o valor da distribuição cumulativa normal correspondente ao valor do limite inferior do *bin*. Ou seja, a título de exemplo, considerem-se as transições a partir do *rating* mais baixo, acima do *default*, ou seja, *rating* de 1. Pela análise da matriz de migração global obtida pelos dados, verifica-se que a $P(1, D) = 8.94\%$. De notar que o limite inferior deste *bin* corresponde a $-\infty$, como anteriormente referido, pelo que $\Phi(-\infty) = 0$, o que significa que o limite superior do respetivo *bin* será igual a $\Phi^{-1}[P(1, D)] = \Phi^{-1}(0.0894) = -1.34$, obtendo-se então $bin_{1,D} = (-\infty; -1.34]$. Por sua vez,

o limite superior deste *bin* coincidirá com o limite inferior do próximo *bin*, sendo o limite superior calculado novamente, da seguinte forma:

$$\Phi^{-1}[P(1,1) + \Phi(x_g^1)] = \Phi^{-1}[0.8406 + 0.0894] = \Phi^{-1}(0.93) = 1.48. \quad (21)$$

O *bin* final ficará então $bin_{1,1} = (-1.34; 1.48]$. O processo repete-se sucessivamente até ao *rating* máximo, onde teremos, neste caso, $bin_{1,7} = (3.62; +\infty)$. Isto repete-se para todos os outros *ratings* iniciais, obtendo os valores dispostos na Tabela III.

TABELA III
ESTIMAÇÃO DOS BINS DA MATRIZ GLOBAL

TTC	7	6	5	4	3	2	1	D								
7	100,00	-1,42	-1,42	-1,89	-1,89	-2,33	-2,33	-2,84	-2,84	-3,06	-3,06	-3,45	-3,45	-3,63	-3,63	-100,00
6	100,00	1,96	1,96	-1,45	-1,45	-2,06	-2,06	-2,47	-2,47	-2,87	-2,87	-3,22	-3,22	-3,60	-3,60	-100,00
5	100,00	2,68	2,68	1,88	1,88	-1,52	-1,52	-2,07	-2,07	-2,53	-2,53	-3,00	-3,00	-3,12	-3,12	-100,00
4	100,00	3,15	3,15	2,50	2,50	1,65	1,65	-1,53	-1,53	-2,09	-2,09	-2,71	-2,71	-2,98	-2,98	-100,00
3	100,00	3,47	3,47	2,87	2,87	2,18	2,18	1,55	1,55	-1,55	-1,55	-2,31	-2,31	-2,69	-2,69	-100,00
2	100,00	3,46	3,46	3,19	3,19	2,59	2,59	2,12	2,12	1,51	1,51	-1,62	-1,62	-2,14	-2,14	-100,00
1	100,00	3,62	3,62	3,44	3,44	2,90	2,90	2,56	2,56	2,18	2,18	1,48	1,48	-1,34	-1,34	-100,00
D	100,00	100,00	100,00	3,62	3,62	3,24	3,24	2,98	2,98	2,70	2,70	2,31	2,31	1,85	1,85	-100,00

De notar que na tabela acima, os limites superiores dos *bins* encontram-se à esquerda e os inferiores à direita, ao contrário do disposto no corpo do texto.

Estes *bins* serão agora empregues na minimização da função objetivo,

$$\min_{Z_t} \sum_G \sum_g \frac{n_{t,g} [P_t(G,g) - \Delta(x_{g+1}^G, x_g^G, Z_t)]^2}{\Delta(x_{g+1}^G, x_g^G, Z_t) [1 - \Delta(x_{g+1}^G, x_g^G, Z_t)]}, \quad (22)$$

onde

$$\Delta(x_{g+1}^G, x_g^G, Z_t) = \Phi\left(\frac{x_{g+1}^G - \sqrt{\rho}Z_t}{\sqrt{1-\rho}}\right) - \Phi\left(\frac{x_g^G - \sqrt{\rho}Z_t}{\sqrt{1-\rho}}\right), \quad (23)$$

através do comando Solver do Excel, sendo que para isso é necessário uma pré seleção de um valor para o rho (ρ), assumidamente positivo, e para a série Z_t . O Solver devolve-nos finalmente os

seguintes valores para a série objetivo, apresentados na Tabela IV que serão usados para a estimação do modelo econométrico:

TABELA IV
SÉRIE Z ESTIMADA

Data de Referência	Série Z
2011/09	0,9749
2011/12	0,1755
2012/03	-0,4422
2012/06	-1,4001
2012/09	-1,8894
2012/12	-1,7976
2013/03	-1,5047
2013/06	-1,1548
2013/09	-0,9938
2013/12	-0,6551
2014/03	-0,2292
2014/06	0,2029
2014/09	0,0157
2014/12	0,0373
2015/03	0,5763
2015/06	0,8296
2015/09	1,0684
2015/12	0,5295
2016/03	0,5123
2016/06	0,1746
2016/09	0,1082
2016/12	0,2880
2017/03	0,5603
2017/06	0,6979
2017/09	0,6752
2017/12	0,6680
2018/03	0,7666
2018/06	1,0076

2.3 SELEÇÃO DO MODELO

Agora que temos a série Z estimada pelo método de Belkin e Suchower, podemos passar ao próximo objetivo desta tese, que passa precisamente por tentar explicar/relacionar o comportamento desta série do fator-z com algumas variáveis macroeconómicas representativas da situação económica do país, como o PIB (taxa de variação homóloga), o *spread* (calculado como a diferença entre a taxa de rendibilidade das obrigações de tesouro e a taxa de uma bond governamental da área euro, a 10 anos), a taxa de desemprego, a taxa de juro, representada pela EURIBOR, o nível de exportação e o Índice de Preços ao Consumidor (IPC) com o intuito de poder fazer previsões da série Z.

Os dados obtidos para as séries das variáveis explicativas são provenientes das várias bases de dados disponíveis ao público: do BPStat (base de dados do Banco de Portugal, que por sua vez obtém os seus valores do INE ou do European Money Market Institute (EMMI), ou de outra fonte) obtiveram-se as séries da taxa de desemprego, do PIB nacional, das exportações e do IPCH (Índice de Preços ao Consumidor Harmonizado); do EMMI retirou-se a série da EURIBOR (EURO InterBank Offer Rate); finalmente a série do *spread* foi obtida indiretamente pelas OTs e Bonds (como já descrito) provenientes do Banco de Portugal e da base de dados do European Central Bank, respetivamente.

2.3.1 ANÁLISE PRELIMINAR

Antes de passar à estimação do modelo, executou-se uma pré seleção das variáveis com base no nível de correlação entre cada uma e a variável dependente, com o intuito de eliminar as variáveis que apresentassem um menor nível de correlação, sendo que foi considerado um *threshold* mínimo de correlação de 75% por corresponder a um nível de exigência considerado, e uma vez que queremos um modelo muito simplista para não perdermos demasiados graus de liberdade.

Pela análise da seguinte Tabela V podemos verificar que as variáveis com menor correlação são as séries do IPCH (-15%), a das exportações (38%) e a da EURIBOR (-29%). Todas as outras variáveis apresentam um nível de correlação acima dos 75%, pelo que nos parece sensato manter essas variáveis para a estimação do modelo.

TABELA V

ANÁLISE DE CORRELAÇÃO DAS VARIÁVEIS

	Z-Factor	R_EURIBOR	PIB_tvh	DESEMPREGO_tx	SPREAD	IPCH_tvh	EXPORT	Absolute Correlation	Variable for Multivariate Analysis
Z-Factor	100%							100%	Y
R_EURIBOR	-29%	100%						29%	N
PIB_tvh	76%	-74%	100%					76%	Y
DESEMPREGO_tx	-80%	54%	-74%	100%				80%	Y
SPREAD	-84%	68%	-92%	94%	100%			84%	Y
IPCH_tvh	-15%	62%	-58%	0%	29%	100%		15%	N
EXPORT	38%	-2%	44%	-29%	-38%	1%	100%	38%	N

2.3.2 ESTIMAÇÃO DO MODELO

Já na fase de estimação, o modelo escolhido, inicialmente, foi um simples modelo ADL (*Augmented Distributed Lags*), considerando todas as variáveis não eliminadas pela análise preliminar, e procedeu-se à típica análise e avaliação de modelos através de testes estatísticos que possibilitem a simplificação do modelo sem nunca comprometer, e até melhorando, a eficiência do mesmo (*trade-off* entre complexidade e eficiência). A este procedimento, que será explicado detalhadamente de seguida, dá-se o nome de GTS (*General-to-Specific*), por partir de um modelo mais geral, com uma sobreparametrização deliberada, para um modelo mais particular, e menos complexo, mas válido estatisticamente.

Como temos uma amostra com uns meros 28 pontos de observação, com periodicidade trimestral (entre Setembro de 2011 e Junho de 2018), não faz sentido escolher um S_{max} muito elevado, para não correr o risco de perder demasiada informação, portanto S_{max} será igual a 2. Temos então, para o modelo inicial de Z_t , um ADL(1,2) da seguinte forma

$$Z_t = \alpha + \phi Z_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \epsilon_t, \quad (24)$$

com $X_{t-i} = (PIB_{t-i}, Desemprego_{t-i}, Spread_{t-i})$, ($i = 0,1,2$) e cada δ_i representa um vetor coluna ($1 \times k$) com k igual ao número de variáveis incluídas em cada desfasamento, ou seja, neste caso, $k = 3$. Correndo este modelo em SAS obtemos os resultados presentes na Figura 4 do Anexo A. Perante estes resultados, a primeira medida a notar é a do uso de dois tipos de teste para autocorrelação dos erros do modelo: o teste de Durbin-Watson e o de Breusch-Godfrey. Neste caso, ambos os testes ditam a mesma conclusão. De facto, de acordo com a estatística de DW, que vem muito próxima de 2 ($d = 2,1351$), não devemos rejeitar a hipótese nula de ausência de

autocorrelação; da mesma forma o teste de Breusch-Godfrey não apresenta evidência estatística suficiente para rejeitar a hipótese de autocorrelação até à ordem 3, visto que o valor-p da estatística $p_{BG(3)} = 0,2452$ é bem maior que 0,05, valor aceite pela comunidade econométrica como a dimensão de referência para rejeição da hipótese nula dos testes econométricos.

No entanto, é preciso notar que, e seguindo Durbin, J. (1970), o teste de Durbin-Watson assenta numa série de hipóteses que caso não se verifiquem levam à invalidação da inferência estatística baseada nesse teste. Dentro dessas hipóteses existe a da exogeneidade estrita dos regressores. No entanto, como é facilmente visível, um dos regressores do modelo ADL é a própria variável dependente, desfasada um período, ou seja, uma variável endógena, muito longe da exogeneidade pretendida, pelo que neste caso devemos dar prioridade ao resultado do teste de Breusch-Godfrey, que coincidentemente é o mesmo.

Assim sendo, garante-se a validade de toda a inferência estatística aplicada futuramente. A conclusão imediata a retirar é, de facto, o elevado grau de evidência de insignificância estatística da variável Z_{t-1} (valor-p de 0.7994 \gg 0.05), tendo em conta que é uma variável muito importante, pois fornece à série uma componente autorregressiva, que neste caso parece ser não significativa. Para além disto, fez-se também um teste-F de significância global do conjunto das variáveis de desfasamento máximo, o qual devolveu um valor-p de 0,2304, muito elevado, levando-nos a acreditar que não existe evidência estatística suficiente para rejeitar a hipótese de que estas variáveis não são estatisticamente significativas em conjunto. De facto, tal como suspeitado, pela quantidade muito reduzida da amostra, a maioria das estatísticas aparece muito longe dos níveis de significância habituais.

Existe, então, a necessidade de simplificar o modelo, retirando um nível de desfasamento e a própria variável desfasada um período, ou seja, formulando agora um DL(1), da seguinte forma

$$Z_t = \alpha + \delta_0 X_t + \delta_1 X_{t-1} + \epsilon_t, \quad (25)$$

com X_{t-i} , ($i = 0,1$), a representar o mesmo conjunto de variáveis, e δ_i o vetor coluna (1×3).

Os resultados desta regressão aparecem na Figura 5 do Anexo A e, de novo, o teste de BG não nos dá evidência estatística suficiente para rejeitar a hipótese de ausência de autocorrelação dos erros até à ordem 3, aliás, a evidência é ainda mais forte ($p_{BG(3)} = 0,5231$). Nota para o facto de

agora a validade da estatística de Durbin Watson ser mais verosímil, uma vez que foi retirada a tal variável endógena. Continuaremos, ainda assim, a dar mais importância ao resultado do teste de Breusch-Godfrey em caso de contrassenso nos mesmos, pela sua maior robustez. Temos, mais uma vez, o teste-F de significância global, agora das variáveis desfasadas um período, que devolveu um valor-p de 0,1033, um valor um pouco intranquilo, devido ao facto de andar muito perto da rejeição a 10%, ainda assim estamos a considerar uma significância de 5%. No entanto, a evidência estatística de que as variáveis desfasadas um período não são significativas conjuntamente não parece ser muito forte, apesar de existir. Este valor pode ser resultante do facto de haver neste conjunto, variáveis não significativas que mais que compensam o efeito das variáveis significativas produzido neste tipo de testes. Desta forma, o próximo passo lógico parece ser o de olhar diretamente para as estatísticas-t de significância individual, em vez de eliminar completamente o desfasamento e obter um mero modelo estático.

Na verdade, notamos que das três variáveis desfasadas apenas uma parece ser estatisticamente significativa ($desemprego_{t-1}$, representado por *desempregolag* na Figura 5). Todas as outras estão muito longe da rejeição dessa hipótese de não significância, algo que vem de certa forma corroborar a ideia que fundamentou o recurso à análise destes valores.

Para resolver este problema é necessário proceder com lógica e de forma gradual sem nunca descurar o bom senso económico. Quer-se com isto dizer que apesar de a variável desfasada do PIB apresentar um valor-p muito alto (0,9238) sabe-se que o PIB é das variáveis que melhor descreve a situação económica do país, pelo que não seria correto retirar o efeito dinâmico desta variável ao modelo. A seguir ao PIB, a variável com menor evidência de significância é a do $spread_{t-1}$, representada por *spreadlag* na Figura 5, com um valor-p de 0,5653.

Ao retirarmos essa variável, obtemos o modelo

$$Z_t = \alpha + \delta_0 X_t + \delta_1 X_{t-1}^* + \delta_3 SPREAD_t + \epsilon_t, \quad (26)$$

onde agora $X_{t-1}^* = (PIB_{t-1}; DESEMPREGO_{t-1})$, ($i = 0,1$) e δ_1 passa a ser um vetor coluna (1×2), mas δ_0 continua (1×3). Os resultados desta regressão, obtidos em SAS, estão apresentados na Figura 6 do Anexo A. Pode-se notar, primeiramente que parece termos finalmente chegado à forma final do modelo em termos de escolha dos regressores e em termos de número de desfasamentos, uma vez que o teste-F de significância conjunta dos regressores de desfasamento

máximo do modelo apresenta um valor-p à volta do usual $\alpha = 0,05$ ($p_{teste-F} = 0,0503$: considera-se este valor suficiente uma vez que retirar mais variáveis ao modelo seria simplifica-lo demais, e o teste anda mesmo muito perto da zona de rejeição). Da mesma forma, o teste de significância global do modelo também produz a mesma conclusão, com evidência ainda mais forte (valor-p < 0,0001) e o próprio teste RESET, um teste à forma funcional do modelo, cuja estatística foi obtida com os quadrados, cubos e quartas potências dos valores ajustados da variável dependente, não rejeita a hipótese de os coeficientes dessas mesmas potências ($\hat{Z}_t^2, \hat{Z}_t^3, \hat{Z}_t^4$), provenientes da regressão auxiliar de teste, serem nulos, o que se traduz em suporte estatístico contra a hipótese de presença de incorreções na forma funcional. De notar ainda a ausência de autocorrelação dos erros, evidenciada pelos valores-p dos testes de Breusch-Godfrey e de Durbin-Watson, que ajuda à validade de todas as conclusões tiradas até aqui.

Finalmente, e seguindo da Silva Lopes, A. (1999), existe ainda um outro problema que advém do facto de as séries macroeconómicas serem muito autocorrelacionadas, ou seja, variarem de forma muito lenta, impondo possíveis problemas de multicolinearidade no modelo e conseqüente inconsistência do estimador do método dos mínimos quadrados (OLS), visto nos encontrarmos num ambiente não estacionário, sendo, portanto, necessário proceder à sua resolução. De facto, é visível através da análise da matriz de autocorrelação os valores próximos de 1, em módulo, entre os pares de variáveis *desemprego_tx* e *desempregolag* (-0,91938) e *PIB_tvh* e *piblag* (-0,74606). É recomendável tentar, pelo menos, suavizar estes problemas, uma vez que é muito complicado, ou até mesmo impossível, eliminar por completo esta condição.

Pode-se mostrar facilmente que o estimador OLS é invariante a transformações no modelo, desde que estas transformações sejam lineares. Desta forma, a transformação mais lógica parece ser a que tem por base a seguinte igualdade:

$$\Delta X_t = X_t - X_{t-1}, \quad (27)$$

Reformulando esta equação obtém-se

$$X_t = \Delta X_t + X_{t-1}. \quad (28)$$

Podemos, então, reescrever a equação (4) da seguinte forma:

$$Z_t = \alpha + \delta_0 \Delta X_t + \beta X_{t-1} + \delta_2 SPREAD_t + \epsilon_t, \quad (29)$$

onde $\Delta X_t = (\Delta PIB_t; \Delta DESEMPREGO_t)$, com δ_0 idêntico ao da equação (4), $X_{t-1} = (PIB_{t-1}; DESEMPREGO_{t-1})$, mas desta vez $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, sendo que $\beta_i = \delta_{0i} + \delta_{1i}$, $i = (1,2)$, onde por sua vez os δ_{0i} representam os coeficientes das variáveis em nível da equação (4), e os δ_{1i} representam os coeficientes das variáveis desfasadas nessa mesma equação. Os resultados são apresentados na Figura 7 do Anexo A, e pode-se notar logo o decréscimo dos valores da matriz de autocorrelações:

$$\begin{aligned} corr(difpib, piblag) &= 0,334247; \\ corr(difdesemprego, desempregolag) &= 0,26178, \end{aligned}$$

suavizando os problemas de colinearidade existentes, que invalidavam os métodos de inferência estatística clássicos. Mais uma vez, reforça-se que todos os modelos intermédios bem como o final estão inseridos no Anexo A.

Agora que parece ter-se chegado à forma funcional e paramétrica final do modelo, representada pela seguinte equação:

$$Z_t = 0,859 + 25,8372\Delta PIB_t + 27,5853PIB_{t-1} + 30,4781\Delta U_t - 4,1613U_{t-1} - 0,001379SPREAD_t \quad (30)$$

podemos proceder para o próximo objetivo, que é o de previsão da série Z_t e, consequentemente, das matrizes de migração e das probabilidades de default, para ajudar o banco em relação às suas expectativas de crédito.

Nesse sentido procurou-se primeiro calcular a série Z, para os períodos passados, através da aplicação do modelo estimado, e para períodos futuros, através da projeção das variáveis do modelo, recorrendo a dados mais uma vez provenientes do INE, do Banco de Portugal e do Banco Central Europeu, e completando manualmente para os restantes períodos seguindo a mesma tendência, para o cenário base. A comparação desta série estimada com a série “original” calculada pelo Solver é apresentada no Gráfico 8 do Anexo B, bem como os valores de projeção da série Z estimada com base nas projeções das variáveis macroeconómicas, na Tabela VI do Anexo B.

Por forma a prever antecipadamente qualquer mudança inesperada no estado económico do país decidiu-se, por fim, criar um cenário pessimista e otimista. De uma forma muito sucinta, para

o cenário otimista considerou-se um choque positivo no valor de 80% do desvio-padrão de cada variável, ou seja, $X_{Otimista} = X_{Base} \pm (0,8 * \sigma_{X_{Base}})$ (o sinal ' \pm ' advém do facto de que o choque positivo depende da relação da variável explicativa com a variável dependente). Para uma perspectiva pessimista, considerou-se o cenário exatamente oposto.

Obtidas as projeções das variáveis para os vários cenários, aplicaram-se esses valores na regressão estimada e estimaram-se as projeções para a série Z, calculando conseqüentemente as respectivas matrizes trimestrais. Dessas matrizes, retiraram-se os valores das PDs, valores da última coluna de cada matriz (a coluna dos *Defaults*) e procedeu-se à análise das curvas de projeção dessas mesmas PDs.

3. DISCUSSÃO

3.1 ANÁLISE DOS RESULTADOS

Ao longo do projeto foram várias as ocasiões em que foi necessário verificar e validar os resultados obtidos, tanto de acordo com a lógica matemática como com a lógica económica.

Os resultados obtidos, principalmente na parte da regressão estatística, equação (30), estão maioritariamente de acordo com o esperado, tendo em conta a teoria económica.

Antes de começar a análise, relembro que os coeficientes das variáveis em nível representam efeitos a curto prazo, e os das variáveis em primeiras diferenças representam efeitos de longo prazo.

O resultado associado ao PIB é um resultado curioso do ponto de vista interpretativo. De notar que neste caso a variável em questão trata-se de um segundo nível de diferenciação, ou seja, representa flutuações na taxa de variação do PIB. Desta forma, um decréscimo desta variável não significa forçosamente um decréscimo do PIB. Este sinal positivo mostra-nos assim que a probabilidade de incumprimento é diretamente proporcional à taxa a que o PIB cresce ou diminui. Ou seja, para períodos em que o aumento na taxa de crescimento não é tão grande, ou em que o decréscimo é mais acentuado, ou até períodos em que passamos de um aumento para um decréscimo do PIB, a probabilidade de incumprimento irá aumentar, mantendo-se todas as outras variáveis explicativas constantes, uma vez que o fator-Z diminuirá. Isto evidencia que a velocidade a que o PIB cresce também tem um impacto positivo na economia do país.

A variável do desemprego não apresenta um sinal tão esperado, no entanto há que ter em conta certos fatores capazes de explicar este resultado. A primeira coisa a notar é a diferença de sinais entre o multiplicador de curto prazo (positivo) e o de longo prazo (negativo). Neste caso, o pressuposto mais óbvio seria o de termos um coeficiente negativo sempre, visto que a taxa de desemprego é um indicador negativo da economia de um país, e consequentemente deveria provocar um aumento da probabilidade de incumprimento, o que acaba por acontecer no longo prazo. No entanto, no curto prazo parece haver um efeito que compensa o efeito negativo de um aumento na taxa de desemprego. Esse efeito dever-se-á em muitos casos, a um comportamento de resposta face à eminência da subida do rácio de incumprimento, o que tornará os lucros da empresa mais voláteis, obrigando o dono a proceder a um corte de capital humano, resultante de uma iniciativa que procurará deixar a empresa financeiramente mais aliviada, e com maior capacidade

de pagar os seus empréstimos, diminuindo portanto a probabilidade de incumprimento. Apesar disto, no longo prazo, um corte no capital humano traduzir-se-á num corte de produção e de rendimento, o que levará ultimamente ao inevitável *default*.

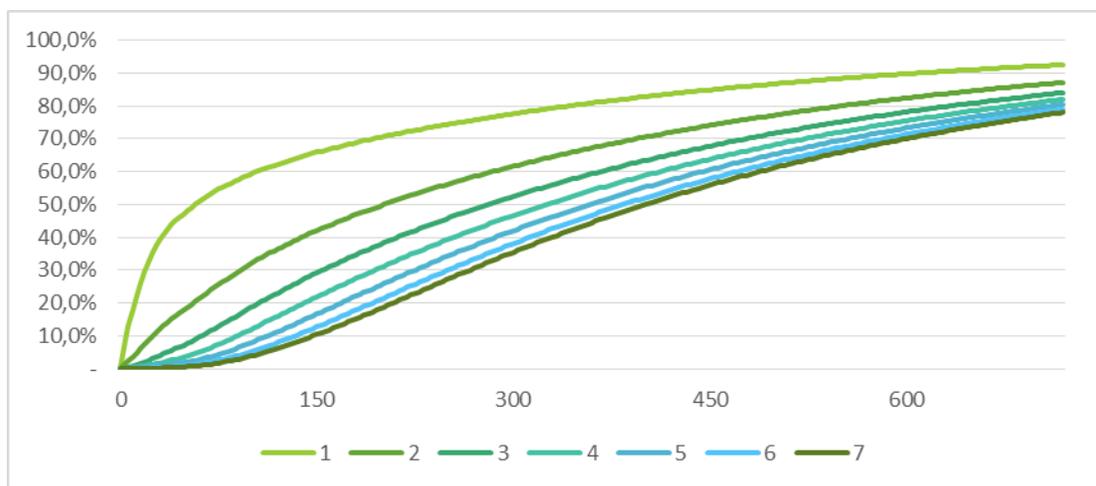
Por último, o *spread* é calculado como a diferença entre a taxa de rendibilidade das obrigações de tesouro e a taxa de uma bond governamental da área euro, a 10 anos. O aumento do spread das Obrigações do Tesouro (OT) face à bond governamental da área euro reflete diretamente no aumento relativo do custo de financiamento da república portuguesa e indiretamente das empresas, que, em última análise, se repercute num aumento do incumprimento por parte das empresas. Antecipa-se o cenário oposto no caso de uma redução do spread das OT face à bond governamental da área euro, daí a sua relação de proporcionalidade inversa.

3.2 APLICAÇÃO DO MODELO

Pela análise das curvas de projeção obtidas, algumas apresentadas de seguida nos Gráfico 1 (curvas por *rating*), Gráfico 3 e Gráfico 2 (curvas por cenário - as restantes no Anexo B, Gráficos 9-13), que mostram a evolução da probabilidade de incumprimento ao longo do tempo, em meses, pode-se retirar também excelentes e expectáveis conclusões.

GRÁFICO 1

CURVA DE PROJEÇÃO DE PDS POR NÍVEL PARA O CENÁRIO BASE



Percebe-se facilmente que na análise entre cenários para cada nível de classificação de crédito, Gráfico 3 e Gráfico 2, as curvas de projeção não se interseam, e tendem todas para o mesmo valor

no longo-prazo. Isto vai de acordo com o esperado porque logicamente que não faria sentido termos PDs maiores para o cenário otimista quando confrontado tanto com o cenário base como com o pessimista, em qualquer momento do tempo. O mesmo acontece para o cenário pessimista, onde é esperado termos sempre a série de PDs mais elevada.

GRÁFICO 2

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O PIOR NÍVEL DE CRÉDITO

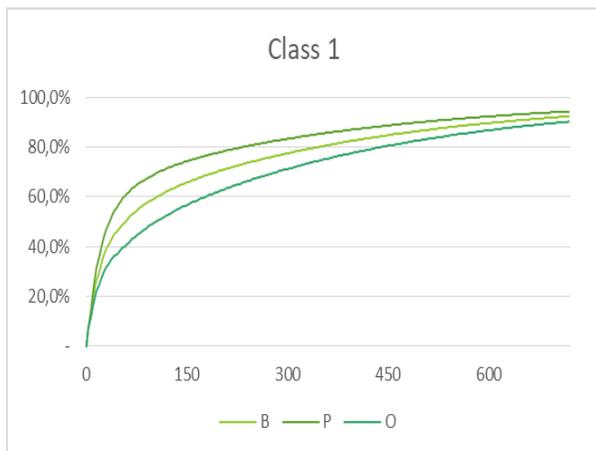
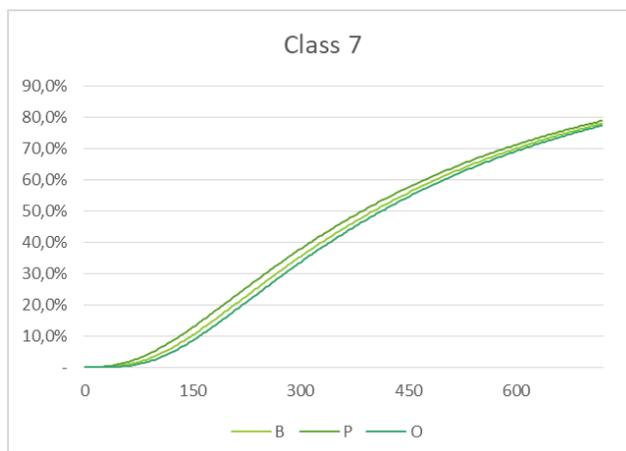


GRÁFICO 3

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O MELHOR NÍVEL DE CRÉDITO



Nota-se também que a definição de cenários é mais impactante em níveis de classificação mais baixos, ou seja, as diferenças entre as curvas de PD consoante os cenários são maiores para *ratings* mais baixos. Isto demonstra que os cenários foram bem definidos pois o efeito negativo, no caso de cenário pessimista, e o efeito positivo, no caso do otimista, acentuam-se bastante em níveis *rating* com grande probabilidade de incumprimento, casos esses que são os mais preocupantes e definidores do risco de um banco. Para os níveis mais elevados, espera-se sempre uma probabilidade de incumprimento muito reduzida, independentemente do cenário.

Na análise por nível de crédito (Gráfico 1) os resultados são igualmente promissores, no sentido em que a níveis de crédito mais elevados, ou seja, com menor probabilidade de entrar em incumprimento, correspondem valores de projeção de PD mais reduzidos. Estas condições mais uma vez verificam-se para todos os cenários em estudo.

De notar, por fim, que todas as curvas tendem para o mesmo valor no longo-prazo. Isto deve-se ao facto de considerarmos que a série-Z assume o valor zero a partir de um certo ponto no

tempo, visto que esse é um valor realmente muito próximo da sua média histórica, e portanto optou-se por considerar essa aproximação para previsões mais longínquas.

Conclui-se assim que o modelo sugerido parece explicar bem o comportamento geral de uma série Z refletora das matrizes de migração e, conseqüentemente da série de PDs ao longo do tempo. Estes valores de PDs seriam depois, num processo futuro e que vai além do âmbito do projeto, incluídas no cálculo do ECL, o que daria uma boa visão geral da situação do banco em termos de risco de crédito, ao fornecer estimativas para todos os cenários da perda esperada para cada cliente ao longo de toda a maturidade do contrato, no caso de clientes com baixo nível de classificação de crédito, ou ao longo do próximo ano, para clientes bem classificados. Logicamente que para os clientes que já entraram em incumprimento, a PD será de 1 e portanto não terá influência no cálculo deste ECL.

3.3 PROBLEMAS E HIPÓTESES DE DESENVOLVIMENTO FUTURO

Ao longo deste projeto foram de facto alguns os contratemplos com que me deparei. Primeiramente, a base de dados encontrava-se bastante incompleta e apresentava algumas incoerências económicas, o que tornava muito difícil a elaboração de matrizes de migração precisas e corretas. Para além da fraca qualidade dos dados, o horizonte temporal era também demasiado pequeno, principalmente tendo em conta os desfasamentos incluídos na regressão, o que impossibilitou a inclusão de mais variáveis macroeconómicas no modelo, sob o risco de perdermos demasiados graus de liberdade.

Outro problema encontrado foi a falta de bibliografia na parte de regressões entre o fator- Z e as componentes macroeconómicas. De facto, a maioria dos estudos e artigos disponíveis nesta vertente utilizam como variável dependente a própria PD e não uma *proxy* desta, apesar das vantagens da *proxy*, recorrendo mais frequentemente a Modelos Logit e Probit, uma vez que nesse caso a variável dependente varia sempre entre 0 e 1. No entanto estes tipos de modelo não fazem sentido quando temos uma série que pode assumir qualquer valor positivo ou negativo. Existem também outras alternativas propostas que não foram tidas em conta neste trabalho, como é o caso do FAVAR (Referência ao autor deste paper).

Em relação a possíveis trabalhos futuros nesta vertente, de destacar o facto de este procedimento ser baseado nas normas da EBA, que estão em constante renovação para fazer face ao desenvolvimento económico, e sendo também que cada banco pode optar por diferentes abordagens no que toca à estimação de PDs, é quase certo que no futuro este processo sofra algumas alterações.

4. REFERÊNCIAS BIBLIOGRÁFICAS

- Akaike, H. (1998). *Information Theory and an Extension of the Maximum Likelihood Principle*. Selected Papers of Hirotugu Akaike, pp. 199–213.
- Altman, E. I. (1968). *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*. The Journal of Finance, 23 (4), pp. 589-609.
- Altman, E. I., Saunders, A. (1998). *Credit risk measurement: Developments over the last 20 years*. Journal of Banking and Finance. Salomon Brothers Center, Leonard Stern School of Business, New York University, New York.
- Aven, T. (2012). *The risk concept—historical and recent development trends*. Reliability Engineering & System Safety, 99, pp. 33–44.
- Beaver, W. H. (1966). *Financial Ratios as Predictors of Failure*. Journal of Accounting Research, Vol. 4, Empirical Research in Accounting: Selected Studies 1966 (1966), pp. 71-111.
- Belkin, B., Suchower, S. & Forest, L. (1998). *A one-parameter representation of credit risk and transition matrices*. CreditMetrics Monitor, pp. 46-55.
- Breusch, T. S. (1978). *Testing for Autocorrelation in Dynamic Linear Models*. Australian Economic Papers. 17: pp. 334–355.
- Durbin, J., Watson, G. S. (1950). *Testing for Serial Correlation in Least Squares Regression, I*. Biometrika. 37 (3–4): pp. 409–428.
- Durbin, J. (1970). *Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables*. Econometrica, 38 (3), pp. 410-421.
- Gilbert, C. L. (1989). *LSE and the British Approach to Time Series Econometrics*, Oxford Economic Papers, 41, 1, pp. 108-128.
- Godfrey, L. G. (1978). *Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables*. Econometrica. 46: 1293–1301.
- Holton, G. A. (2004). *Defining Risk*. Financial Analysts Journal, Volume 60, 2004.

IFRS 9 Financial Instruments, International Accounting Standards Board (IASB)

Keynes, J. M. (1921). *Treatise on Probability*, Macmillan & Co., Londres

Knight, F. H. (1921). *Risk, Uncertainty and Profit*, Reprints of Economic Classics. Augustus M. Kelley, Bookseller. New York 1964.

Lopes, A.S. (1999). *Modelos DL e ADL, Raízes Unitárias e Cointegração: uma Introdução*. Econometria II. Universidade Técnica de Lisboa. Instituto Superior de Economia e Gestão: pp. 8-12

Ramsey, J. B. (1969). *Tests for Specification Errors in Classical Linear Least Squares Regression Analysis*. Journal of the Royal Statistical Society, Series B. 31 (2): pp. 350–371.

Sakai, Y. (2016). *J.M. Keynes and F.H. Knight: How to Deal with Risk, Probability and Uncertainty*. CRR Discussion Paper Series, Discussion Paper No. A-15. Center for Risk Research, Faculty of Economics, SHIGA University, Japan.

Stigler, G. J. (1987). *The Theory of Price*, Fourth Edition. New York: Macmillan.

5. ANEXO

5.1 ANEXO A

Resultados da Seleção do Modelo

FIGURA 4

MODELO ADL(1,2) DA SÉRIE Z, COM TESTES DE ESPECIFICAÇÃO

The AUTOREG Procedure					
Ordinary Least Squares Estimates					
SSE	2.36225	DFE			15
MSE	0.15748	Root MSE			0.39684
SBC	47.26333	AIC			33.42427
MAE	0.22388	AICC			52.28141
MAPE	555.92746	HQC			37.40942
Durbin-Watson	2.1351	Total R-Square			0.9134
Godfrey's Serial Correlation Test					
Alternative	LM	Pr > LM			
AR(1)	1.0535	0.3047			
AR(2)	3.3458	0.1877			
AR(3)	4.1546	0.2452			
Ramsey's RESET Test					
Power	RESET	Pr > F			
2	4.2304	0.0588			
3	3.0721	0.0808			
4	1.9693	0.1725			
Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	0.8819	0.9121	0.97	0.3489
z_factorlag	1	-0.0610	0.2359	-0.26	0.7994
PIB_tvh	1	33.2019	12.5778	2.64	0.0186
piblag	1	-10.5783	15.8062	-0.67	0.5135
piblag2	1	7.3249	11.9013	0.62	0.5475
DESEMPREGO_tx	1	21.9033	13.9831	1.57	0.1381
desempregolag	1	2.8856	21.5961	0.13	0.8955
desempregolag2	1	-27.0343	14.8677	-1.82	0.0890
SPREAD	1	-0.00319	0.001372	-2.33	0.0343
spreadlag	1	0.001897	0.001824	1.04	0.3149
spreadlag2	1	-0.001169	0.001223	-0.96	0.3543

Test 1				
Source	DF	Mean Square	F Value	Pr > F
Numerator	30.252531		1.600	0.2304
Denominator	150.157483			

FIGURA 5

MODELO DL (1) DA SÉRIE Z, COM TESTES DE ESPECIFICAÇÃO E VALIDAÇÃO (TEST 1 É UM TESTE-F AOS COEFICIENTES DE DESFASAMENTO)

The AUTOREG Procedure

Ordinary Least Squares Estimates		
SSE	4.1858485	DFE 20
MSE	0.20929	Root MSE 0.45748
SBC	49.3620981	AIC 40.29124
MAE	0.27408325	AICC 46.1859769
MAPE	307.372787	HQC 42.9884817
Durbin-Watson	1.9954	Total R-Square 0.8487

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.9954	0.2220	0.7780

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Godfrey's Serial Correlation Test		
Alternative	LM	Pr > LM
AR(1)	0.2666	0.6056
AR(2)	1.9095	0.3849
AR(3)	2.2452	0.5231

Ramsey's RESET Test		
Power	RESET	Pr > F
2	0.0145	0.9053
3	0.2623	0.7721
4	0.1654	0.9182

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	0.9611	0.8385	1.15	0.2652
PIB_tvh	1	24.2647	12.5959	1.93	0.0684
piblag	1	1.2172	12.5611	0.10	0.9238
DESEMPREGO_tx	1	30.9452	14.2947	2.16	0.0427
desempregolag	1	-35.2857	13.9535	-2.53	0.0200

SPREAD	1	-0.000864	0.001097	-0.79	0.4399
spreadlag	1	-0.000703	0.001202	-0.58	0.5653

Test 1				
Source	DF	Mean Square	F Value	Pr > F
Numerator	30.491344		2.350	0.1033
Denominator	200.209292			

FIGURA 6

MODELO DL (1) SEM SPREADLAG, TABELA DE AUTOCORRELAÇÕES COM TESTES DE ESPECIFICAÇÃO E VALIDAÇÃO (TEST 1 É UM TESTE-F AOS COEFICIENTES DE DESFASAMENTO, TEST 2 É UM TESTE DE SIGNIFICÂNCIA GLOBAL

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	4.25739822	DFE	21
MSE	0.20273	Root MSE	0.45026
SBC	46.5238787	AIC	38.7488575
MAE	0.27654485	AICC	42.9488575
MAPE	399.439299	HQC	41.0607789
Durbin-Watson	1.8830	Total R-Square	0.8461

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.8830	0.1681	0.8319

Godfrey's Serial Correlation Test		
Alternative	LM	Pr > LM
AR(1)	0.0763	0.7824
AR(2)	1.6817	0.4313
AR(3)	1.9145	0.5904

Ramsey's RESET Test		
Power	RESET	Pr > F
2	0.0008	0.9778
3	0.4027	0.6741
4	0.3497	0.7898

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	0.8590	0.8072	1.06	0.2993
PIB_tvh	1	25.8372	12.1110	2.13	0.0449
piblag	1	1.7481	12.3304	0.14	0.8886
DESEMPREGO_tx	1	30.4781	14.0469	2.17	0.0416
desempregolag	1	-34.6394	13.6900	-2.53	0.0195
SPREAD	1	-0.001379	0.000643	-2.14	0.0439

Correlation of Parameter Estimates						
	Intercept	PIB_tvh	piblag	DESEMPREGO_tx	desempregolag	SPREAD
Intercept	1	-0.1057	-0.47996	-0.1792	-0.20869	-0.53039
PIB_tvh	-0.1057	1	-0.74606	0.255632	-0.26154	0.351515
piblag	-0.47996	-0.74606	1	-0.00859	0.22074	0.132241
DESEMPREGO_tx	-0.1792	0.255632	-0.00859	1	-0.91938	-0.13299
desempregolag	-0.20869	-0.26154	0.22074	-0.91938	1	0.262474
SPREAD	-0.53039	0.351515	0.132241	-0.13299	0.262474	1

Test 1				
Source	DF	Mean Square	F Value	Pr > F
Numerator	20	701242	3.46	0.0503
Denominator	210	202733		

Test 2				
Source	DF	Mean Square	F Value	Pr > F
Numerator	54	681290	23.09	<.0001
Denominator	210	202733		

FIGURA 7

MODELO DL (1) FINAL, COM TRANSFORMAÇÕES LINEARES E RESPECTIVOS TESTES DE ESPECIFICAÇÃO E VALIDAÇÃO (TEST 1 É UM TESTE DE SIGNIFICÂNCIA GLOBAL

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	4.25739822	DFE	21
MSE	0.20273	Root MSE	0.45026
SBC	46.5238787	AIC	38.7488575
MAE	0.27654485	AICC	42.9488575
MAPE	399.439299	HQC	41.0607789
Durbin-Watson	1.8830	Total R-Square	0.8461

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.8830	0.1681	0.8319

Godfrey's Serial Correlation Test		
Alternative	LM	Pr > LM
AR(1)	0.0763	0.7824
AR(2)	1.6817	0.4313
AR(3)	1.9145	0.5904

Ramsey's RESET Test		
Power	RESET	Pr > F
2	0.0008	0.9778
3	0.4027	0.6741
4	0.3497	0.7898

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	0.8590	0.8072	1.06	0.2993
difpib	1	25.8372	12.1110	2.13	0.0449
piblag	1	27.5853	8.7117	3.17	0.0047
difdesemprego	1	30.4781	14.0469	2.17	0.0416
desempregolag	1	-4.1613	5.5799	-0.75	0.4641
SPREAD	1	-0.001379	0.000643	-2.14	0.0439

Correlation of Parameter Estimates						
	Intercept	difpib	piblag	difdesemprego	desempregolag	SPREAD
Intercept	1	-0.1057	-0.82628	-0.1792	-0.96314	-0.53039
difpib	-0.1057	1	0.334247	0.255632	0.001872	0.351515
piblag	-0.82628	0.334247	1	0.343219	0.738515	0.675851
difdesemprego	-0.1792	0.255632	0.343219	1	0.26178	-0.13299
desempregolag	-0.96314	0.001872	0.738515	0.26178	1	0.309183
SPREAD	-0.53039	0.351515	0.675851	-0.13299	0.309183	1

Test 1				
Source	DF	Mean Square	F Value	Pr > F
Numerator	5	4.681290	23.09	<.0001
Denominator	210	0.202733		

5.2 ANEXO B

Resultados da Previsão

GRÁFICO 8

SÉRIE Z ESTIMADA VS CALCULADA

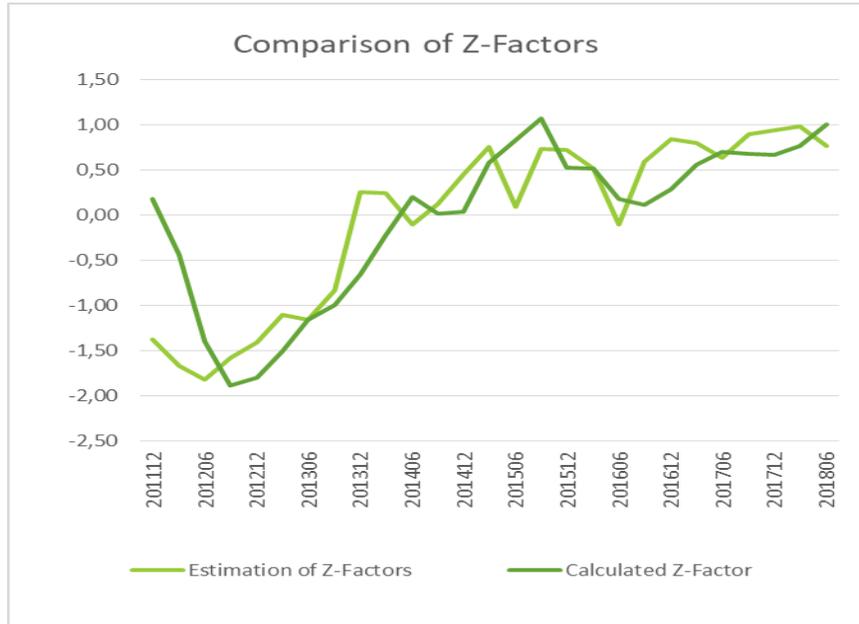


TABELA VI

ESTIMATIVAS DAS PROJEÇÕES PARA A SÉRIE Z

Reference Date	Estimation of Z-Factors
201809	1,08925
201812	1,07586
201903	0,95603
201906	0,96610
201909	0,94859
201912	0,95516
202003	0,98634
202006	0,99155
202009	0,99675
202012	1,00195
202103	1,00715
202106	1,01235
202109	1,01755
202112	1,06085

GRÁFICO 9

CURVA DE PROJEÇÃO DE PDS POR NÍVEL PARA O CENÁRIO OTIMISTA

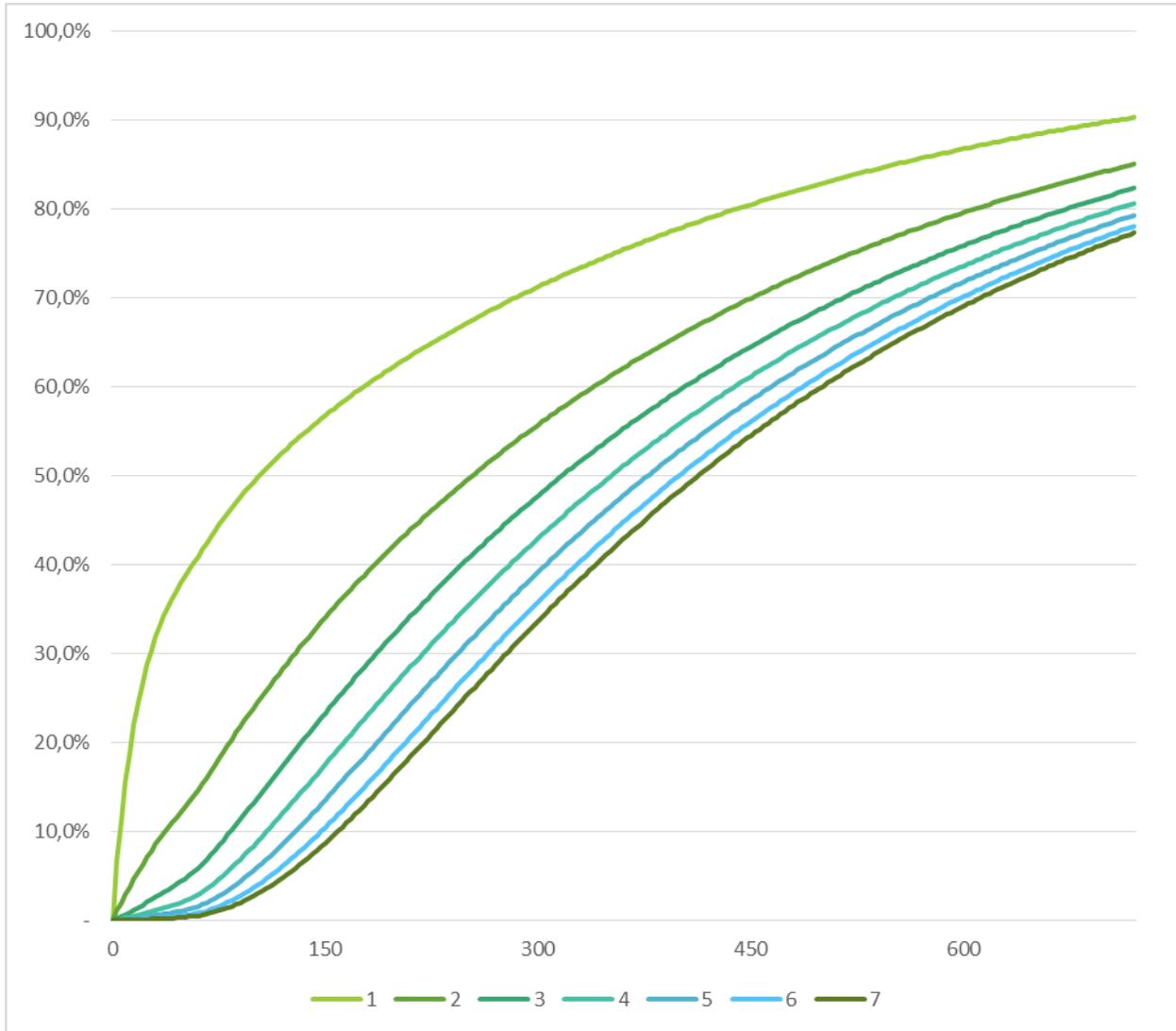


GRÁFICO 10

CURVA DE PROJEÇÃO DE PDS POR NÍVEL PARA O CENÁRIO PESSIMISTA

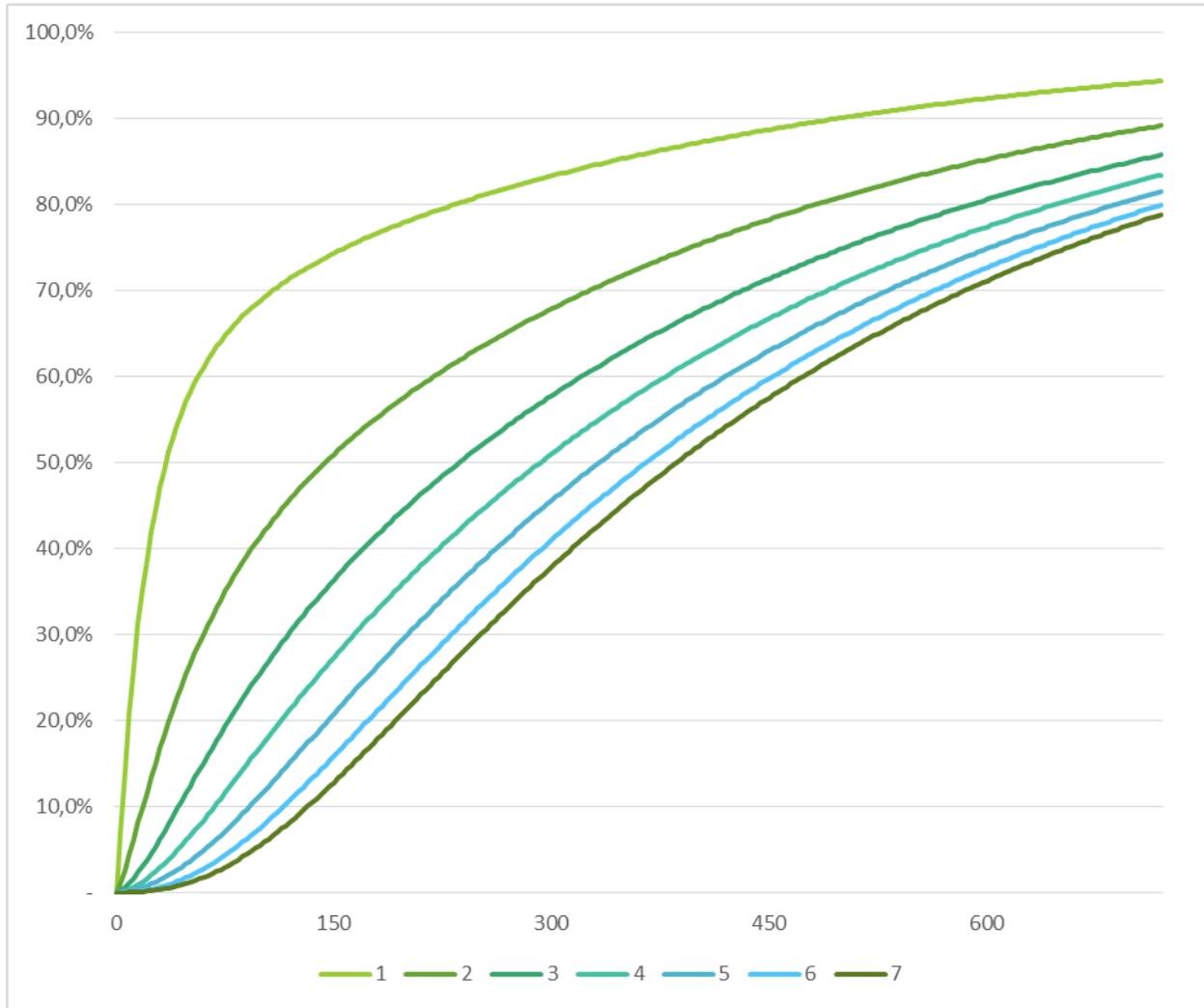


GRÁFICO 11

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O NÍVEL 2 DE CRÉDITO

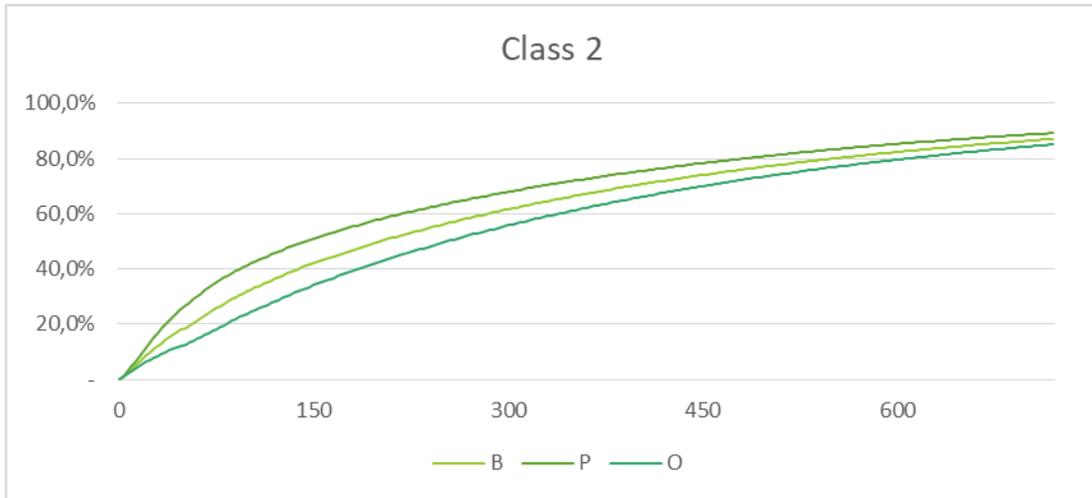


GRÁFICO 12

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O NÍVEL 3 DE CRÉDITO

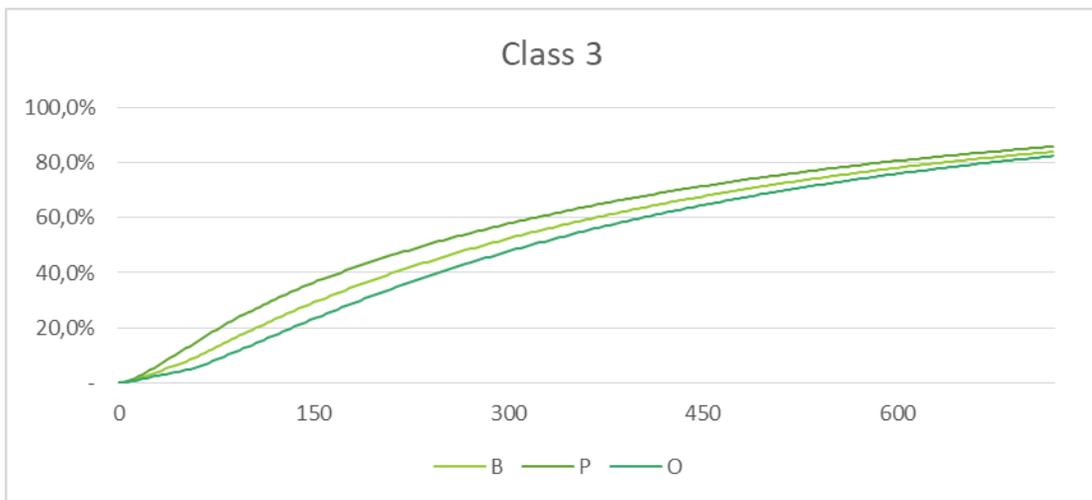


GRÁFICO 13

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O NÍVEL 4 DE CRÉDITO

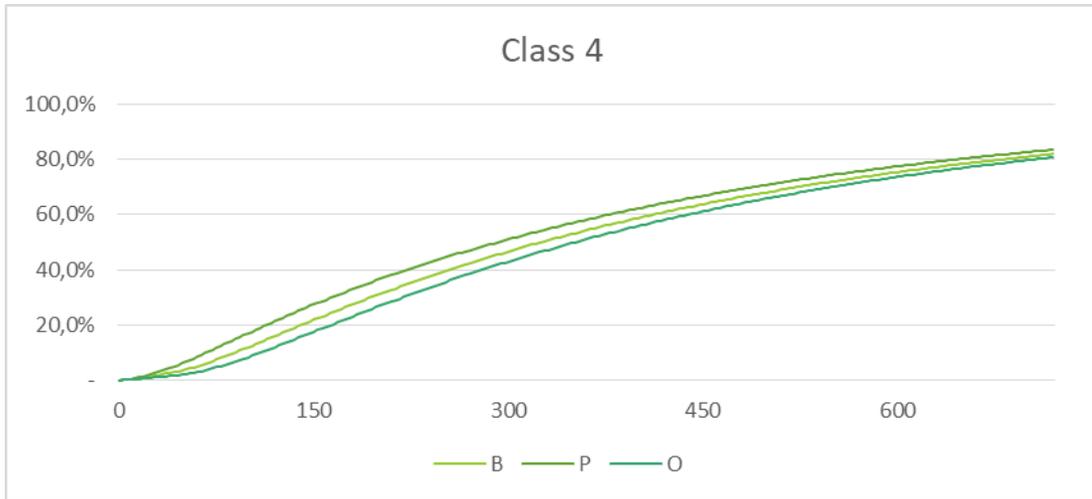


GRÁFICO 14

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O NÍVEL 5 DE CRÉDITO

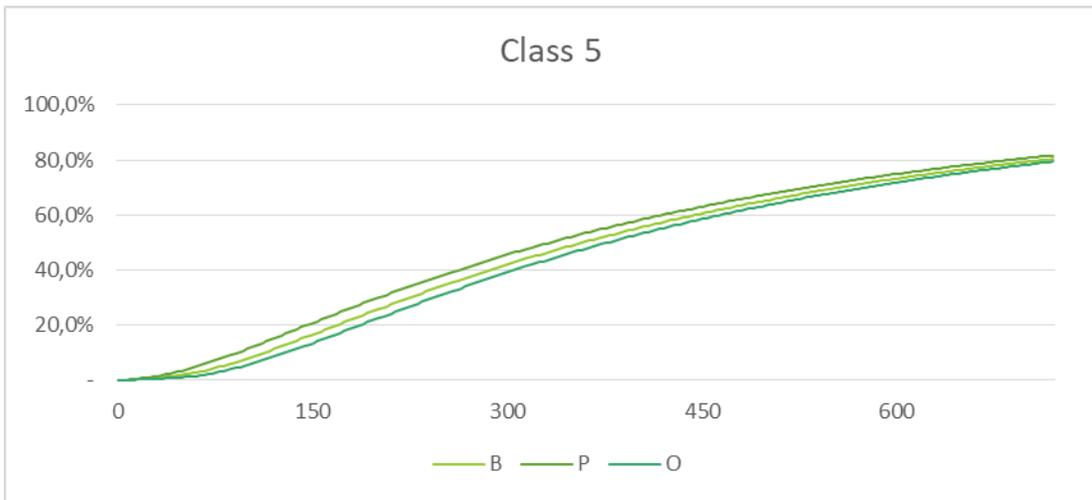


GRÁFICO 15

CURVA DE PROJEÇÃO DE PDS POR CENÁRIO PARA O NÍVEL 6 DE CRÉDITO

