

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Fatores determinantes do *rating* soberano português

Miguel Cortes Rodrigues

Mestrado em Matemática Aplicada à Economia e Gestão

Trabalho de Projeto orientado por:
Maria Teresa dos Santos Hall de Agorreta de Alpuim
Patrícia Alexandra Henriques Silva

2017

Agradecimentos

Dedico este espaço a quem foi importante para que a execução deste trabalho fosse possível.

Primeiramente, aos meus pais por todo o esforço que tiveram comigo durante o meu percurso académico, paciência e pela força que sempre me transmitiram de forma a superar as contrariedades que foram surgindo.

À minha namorada, Inês Graça. Um grande obrigado por tudo.

Às minhas orientadoras, Teresa Alpuim e a Patrícia Silva pela ajuda, disponibilidade e partilha de interesse em fazer mais e melhor. Obrigado por todas as reuniões de *brainstorming* e pelo companheirismo. Também agradeço à Professora Raquel Fonseca pela revisão da dissertação e discussão de alguns aspetos económicos.

Aos meus amigos, em especial ao Tomé Salgueiro que me ajudou a clarificar alguns temas macroeconómicos, e ainda aos meus colegas do Novo Banco que foram também muito importantes nesta etapa.

Por fim, à Faculdade de Ciências da Universidade de Lisboa por me ter permitido desenvolver novas ferramentas para atacar os problemas do mundo real.

Resumo

O risco soberano é a probabilidade de uma entidade soberana incumprir as suas obrigações contratuais. Esta probabilidade é avaliada e calculada por diferentes agências de atribuição de *rating*, que têm em conta não só diferentes variáveis macroeconómicas, mas como outros fatores qualitativos. Cada uma das referidas agências utiliza diferentes estimativas e estimadores para atribuição de uma notação alfanumérica contida na categoria de investimento ou na categoria de notação especulativa.

Este trabalho visa encontrar as variáveis determinantes para a atribuição do *rating* soberano no caso português nos últimos 21 anos (1996-2015) através de uma seleção inicial de variáveis tendo por base racional económico no caso português e a revisão bibliográfica. Maximizando a redução de situações de multicolinearidade através dos fatores da inflação da variância e a análise de componentes da variância e ainda aplicando os métodos de seleção de variáveis mais conhecidos - *backwards*, *stepwise* e *forward* linear - chegou-se à conclusão que os fatores determinantes são a variação do PIB (a preços constantes, base de 2011), a taxa Euribor, o endividamento de particulares, o rácio de Importações e Exportações sobre o PIB e ainda as Reservas Totais. Além disso, assegurou-se a qualidade da regressão com a validação da normalidade dos resíduos.

Em seguida, foi estudada a influência do ciclo económico nos fatores determinados selecionados e chegou-se à conclusão que, além do reduzido significado estatístico devido à redução da amostra inicial, poucas são efetivamente significativas em cada uma das fases do ciclo económico.

Por fim, estudou-se o efeito do *rating* no juro da dívida soberana portuguesa e chegou-se à conclusão que não se incorporavam as forças da procura do mercado nem os fatores qualitativos neste ajustamento, pelo que se decidiu replicar a mesma análise para um prémio de mercado, que consistiu na diferença entre o juro da dívida portuguesa e o juro da dívida alemã. Com efeito, concluiu-se que o *rating* era significativo na explicação do prémio de mercado, aliado dos três dos cinco fatores determinantes previamente selecionados, a variação do PIB, a taxa Euribor e ainda as Reservas Totais.

Palavras-chave: risco soberano, variáveis macroeconómicas, multicolinearidade, ciclo económico

Abstract

The sovereign risk is the probability of a sovereign entity defaulting its contractual obligations. This probability is evaluated and calculated by different rating agencies considering macroeconomic variables and qualitative factors. Each of these agencies uses different estimation methodologies to attribute an alphanumeric grade that can range from “investment grade” to “speculative/non-investment grade”.

This project aim is to find the determinant factors in the sovereign rating of the Portuguese Republic for the period between 1996 and 2015. After extensive literature review and some economic expert consultation, several initial variables, applicable to the Portuguese case, were set. To reach significant variables in determining the sovereign risk level, some mathematical methods were used, including multicollinearity cleansing (through VIF and ACV) and variable selection methods (backwards, stepwise, linear forward). These processes led to the selection of the following variables: “GDP variation (constant prices, 2011)”, “Euribor”, “Private Indebtedness”, “(Imports+Exports)/GDP” and “Total Reserves”. In addition, the regression quality was verified by the normal distribution of the residuals.

Afterwards, the influence of the economic cycle on the selected determinants was studied. Few of the selected variables seem to be statistically significant in each of the economic cycle phases. However, the low statistical significance of this study, due to the reduction of the initial sample, should be considered.

Finally, the effect of the rating on Portuguese sovereign debt interest was studied and it was concluded that neither the demand forces of market nor the qualitative factors in this adjustment were incorporated. To fix this issue, an analysis of the premium market between Portuguese and German debt interest was conducted. Everything considered, it was concluded that the rating was significant in explaining the market premium, together with three of the five variables previously selected: the GDP variation, the Euribor rate and the Total Reserves.

Keywords: sovereign risk, macroeconomic variables, multicollinearity, economic cycle

Índice geral

Agradecimentos.....	i
Resumo	iii
Abstract.....	iv
1. Introdução.....	1
2. Revisão de Literatura.....	2
2.1 Risco soberano	2
2.2 Perspetiva Histórica do risco soberano	3
2.2.1 O caso português.....	5
2.3 Sistemas de rating das principais agências internacionais	7
2.4 Determinantes do risco soberano	10
3. Metodologia de Investigação e Conceitos Teóricos.....	14
3.1 Modelos Matemáticos	14
3.1.1 Regressão linear múltipla	15
3.1.2 Método dos Mínimos Quadrados.....	16
3.1.3 Intervalo de confiança e intervalo de predição	17
3.1.4 Testes nos coeficientes de regressão	19
3.1.5 Análise do R^2 e S^2	19
3.1.6 Tabela ANOVA.....	21
3.1.7 Análise de resíduos	23
3.2 Estudo da multicolinearidade entre variáveis	24
3.2.1 Multicolinearidade	24
3.2.2 Método VIF – Fatores de inflação das Variâncias	25
3.2.3 Números condição e componentes da variância	26
3.2.4 Métodos de seleção de variáveis	27
3.3 Tratamento da informação e construção do modelo estatístico	29
3.3.1 Variável dependente	29

3.3.2	Variáveis independentes.....	30
3.3.3	Modelação e estimação dos parâmetros da regressão linear múltipla.....	31
3.3.4	Análise aos fatores determinantes do risco soberano nos períodos antes e durante/depois da crise.....	32
3.3.5	Análise à taxa de juro soberano e prémio de mercado.....	33
4.	Análise dos determinantes do rating	34
4.1	Análise da variável resposta, rating.....	34
4.2	Análise das variáveis independentes	37
4.3	Estimação do modelo de regressão linear múltipla.....	38
4.3.1	Análise de multicolinearidade das variáveis independentes.....	39
4.3.2	Método de seleção de variáveis – stepwise, backwards, linear forward	42
4.3.3	Avaliação da qualidade do modelo.....	47
4.3.3.1	Teste F e tabela ANOVA.....	47
4.3.3.2	Análise da relação entre as variáveis independentes e o rating	47
4.3.3.3	Análise dos resíduos.....	48
4.3.3.4	Intervalo de confiança e análise ao conservadorismo do modelo.....	50
4.4	Análise aos fatores determinantes do risco soberano nos dois períodos de análise.....	52
4.5	Impacto do rating e dos fatores determinantes do risco português na taxa de juro da dívida portuguesa e no prémio de mercado	54
5.	Conclusões	59
6.	Sugestões de trabalho futuro	61
	Referências Bibliográficas.....	62
	Anexos.....	64
A.	Descrição do set inicial de variáveis.....	64
B.	Análise VIF preliminar ao set inicial de variáveis	66

Índice de Figuras

Figura 1 - Variação do PIB e desemprego anual (%) entre 1996 e 2016 em Portugal [15] [16].....	6
Figura 2 - Passos na atribuição do rating por parte de uma agência das “Big Three” [2] [5].	8
Figura 3 - A influência do rating e do mercado nos juros da dívida soberana [5].	10
Figura 4 - Resíduos vs variável explicativa – caso de concordância com o modelo linear	23
Figura 5 - Metodologia de estudo do juro face ao rating soberano português.....	33
Figura 6 - Distribuição do rating do risco soberano português no período em análise: 1996-2015 ...	37
Figura 7 - Evolução das variáveis: Importações, Exportações e PIB (M€)	46
Figura 8 - Variáveis independentes contra o rating.	48
Figura 9 - Representação dos resíduos vs variáveis independentes.....	49
Figura 10 - Histograma dos resíduos.....	50
Figura 11 - Rating observado, rating previsto e intervalo de previsão.	51
Figura 12 – Diferença dos juros entre o mês m e o mês m-1 e o rating (análise: 01/1995-04/2017).	55
Figura 13 - Juro da dívida portuguesa, prémio de mercado e o rating para o período de análise (45 observações)	56

Índice de tabelas

Tabela 1 - Impacto da crise económica nos países "PIIGS" (setembro 2008 - setembro 2011) [11]... 4	4
Tabela 2 - Comparação do juro da dívida portuguesa a 10 anos e a notação da dívida soberana portuguesa num período 12 meses (entre fevereiro de 2012 a fevereiro de 2013)..... 9	9
Tabela 3 - Tabela ANOVA para aferir a significância do modelo. [21]..... 21	21
Tabela 4 - Tabela ANOVA para testar igualdade de médias entre populações. [23] 22	22
Tabela 5 - Classificação da magnitude da correlação entre variáveis. [28]..... 25	25
Tabela 6 - Tabela com valores próprios, números condição e componentes da variância. [26]..... 27	27
Tabela 7- Métodos de seleção de variáveis [21] 28	28
Tabela 8 -Master Scale (S&P's, Moody's e Fitch) [30] [19] [31] 30	30
Tabela 9 - Ratings atribuídos pelas três agências (1996-2015)..... 34	34
Tabela 10 - Diferença de notches entre os três pares de CRA (valores absolutos) 35	35
Tabela 11 - Diferença de notches, valores relativos, em módulo..... 35	35
Tabela 12 - Correlação entre os três pares de CRA. 35	35
Tabela 13 - Teste ANOVA para igualdade de médias entre populações..... 36	36
Tabela 14 - Características amostrais do rating do risco soberano português no período em análise. 36	36
Tabela 15 - Descrição sumária estatística das variáveis em análise (79 observações, entre 1996-2015)..... 38	38
Tabela 16 - Seleção de variáveis através de sucessivas iterações com o método VIF 40	40
Tabela 17 - Análise às componentes da variância para as 9 variáveis resultantes do método VIF... 40	40
Tabela 18 - Matriz de correlações: Poupanças das Famílias, Endividamento Particulares, (Importações+Exportações) /PIB e Reservas Totais..... 40	40
Tabela 19 - Matriz de correlações: Custos unitários laborais, Endividamento Particulares, (Importações+Exportações) /PIB e Reservas Totais..... 41	41
Tabela 20 - Matriz de correlações: Leverage banking, Endividamento Particulares, (Importações+Exportações) /PIB e Reservas Totais..... 42	42
Tabela 21 - Construção do modelo pelo método backwards 42	42
Tabela 22 - Construção do modelo pelo método stepwise. 42	42
Tabela 23 - Construção do modelo pelo método forward linear..... 43	43

Tabela 24 – Modelo obtido através dos três diferentes métodos de seleção de variáveis.....	43
Tabela 25 - Tabela ANOVA para o modelo sugerido.	47
Tabela 26 - Principais características amostrais dos resíduos e valores necessários para construção do histograma.....	49
Tabela 27 - Resultados do teste do Qui-Quadrado.	50
Tabela 28 – Correlações entre as diferentes agências para dois períodos de análise.	52
Tabela 29 – Características dos ratings atribuídos pelas 3 CRA nos dois períodos de análise.	52
Tabela 30 – Resumo relativo à estatística F e respetivo p-value para ambos os períodos em análise.	52
Tabela 31 – Ajuste pelo mínimo dos métodos quadrados e significância estatística de cada variável (período 1996-2008).....	53
Tabela 32 – Ajuste pelo mínimo dos métodos quadrados e significância estatística de cada variável (período 2009-2015).....	53
Tabela 33 – Análise rating vs juro da dívida soberana portuguesa.	54
Tabela 34 – Análise rating vs prémio de mercado.	55
Tabela 35 – Análise variáveis vs juro ou prémio de mercado.	56
Tabela 36 – Análise das variáveis para cada uma das taxas em estudo (com rating).....	57
Tabela 37 – Análise das variáveis para cada uma das taxas em estudo (sem rating).	57
Tabela 38 – Análise das variáveis para cada o prémio de mercado (com rating e sem endividamento de particulares).....	58
Tabela 39 – Qualidade da regressão linear através da seleção de variáveis pelo método backwards.	58
Tabela 40 - Variáveis, descrição e fonte de dados.....	65
Tabela 41 – Análise VIF (1996-2015).	66

Simbologia e Notações

€	Euro – Moeda usada na União Europeia
ACV	Análise das Componentes da Variância
BCE	Banco Central Europeu
BLUE	<i>Best Linear Unbiased Estimators</i>
CDS	<i>Credit Default Swap</i>
CRA	<i>Credit Rating Agency</i>
DBRS	Agência canadiana que atribui <i>rating</i> ao risco soberano. É uma das 4 <i>credit rating agencies</i> reconhecida pelo BCE.
EAD	<i>Exposure at Default</i>
EL	<i>Expected Loss</i> . Dado pela seguinte fórmula: $EL = PD \times LGD \times EAD$
EMQ	Estimadores de Mínimos Quadrados
EU	União Europeia
FMI	Fundo Monetário Internacional
IDH	Índice de Desenvolvimento Humano
LGD	<i>Loss given default</i> . Dado pela seguinte fórmula: $LGD = 1 - RR$, onde RR é Recovery Rate
NPL	<i>Non Performing Loans</i>
Pb	Pontos base
PD	Probabilidade de <i>default</i>
PIB	Produto Interno Bruto
PIIGS	Acrónimo usado para os países com maior instabilidade económica na União Europeia (<i>Portugal, Irland, Italy, Greece, Spain</i>)
VIF	<i>Variance Inflation Factors</i>

1. Introdução

O objetivo deste projeto é encontrar e analisar os principais determinantes do *rating* soberano português no período referente aos últimos 21 anos, o qual compreende todos os *ratings* atribuídos entre 1996 e 2015 pelas três principais agências, geralmente designadas pelo grupo das “Big Three”: a Standard & Poor’s (S&P’s), a Fitch e a Moody’s. Para tal, recorreu-se ao ajustamento de um modelo de regressão linear múltipla, utilizando diferentes métodos de seleção de variáveis.

Tendo em conta os seus principais objetivos, o presente trabalho desenvolve-se de acordo com os seguintes eixos de análise:

- 1) Determinar as variáveis que melhor explicam a variação observada no *rating* soberano português;
- 2) Avaliar o ajustamento do modelo, comparando os resultados previstos com os *ratings* efetivamente atribuídos pelas três agências;
- 3) Avaliar a robustez do modelo em dois períodos distintos: antes e após a crise financeira de 2008;
- 4) Avaliar o impacto do *rating* e outras variáveis na taxa de juro da dívida portuguesa e no prémio de mercado.

Assim, de forma a atingir os objetivos propostos, o capítulo 2 apresenta uma revisão da literatura sobre o risco soberano, a sua importância, o contexto histórico do risco de crédito dos estados soberanos, com enfoque no caso português. Seguidamente, descreve-se sucintamente o processo de atribuição de *rating*, bem como os determinantes frequentemente referidos pelos vários estudos empíricos realizados e sua relação com a probabilidade de incumprimento soberano. Por fim, apresentam-se, sumariamente, os principais efeitos da publicação dos *ratings* soberanos no mercado de capitais.

No capítulo 3, elabora-se a explicação matemática que suporta a metodologia utilizada para obter uma estimação do *rating* da entidade soberana portuguesa. A metodologia descrita neste capítulo contempla os processos de tratamento da informação, a metodologia de seleção de variáveis e construção (avaliação e validação) do modelo de regressão linear e ainda os pressupostos metodológicos adotados na análise do impacto do *rating* na taxa de juro da dívida portuguesa e no prémio de mercado.

Por fim, nos capítulos 4 e 5, respetivamente, descrevem-se e discutem-se os resultados e as principais conclusões do presente trabalho, i.e., os fatores determinantes do *rating* português, as alterações observadas antes e após a crise financeira na significância destes determinantes, o *rating* previsto pelo modelo e a sua diferença com o observado e, por último, as consequências da publicação de uma notação na taxa de juro da dívida portuguesa e no prémio de mercado.

2. Revisão de Literatura

2.1 Risco soberano

O risco soberano pode ser entendido como a probabilidade de uma entidade soberana incumprir nas suas obrigações contratuais, i.e., no reembolso da dívida ou no pagamento de juros nos prazos e condições estabelecidas na data de celebração do contrato [1].

O incumprimento não apresenta uma definição geral, pelo que, por exemplo, as situações de *swaps* ou reestruturações “involuntárias” de dívida, quando estas reduzem substancialmente o retorno do título de dívida [2], podem também ser consideradas casos de incumprimento da dívida soberana. Alguns autores [3] consideram ainda como incumprimento, casos de ajuda externa à sua Balança de Pagamentos atribuídas pelo FMI ou por outras instituições multilaterais, dado que, em princípio, estes não teriam tido capacidade de cumprir as suas obrigações externas sem este suporte financeiro [4].

Contrariamente aos agentes privados (por exemplo, empresas e particulares), o enquadramento jurídico dos Estados soberanos não permite que exista um mecanismo que os obrigue a cumprir qualquer condição alheia à sua vontade, ou seja, tomar alguma ação que possa pôr em risco a sua soberania. Desta forma, a avaliação do risco soberano incorpora ainda essa “disposição” para cumprir as suas obrigações contratuais. Neste sentido, a avaliação das várias agências de *rating* privadas, efetuada periodicamente ao risco soberano, contempla não só a capacidade, mas também a disposição para cumprir com as obrigações contratuais. Para tal, estas agências classificam o risco de cada Estado com base não só em fatores económicos, mas também sociais e políticos [5].

Na perspetiva da entidade soberana, existem diversas teorias que defendem como positiva a acumulação de dívida pública e que, nesse sentido, podem levar estas entidades a não reduzirem a sua dívida [4]. Um Estado pode, por exemplo, optar por uma estratégia de *tax smoothing*, a qual consiste em manter uma taxa de impostos relativamente estável ao longo do tempo com o intuito de evitar o desincentivo económico, em detrimento de ajustar a taxa de imposto consoante o ciclo económico. No entanto, em períodos de recessão, onde se assiste à redução da atividade económica, o défice orçamental tende a ser mais pronunciado, o que leva ao aumento da dívida pública nesta fase do ciclo [6]. Outra fundamentação dada à opção de acumulação de dívida designa-se por “ilusão fiscal”, a qual defende a existência de uma tendência sistemática dos eleitores a realçar os benefícios dos gastos governamentais em detrimento dos custos da dívida.

2.2 Perspetiva Histórica do risco soberano

Em termos históricos, o primeiro evento de incumprimento deu-se no século IV A.C, na Grécia. Até à atualidade, contabilizaram-se 97 casos de *default* soberanos [3] (tendo em conta a metodologia mais moderna de definição e contabilização de eventos de incumprimento soberano), sendo que Portugal constitui um dos países desta lista [4]. O contexto económico mundial, segundo vários autores, assume um papel relevante nestes eventos de *default*. Estes autores sugerem que o incumprimento soberano ocorre em *clusters*, podendo ser resultante de uma fase de recessão do ciclo económico, de tendências nos mercados internacionais ou ainda de outro tipo de eventos globais [4].

Neste contexto, o século XX foi um período marcado por vários saltos tecnológicos, mudanças de padrões monetários, diferentes ciclos económicos (crescimento sustentado *versus* estagnação), alterações nos padrões monetários e, sobretudo, pelo aumento da globalização [7]. O final do século é marcado também pela alteração do paradigma de pensamento da sociedade, em que a informação e o conhecimento passam a ter um papel fundamental.

Na primeira década do século XXI, os Estados, em geral, beneficiam de relativa prosperidade económica, que termina com a crise financeira de 2008. Este período de crise, que medeia entre 2007 e 2012 [7] [8] [9] e que apresenta algumas semelhanças com a crise económica de 1929 (por exemplo, produção de crédito sem colaterais e consequente *crash* na bolsa de Nova Iorque¹), pode ser descrito pelos seguintes principais eventos:

1. Agosto de 2007, o BNP Paribas anuncia o fim de atividade em três *hedge funds*² especializados em dívida hipotecária (rondando dezenas de mil biliões de euros), nomeadamente hipotecas de alto risco (também designadas por *subprime*). Esta ação contribuiu para reduzir a confiança no sector bancário e nos negócios entre diferentes instituições bancárias;
2. Setembro de 2008, o banco de investimento Lehman Brothers entra em falência, após a nacionalização do banco inglês Northern Rock e da venda do banco norte americano Bear Stearns. Este acontecimento colocou em causa a premissa de que estes bancos eram demasiado grandes para falirem. Já na Europa, observou-se o colapso da economia islandesa, após três dos principais bancos privados entrarem em *default*;
3. Novembro de 2008, ocorre uma reunião das nações pertencentes ao G20³, tendo em vista a capitalização do FMI e outras instituições globais para aumento da criação de postos de trabalho, crescimento económico e reforma bancária. Adicionalmente, as taxas de juro foram

¹ A crise levou a uma recessão de 15% do PIB mundial entre 1929 e 1932, aumentando o risco soberano para as principais entidades afetadas [34].

² Designam-se por *hedge funds* (Fundos de cobertura) os investimentos especulativos sobre o mercado, no sentido de obterem um elevado retorno para os seus investidores. São considerados fundos de investimento com elevado grau de risco (e alvo de reduzida regulamentação no período anterior à crise económica de 2008).

³ Grupo formado pelos ministros de finanças e chefes de bancos centrais das 19 maiores economias do mundo e a UE. Foi criado para dar respostas às sucessivas crises financeiras e favorecer a negociação internacional.

reduzidas, adotando-se uma política monetária expansionista, com a introdução de medidas de *quantitative easing*.

4. Maio de 2010, ocorre o resgate económico à Grécia por parte do FMI e a EU. Com este episódio, soma-se ao risco de insolvência das instituições bancárias, o risco de insolvência de alguns estados soberanos. Em resultado, foram implementadas medidas de austeridade aos países pertencentes à Zona Euro, que se consubstanciavam numa política fiscal restritiva;
5. Agosto de 2011, o *rating* soberano dos EUA foi revisto em baixa pela S&P's, de AAA para AA+, ou seja, um *downgrade* de um *notch*. De notar que esta notação corresponde, ainda assim, à melhor notação da categoria de "High grade".

Durante a crise financeira, os países europeus designados PIIGS – Portugal, Irlanda, Itália, Grécia e Espanha - necessitaram de ajuda externa para cumprirem as suas obrigações contratuais, uma vez que, por diferentes causas económicas e políticas, foi-lhes retirado o acesso aos mercados. O primeiro país a receber ajuda externa foi a Grécia em maio de 2010, seguido pela Irlanda em novembro de 2010 e Portugal em abril de 2011. Posteriormente, Espanha recebeu ajuda oficial em junho de 2012 [10].

O *downgrade* do *rating* destes países foi uma das consequências desta intervenção. Ao mesmo tempo, verificou-se a subida dos *spreads* no mercado. Numa breve análise comparativa dos *spreads* e *ratings* antes e durante a crise financeira (2008-2011), constata-se, na Tabela 1, que tanto os *spreads* como os CDS⁴ (*Credit Default Swap*) sofreram um aumento significativo durante este período de crise em todos os países "PIIGS". Ainda assim, este aumento não foi linear entre estes, revelando-se mais acentuado nos países com maior *downgrade* no *rating*, nomeadamente a Grécia com uma variação de 13 *notches*, seguido de Portugal com uma variação de 7 *notches* e ainda a Irlanda, com uma variação de 6 *notches*. Assim, Portugal e Grécia passam a ser incluídos na categoria de "Non investment grade".

PIIGS	Spread (títulos) ⁵			Credit Default Swap (títulos)			Rating		
	09/08	09/11	Δ pb	09/08	09/11	Δ pb	09/08	09/11	Δ notches
Portugal	-70,3	1192,5	1262,8	37,7	994,4	956,7	5	12	7
Irlanda	-59,3	760,0	819,3	29,8	822,5	792,7	2	8	6
Itália	-26,1	551,8	577,9	40,1	378,7	338,6	5	6	1
Grécia	11,4	3186,6	3175,2	50,7	2989,9	2939,2	7	20	13
Espanha	-83,1	475,1	558,2	37,5	372,1	334,6	3	4	1

Tabela 1 - Impacto da crise económica nos países "PIIGS" (setembro 2008 - setembro 2011) [11]

Em suma, tal como referido por alguns autores, a crise financeira de 2008 parece ter contribuído substancialmente para o risco de incumprimento soberano em vários países europeus. Por sua vez, as dificuldades de financiamento e o aumento dos juros da dívida soberana de cada país poderão ter contribuído para exacerbar o efeito negativo da crise financeira (por exemplo, em Portugal o aumento no

⁴ O *Credit Default Swap* é um derivado que resulta na transferência do risco de crédito de títulos no caso de incumprimento (*default*). Neste caso específico, em caso de *default*, o comprador do CDS recebe o valor acordado e o valor de juro respetivo, sendo que esta última parcela é proporcional ao risco de falência do vendedor do CDS.

⁵ Corresponde à diferença em pb (pontos base) face a uma taxa de juro de mercado a três meses.

spread em 2010 e 2011 traduziu-se num acréscimo de custos anuais de 3,1 mil milhões de euros, que equivalem a 1,8% do PIB).

2.2.1 O caso português

Como referido anteriormente, Portugal é um dos países alvo de maior vigilância no que respeita ao risco de incumprimento, fazendo parte dos 5 países que constituem o “PIIGS”, caracterizados por apresentarem histórico de incumprimento nos últimos 200 anos. Neste intervalo temporal, Portugal entrou em *default* quatro vezes, com a última ocorrência no fim do século XIX [3].

No século XX, Portugal foi caracterizado por maior instabilidade política, especialmente após a queda do regime monárquico em 1910 e o regime ditatorial que terminou em 1974. No entanto, após este período, e especialmente após a entrada de Portugal na UE, em 1986, o PIB cresceu a uma média de 6,2% ao ano até ao final da década de 80. A partir do início da década de 90 e até 2001 o crescimento do PIB desacelera para metade do ritmo de crescimento europeu, sendo sustentado, sobretudo, pela procura interna, alavancada pelo recurso ao crédito. Posteriormente, após a adesão de Portugal à moeda única da Zona Euro, em 2001, o crescimento estagna e o nível de desemprego começa a subir, atingindo, pela primeira vez, níveis superiores à média europeia. Com a desaceleração da procura interna e a redução dos níveis de produtividade decorrentes da década de 90, o país acumula desequilíbrios externos, provocados por consecutivos défices comerciais, aumentando significativamente a sua dívida pública e privada. Neste contexto, poder-se-á concluir que, embora a entrada de Portugal na UE tenha favorecido a economia portuguesa, que prosperou até ao final do século XX e meados da primeira década do século XXI, (o PIB per capita a preços constantes, base 2011, em 1986 era de 9490 € e em 2007 de 17191 € [12]), o elevado nível de desemprego e os encargos com a dívida externa acumulada podem colocar desafios significativos ao crescimento do PIB.

Por outro lado, a crise financeira de 2008 (também designada crise *subprime*) originada no mercado financeiro agravou a situação económica do país, afetando significativamente a estabilidade das contas públicas, pela contração significativa do PIB e pelo já elevado nível de endividamento público, resultante da acumulação de défices públicos excessivos. A contração do PIB foi devida, em grande parte, à redução significativa do consumo privado, tendo em conta o seu elevado nível de endividamento, o qual é consequência da facilidade de acesso ao crédito nas décadas anteriores, sobretudo durante o *boom* do setor imobiliário.

No período subsequente a 2008, no contexto do resgate à Grécia e da crise na Irlanda em 2010, Portugal procurava soluções para evitar um *bailout* (em novembro de 2010, os juros a 10 anos ultrapassavam os 7% e a dívida pública ultrapassava os 90%). O Estado procurava suporte nas instituições financeiras nacionais, que pouco responderam face aos efeitos de redução de liquidez provocados pela crise financeira. O Governo é pressionado pela UE a adotar medidas severas de redução da despesa pública, apresentadas no PEC IV e chumbadas pela maioria parlamentar em março de 2011. No mês

seguinte e após a reação negativa dos mercados a este evento, Portugal é obrigado a pedir ajuda externa, com a assinatura de um acordo com a Troika (BCE, CE, FMI). Em resultado, o *rating* soberano português sofre um *downgrade* acentuado, que se deveu não somente aos fatores económicos de elevado endividamento público e reduzido desempenho macroeconómico, mas também à instabilidade política, gerada, sobretudo, após a demissão do primeiro-ministro em abril de 2011 [13].

Com efeito, analisando duas das variáveis macroeconómicas mais importantes nos últimos 21 anos (Figura 1), verifica-se uma tendência evolutiva oposta entre a variação do PIB (a preços constantes) e a taxa de desemprego.

Da análise evolutiva do PIB, verifica-se que o crescimento positivo do PIB começou a reduzir-se no início do século XXI, tendo este estagnado (registando-se variações próximas de 0%) até 2008. Durante o mesmo período, a taxa de desemprego aumenta paulatinamente. Desde o início da crise *subprime* até ao pico da crise em Portugal, a taxa de desemprego duplica (dezembro de 2007: 8% e dezembro de 2013: 16,2%). Contrariamente, o PIB varia negativamente aproximadamente de -4% em 2012 (período de recessão económica). Ainda assim, no período subsequente à introdução das medidas de austeridade económica acordadas com a Troika, verifica-se uma ligeira melhoria destes indicadores, que se aproximam dos valores registados antes da crise financeira (taxa de desemprego em 2016 de 10,3%). Durante a crise financeira, S&P's, Moody's e a Fitch reduzem o *rating* de Portugal de “*Investment grade*” para “*non investment grade*”. Apenas a agência a DBRS atribui um “*Investment grade*” de BBB (baixo), que possibilita ao Estado português aceder ao programa de compra de títulos de dívida lançado pelo BCE em 2015 e à utilização destes títulos como colateral nos financiamentos europeus aos bancos portugueses [14]. O acesso a este programa está, no entanto, condicionado a que o Estado soberano apresente um “*investment grade*” de pelo menos uma agência de *rating*, ou seja, pelo menos uma notação superior a “lixo”.

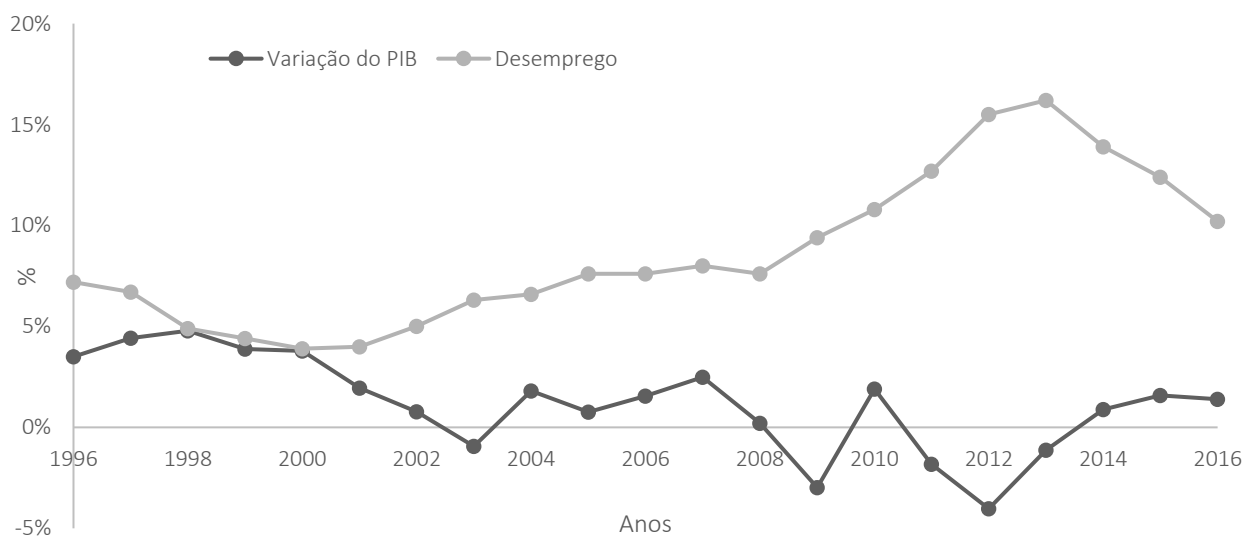


Figura 1 - Variação do PIB e desemprego anual (%) entre 1996 e 2016 em Portugal [15] [16]

2.3 Sistemas de *rating* das principais agências internacionais

As agências de *rating* que atribuem uma classificação ao risco soberano português são a Moody's, a S&P's, a Fitch e a DBRS. A terceira começou a atribuir *rating* em 1994 e as duas primeiras avaliam o risco soberano português desde a sua entrada na UE. A agência canadiana DBRS começou a atribuir *rating* posteriormente às *Big Three*. As agências avaliam a capacidade de pagamento e vontade de uma entidade soberana pagar a sua dívida numa determinada data de vencimento (ou maturidade) e de acordo com as condições contratuais estabelecidas. Esta avaliação é resumida num *rating*, que serve como estimador para a probabilidade de incumprimento (PD) de uma entidade soberana. Esta avaliação não contempla créditos bilaterais ou dívidas contratadas com entidades tais como o Banco Mundial, o BCE ou mesmo o FMI. O *rating*, expresso numa escala alfanumérica e simbólica, contém um significado para cada *range* de *ratings* e ainda a divisão em *investment grade* (notação de investimento) ou *speculative grade* (grau de especulação). Cada notação é atribuída consoante os resultados de cada país nos vários fatores do risco soberano, sendo esta informação de *rating* tipicamente utilizada nas tomadas de decisão dos investidores de dívida pública.

As principais características de cada agência (das “Big Three”) são fundamentalmente as seguintes [2]:

- **Moody's:** os *ratings* são indicadores de EL (*Expected Loss*), *outcome* da PD e da perda esperada pela entidade em *default*. Este *rating* é dado por uma escala alfanumérica, onde se diferenciam as entidades devedoras através de letras correspondentes a diferentes categorias de risco e de uma componente numérica, no caso de pertencerem à mesma categoria.
A escala vai de “AAA” (melhor) até “C” (pior). Para distinguir duas entidades consideradas como *upper medium grade*, é possível uma entidade ter uma notação de A1, caso se encontre perto da categoria acima (melhor) ou de A3 caso se considere no fim desta categoria (e, portanto, perto da categoria abaixo).
- **Fitch's:** apenas avalia a PD antes que o *default* ocorra e avalia a entidade com base na recuperação do principal (conseguindo, deste modo, estimar o LGD, *Loss Given Default*). A escala começa em AAA e termina em D, diferenciando as entidades de modo semelhante à Moody's, com a diferença da utilização de símbolos (+ ou -) em vez de números.
- **S&P's:** apenas reflete a PD e não a dimensão do *default*, o período que a entidade se encontrará em *default* nem os termos de possível renegociação. Partilha a mesma escala da Fitch's.

A atribuição do *rating* passa pelos seguintes três passos, tal como descrito na Figura 2.

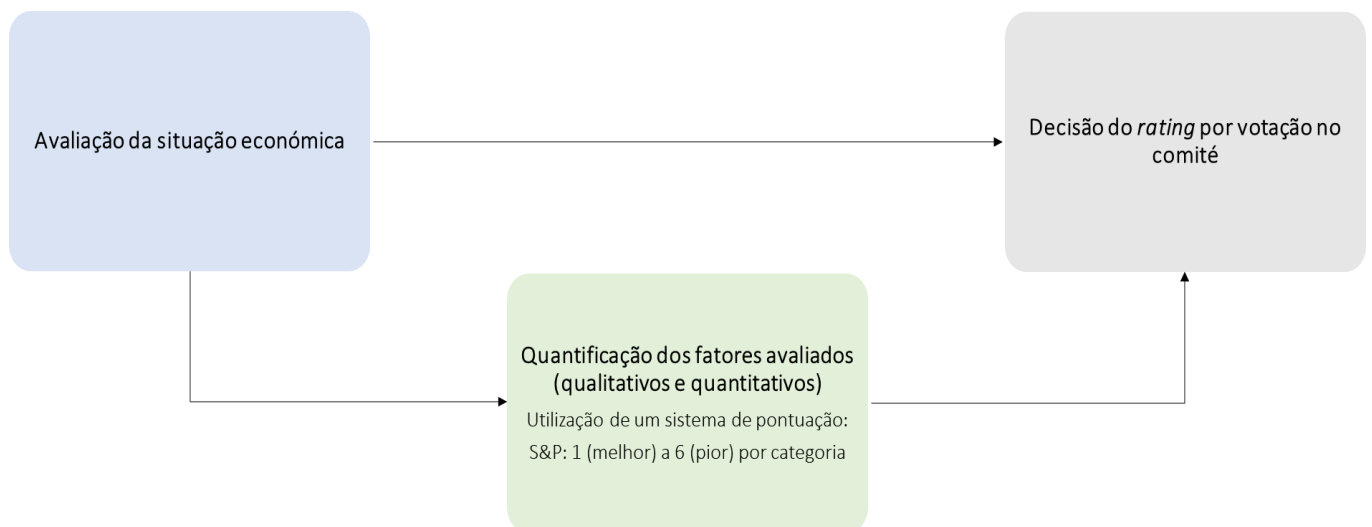


Figura 2 - Passos na atribuição do *rating* por parte de uma agência das “Big Three” [2] [5].

Associado ao *rating*, cada agência também fornece a indicação do *outlook* da direção do *rating*: positivo, negativo, estável ou em desenvolvimento, sendo o último referente ao caso que está dependente da ocorrência (ou não) de um acontecimento específico.

O *rating* é o resultado de um trabalho interdisciplinar que combina parâmetros qualitativos e quantitativos e não o resultado de um modelo específico estatístico que determine quantitativamente a PD. As categorias avaliadas são 10 no caso da S&P e 14 no caso da Fitch. Contudo, é possível consolidar as categorias em 5 mais gerais: risco político, civil e institucional; setor real⁶; o setor monetário e financeiro; o setor externo e ainda o setor fiscal.

Os primeiros dois passos contemplam ainda a visita dos analistas ao país, a qual inclui reuniões com os principais elementos do governo, analistas do setor privado, jornalistas, investigadores, membros da oposição e ainda contacto com os outros setores. Estas visitas permitem fidelizar a informação obtida por indicadores oficiais, com a compilação de informação mais detalhada sobre políticas fiscais e monetárias e a avaliação da sua adequabilidade face ao reportado pelas diferentes entidades.

As entidades soberanas que recebem *rating* por estas agências têm preferência por parte dos investidores às que não são alvo de avaliação de risco, dado que a mesma indica maior grau de transparência, permitindo reduzir os custos de agência. Os *ratings* são utilizados pelos investidores para determinar preços (taxa de juro da dívida) no que toca à compra e venda de títulos de dívida soberana. Representam também um teto superior para as entidades que operam num país (por exemplo, grandes empresas), embora possam existir exceções sob o cumprimento de requisitos específicos.

⁶É neste setor onde ocorrem operações de geração de bens (produtos tangíveis) e serviços não financeiros (produtos intangíveis tais como transporte, comércio, educação, comunicações, etc.)

Embora o *rating* possa contribuir para a taxa de juro da dívida soberana, outras forças de mercado influenciam o seu preço. Por exemplo, a conjuntura económica mundial influencia a confiança dos investidores, i.e., em momentos de recessão económica a avaliação relativa dos fatores determinantes do risco soberano tende a ser mais conservadora. Assim, embora uma entidade possa apresentar a mesma notação em momentos diferentes ao longo de um determinado período temporal, a taxa de juro da dívida pública poderá divergir (vide Figura 3). A título de exemplo, relativamente ao caso português, fazendo uma avaliação *ceteris paribus*, na Tabela 2, observa-se um período de doze meses onde, embora o *rating* permaneça constante, regista-se uma variação de 641 pb. Por outro lado, se se retirar o efeito da oferta e da procura, é possível identificar um impacto significativo do *downgrade* do *rating* entre 2010 e 2011 no *spread* da dívida portuguesa que aumentou 440pb relativamente ao *spread* da dívida alemã.

Vários estudos sugerem que as entidades soberanas que apresentam uma notação “lixo” têm custos de financiamento mais elevados quando acedem aos mercados internacionais comparativamente às entidades com *ratings* acima deste nível [17]. Por outro lado, tal como descrevem a Tabela 2 e a Figura 3, podem existir discrepâncias entre a avaliação de risco do mercado e a avaliação de risco contida no *rating* publicado pelas agências, uma vez que ao *spread* é subtraído o valor dos ativos, sujeitos à lei da oferta e procura. Adicionalmente, estes preços de mercado (*spread*) também são mais sensíveis ao ciclo económico, isto é, respondem rapidamente a variações de curto prazo no ambiente económico, contrariamente às agências de *rating* que avaliam o risco numa perspetiva de longo prazo (em geral, utilizando previsões a três anos dos principais indicadores económicos do país).

Data (Ano/mês)	Juro da dívida a 10 anos (%)	Rating “numérico” ⁷
2012/02	12,81	13
2012/03	13,01	13
2012/04	12,01	13
2012/05	11,59	13
2012/06	10,56	13
2012/07	10,49	13
2012/08	9,89	13
2012/09	8,62	13
2012/10	8,17	13
2012/11	8,32	13
2012/12	7,25	13
2013/01	6,24	13
2013/02	6,40	13

Tabela 2 - Comparação do juro da dívida portuguesa a 10 anos e a notação da dívida soberana portuguesa num período 12 meses (entre fevereiro de 2012 a fevereiro de 2013)

Um dos fatores que influenciam também os *spreads* são as políticas monetárias adotadas pelo sistema financeiro central. Um dos exemplos, é o efeito da medida de *quantitative easing*⁸ adotada pelo

⁷ A transformação em *rating* numérico, proveniente da metodologia usada neste projeto, é explicada na subseção 3.2.1. Seria o equivalente a um BB- ou um Ba3, dependente da agência em causa.

⁸ Programa de compra de grandes volumes de ativos públicos e privados por parte de Bancos Centrais. O programa de Compra de Ativos do BCE com início em março de 2015 e setembro de 2016 visava a aquisição de 1,14 trilhões de títulos públicos e privados, os quais representavam aproximadamente 11% do PIB dos países da zona euro.

BCE a junho de 2015. Entre setembro de 2014 e outubro de 2015, a média da taxa de juro a 10 anos dos títulos de dívida soberana dos principais países da zona euro diminuiu em aproximadamente 69 pontos base, tendo os países mais vulneráveis sido os mais beneficiados desta medida (por exemplo, Portugal observou uma diminuição de 106 pontos base). [18]

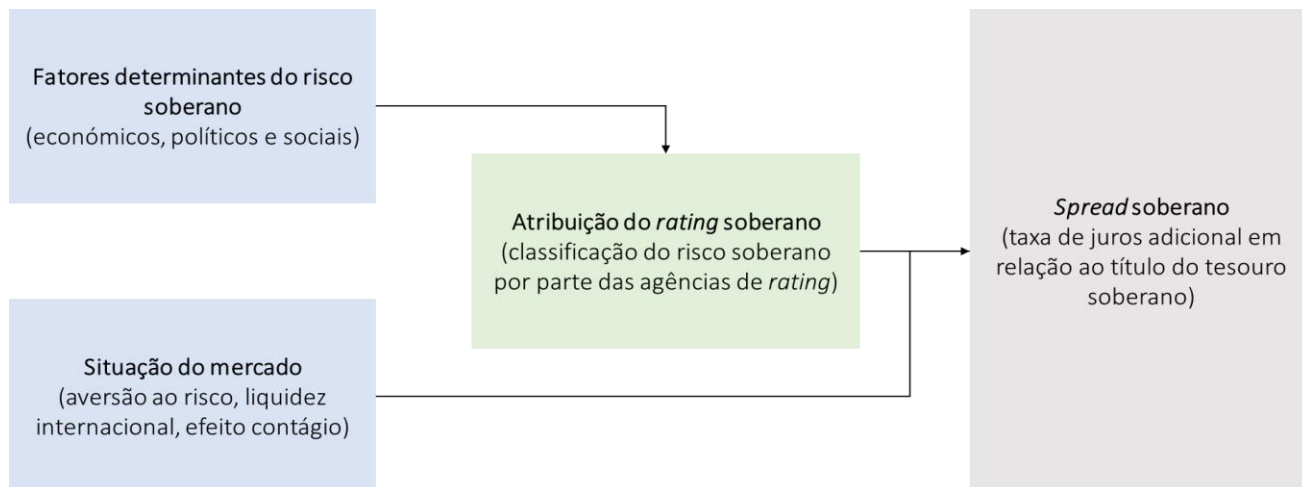


Figura 3 - A influência do *rating* e do mercado nos juros da dívida soberana [5].

2.4 Determinantes do risco soberano

O risco soberano é avaliado por várias agências que têm em conta a evolução de um conjunto de variáveis relevantes para a atribuição de uma notação ou um *outlook*. De um modo geral, as variáveis mais utilizadas englobam aspetos macroeconómicos, políticos, sociais, podendo ser do tipo quantitativas ou qualitativas [2] [5] [19].

As agências não divulgam os pesos a atribuir às diferentes variáveis ou qual a influência dos fatores qualitativos na atribuição de *rating*, no entanto divulgam os setores mais determinantes. Tipicamente, o PIB é considerado um bom indicador na avaliação de risco, observando-se uma relação inversa entre o crescimento do PIB e o risco soberano. Também a inflação é considerada um fator importante, uma vez que contém aspetos como a estabilidade política, social, institucional e a política monetária seguida pela entidade soberana a ser avaliada [2]. É sugerido que 80% da variação dos *ratings* se deve aos fatores PIB *per capita*, dívida externa, comportamento da taxa de inflação, histórico de *default* e Índice de Desenvolvimento Humano. [20]

Analogamente ao que foi efetuado pelos diferentes estudos, recolheram-se as variáveis consideradas relevantes para determinar o risco soberano. Após uma análise ao contexto português, determinou-se uma

lista⁹ de variáveis que se descrevem seguidamente e cujo detalhe pode ser consultado na tabela em Anexo (vide Anexo A):

- **PIB per capita:** reflete o nível de rendimento médio da população, sendo também um bom indicador do nível de estabilidade política e de outros fatores económicos. Uma variação positiva sugere maior capacidade de pagar a dívida, uma vez que a base tributável é superior [20];
- **Variação do PIB:** uma variação positiva e elevada sugere uma maior capacidade para pagar o serviço da dívida [20];
- **Inflação:** reflete a evolução do preço médio dos bens e serviços; uma taxa de inflação elevada poderá indiciar problemas estruturais nas finanças públicas;
- **Desemprego:** uma variação negativa permite não só reduzir os custos do Estado com benefícios sociais, mas também aumentar a base de tributação, contribuindo para a estabilidade orçamental e, conseqüente, diminuição no risco de incumprimento;
- **Despesas estado social/PIB:** reflete uma das conseqüências do desemprego – o aumento em encargos do Estado com benefícios sociais. À semelhança da variável anterior, uma variação negativa desta variável diminuirá o risco de incumprimento. A incorporação desta variável deve-se à análise efetuada aos custos do desemprego após a crise financeira, tendo-se concluído que este aumento resultou num acréscimo extraordinário nas despesas orçamentais;
- **Reservas totais:** maiores reservas permitem salvaguardar o Estado do risco de incumprir nas suas obrigações em moeda estrangeira. Quando existe controlo da moeda por parte do Banco Central a constituição de um valor suficiente de reservas permite transmitir confiança às agências de *rating* e ao mercado de capitais (por exemplo, até à entrada na moeda única em janeiro de 2001, o Banco de Portugal detinha uma maior percentagem de reservas em proporção do PIB comparativamente aos níveis posteriores à entrada no euro);
- **Balança Corrente/PIB:** um aumento no défice da balança corrente poderá refletir um excesso de consumo, pondo em causa a sustentabilidade a longo prazo. Em contrapartida, o mesmo défice pode refletir a acumulação de investimento fixo, que deverá contribuir para o crescimento económico e maior sustentabilidade no médio prazo. Deste modo, esta variável pode ter, portanto, uma variação positiva ou negativa no risco de incumprimento, consoante o racional para o défice observado;
- **Balança Comercial:** uma variação positiva desta variável está associada a uma maior capacidade produtiva, i.e., a aumento dos produtos exportados comparativamente aos importados do exterior. Esta variável tem um comportamento semelhante à balança corrente;
- **Euribor:** considerada como a taxa de juro dos empréstimos entre bancos europeus. Esta taxa é utilizada por exemplo como taxa de referência para crédito à habitação e contas poupança;

⁹ Esta lista também se encontra em Anexo (Anexo A), onde se poderá obter mais informação relativa à descrição das mesmas, e a fonte de onde foram retirados os dados.

- **Leverage banking:** traduz-se num rácio entre o capital próprio e os ativos de uma instituição bancária, i.e., o grau de alavancagem do setor financeiro. Uma variação positiva nesta variável significa que o setor, em princípio, terá maior capacidade e solidez financeira para dar suporte aos agentes económicos de um país, o que por sua vez tenderá a diminuir o risco de incumprimento da dívida soberana. O impacto indireto desta variável no risco soberano explica-se, por um lado, na redução do risco de ajuda estatal a estas entidades, e por outro, na capacidade destas entidades garantirem maior liquidez à Administração Pública. Note-se que a decisão de incorporar este rácio prende-se também com os constrangimentos que a mesma provoca no acesso aos estímulos da política monetária expansionista adotada pelo BCE (por exemplo, através de *spreads* elevados e critérios rígidos na contratação de empréstimos);
- **Bank nonperforming loans to total gross loans:** reflete a proporção de ativos (créditos) em incumprimento no total de ativos da instituição financeira: uma variação positiva nesta variável significa um setor financeiro mais frágil e com maiores riscos de perdas, que resultam na menor disponibilidade para financiar a economia;
- **PIB/Hora trabalhada:** medida de produtividade; mede a eficiência da utilização do tempo laboral combinado com outros fatores de produção; por outras palavras, mede o nível de investimento tecnológico, a existência de recursos especializados como a utilização de processos/métodos mais atuais que permitam a agilização e eficiência nos processos de produção. Geralmente, uma variação positiva indica uma melhoria na dinâmica da economia, o que se poderá traduzir na redução do risco de incumprimento;
- **Custos laborais unitários:** mede a competitividade dos preços dos produtos; define-se como a média dos custos de trabalho (rendimentos do trabalhador) por unidade de bem produzido. Uma variação positiva poderá estar relacionada com o aumento dos rendimentos do trabalhador, o que, embora leve ao aumento dos custos dos produtos, traduzir-se-á, em princípio, numa melhoria da capacidade económica (rendimento disponível) dos mesmos para adquirirem bens e serviços, contribuindo para o crescimento económico;
- **Endividamento de particulares:** representa o nível de dívida dos particulares, isto é, o peso dos créditos obtidos no total de rendimento das famílias. Em princípio, uma variação positiva nesta variável representa um aumento no consumo privado, e deste modo, um aumento do PIB, o que contribui para uma apreciação positiva do risco soberano; no entanto, se o nível de endividamento observado for elevado, uma variação positiva pode representar maiores riscos na sustentabilidade da procura interna, e conseqüentemente, no crescimento do PIB a longo prazo. Nesta situação, uma variação positiva poderá aumentar o risco de incumprimento soberano;
- **Poupanças das famílias:** representa a parte do rendimento das famílias que não é gasto em bens ou serviços; no sentido oposto à variável anterior, uma variação positiva poderá contribuir para a redução do risco de incumprimento soberano, se se verificar que o nível de endividamento se encontra num nível não sustentado;

- **Formação bruta de capital fixo:** representa o nível de investimento público e privado, isto é, a tendência evolutiva da capacidade produtiva do país, e, num sentido lato, o nível de confiança dos empresários no crescimento económico futuro. Uma variação positiva está tipicamente associada a uma diminuição do risco de soberano;
- **Défice/excedente público:** reflete a sustentabilidade orçamental da entidade soberana; na presença de défices elevados, as poupanças domésticas são absorvidas pelo Estado, sugerindo desequilíbrios macroeconómicos; a persistência de défices elevados pode também sinalizar problemas no funcionamento das instituições públicas; em períodos de recessão económica, um défice acentuado e persistente poderá conduzir a um aumento do risco de incumprimento [20];
- **Dívida externa/PIB:** reflete a sustentabilidade da dívida total do país (i.e., do Estado e de entidades privadas) comparativamente à dimensão da economia; uma dívida elevada aumenta o risco de necessidades de financiamento quer de entidades privadas ou do setor público e, deste modo, o risco de incumprimento soberano;
- **Dívida externa/Exportações:** reflete o peso da dívida total sobre o valor dos bens ou serviços vendidos ao exterior; uma variação positiva neste rácio indica que a geração de divisas através de exportações é insuficiente para liquidar as necessidades de financiamento dos agentes públicos e privados, criando problemas de sustentabilidade macroeconómica a médio e longo prazo; neste sentido, esta variação tende a aumentar o risco de incumprimento;
- **Dívida pública bruta/PIB:** avalia a dívida das Administrações Públicas relativamente à dimensão da economia (PIB); uma variação positiva resulta num aumento da PD;
- **Impostos sobre o rendimento e património:** reflete a evolução da tributação sobre o rendimento do trabalho de empresas e particulares e sobre o património destas entidades. Um aumento deste indicador é normalmente resultante de uma situação de ausência de liquidez por parte da entidade soberana, o que poderá indiciar maiores riscos de incumprimento;
- **(Importações + Exportações) / PIB:** esta variável é tipicamente utilizada para avaliar o nível de abertura da economia ao exterior ao refletir o peso das transações comerciais sobre o total do PIB do país. Numa economia desenvolvida, uma variação positiva poderá indiciar contração económica, pois indica o aumento do peso das transações comerciais com o exterior no PIB e a diminuição das restantes componentes do mesmo, tais como Investimento, Gastos Governamentais e o Consumo Privado.

3. Metodologia de Investigação e Conceitos Teóricos

3.1 Modelos Matemáticos

Neste subcapítulo, serão abordados os seguintes tópicos:

- 1) Regressão linear múltipla:
 - i. Tópicos sobre regressão linear múltipla;
 - ii. Método dos Mínimos Quadrados;
 - iii. Intervalo de confiança e predição;
 - iv. Testes nos coeficientes de regressão;
 - v. Medidas de ajustamento: R^2 e S^2 ;
 - vi. Tabela ANOVA;
 - vii. Análise de resíduos.
- 2) Estudo da multicolinearidade entre variáveis:
 - i. Definição;
 - ii. Método VIF;
 - iii. Números condição e componentes da variância.
- 3) Seleção de variáveis e escolha do modelo:
 - i. Método *backwards*;
 - ii. Método *stepwise*;
 - iii. Método *forward* linear.

Uma vez que o âmbito deste trabalho não engloba modelos de *scoring* (e, conseqüentemente, de atribuição de *rating*), não são abordados temas como a transformação de variáveis, a atribuição de diferentes pesos a cada variável ou a classificação de uma determinada variável por *buckets* de *score*.

3.1.1 Regressão linear múltipla

A regressão linear simples é um modelo estatístico que procura escrever uma variável resposta Y , a variável dependente, como função linear de uma outra variável X , a variável independente. Assim, para um conjunto de pares de observações destas variáveis, (x_i, y_i) , estas dispõem-se em torno de uma reta,

$$y_i = b_0 + b_1x_i + \varepsilon_i,$$

em que os ε_i são termos de erro, de média nula. No entanto, em muitos casos, a variável resposta não depende só de uma variável independente, mas sim de várias. A regressão linear múltipla procura ajustar uma relação linear entre a variável resposta e um conjunto de p variáveis independentes. Assim, para um conjunto de observações y_i , no nosso caso os *ratings*, e as correspondentes observações das variáveis independentes, x_{ij} , o modelo de regressão linear múltipla pressupõe válida a relação:

$$y_i = \sum_{j=1}^p x_{ij}b_j + \varepsilon_i,$$

onde $i \in \{1, \dots, n\}$, $j \in \{1, \dots, p\}$, em que os b_j , os chamados coeficientes de regressão, são parâmetros a estimar e ainda os ε_i 's são variáveis aleatórias que verificam as condições de Gauss-Markov, i.e., de média nula, variância constante e não correlacionadas:

- 1) $E(\varepsilon_i) = 0$;
- 2) $\text{Var}(\varepsilon_i) = \sigma^2$, para $i \in \{1, \dots, n\}$;
- 3) $E(\varepsilon_i\varepsilon_j) = 0$ se $i \neq j$.

Para além destas condições, é usual exigir que os termos de erro, as variáveis ε_i , tenham distribuição normal, $N(0, \sigma^2)$.

À semelhança do que sucede na regressão linear simples, é usual o modelo de regressão linear múltipla conter um termo constante. Este termo pode ser visto como o coeficiente de uma variável independente cujos valores são identicamente iguais a 1, isto é, $x_{i1}=1$, para $i=1, \dots, n$.

Neste estudo, é pressuposto que o *rating* do risco soberano português segue um modelo de regressão múltipla, isto é, admite a representação:

$$rating_i = b_1x_{i1} + b_2x_{i2} + \dots + b_jx_{ij} + \varepsilon_i,$$

em que o primeiro termo corresponde ao termo constante, e em que as restantes variáveis descrevem aspetos relacionados com a situação económica, social e financeira do país. Este modelo pode ainda ser escrito na forma matricial:

$$rating = Xb + \varepsilon,$$

em que *rating* representa um vetor coluna $n \times 1$, que agrupa todas as observações da variável resposta e cujo elemento i é *rating_i*, isto é, o *rating* no momento i ,

$$rating = \begin{bmatrix} rating_1 \\ \dots \\ rating_n \end{bmatrix}.$$

Já a matriz X , de dimensão $n \times (p+1)$, também chamada matriz de planeamento, é definida por:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{1j} \\ 1 & X_{21} & X_{22} & X_{2j} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & X_{n2} & X_{nj} \end{bmatrix}.$$

O vetor dos declives, b , é dado pela agregação do termo constante com os coeficientes de todas as variáveis explicativas, num vetor coluna de dimensão $(p+1) \times 1$:

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_p \end{bmatrix}.$$

Por fim, agrupando todos os termos de erro num vetor coluna de dimensão $n \times 1$, tem-se o vetor dos erros:

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}.$$

3.1.2 Método dos Mínimos Quadrados

A estimativa dos valores constituintes do vetor declive é usualmente obtida através do Método dos Mínimos Quadrados. Facilmente se observa que a função que define o valor de *rating* obedece às seguintes condições [21]:

- 1) Cada variável explicativa incluída na função é multiplicada por um parâmetro desconhecido;
- 2) Há no máximo 1 parâmetro desconhecido sem variável explicativa correspondente;
- 3) Todos os termos individuais são somados de forma a produzir o valor final da função.

Os Estimadores de Mínimos Quadrados (EMQ) - estimadores dos parâmetros do vetor b são aqueles que minimizam a soma dos quadrados das distâncias das observações ao espaço linear gerado pelas variáveis independentes.

Designando por B^T a transposta do vetor b e X_i o vetor das variáveis independentes relativas à observação i , o modelo de regressão múltipla permite escrever esta última como

$$rating_i = B^T X_i + \varepsilon_i.$$

Com esta notação, a estimativa dos coeficientes de regressão é então obtida pela minimização da seguinte expressão:

$$\sum_{i=1}^n (\varepsilon_i)^2 = \sum_{i=1}^n (rating_i - B^T X_i)^2$$

Para obter o mínimo desta expressão basta derivar em ordem a cada um dos elementos do vetor B^T e igualar a zero, obtendo-se o sistema constituído por $j+1$ equações a $j+1$ incógnitas:

$$\sum_{i=1}^n (rating_i - B^T X_i) X_{ij} = 0$$

$j=1, \dots, p$, ou, em notação matricial

$$X'Y - X'XB = 0.$$

Assim, o conjunto de EMQ para os coeficientes de regressão, \hat{B} , vem dado por:

$$\hat{B} = (X^T X)^{-1} X^T rating.$$

Em suma, o preditor (também denominado por valor ajustado) de $rating_i$, \widehat{rating}_i , toma a seguinte forma:

$$\widehat{rating}_i = \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2} + \dots + \hat{B}_j X_{ij}$$

Denominam-se por resíduos, e_i , os valores dados pela diferença entre a observação do $rating_i$ e o respetivo valor ajustado, ou seja:

$$e_i = rating_i - \widehat{rating}_i.$$

Tal como já referido anteriormente, a soma de todos os resíduos deverá ser igual a zero, uma vez que a equação que se obtém derivando a soma de quadrados em ordem ao termo constante e igualando a zero corresponde exatamente à restrição de que a soma dos resíduos deve ser nula.

3.1.3 Intervalo de confiança e intervalo de predição

O Método dos Mínimos Quadrados produz estimadores para os coeficientes de regressão com propriedades estatísticas que, em certo sentido, são ótimas. Sob a validade das condições de Gauss-Markov pode-se provar que estes estimadores são lineares centrados de variância mínima, BLUE (*Best Linear Unbiased Estimators*) com matriz de covariâncias dada por:

$$cov(\hat{B}) = \sigma^2 (X'X)^{-1}$$

Deste modo é possível encontrar um estimador linear centrado de variância mínima para qualquer combinação linear dos coeficientes de regressão $\theta = a^T b$, em que $a^T = [a_1 \dots a_p]$ é qualquer vetor de constantes, que será dado por $\hat{\theta} = a^T \hat{b}$. Este estimador é também linear centrado, uma vez que:

$$E(\hat{\theta}) = a^T E(\hat{b}) = a^T b = \theta.$$

Além disso, de entre os lineares centrados tem variância mínima (BLUE) que é dada por:

$$\text{var}(\hat{\theta}) = a^T \text{cov}(\hat{b})a = \sigma^2 a^T (X^T X)^{-1} a.$$

Um caso particular importante ocorre quando o vetor de constantes, a , corresponde a valores não observados das variáveis independentes:

$$a^T = [1, X_1^*, X_2^*, \dots, X_p^*]^T.$$

Do que foi dito anteriormente, decorre que o valor ajustado para estes valores das variáveis independentes, $a^T \hat{b}$, é um estimador BLUE para o valor médio da variável dependente. Além disso, pode-se demonstrar que o vetor dos EMQ, quando os resíduos seguem uma distribuição normal, têm distribuição normal multivariada. Deste modo pode-se construir um intervalo de confiança para o valor médio da variável dependente a partir da variável fulcral:

$$\frac{\hat{\theta} - \theta}{S \sqrt{a^T (X^T X)^{-1} a}},$$

que tem distribuição *t-student* com $n-p$ graus de liberdade. Esta variável conduz, então, ao intervalo de $(1-\alpha)100\%$ de confiança para o parâmetro θ :

$$\hat{\theta} \pm t_{n-p}^{1-\frac{\alpha}{2}} S \sqrt{a^T (X^T X)^{-1} a}.$$

De maior interesse prático é a construção de um intervalo de confiança para a própria variável dependente, y^* , correspondente às variáveis dependentes não observadas, $X_1^*, X_2^*, \dots, X_p^*$ a partir do preditor $\hat{y}^* = \sum_{j=1}^p \hat{b}_j x_j^*$, designado usualmente por intervalo de predição. O erro quadrático médio (EQM) deste preditor é dado por:

$$E[(\hat{y}^* - y^*)^2] = E[(\hat{y}^* - y^*)(\hat{y}^* - y^*)^T] = \sigma^2 [X^{*T} (X^T X)^{-1} X^* + 1].$$

Atendendo a que o erro de predição é uma combinação linear do vetor dos EMQ com distribuição normal multivariada de média nula e variância igual ao EQM de \hat{y}^* , tem-se a seguinte variável fulcral:

$$\frac{\hat{y}^* - y^*}{S \sqrt{X^{*T} (X^T X)^{-1} X^* + 1}}$$

Assim, com distribuição *t-student* com $n-p$ graus de liberdade, o intervalo de predição de $(1-\alpha)100\%$ confiança é dado por:

$$\hat{y}^* \pm t_{n-p}^{1-\frac{\alpha}{2}} S \sqrt{X^{*T} (X^T X)^{-1} X^* + 1}.$$

No fundo, um bom modelo de regressão linear múltipla deve produzir um bom compromisso entre o valor de S e o valor de $\sqrt{X^{*T} (X^T X)^{-1} X^* + 1}$. O objetivo é minimizar este produto a fim de se obter o menor intervalo de predição, o que garante melhor qualidade na previsão.

3.1.4 Testes nos coeficientes de regressão

Após a estimação dos parâmetros do modelo de regressão linear, surge a necessidade de avaliar a significância de cada variável no modelo. Pretende-se saber se a contribuição de uma variável para a explicação da variável dependente é significativa, o que se pode fazer rigorosamente através de um teste às hipóteses:

$$H_0: b_j = 0 \text{ vs } H_1: b_j \neq 0$$

A rejeição da hipótese nula significa que esta variável deverá ser incluída no modelo, pois significa que o coeficiente de regressão associado não é igual a zero (logo é existente). A estatística de teste para estas hipóteses é dada por:

$$\hat{t} = \frac{\hat{b}_j}{\hat{\sigma}(\hat{b}_j)},$$

que, sob a validade da hipótese nula tem distribuição *t-student* com $n-p$ graus de liberdade. Deste modo, a região de rejeição é dada por:

$$R = \left\{ \frac{\hat{b}_j}{\hat{\sigma}(\hat{b}_j)} > t_{n-p}^{1-\frac{\alpha}{2}} \right\}.$$

A estimativa do desvio-padrão do coeficiente de regressão, $\hat{\sigma}(\hat{b}_j)$, é dada por Sz_{jj} , em que S é o estimador do desvio-padrão dos erros e z_{jj} é o j -ésimo elemento da diagonal da matriz $(X'X)^{-1}$.

3.1.5 Análise do R^2 e S^2

A avaliação da qualidade do ajustamento do modelo pode e deve partir da análise de duas estatísticas, o R^2 e o S^2 . Analise-se então a decomposição da soma dos quadrados dos resíduos:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{y}_i^2 = \left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 \right) - \left(\sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 \right).$$

Esta igualdade verifica-se devido à ortogonalidade entre os resíduos e os valores ajustados, i.e., os resíduos não contêm qualquer informação que possa contribuir para melhorar a qualidade do ajustamento do modelo. Somando e subtraindo os correspondentes valores ajustados a cada observação obtém-se:

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 + \sum_{i=1}^n \hat{y}_i^2$$

Sendo que, na presença de um termo constante no modelo, se observa que:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (\hat{y}_i - \bar{y})^2,$$

porque a soma dos resíduos é nula, ou seja,

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (y_i) - \sum_{i=1}^n (\hat{y}_i) = 0,$$

o que resulta em que a soma das observações é igual à soma dos valores ajustados. Por outras palavras, a média das observações assume o mesmo valor que a média dos valores ajustados. Portanto, a igualdade:

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2$$

permite concluir-se que a decomposição da soma de quadrados poderá ser escrita na seguinte forma:

$$SQ_{total} = SQ_e + SQ_{Reg},$$

substituindo o resultado anterior na decomposição da soma dos quadrados dos resíduos.

Com base nesta decomposição, é possível definir o coeficiente de determinação múltipla, R^2 . Este coeficiente mede a percentagem de variação da amostra explicada pelo modelo de regressão, ou seja, o quadrado do coeficiente de correlação entre as observações e os respetivos valores ajustados.

$$R^2 = \frac{SQ_{Reg}}{SQ_{Tot}} = 1 - \frac{SQ_e}{SQ_{Tot}} \leftrightarrow R^2 = \left[\frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}} \right]^2.$$

Este poderá assumir um valor compreendido entre [0;1]. Quanto mais próximo de 1, melhor o ajustamento do modelo. Por outras palavras, num caso extremo, quando o R^2 é igual a 1, todos os resíduos são nulos.

Por outro lado, partindo do termo da soma dos quadrados de resíduos, SQ_e , pode-se escrevê-lo da seguinte forma:

$$SQ_e = \sum_{i=1}^n e_i^2 = e^T e = \varepsilon^T M^T M \varepsilon = \sum_{i=1}^n m_{ii} \varepsilon_i^2 + \sum_{i \neq j} m_{ii} \varepsilon_i \varepsilon_j,$$

onde $M = (I_n - H)$, I_n é a matriz identidade $n \times n$ e H , *hat matrix*, é definida por $H = X(X^T X)^{-1} X^T$. Aplicando os valores médios e tendo em conta que os erros não são correlacionados e que apresentam variância comum σ^2 , tem-se que:

$$E \left(\sum_{i=1}^n e_i^2 \right) = \sigma^2 \sum_{i=1}^n m_{ii} = \sigma^2 \text{tr}(M) = \sigma^2 \text{tr}(I_n - H) = \sigma^2 (n - \text{tr}(H)) = \sigma^2 (n - p)$$

Deste modo, chega-se ao estimador centrado da variância, S^2 , com a seguinte expressão:

$$S^2 = \frac{1}{n - p} \sum_{i=1}^n e_i^2$$

Conclui-se a partir da expressão suprarreferida que o valor de S^2 é menor quanto menor for o valor de todos os resíduos, resultando num menor erro de previsão por parte do modelo obtido.

3.1.6 Tabela ANOVA

No que toca à avaliação da qualidade da regressão linear obtida, i.e., à aferição da necessidade de ajustamento de um modelo linear ao conjunto de observações, também se pode proceder à análise das variâncias (através de uma tabela ANOVA), o que equivale a testar a hipótese de que todos os p coeficientes, exceto o termo constante, são nulos:

$$H_0: b_2 = b_3 \dots = b_p = 0 \text{ contra } H_1: b_j \neq 0, \text{ para algum } j=1, \dots, p$$

A estatística para este teste baseia-se na decomposição da soma de quadrados apresentada anteriormente, isto é, identifica as diferentes fontes de variação das observações, nomeadamente, devido à regressão, aos resíduos e o total. Adicionalmente, a variável fulcral usada como estatística de teste apresenta distribuição F com $p-1$ graus de liberdade no numerador e $n-p$ graus de liberdade no denominador:

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / (p - 1)}{\sum_{i=1}^n e_i^2 / (n - p)} = \frac{\frac{SQ_{reg}}{p - 1}}{\frac{SQ_e}{n - p}} = \frac{MQ_{reg}}{MQ_e}$$

De notar que n representa o número de observações e p o número de variáveis do modelo. Por outro lado, aplica-se a seguinte região de rejeição, R , no caso desta ser utilizada como variável teste para a hipótese de igualdade de médias, onde $F_{p-1, n-p}^{1-\alpha}$ representa o quantil de ordem $1-\alpha$ da distribuição F com os graus de liberdade acima referidos:

$$R = \{F > F_{p-1, n-p}^{1-\alpha}\}.$$

É usual apresentar os resultados deste teste numa tabela ANOVA, que se constrói como se apresenta na Tabela 3.

Fonte	Soma de Quadrados	Graus de Liberdade	Média de Quadrados
Regressão	$SQ_{Reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$p-1$	$MQ_{Reg} = \frac{SQ_{ext}}{p - 1}$
Residual	$SQ_e = \sum_{i=1}^n e_i^2$	$n-p$	$MQ_e = \frac{SQ_e}{n - p}$
Total	$SQ_{Tot} = \sum_{i=1}^n (y_i - \bar{y})^2$	$n-1$	$F = MQ_{Reg}/MQ_e$

Tabela 3 - Tabela ANOVA para aferir a significância do modelo.

A tabela ANOVA, apresentando a construção da estatística de teste para verificar se pelo menos alguma das variáveis no modelo de regressão é significativa, toma o seu nome a partir do mesmo tipo de tabela que é utilizada com um objetivo semelhante, mas considerando um modelo de Análise da Variância

[24]. Nesta análise, comparam-se os valores médios de I populações independentes com igual variância e todas com distribuição normal. Para tal, observam-se I amostras, uma de cada população e com dimensões que podem ser diferentes, pressupondo que o conjunto das observações segue o modelo

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

em que $i=1, \dots, I$; $j=1, \dots, n_i$ e os termos de erro ε_{ij} são independentes, com distribuição normal de média nula e variância σ^2 . Assim, o teste de igualdade de médias das I populações pode ser formalizado como:

$$H_0: \mu_1 = \dots = \mu_I \text{ vs } H_1: \mu_i \neq \mu_j, \text{ pelo menos para um par } i, j, \text{ onde } i=1, \dots, I$$

Pode-se provar que, quando a hipótese nula é verdadeira, a estatística em baixo tem distribuição F com I-1 graus de liberdade no numerador e N-I graus de liberdade no denominador:

$$F = \frac{\sum_{i=1}^I \frac{n_i(\bar{X}_i - \bar{X})^2}{I-1}}{\sum_{i=1}^I \frac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{N-I}} = \frac{SQ_{ext}}{I-1} = \frac{MQ_{ext}}{MQ_e}$$

em que \bar{X}_i representa a média de cada amostra i , $i=1, \dots, I$ e \bar{X} representa a média de todas as observações. De notar que N representa o número total de observações e I o número de populações. Assim pode-se definir a região de rejeição para as hipóteses acima especificadas como:

$$R = \{F > F_{I-1, N-I}^{1-\alpha}\},$$

onde $F_{I-1, N-I}^{1-\alpha}$ representa o quantil de ordem $1-\alpha$ da distribuição F com os graus de liberdade acima referidos. Os passos para a construção da estatística de teste são usualmente apresentados numa tabela ANOVA, semelhante à que já foi apresentada para o caso da regressão múltipla, e com a configuração que se mostra na Tabela 4.

Fonte	Soma de Quadrados	Graus de Liberdade	Média de Quadrados	Estatística F
Entre grupos	$SQ_{ext} = \sum_{i=1}^I n_i(\bar{X}_i - \bar{X})^2$	I-1	$MQ_{ext} = \frac{SQ_{ext}}{I-1}$	$F = MQ_{ext}/MQ_e$
Erro	$SQ_e = \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$	N-I	$MQ_e = \frac{SQ_e}{N-I}$	
Total	$SQ_{Tot} = \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$	N-1	$MQ_{Tot} = \frac{SQ_{Tot}}{N-1}$	

Tabela 4 - Tabela ANOVA para testar igualdade de médias entre populações.

É importante referir ainda que um modelo de Análise da Variância pode ser escrito como um modelo de regressão múltipla, recorrendo a I variáveis indicatrizes representativas de cada população, isto é, que tomam apenas os valores "1" ou "0" consoante a observação pertence ou não à população representada pela variável. Daqui resulta que a tabela ANOVA para um modelo de regressão construído deste modo é exatamente a mesma do que a Tabela ANOVA para o modelo de análise da variância.

3.1.7 Análise de resíduos

A análise de resíduos é utilizada para validar o modelo linear e as condições de Gauss-Markov. Com efeito, esta análise permite obter indicações a fim de melhorar o poder preditivo da variável explicativa (e, conseqüentemente, do modelo). Primeiramente, entenda-se resíduos como as diferenças entre os valores observados e os valores ajustados, tal como já foi anteriormente definido.

Neste estudo, serão abordados dois procedimentos dentro de uma panóplia de possibilidades:

- Representação dos resíduos contra cada variável X_j , $j=1, \dots, p$. A ausência de padrão indica a concordância com o modelo linear. Poderá também ocorrer uma relação quadrática ou indicação de que a variância não é constante. No caso de uma relação quadrática, bastaria uma transformação potência da variável;

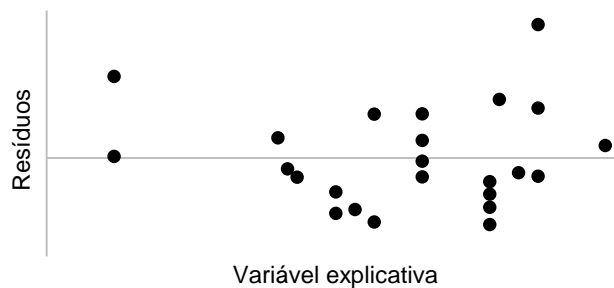


Figura 4 - Resíduos vs variável explicativa – caso de concordância com o modelo linear

- Análise dos resíduos eles próprios: histograma dos resíduos e teste do qui-quadrado para testar a normalidade dos resíduos. Primeiramente, constrói-se um histograma dos resíduos, tendo em conta a regra de Sturges para definir o número de classes (N_c) [22]:

$$2^{N_c} > n \leftrightarrow N_c = 1 + [\ln n / \ln 2]$$

onde $[\ln n / \ln 2]$ é referente à parte inteira deste quociente. Definido então o número de classes, falta apenas definir a amplitude das mesmas, h :

$$h = \frac{X_{n:n} - X_{1:n}}{N_c}$$

em que $X_{n:n}$ é referente ao máximo e $X_{1:n}$ é o mínimo observado dos resíduos obtidos resultantes da regressão.

Deste modo, é possível obter-se indicações relativamente à distribuição através da construção do histograma. No entanto, uma análise mais rigorosa pode passar por testar a hipótese de que a população tem distribuição $F_0(x; \theta_1, \dots, \theta_r)$ utilizando a seguinte estatística de teste X^2 , para j classes:

$$X^2 = \sum_{j=1}^k (N_j - n\hat{p}_j)^2 / n\hat{p}_j$$

Ora, como o objetivo da análise de resíduos passa pela verificação da normalidade dos termos de erro e \hat{p}_j é a frequência esperada para cada classe j , \hat{p}_j corresponde à diferença dos valores da função de distribuição normal nos extremos de cada intervalo. A estatística X^2 , sob a hipótese nula, tem distribuição qui-quadrado com $k-r-1$ graus de liberdade, em que r é o número de parâmetros estimados. Deste modo, é possível validar a normalidade dos resíduos, através da seguinte região de rejeição:

$$R = \{X^2 > \chi_{(k-r-1)}^{1-\alpha}\}$$

3.2 Estudo da multicolinearidade entre variáveis

Um bom modelo de regressão não deve incluir demasiadas variáveis. Por um lado, a inclusão de muitas variáveis no modelo tende a aumentar a variância dos estimadores dos coeficientes de regressão e, por outro, acontece com frequência que, de entre um conjunto bastante vasto de potenciais variáveis independentes, algumas estão linearmente relacionadas entre si, o que também provoca um aumento, que pode ser considerável, nas variâncias dos EMQ dos coeficientes e na amplitude dos intervalos de previsão. Estas variáveis não trazem informação relevante para a explicação da variável dependente, mas provocam uma perda de precisão no ajustamento do modelo. Nesta secção vamos ver alguns métodos que permitem identificar relações lineares entre as variáveis independentes.

3.2.1 Multicolinearidade

A multicolinearidade é uma associação linear entre duas variáveis explicativas, ao passo que multicolinearidade é referente a uma situação de presença de duas ou mais variáveis explicativas num modelo de regressão linear múltipla altamente correlacionadas entre si e/ou a variável resposta [23].

É possível inferir relativamente à multicolinearidade entre variáveis através do cálculo da correlação entre duas variáveis. Analise-se o caso de correlação entre a variável explicativa X e a variável resposta Y :

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}$$

onde x_i e y_i são os valores observados de ambas as variáveis, onde $i=1, \dots, n$ e \bar{x} e \bar{y} são as médias aritméticas de ambas as variáveis. O sinal indica a direção: se a correlação é positiva (+) ou negativa (-) e o valor de ρ , compreendido entre $[-1;1]$, indica a força da correlação. Quanto mais próximo de 1 (em módulo), mais linear será a relação entre as variáveis. Em suma, o número de variáveis potenciais para uma regressão linear múltipla deve ter em conta um *threshold* de correlação, uma vez que a presença de muitas variáveis poderá trazer problemas de multicolinearidade, que se explicará a seguir. No presente trabalho, teve-se em conta que a magnitude do coeficiente de correlação deve ser, pelo menos, considerada fraca (vide Tabela 5).

Magnitude do coeficiente de correlação	Limite inferior	Limite superior
Negligenciável	0	0,2
Fraca	0,2	0,4
Moderada	0,4	0,6
Forte	0,6	0,8
Muito forte	0,8	1

Tabela 5 - Classificação da magnitude da correlação entre variáveis. [24]

A multicolinearidade inflaciona a variância dos coeficientes estimados por uma regressão linear, resultando em valores observados para a estatística t erradamente próximos de zero. Assim, isto poderá resultar em algumas variáveis que, avaliando pelo quociente t , não são estatisticamente significativas, quando, na verdade, deveriam ter significância estatística. [23]

3.2.2 Método VIF – Fatores de inflação das Variâncias

Tal como já referido anteriormente, a multicolinearidade verifica-se na presença de dependência forte entre uma variável explicativa e outras duas ou mais variáveis explicativas (relações aproximadamente lineares entre as variáveis em questão).

O método dos fatores de inflação das variâncias, abreviadamente VIF, tem como principal objetivo determinar o quão “inflacionada” se encontra a variância do coeficiente estimado para o modelo de regressão linear em comparação ao que seria expectável se não houvesse qualquer correlação entre qualquer variável no modelo, baseado numa tolerância pré-definida que define a instabilidade da estimação dos coeficientes [25]. Na expressão de cálculo do VIF:

$$VIF = \frac{1}{1-Tol_j} = \frac{1}{1-R_j^2},$$

a tolerância é definida por $(1-R_j^2)$ e quanto maior este valor, maior a presença de multicolinearidade. Considerando uma variável cujo R^2 seja 0,90, a tolerância em questão será 0,10. Adicionalmente, o VIF será igual a 10.

O critério de decisão (incluir ou não incluir a variável no modelo) passará então pela análise do VIF ou do valor de tolerância. Considera-se a existência de um problema de multicolinearidade se existir um VIF maior ou igual a 10. Um critério alternativo poderá passar pela comparação entre o R^2 da variável e o R^2 da regressão linear do modelo. Ou seja, se o VIF da variável for superior ao VIF utilizando o R^2 da regressão, é considerado que a multicolinearidade é estatisticamente significativa [25].

Contudo, uma vez que este trabalho trata variáveis macroeconómicas, o critério de seleção das variáveis admitirá algum grau de multicolinearidade, uma vez que é normal a presença de multicolinearidade entre as mesmas.

3.2.3 Números condição e componentes da variância

Uma forma alternativa de detetar multicolinearidade entre um conjunto de potenciais variáveis dependentes é através da chamada análise das componentes da variância. Para fazer esta análise, considere-se o modelo equivalente ao $Y=Xb+\varepsilon$, dado por:

$$Y = X_{(s)}b_{(s)} + \varepsilon,$$

em que $X_{(s)}$ representa a matriz de planeamento em que as suas colunas, isto é, os valores observados da cada variável, são reduzidas à mesma escala, isto é,

$$X_{(s)} = XD_{(s)}^{-1},$$

em que $D_{(s)}$ é a matriz diagonal dada por:

$$D_{(s)} = \text{diag}(\|x_{[1]}\|, \dots, \|x_{[p]}\|).$$

Uma vez que se procura a existência de variáveis para as quais existem combinações lineares próximas de zero, com pelo menos dois dos coeficientes diferentes de zero, $\sum c_i x_{[i]} \approx 0$, a utilização da matriz $X_{(s)}$ evita confundir a situação em que uma combinação linear é aproximadamente nula porque algumas das variáveis tomam valores muito pequenos com a existência de uma relação linear entre elas com valor aproximadamente nulo.

Outra vantagem de utilizar a matriz $X_{(s)}$ reside no facto de cada elemento na diagonal de $X'_{(s)}X_{(s)}$ ser unitário e, por consequência, a soma dos valores próprios de uma matriz é igual ao seu traço, ou seja, p , onde p é o número de colunas presentes na matriz de planeamento.

Se existir uma relação linear entre as colunas da matriz $X_{(s)}$, então um dos valores próprios de $X'_{(s)}X_{(s)}$ estará próximo de 0. De forma a avaliar a grandeza de um valor próprio relativamente aos restantes, recorre-se ao cálculo dos números condição, n_j , definido por:

$$n_j = \sqrt{\frac{\lambda_{\max}}{\lambda_j}},$$

em que $\lambda_1, \dots, \lambda_p$ são os valores próprios de $X'_{(s)}X_{(s)}$ e λ_{\max} é o máximo de todos os valores próprios. Por definição, $n_j \geq 30$ indica a presença de uma relação linear entre as variáveis da matriz de planeamento. De forma a detetar as variáveis a ser descartadas da construção do modelo, é necessário analisar a matriz dos vetores próprios de $X'_{(s)}X_{(s)}$. Atendendo à simetria da referida matriz $X'_{(s)}X_{(s)}$, é possível proceder à seguinte diagonalização:

$$X'_{(s)}X_{(s)} = \Gamma D_\lambda \Gamma',$$

em que D_λ é uma matriz diagonal dada por:

$$D_{(\lambda)} = \text{diag}(\lambda_1, \dots, \lambda_p)$$

Γ é a matriz ortogonal cujas colunas são os vetores próprios de $X'(s)X(s)$:

$$\Gamma = \begin{bmatrix} \gamma_{11} & \dots & \gamma_{1p} \\ \gamma_{21} & \dots & \gamma_{2p} \\ \dots & \dots & \dots \\ \gamma_{p1} & \dots & \gamma_{pp} \end{bmatrix}$$

Como a matriz de covariâncias dos EMQ dos coeficientes é dada por:

$$\text{cov}(\hat{b}_{(s)}) = \sigma^2(X'X)^{-1} = \sigma^2\Gamma D_{\lambda}^{-1}\Gamma',$$

vem que a variância do estimador de um dado coeficiente b_j é:

$$\text{var}(\hat{b}_j^{(s)}) = \sigma^2 \sum_{k=1}^p \lambda_k^{-1} \gamma_{jk}^2.$$

Deste modo, aos valores $(\lambda_k^{-1}\gamma_{jk}^2)$ é dado o nome de componentes da variância de $\hat{b}_k^{(s)}$, $k=1, \dots, p$, e os coeficientes $\Phi_{kj} = \frac{(\lambda_k^{-1}\gamma_{jk}^2)}{(\sum_{l=1}^p \lambda_l^{-1}\gamma_{jl}^2)}$ medem a proporção da variância que é explicada pelo k-ésimo valor próprio. A um valor próprio próximo de zero corresponde uma só relação linear entre algumas das variáveis independentes e a análise das componentes da variância permite identificar quais as variáveis envolvidas nessa relação.

Valores próprios	Número Condição	Proporção de Var (\hat{b}_1)	...	Proporção de Var (\hat{b}_p)
λ_1	n_1	$\sigma^2\gamma_{11}^2/\lambda_1 \text{Var}(\hat{b}_1)$...	$\sigma^2\gamma_{p1}^2/\lambda_1 \text{Var}(\hat{b}_p)$
λ_2	n_2	$\sigma^2\gamma_{12}^2/\lambda_2 \text{Var}(\hat{b}_1)$...	$\sigma^2\gamma_{p2}^2/\lambda_2 \text{Var}(\hat{b}_p)$
...
λ_p	n_p	$\sigma^2\gamma_{1p}^2/\lambda_p \text{Var}(\hat{b}_1)$...	$\sigma^2\gamma_{pp}^2/\lambda_p \text{Var}(\hat{b}_p)$

Tabela 6 - Tabela com valores próprios, números condição e componentes da variância.

Resumidamente, realiza-se então uma análise com base na Tabela 6, onde o objetivo é detetar as variáveis que estão envolvidas em relações lineares e, através da sua eliminação, diminuir a variância dos estimadores dos coeficientes da regressão. Tal é possível através da inspeção das linhas com número condição maior ou igual a 30, valores próprios quase nulos, e, nestas linhas, detetar as colunas cuja proporção da variância é consideravelmente superior às restantes.

3.2.4 Métodos de seleção de variáveis

Em geral, a construção de um modelo de regressão linear múltipla inicia-se com um conjunto bastante alargado de potenciais variáveis independentes e, de entre estas, escolhem-se aquelas que melhor explicam a variável resposta. Como já foi referido, não é conveniente incluir variáveis não significativas ou que provocam problemas de multicolinearidade. Uma possibilidade de escolha das variáveis é ajustar todos os possíveis conjuntos de variáveis e escolher a que produz o melhor modelo. No entanto, o número de possíveis regressões aumenta exponencialmente com o número de variáveis (p), pois é dado por 2^p-1 . Por isso, é comum recorrer-se a diferentes métodos de seleção de variáveis que

procuram eliminar ou adicionar as variáveis ao modelo, com base na análise dos efeitos produzidos por cada variável.

A Tabela 7 apresenta os três métodos mais comuns na literatura. Resumidamente, o método de seleção *forward* inicia o modelo sem variáveis e vai juntando, uma a uma, aquela que produz maior aumento do R^2 (ou maior diminuição na variância dos resíduos) até que a próxima variável a entrar no modelo não seja significativa. O método *stepwise* constitui uma variante deste método em que, quando se junta uma nova variável ao modelo, todas as variáveis já incluídas são testadas com o intuito de verificar se alguma deixa de ser significativa com a entrada da nova variável. Finalmente, o método de seleção *backwards* inicializa o modelo com todas as variáveis e vai retirando, uma a uma, aquela com menor valor da estatística t ou (por outras palavras) com o maior p -value para o teste t , até que todas as variáveis incluídas no modelo sejam significativas para um determinado nível de significância, definido à partida.

PASSOS	<i>Backward Selection</i>	<i>Forward Selection</i>	<i>Stepwise</i>
1	Iniciar o algoritmo com todas as p variáveis no modelo	Iniciar o algoritmo com 0 variáveis no modelo ($p=1$)	
2	Ajustar o modelo com as p variáveis e calcular os quocientes t para cada variável	Considerar a variável que provoca maior aumento no R^2 (i.e., redução no SQe)	
3	Averiguar a significância do quociente t mais pequeno. Se não for significativo, elimina-se a variável com maior p -value. Regressar ao passo 2, com $p-1$ variáveis	Testar, através de um teste t , se a variável é significativa	
4	Se se verificar que todos os quocientes t são significativos, adota-se o modelo atual	Se a variável em 3 for significativa, juntar a variável e voltar ao passo 2. Caso contrário, ir para o passo 5	Se a variável em 3 for significativa, juntar a variável e voltar ao passo 2. Caso contrário, ir para o passo 5
4A			Depois de incluída nova variável, testar a significância das restantes incluídas (teste t) e retirar as não significativas. Ir para o passo 2.
5	N/A	Adota-se o modelo com p variáveis	

Tabela 7- Métodos de seleção de variáveis.

3.3 Tratamento da informação e construção do modelo estatístico

Este capítulo pretende explicar o tratamento necessário de informação relativamente às variáveis em estudo e a construção do modelo de regressão linear múltipla cuja variável dependente é um valor numérico ordinal, representante da posição de cada grau de risco da escala de *rating*.

Para a construção do modelo foi necessário ter em conta os aspetos relacionados com o tratamento da informação e o processo de escolha de variáveis, no sentido de se obter uma sistematização adequada.

Seguidamente apresentam-se os passos adotados no presente trabalho e que conduziram à construção do modelo.

3.3.1 Variável dependente

Este estudo tem como objetivo determinar as variáveis que melhor explicam a variação observada no *rating* português. Deste modo, para a construção da variável dependente (*rating*) procederam-se às seguintes etapas descritas em seguida:

1. Recolheram-se as observações (atribuição de *rating*) para o período de 1996 a 2015¹⁰, por parte de três agências: Moody's, S&P's e Fitch, tomando-se as seguintes opções:
 - a. Considerou-se que as observações de *rating* das três agências integram uma única variável dependente. Esta opção decorre dos resultados obtidos nos seguintes testes estatísticos:
 - i. Tabela ANOVA: permitiu verificar que as médias dos *ratings* de cada uma das agências não são estatisticamente diferentes entre si;
 - ii. Correlação: permitiu avaliar se os *ratings* publicados pelas agências apresentam um comportamento similar.
 - b. Optou-se pelo histórico mínimo de 1996 devido às falhas de informação no período anterior num grande número de variáveis.
 - c. Optou-se por incluir apenas as 3 agências com histórico desde 1996, isto é, ignorando os *ratings* mais recentes da agência DBRS;
 - d. Para cada ano, utilizaram-se todos os *ratings* publicados por cada agência. No caso de uma agência não ter publicado uma notação de *rating* para uma dada posição anual, atribuiu-se o valor da notação imediatamente anterior. Assim, em cada ano existem pelo menos 3 posições diferentes.

¹⁰ Não foi possível, à data de realização deste trabalho, obter informação para o ano de 2016. (Data de referência: fevereiro de 2017)

2. Transformou-se a notação dada por cada agência (uma notação em forma de letras para uma notação numérica ordinal, i.e., de 1 a 22, onde 1 representa o menor risco e 22 o maior risco) através de uma *master scale*, tal como se pode observar na Tabela 8.

#	S&P'S	MOODY'S	FITCH	DESCRIÇÃO DO RATING	CLASSIFICAÇÃO	
1	AAA	Aaa	AAA	<i>Prime</i>	Investment Grade	
2	AA+	Aa1	AA+	<i>High grade</i>		
3	AA	Aa2	AA			
4	AA-	Aa3	AA-			
5	A+	A1	A+	<i>Upper medium grade</i>		
6	A	A2	A			
7	A-	A3	A-			
8	BBB+	Baa1	BBB+	<i>Lower medium grade</i>		
9	BBB	Baa2	BBB			
10	BBB-	Baa3	BBB-			
11	BB+	Ba1	BB+	Non-investment grade speculative	Speculative Grade	
12	BB	Ba2	BB			
13	BB-	Ba3	BB-			
14	B+	B1	B+			<i>Highly speculative</i>
15	B	B2	B			
16	B-	B3	B-			
17	CCC+	Caa1	CCC+			<i>Substantial risks</i>
18	CCC	Caa2	CCC			
19	CCC-	Caa3	CCC-			
20	CC	Ca	CC			<i>In or near default, with possibility of recovery</i>
21	C		C			
22	SD	C	DDD	<i>Default, with little chance of recovery</i>		

Tabela 8 -Master Scale (S&P's, Moody's e Fitch) [26] [19] [27]

3.3.2 Variáveis independentes

1. Após a revisão bibliográfica e a escolha das variáveis que se consideram relevantes para a modelação do risco soberano, tomaram-se as seguintes opções de tratamento das mesmas:
- Se uma variável X não apresentar um valor para uma determinada posição anual j, calcule-se o mesmo através de uma interpolação simples tal como apresentada em (1):

$$X_j = X_{j-1} + \frac{X_{j+1} - X_{j-1}}{j + 1 - (j - 1)} \times (j + 1 - j) = X_{j-1} + \frac{X_{j+1} - X_{j-1}}{2} \quad (1)$$

Exemplo: Para a variável “Despesas em atividades de Investigação e Desenvolvimento” (D), registou-se a ausência do valor para o ano 1998:

ANO	D (M€)
1997	577
1998	N/A
1999	815

Logo, através da expressão (1):

$$D_{1998} = D_{1997} + \frac{D_{1999} - D_{1997}}{2} = 577 + \frac{815 - 577}{2} = 696 \text{ M€}$$

- a. Se as variáveis recolhidas estiverem em diferentes ordens de grandeza como por exemplo: Despesa (€), Receita (m€) e Investimento (M€), a ordem de grandeza a adotar será sempre a maior, uma vez que a análise é efetuada para um contexto macroeconómico. Neste caso, as variáveis foram analisadas em mM€;
- b. Numa variável com registos mensais em vez de registos anuais, considera-se o último mês do ano como o valor representativo do ano. Por exemplo, no caso da Euribor, é possível recolher a partir do BCE valores numa base mensal, no entanto, por coerência com as restantes variáveis, optou-se por considerar apenas o último registo mensal, por ano.

3.3.3 Modelação e estimação dos parâmetros da regressão linear múltipla

Com o conjunto de variáveis já selecionadas (e tratadas) como determinantes do *rating* soberano português, definiu-se a metodologia, que seguidamente se apresenta, para identificar as variáveis que melhor se adequam ao modelo. Refira-se que muitas variáveis são macroeconómicas, i.e., tendem a seguir um padrão de comportamento similar durante um ciclo económico, o que naturalmente, gera correlações elevadas entre elas.

1. Numa primeira fase estudou-se a correlação das variáveis:
 - a. Entre cada variável independente e a variável dependente. O *threshold* (em módulo) definido para correlação mínima foi fixado em 0,20, o que equivale a uma correlação fraca: $R^2 \geq 0,04$;
 - b. Numa segunda fase, avaliaram-se as correlações entre as variáveis independentes, de acordo com os seguintes métodos:
 - i. **VIF (Método dos Fatores de Inflação da Variância):** Este método foi utilizado para identificar as principais fontes de multicolinearidade. Para esta análise, considerou-se o valor máximo de VIF igual a 10 (ou seja, é o equivalente a afirmar um R^2 de 0,90, tal como foi explicado anteriormente na secção 3.1). Devem ser consideradas as seguintes situações:
 1. Caso se verifique que a matriz tenha uma diagonal com valores positivos muito elevados, trata-se de uma situação de multicolinearidade. Para se resolver esta situação, remove-se a variável referente ao máximo da diagonal do VIF. Após a remoção desta, efetua-se a análise novamente sem a referida variável. Repete-se este procedimento até que finalmente não se verifique nenhum valor na diagonal maior ou igual a 10. Quando tal se verifica, parte-se para o passo (ii), com o intuito de identificar relações lineares entre as variáveis, caso ainda existam;
 2. Pode acontecer que exista um problema de multicolinearidade tão acentuado que provoque erros de arredondamento graves, a ponto de os

métodos numéricos de inversão de matrizes produzirem uma matriz inversa da matriz de correlações com uma diagonal com valores negativos; na verdade, a matriz é, na prática, não invertível, prosseguindo-se para o passo seguinte (ii), de forma a fazer-se uma melhor triagem às variáveis que possam estar a provocar sérios problemas de multicolinearidade.

- ii. **Método da análise das componentes da variância:** Este método apresenta uma componente de análise mais subjetiva, sendo necessário decidir qual das variáveis presentes nas linhas com os maiores valores da componente da variância deve ser escolhida entre as diferentes variáveis envolvidas numa mesma relação linear. A análise passa por inspecionar todas as linhas que apresentem número condição maior ou igual a 30 para escolher, de entre as variáveis nessa linha com componentes da variância elevadas, qual deverá ser a excluída. Identificadas as variáveis que respeitem as condições suprarreferidas, procede-se ao cálculo de uma matriz de correlações entre as mesmas. A variável que apresentar maior correlação com as restantes é excluída do conjunto de variáveis. De notar que este pressuposto tem em consideração que poderão existir outras variáveis que possam estar a contribuir para outra(s) relação/relações linear(es), mas que se consideram apenas os casos mais evidentes para análise;
 - iii. A análise só termina quando se obtém uma análise VIF com a diagonal a respeitar as condições em (i.1) e quando a matriz de correlações mencionada em (ii) revelar que as correlações entre as variáveis selecionadas não são demasiado fortes.
2. Após a identificação dos casos mais evidentes de multicolinearidade e eliminação das variáveis que os provocam, utilizam-se os três métodos de seleção de variáveis para construção do modelo de regressão linear múltipla: seleção regressiva, seleção progressiva e seleção regressiva e regressiva (*stepwise*), tal como apresentados anteriormente. Para a decisão de inclusão ou exclusão de variáveis usou-se um nível de significância de 5%.

3.3.4 Análise aos fatores determinantes do risco soberano nos períodos antes e durante/depois da crise

Posteriormente à análise das variáveis que melhor explicam o risco soberano no período em análise, procurou-se avaliar a robustez dos resultados nas duas fases do ciclo económico, nomeadamente:

- **Antes da crise:** período compreendido entre 1995 e 2008;
- **Durante/depois da crise:** período compreendido entre 2009 e 2015.

Esta avaliação consubstanciou-se nas seguintes análises estatísticas:

1. Verificação de que as publicações de *rating* por parte das três agências pertencem à mesma população:

- a. Tabela ANOVA;
 - b. Correlação entres pares de publicações (Moody's e S&P's, Fitch e Moody's, Fitch e S&P's).
2. Verificação de que as variáveis selecionadas em 3.2.3 são (ou não) determinantes na atribuição de *rating* para os dois períodos de análise. Tal é efetuado através da estimação dos parâmetros da regressão pelo Método dos Mínimos Quadrados e consequente utilização da estatística de teste *t* para determinar se a variável é significativa para os níveis de significância usuais (1, 5 e 10%). Adicionalmente, avalia-se o poder preditivo de cada variável.

3.3.5 Análise à taxa de juro soberano e prémio de mercado

A taxa de juro da dívida soberana portuguesa é recolhida numa base mensal (última posição do mês), tendo sido necessário transformar o *rating* soberano numa informação também mensal.

Com o intuito de melhor compreender a relação entre o *rating* e a taxa de juro, definiram-se dois *drivers* de análise distintos, a taxa de juro e o prémio de mercado. Na construção das amostras, é importante ter em consideração o seguinte:

- Na ausência de publicações de *rating* num determinado mês, considera-se que não se verificaram alterações na avaliação ao risco soberano, e deste modo, assume-se a última publicação disponível;
- O *rating* em vigor para uma posição mensal é dado pela última publicação nesse mês, ou seja, desprezam-se as publicações que possam ter sido efetuadas naquele mês (quando se trata da mesma agência).

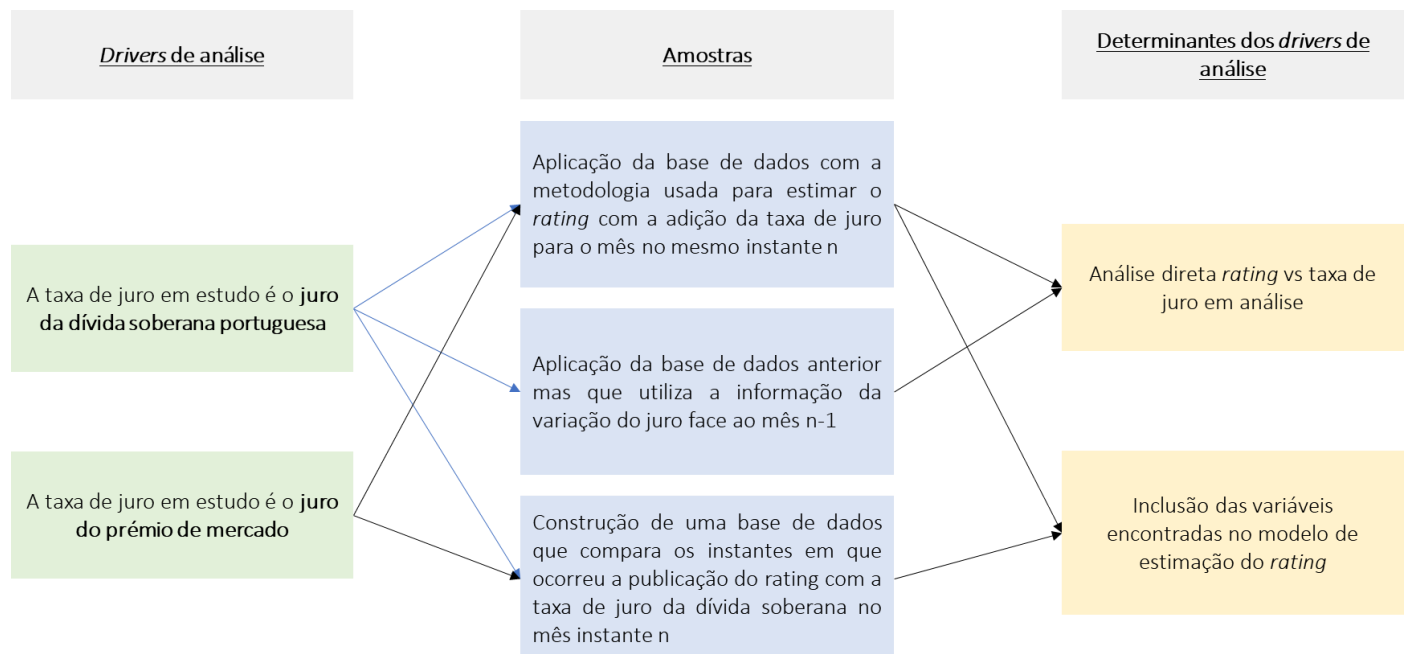


Figura 5 - Metodologia de estudo do juro face ao *rating* soberano português.

As amostras obtidas respeitam o esquema apresentado na Figura 5, onde o prémio de mercado¹¹ é dado pela seguinte expressão:

$$prémio_{mercado} = bond_{Portugal} - bond_{Alemanha}$$

4. Análise dos determinantes do *rating*

4.1 Análise da variável resposta, *rating*

No período em análise, 1996-2015, observaram-se 46 ratings relativos ao risco soberano português. Estas observações foram transformadas em *rating* numérico, através da escala presente na Tabela 9. Numa primeira fase, procurou-se avaliar se os *ratings* das três agências eram similares, i.e., se assumindo a sua independência, não diferiam estatisticamente. Para o efeito, recorreu-se a uma tabela ANOVA, assumindo-se os seguintes pressupostos na construção da amostra:

- i. As publicações de *rating* anteriores são iguais aos anos seguintes na ausência de informação desses anos;
- ii. Preenche-se o valor do *rating* de uma dada agência à data de publicação para um dado ano, mantendo o valor numérico deste para as restantes agências.

Assim, tendo em conta os pressupostos assumidos, a amostra contempla 90 observações de *rating* (em vez de 46), distribuídas por 30 *ratings* de cada agência. Na Tabela 9, apresentam-se as principais características amostrais por agência. Da análise da mesma, é sugerido que a Moody's apresenta uma perspetiva, em média, ligeiramente mais conservadora do que a Fitch (+0.2), embora menos que a S&P (-0.5). No entanto, esta agência aparenta um maior grau de volatilidade nas suas publicações, registando o maior valor de variância, o que poderá dever-se a uma abordagem mais conservadora em tempos de crise, relativamente às restantes CRA. No mesmo sentido, outros autores referem que esta CRA tem maior probabilidade de ser mais conservadora. [28] [29] [30]

Agências	Contagem	Soma	Média	Variância
Moody's	30	195	6,5	17,8
Fitch	30	190	6,3	13,3
S&P	30	209	7,0	13,9

Tabela 9 - Ratings atribuídos pelas três agências (1996-2015)

Por outro lado, a Tabela 10 procura quantificar as diferenças de *rating* no mesmo período de referência, tendo-se calculado a diferença de *rating*, em *notches*, entre os três pares distintos de CRA.

¹¹ Considera-se que a dívida alemã é representativa de uma taxa de juro sem risco, tendo em conta que desde 1983 mantém o *rating* de AAA e apresenta uma situação económica sólida no contexto mundial. Esta metodologia visa eliminar o efeito das flutuações do mercado no valor do *spread* da dívida portuguesa.

Notches	Moody's vs Fitch	Moody's vs S&P	Fitch vs S&P
-4	0	2	0
-3	0	0	1
-2	2	3	4
-1	3	8	10
0	17	10	13
1	4	6	2
2	4	1	0
Total	30	30	30

Tabela 10 - Diferença de *notches* entre os três pares de CRA (valores absolutos)

Por observação da Tabela 10, conclui-se que a Fitch tem tendência a ser a CRA menos conservadora e a S&P a mais conservadora. Por outro lado, verifica-se maior alinhamento entre a Moody's e Fitch. De facto, a Tabela 11, avalia as semelhanças entre as agências, procedendo-se ao módulo das diferenças observadas na tabela anterior. Deste modo, obteve-se a seguinte matriz:

Notches	Moody's vs Fitch	Moody's vs S&P	Fitch vs S&P
0	56,7%	33,3%	43,3%
1	23,3%	46,7%	40,0%
2	20,0%	13,3%	13,3%
3	0,0%	0,0%	3,3%
4	0,0%	6,7%	0,0%

Tabela 11 - Diferença de *notches*, valores relativos, em módulo.

Embora se observe maior discordância entre a Moody's e a S&P, a Tabela 11 mostra que essa diferença é igual ou inferior a 1 *notch* em cerca de 80% das publicações. Adicionalmente, também é possível verificar que entre as "Big Three", pelo menos 80% das publicações são exatamente a mesma nota ou uma nota com a diferença de um *notch*. Contudo, é relevante avaliar se os restantes 20% resultam numa diferença estatisticamente significativa entre as três agências.

Uma maneira de avaliar rapidamente se os *ratings* das três agências não são independentes passa por calcular a correlação entre os três pares. Aplicando a fórmula da correlação explicada na subsecção 3.2.1, obtêm-se os resultados da Tabela 12, os quais indicam uma magnitude de correlação muito forte.

Moody vs Fitch	Moody vs S&P	Fitch vs S&P
97,7%	94,7%	96,9%

Tabela 12 - Correlação entre os três pares de CRA.

Procedendo a uma análise mais robusta do ponto de vista estatístico, efetuou-se um teste de igualdade de valores médios, isto é, uma análise da variância. Assim, à semelhança da Tabela 4, obteve-se a seguinte tabela ANOVA (Tabela 13), que resulta do seguinte teste de hipóteses:

$$H_0: \mu_i = \dots = \mu_p \text{ vs } H_1: \mu_i \neq \mu_p, \text{ pelo menos para um par } i, p, \text{ onde } i=1, \dots, p$$

Fonte	Soma de Quadrados	Graus de Liberdade	Média de Quadrados	Estatística F
Entre grupos	6,5	2	3,2	0,22
Dentro de grupos	1305,1	87	15,0	
Total	1311,6	89	14,9	

Tabela 13 - Teste ANOVA para igualdade de médias entre populações.

Tendo por base a tabela ANOVA e a região de rejeição, R, dada por:

$$R = \{F > F_{I-1, N-1}^{1-\alpha}\} \leftrightarrow \{0,22 > 3,10\}$$

Conclui-se que a hipótese nula é verdadeira: 0,22 é inferior ao $F_{\text{crítico}}$. Assim, não se rejeita a igualdade de valores médios entre populações. Por outras palavras, o *p-value* para o teste de hipóteses suprarreferido é de 0,81, o que equivale a não rejeitar a hipótese nula para qualquer valor usual de significância (1, 5 e 10%). Deste modo, poder-se-á concluir que as publicações das três agências pertencem à mesma população.

Atendendo à metodologia de construção da amostra de *ratings*, tal como descrito na secção 3.3.1 (1d), descreve-se sucintamente as características amostrais do *rating* do risco soberano português. Na Tabela 14, observa-se o efeito da crise *subprime* em 2008, dado que a média se posiciona acima da mediana (6,2 vs 4), comportamento que se observa, com maior clareza, na Figura 6.

Observações	Mínimo	Máximo	Média	Mediana	Desvio-Padrão
79	3	13	6,20	4	3,72

Tabela 14 - Características amostrais do *rating* do risco soberano português no período em análise.

Da análise da distribuição dos *ratings* ao longo do tempo (que compreende uma amostra de 79 observações) destaca-se a elevada amplitude (10 *notches*) entre o melhor e o pior *rating*, sendo a mediana inferior à média em 2,2, isto é, um enviesamento à direita da distribuição. Por outro lado, verifica-se o aumento da variabilidade a partir de 2009, sugerindo um maior grau de incerteza na avaliação de risco por parte das agências de *rating*.

Tal como referido, a distribuição do *rating* centrada à direita deve-se essencialmente à existência de um período de crise historicamente acentuado (contido entre 2008 e 2015), onde o *rating* assumiu 10 *notches* acima do período de prosperidade económica registado entre 1998 e 2005. É importante referir que embora se tenham verificado momentos de estagnação da economia entre 2001 e 2005, estes não resultaram num *downgrade* do *rating* soberano [31]. Isto sugere que a recessão económica provocada pela crise financeira de 2008 não reflete o resultado expectável de uma avaliação de *rating* num período normal de recessão económica, mas um evento severo com características específicas. Recorde-se que a última recessão profunda, com repercussões semelhantes nos indicadores macroeconómicos (por exemplo, variação no PIB superior a -4%), ocorreu em 1975, tendo sido originada pela crise do petróleo (1973) e pela instabilidade política instalada após o 25 de abril de 1974. Tal como em 2011, o país foi obrigado a solicitar a intervenção do FMI em 1977. Tendo em conta estes aspetos, os resultados do presente trabalho estão limitados pelo histórico económico atípico. Adicionalmente, tal como já foi referido também no

capítulo 2 e, em linha com o que ocorreu na crise do petróleo, a crise macroeconómica foi aumentada pela instabilidade política, uma vez que se verifica uma degradação acelerada do *rating* em 2011 e 2012. Ainda assim, no período mais recente de 2015, a recuperação dos principais indicadores macroeconómicos e a redução da instabilidade política, permitiu a melhoria do *rating* que se posiciona na primeira notação de “lixo” equivalente a BB+ na escala de *rating* da S&P), isto é, dois *notches* acima do *rating* de 2012.

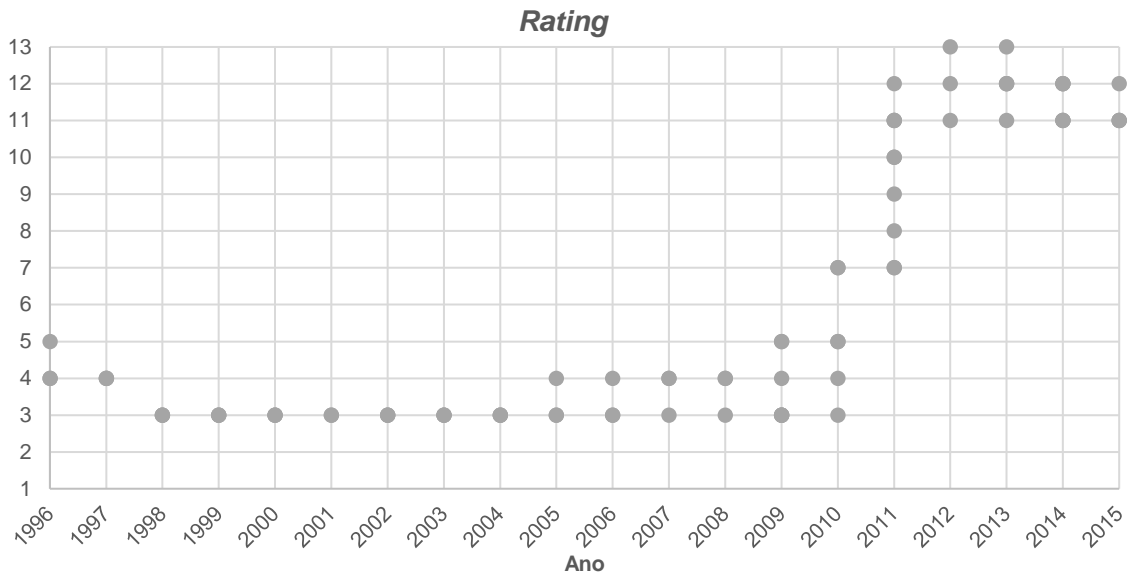


Figura 6 - Distribuição do *rating* do risco soberano português no período em análise: 1996-2015

4.2 Análise das variáveis independentes

A atribuição de *rating* contempla a análise de diferentes tipos de variáveis, tais como sociais, macroeconómicas, políticas, qualitativas, entre outras. Neste sentido, após um levantamento das variáveis mais relevantes, as quais foram sumarizadas na subsecção 2.4, apresenta-se na Tabela 15, uma breve descrição estatística das mesmas, i.e., média, mediana, desvio-padrão, mínimo, máximo, bem como a correlação e o R^2 com a variável dependente *rating*. As variáveis encontram-se ordenadas da maior para a que apresenta menor correlação com a variável resposta.

É possível concluir que as variáveis não são centradas, i.e., todas apresentam uma média maior ou menor do que a mediana (a inflação tem a média ligeiramente inferior à mediana, se os resultados forem apresentados com três casas decimais).

No que se refere à correlação com o *rating*, conclui-se que, em módulo, todas as variáveis apresentam correlação superior a 0,20 e que um número considerável de variáveis (13 das 22 analisadas), explica mais do que 50% do *rating*. Se este exercício se tratasse de uma situação de regressão linear simples e o único critério de seleção de variáveis fosse a maximização do R^2 , concluir-se-ia que o endividamento do setor bancário português seria o principal determinante na atribuição de *rating*, nomeadamente através do rácio de *nonperforming loans* dos bancos sobre o empréstimo total. No entanto, apesar desta variável poder

influenciar o *rating* de forma indireta, quer através da redução de liquidez à atividade económica quer pelo aumento do risco de injeção de fundos estatais para suporte ao capital destas instituições, ela não explica por si só o *downgrade* verificado no *rating*. De facto, devemos ter em consideração que o *rating* avalia o risco de incumprimento soberano que depende sobretudo do desempenho macroeconómico, da administração da dívida pública e dos aspetos políticos e sociais do país.

Salienta-se ainda a elevada correlação das variáveis macroeconómicas “dívida bruta em % do PIB”, “taxa de desemprego”, “Formação Bruta de Capital Fixo” e “Balança Comercial”. De facto, dada a sua elevada capacidade explicativa, seria possível construir um bom modelo utilizando apenas uma destas variáveis e conjugando-a com uma variável com menor correlação, mas independente da anterior.

Variáveis	Média	Mediana	σ^2	Mínimo	Máximo	Correlação	R ²
Banks NPL/Empréstimo (%)	5,628	4,82	3,51	1,300	11,925	0,931	86,6%
Dívida bruta em % do PIB	86,970	83,60	29,47	50,300	130,600	0,926	85,8%
Desemprego %	9,538	9,40	3,71	3,900	16,200	0,899	80,9%
Formação bruta de capital fixo (% PIB)	21,159	21,10	4,22	14,800	28,000	-0,888	78,9%
Balança Comercial (Importações+Exportações) / PIB	-8077,872	-10140,00	6191,67	-16769,8	3132,200	0,887	78,6%
Balança Corrente/PIB	0,690	0,68	0,07	0,588	0,808	0,870	75,8%
Despesas Estado Social/PIB	-0,066	-0,08	0,04	-0,121	0,015	0,844	71,2%
PIB/Hora trabalhada	0,058	0,06	0,02	0,033	0,081	0,773	59,7%
Poupanças das famílias %	30,145	30,44	2,00	26,109	32,572	0,769	59,1%
Impostos sobre o rendimento e património	1,004	1,34	2,60	-4,009	5,073	-0,768	59,0%
Euribor	14690,633	15140,90	3224,30	8247,600	19438,300	0,748	55,9%
Dívida Externa/PIB	2,253	2,03	1,42	0,060	5,054	-0,737	54,3%
Custos laborais unitários	0,643	0,81	0,34	-0,021	1,046	0,725	52,5%
PIB per Capita	1,157	1,78	2,35	-3,174	4,355	-0,685	46,9%
Reservas Total	24134,493	26496,18	4393,77	14903,322	29717,694	0,648	41,9%
Leverage banking %	14,72	15,02	4,01	8,443	19,807	0,623	38,9%
Divida externa / Exportações PIB %	13,581	13,29	3,76	8,320	20,400	0,579	33,5%
Endividamento Particulares %	1,944	2,42	0,91	-0,079	3,057	0,476	22,6%
Inflação	0,643	0,91	2,40	-4,028	4,792	-0,434	18,8%
Défice público (M€)	129,814	141,32	26,91	58,707	154,364	0,370	13,7%
	0,020	0,02	0,02	-0,008	0,044	-0,273	7,5%
	-9764,839	-8245,20	4916,75	-20100,2	-3625,800	-0,226	5,1%

Tabela 15 - Descrição sumária estatística das variáveis em análise (79 observações, entre 1996-2015).

4.3 Estimação do modelo de regressão linear múltipla

Tal como referido anteriormente, um modelo de atribuição de *rating* utiliza diferentes variáveis para estimar o risco soberano, no entanto, para o histórico de 21 anos em análise, seria suficiente a utilização da variável “dívida bruta em % do PIB” para explicar 85,8% do *rating*. Porém, se se pretender um maior nível de precisão na estimação e previsão do *rating*, é premente recorrer a outras variáveis explicativas.

O processo de seleção de variáveis, tal como referido na subsecção metodológica, foi desenvolvido de acordo com as seguintes etapas:

1. Verificação do *threshold* de correlação com a variável dependente;
2. Tratamento das variáveis – eliminação de situações de multicolinearidade a partir da análise VIF e da análise das componentes da variância;
3. Seleção das variáveis para o modelo (*backward, stepwise, forward*).

Uma vez que a primeira etapa corresponde às variáveis da Tabela 15 (que apresentam, em módulo, uma correlação com a variável resposta superior a 0,20), apresentam-se os resultados das etapas 2 e 3.

4.3.1 Análise de multicolinearidade das variáveis independentes

A análise de multicolinearidade pode ser efetuada com recurso a diferentes métodos. Neste trabalho abordam-se dois métodos, tal como descritos metodologicamente: VIF e a análise das componentes da variância. Na análise VIF é necessário o cálculo das seguintes matrizes:

- Matriz de correlações entre as variáveis independentes;
- Matriz inversa da matriz de correlações, cuja diagonal contém os valores de VIF.

Em seguida, analisa-se a diagonal da segunda matriz e observa-se qual a variável correspondente à coluna com maior valor de VIF. Para a base de análise inicial (vide Anexo B), concluiu-se que a variável a ser removida é a “Balança Comercial”, uma vez que apresenta o maior valor da diagonal, 3×10^{15} . É de notar que este valor é muito superior a 10, o valor que se fixou como *threshold*. Tal se deve ao facto de se iniciar com uma base de 22 variáveis que, na sua maioria são macroeconómicas, isto é, encontram-se muito correlacionadas entre si.

O processo iterativo para quando se obtém uma diagonal com todos os elementos inferiores a 10. Em cada iteração calcula-se a matriz cuja diagonal corresponde aos valores VIF, removendo-se da base a variável que contém o valor máximo da diagonal e voltando a calcular a matriz com p-1 variáveis. A Tabela 16 apresenta um quadro resumo de todas as iterações realizadas até se obter uma base com o valor de VIF inferior ao *threshold* definido.

Da análise da Tabela 16, conclui-se que as variáveis com maior correlação com as restantes, ou seja, que geram maiores problemas de multicolinearidade são: o “rácio entre a dívida externa e o PIB”, o “rácio entre dívida externa e exportações” e ainda o “rácio entre PIB e horas despendidas”.

Iteração	Variáveis a remover	VIF
2	Dívida externa / PIB	6,3 x 10 ¹⁴
3	Dívida Externa / Exportações	2,1 x 10 ¹⁴
4	PIB/Hora	3952,6
5	Despesas estado social / PIB	835,2
6	Dívida Bruta (% PIB)	514,0
7	PIB per capita	389,0
8	Formação bruta de capital fixo (% PIB)	285,3
9	Bank NPL / Empréstimo total	169,2
10	Impostos sobre o rendimento e património	100,7
11	Desemprego (%)	27,3
12	Défice público (M€)	21,9
13	Balança Corrente / PIB	16,2

Tabela 16 - Seleção de variáveis através de sucessivas iterações com o método VIF

Com efeito, a variável “Poupança das famílias %” apresenta um VIF de 8, ou seja, inferior a 10, de modo que não se deve eliminar esta variável. No entanto, para se confirmar que a magnitude da correlação desta variável é forte com alguma(s) das restantes, procedeu-se à análise das componentes da variância utilizando as variáveis resultantes do procedimento anterior. Na Tabela 17, verifica-se a presença de duas linhas com número condição superior a 30 e, conseqüentemente, valores próprios muito próximos de zero. Observa-se também a presença de duas relações lineares, tanto na primeira como na penúltima linha, entre duas variáveis.

Valores próprios	Inflação	PIB %	Lev. banking %	Custos unitários laborais	Euribor	Poup. Famílias %	Endiv. Partic. %	(Import. + Export.) / PIB	Reservas Total	Número Condição
0,00	0,00	0,01	0,00	0,07	0,02	0,50	0,16	0,79	0,05	97,8
6,74	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,0
1,92	0,00	0,03	0,00	0,02	0,00	0,02	0,00	0,00	0,00	1,9
0,38	0,13	0,02	0,00	0,06	0,01	0,02	0,00	0,00	0,00	4,2
0,64	0,00	0,24	0,00	0,02	0,00	0,04	0,00	0,00	0,00	3,2
0,06	0,28	0,01	0,00	0,21	0,36	0,03	0,01	0,00	0,02	10,4
0,21	0,03	0,06	0,00	0,14	0,03	0,20	0,00	0,00	0,01	5,6
0,01	0,00	0,45	0,52	0,05	0,05	0,00	0,02	0,02	0,08	24,5
0,00	0,41	0,12	0,46	0,43	0,08	0,16	0,77	0,18	0,74	52,4
0,03	0,14	0,06	0,01	0,00	0,45	0,04	0,04	0,00	0,11	15,2

Tabela 17 - Análise às componentes da variância para as 9 variáveis resultantes do método VIF.

No sentido de se esclarecer qual das quatro variáveis marcadas a escuro, se encontra mais correlacionada com as restantes, recorreu-se a uma matriz de correlações entre as mesmas.

Correlações	Poup. Famílias %	Endiv. Partic. %	(Import.+Export.)/PIB	Reservas Total
Poupanças Famílias %	1,000	-0,623	-0,896	-0,349
Endividamento Particulares %		1,000	0,469	0,135
(Importações+Exportações) / PIB			1,000	0,394
			Reservas Total	1,000

Tabela 18 - Matriz de correlações: Poupanças das Famílias, Endividamento Particulares, (Importações+Exportações) / PIB e Reservas Totais.

Concluiu-se, através da Tabela 18, que a variável “Poupanças das famílias em %” apresenta uma correlação negativa muito forte com a variável “(Importações+Exportações) / PIB” e com o “Endividamento

de particulares em %”. De facto, do ponto de vista económico, é expectável que um maior nível de endividamento dos particulares esteja, geralmente, associado a um menor nível das poupanças das famílias. Assim, tendo em conta os resultados das duas análises, optou-se por removê-la, em detrimento da variável “Endividamento de Particulares (%)”. Já as restantes variáveis têm uma magnitude de correlação menos forte entre si (fraca ou moderada).

Procedendo a uma nova iteração (sem a variável referida acima), obteve-se a um VIF de 7,98 para a variável “Endividamento de particulares em %”. No entanto, face à iteração anterior, concluiu-se que a variação no valor de VIF é reduzida, pelo que se recorreu, novamente, a uma análise das componentes da variância. Da referida matriz, podem-se visualizar duas linhas com número de condição superior a 30:

- A primeira linha tem o número de condição igual a 72,7, sugerindo a relação linear entre as variáveis e as suas respetivas componentes da variância: “Custos unitários laborais” (0,61), “Endividamento dos particulares em %” (0,56) e “Reservas Totais” (0,53);
- A penúltima linha da matriz tem o número condição igual a 40,2 e apresenta a variável “(Importações+Exportações) / PIB” com um valor de componente da variância relativamente elevado face às restantes (0,57).

Correlações	Custos unitários laborais	Endiv. Partic. %	(Import.+Export.) / PIB	Reservas Total
Custos unitários laborais	1,000	-0,618	-0,669	-0,608
Endividamento Particulares %		1,000	0,469	0,135
(Importações+Exportações)/ PIB			1,000	0,394
			Reservas Total	1,000

Tabela 19 - Matriz de correlações: Custos unitários laborais, Endividamento Particulares, (Importações+Exportações) / PIB e Reservas Totais.

Procedendo a uma análise semelhante à anterior (Tabela 19), chegou-se à conclusão que a variável “Custos Unitários Laborais” apresenta uma forte correlação com as restantes variáveis, ao passo que “Endividamento de Particulares em %” apresenta uma correlação com as restantes negligenciável e/ou moderada, não sendo, deste modo a principal responsável por situações de multicolinearidade. Deste modo, optou-se por remover a variável “Custos Unitários Laborais” da base, restando apenas 7 variáveis.

Nesta iteração, voltou-se a recalcular a matriz VIF, registando-se um valor inferior, de 6,7 para *leverage banking*. De forma a verificar se esta ou outra variável continuava a causar problemas de multicolinearidade, voltou-se a proceder a uma análise às componentes da variância, surgindo duas linhas que necessitam de uma maior análise:

- A penúltima linha apresenta um número condição de 37,6 com a seguinte relação linear sugerida: “Endividamento de particulares em %” (0,83), “Reservas totais” (0,61) e “Leverage banking” (0,52);
- A última linha apresenta um número condição de 47,3 com a variável “(Importações+Exportações) / PIB” a apresentar a componente de variância de 0,67.

Por fim, procedeu-se à análise da matriz de correlações entre as referidas variáveis (Tabela 20), com a qual se concluiu que a variável “*Leverage banking*” deveria ser eliminada, uma vez que apresentava duas correlações fortes e uma correlação moderada com as restantes variáveis.

Correlações	Leverage Banking	Endiv. Partic. %	(Import.+Export.) / PIB	Reservas Total
Leverage Banking	1,000	0,636	0,466	0,695
Endividamento Particulares %		1,000	0,469	0,135
(Importações+Exportações) / PIB			1,000	0,394
			Reservas Total	1,000

Tabela 20 - Matriz de correlações: *Leverage banking*, Endividamento Particulares, (Importações+Exportações) / PIB e Reservas Totais.

Na última iteração concluiu-se que o maior valor da diagonal da matriz VIF nas 6 variáveis restantes é de 4,3 (variável “Euribor”). A análise das componentes da variância, permitiu ainda verificar que só existe uma linha com número condição de 42,9, com a variável de “(Importações+Exportações) / PIB” com uma componente de variância de 0,81. Deste modo, não existe informação suficiente relativamente à presença de multicolinearidade entre as restantes variáveis. A título adicional, também é de referir que são conhecidas as relações de multicolinearidade com as restantes variáveis com maior valor de componente de variância (vide Tabela 20) – moderada com o “Endividamento de particulares em %” e fraca com a variável “reservas totais” – pelo que se decidiu manter esta variável na base.

4.3.2 Método de seleção de variáveis – *stepwise, backwards, linear forward*

A base final contempla 6 variáveis independentes, uma vez que se procurou garantir, na etapa anterior, a inexistência de situações de forte multicolinearidade. Partindo desse pressuposto, é expectável que a construção do modelo, independentemente do método de seleção de variáveis, resulte na escolha das mesmas variáveis e na estimação de parâmetros de regressão similares.

Considerando um nível de significância de 5%, apresentam-se os resultados obtidos pelos três métodos de seleção de variáveis nas tabelas abaixo:

R ²	Inflação	PIB%	Euribor	Endiv. Partic. %	(Import. + Export.) / PIB	Reservas Total	Constante
0,927	0,103	0,000	0,000	0,000	0,000	0,002	0,000
0,924	-	0,000	0,000	0,000	0,000	0,001	0,000

Tabela 21 - Construção do modelo pelo método *backwards*

R ²	Inflação	PIB%	Euribor	Endiv. Partic. %	(Import. + Export.) / PIB	Reservas Total	Constante
0,758					0,000		0,000
0,851					0,000	0,000	0,000
0,886			0,000		0,000	0,000	0,000
0,898			0,000	0,005	0,000	0,000	0,000
0,924		0,000	0,000	0,000	0,000	0,001	0,000
0,927	0,103	0,000	0,000	0,000	0,000	0,002	0,000

Tabela 22 - Construção do modelo pelo método *stepwise*.

R ²	Inflação	PIB%	Euribor	Endiv. Partic. %	(Import. + Export.) / PIB	Reservas Total	Constante
0,758					0,000		0,000
0,851					0,000	0,000	0,000
0,886			0,000		0,000	0,000	0,000
0,898			0,000	0,005	0,000	0,000	0,000
0,924		0,000	0,000	0,000	0,000	0,001	0,000
0,927	0,103	0,000	0,000	0,000	0,000	0,002	0,000

Tabela 23 - Construção do modelo pelo método *forward linear*.

Sumariamente, tem-se o seguinte:

- O método *backwards* parte de um modelo com todas as variáveis, eliminando a “Inflação” na primeira iteração, pois o seu *p-value* é de 0,103, ou seja, rejeita-se para o nível de significância considerado (Tabela 21);
- O método *stepwise* vai selecionando as variáveis uma a uma, terminando quando identifica uma variável não significativa para o nível de significância definido. Este método rejeita também a variável “Inflação” (Tabela 22);
- O método *forward linear* avalia qual a variável que provoca maior aumento no R², à semelhança do método *stepwise*, até que se conclua que, na adição de uma variável ao modelo, que não é significativa. Tal como nos métodos anteriores, a “Inflação” não é significativa (Tabela 23).

Os três métodos conduziram à mesma escolha de variáveis, apresentadas na Tabela 24, juntamente com os coeficientes obtidos para o modelo de regressão:

	b ₀	PIB %	Euribor	Endiv. Partic. %	(Import. + Export.) / PIB	Reservas Total
EMQ	-15,875	-0,405	-0,619	-0,042	0,395	0,135
Desvio-Padrão	1,705	0,080	0,129	0,007	0,023	0,039
R² e SRM	92,4%	1,058				
Estatística F Graus de Liberdade	178,019	73				
Soma Regressão Soma Residual	996,993	81,767				
Estatística T	-9,309	-5,062	-4,783	-5,848	17,373	3,496
p-value	4,92x10 ⁻¹⁴	3 x10 ⁻⁶	8,75 x10 ⁻⁶	1,3 x10 ⁻⁷	1,20 x10 ⁻²⁷	0,001

Tabela 24 – Modelo obtido através dos três diferentes métodos de seleção de variáveis.

É possível visualizar pelas Tabela 21, Tabela 22 e Tabela 23 que a inflação é a variável menos significativa e, aos níveis de significância usuais (1, 5 e 10%), é retirada do modelo, pois apresenta um *p-value* de 0,103 em todos os casos. Adicionalmente, e tal como observado na subsecção 4.2, esta variável é também a menos correlacionada com a variável resposta, dentro do conjunto final das 6 variáveis em análise. Do ponto de vista económico, é expectável o reduzido contributo desta variável na explicação da variação ocorrida nos *ratings*, tendo em conta que Portugal, como membro da UE, garante, em princípio, a estabilidade dos preços. Recorde-se que, tipicamente, uma inflação elevada está associada a problemas estruturais nas finanças públicas, os quais ocorrem, sobretudo, em países em desenvolvimento. Ainda assim, poder-se-á concluir, no caso de Portugal, que o *rating* melhora quando a inflação é mais elevada,

uma vez que esse aumento está associado a períodos de expansão económica. De facto, os valores de inflação encontravam-se mais elevados no início do século, atingindo mínimos no ano de 2009 e mantendo-se relativamente baixa (próxima de 0%) desde esse período.

Por fim, observa-se que o valor do *p-value* correspondente à estatística T (*t-student*) para as variáveis selecionadas permite rejeitar a hipótese nula ($H_0: b_j=0$), i.e., as variáveis são significativas para qualquer nível usual de significância, pois os três métodos de seleção de variáveis usados têm essa finalidade. Assim, as selecionadas são: “Variação do PIB”, “Euribor”, “Endividamento de Particulares em %”, “(Importações+Exportações) / PIB” e “Reservas Totais”.

Para avaliar a qualidade do modelo, pode-se considerar os seguintes parâmetros:

- Poder preditivo: R^2 de 92,4% e um R^2 ajustado de 91,9%¹² indicam um bom poder preditivo;
- Erro padrão: S^2 de 1,058 representa uma margem de erro relativamente reduzido.

Analisando quantitativamente o modelo, existem alguns aspetos que valem a pena realçar:

- O termo constante apresenta um valor negativo (-15,875). Em princípio, este seria o valor mínimo numérico que o *rating* de Portugal poderia estar sujeito, no entanto, sendo negativo nada se pode concluir. O modelo linear ajusta-se bem no domínio de variação das variáveis que ocorreu no período em análise. Para um período mais longo, com um intervalo de variação maior para as diferentes variáveis, é possível que o modelo já não fosse linear, principalmente quando se aproxima de casos extremos (por exemplo, todas as variáveis serem próximas de 0). Com efeito, o caso em que o valor de todas as variáveis é 0 é meramente teórico, pois não existem países com valor nulo de PIB. Deste modo, o termo constante tende a bonificar o *rating*;
- O contributo positivo na melhoria do *rating* das variáveis “variação do PIB”, “Euribor” e “Endividamento de particulares em %”. Efetuando uma análise *ceteris paribus*:
 - Uma variação positiva de 1% no PIB resulta numa melhoria de 0,405 no *rating*. Por outras palavras, uma variação positiva de 1%, sendo um indicador de expansão económica, implica que pelo menos uma das agências deverá melhorar o *rating* em 1 *notch*;
 - Uma variação positiva de 1 p.p. na Euribor resulta numa melhoria de 0,619 no *rating*. Tal como a variável anterior, uma variação positiva implica que pelo menos uma agência deverá melhorar o *rating* em 1 *notch*. De notar que esta relação com o *rating* se encontra condicionada ao histórico subjacente, dado que o valor da taxa Euribor encontrava-se mais elevado nos momentos de crescimento económico, onde se verificaram os melhores *ratings*. De facto, períodos de expansão económica estão associados a taxas de juro mais elevadas, apenas na presença de condições de sustentabilidade no nível de endividamento público e privado. Deste modo, face ao elevado nível de endividamento

¹² $R^2_{ajustado} = 1 - \left(\frac{n-1}{n-(p+1)} \right) (1 - R^2_p)$, onde n é referente a observações e p a variáveis.

público e privado atual, o aumento da taxa Euribor poderá ter efeitos contrários aos esperados. Uma vez que por um lado, este aumento levará ao acréscimo dos encargos com os juros da dívida atual e, por outro, a maiores dificuldades no acesso ao crédito, dadas as fragilidades do setor financeiro. Estas limitações poderão levar à contração no consumo interno, e, por conseguinte, no PIB do país;

- Uma variação positiva de 1 p.p. no endividamento dos particulares resulta numa melhoria do *rating* em 0,042. A título de exemplo, uma variação de 24 p.p. levaria a uma melhoria do *rating* em 1 notação. Tal como a variável anterior, o efeito desta variável no *rating* reflete um período histórico em que o crescimento económico foi sustentado pela procura interna, ancorada, em grande medida, pelo recurso ao crédito. No entanto, face ao elevado nível de endividamento atual e ao seu efeito limitador na capacidade de consumo dos agentes privados, é expectável que o verdadeiro impacto no *rating* possa ser negativo.
- A variação positiva nas variáveis “(Importação+Exportação) / PIB” e “Reservas Totais” contribui negativamente para o *rating* (degradação da PD). Efetuando uma análise *ceteris paribus*:
 - Uma variação positiva de 1mM€ de reservas totais resulta numa deterioração do *rating* em 0,135 *notches*. Esta relação sugere a intervenção do Banco de Portugal em períodos de crise, procurando aumentar o nível de reservas, no sentido de salvaguardar o país do risco de incumprir nas suas obrigações em moeda estrangeira;
 - Uma variação positiva de 1 p.p. no rácio “(Importação+Exportação) / PIB” reflete uma deterioração do *rating* em 0,395 *notches*. Deste modo, o modelo indica que é provável que uma agência reveja o seu *rating* em baixa com o aumento deste rácio. Tal como referido anteriormente, sendo Portugal uma economia desenvolvida, a avaliação do comportamento desta variável deve incorporar uma análise mais detalhada sobre as variações observadas nas componentes do PIB. Assim, seguidamente apresenta-se uma análise deste coeficiente, onde se procurou dividi-lo em três parcelas.

O denominador é dado pelas variáveis constituintes do PIB:

$$PIB = C + I + G + X - M$$

em que

- C= Consumo privado;
- I= Total de investimentos realizados;
- G= Gastos governamentais (ou Consumo Público);
- X= Total Exportações;
- M= Total Importações.

Pelo que o rácio (Importações + Exportações) / PIB poderá ser reescrito da seguinte forma:

$$\frac{X + M}{PIB} = \frac{X + M}{C + I + G + X - M}$$

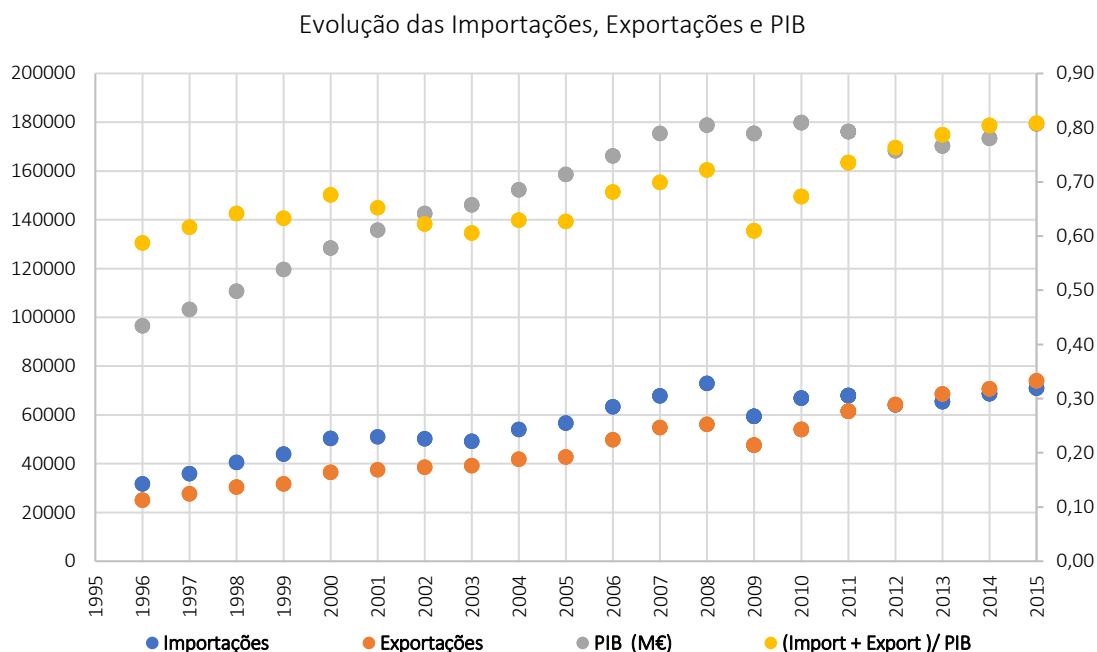


Figura 7 - Evolução das variáveis: Importações, Exportações e PIB (M€)

Como se pode analisar pela expressão que determina o rácio suprarreferido e pela Figura 7, durante o período anterior à crise de 2008 o crescimento das importações e das exportações acompanhou o crescimento do PIB, embora apresentando sempre um défice comercial. Entre 2001 e 2004, as transações económicas diminuem, embora o PIB se mantivesse em crescimento. No entanto, desde essa data até ao início da crise verificou-se um aumento significativo das transações económicas relativamente ao PIB. Entre os anos de 2008 a 2010, assiste-se a uma contração significativa das transações económicas, levando este rácio a atingir os valores mais baixos. No período pós-crise, verifica-se a inversão entre as importações e as exportações: o volume de exportações, a partir de 2012, torna-se superior ao de importações. Este aumento nas exportações justifica o acréscimo do numerador, apesar da evolução constante das importações. Contrariamente, no mesmo período, o PIB mantém-se estagnado. Este comportamento indica que, no período da crise financeira, as empresas nacionais tiveram necessidade de exportar os seus produtos, devido à quebra no consumo interno (consumo público e gastos do estado). Embora o aumento das exportações favoreça o crescimento da economia, tendo em conta que o saldo da balança comercial está próximo de zero, o seu contributo para o crescimento do PIB é praticamente nulo. Assim, dado que o valor do PIB se manteve constante, conclui-se que as restantes componentes do PIB decresceram (Consumo e Investimento). Deste modo, o aumento deste rácio poderá ser um indicador de

debilidade na economia interna, que não é compensado pelo aumento das exportações, levando as agências a concluírem a existência de maiores riscos de incumprimento e, portanto, a penalizarem o *rating*.

4.3.3 Avaliação da qualidade do modelo

No sentido de verificar a qualidade do modelo, isto é, o bom ajustamento do modelo aos dados em observação, realizou-se o teste F a partir da tabela ANOVA, analisou-se a relação entre cada uma das variáveis independentes e o *rating* (para verificar eventuais necessidades de transformação das variáveis) e procedeu-se a uma análise à normalidade dos resíduos. Adicionalmente, para verificar o grau de precisão do modelo construiu-se um intervalo de confiança/predição, o qual não deverá ser excessivamente grande.

4.3.3.1 Teste F e tabela ANOVA

Como referido, após a construção do modelo, e no sentido de validar a qualidade do mesmo, é importante testar a hipótese de que todos os coeficientes, com exceção do termo constante, são nulos ou não, ou seja, se são relevantes ou não para explicar as variações observadas na variável dependente (*rating*). Por outras palavras, teste-se as seguintes hipóteses:

$$H_0: b_1 = b_2 \dots = b_5 = 0 \text{ contra } H_1: b_j \neq 0, \text{ para algum } j=1, \dots, 5.$$

Fonte	Soma de Quadrados	Graus de Liberdade	Média de Quadrados	Estatística F
Regressão	996,99	5	199,40	178,02
Residual	81,77	73	1,12	
Total	1078,76	78		

Tabela 25 - Tabela ANOVA para o modelo sugerido.

Conclui-se, a partir do valor da estatística F, que o valor do *p-value* é de $1,9 \times 10^{-39}$, o que permite rejeitar a hipótese nula com toda a segurança. Ou seja, as variáveis incluídas no modelo contribuem significativamente para a explicação da variação do *rating*.

4.3.3.2 Análise da relação entre as variáveis independentes e o *rating*

Uma das formas de melhorar o modelo é perceber a forma como cada variável independente se relaciona com a variável resposta e, no caso de se identificarem relações não lineares, proceder à transformação dessas variáveis, no sentido de melhorar a correlação com a variável resposta. Pela análise das cinco variáveis do modelo, verifica-se que não existe nenhuma variável que apresente um comportamento aproximadamente quadrático, logarítmico ou exponencial, pelo que não se efetuou quaisquer transformações às variáveis.

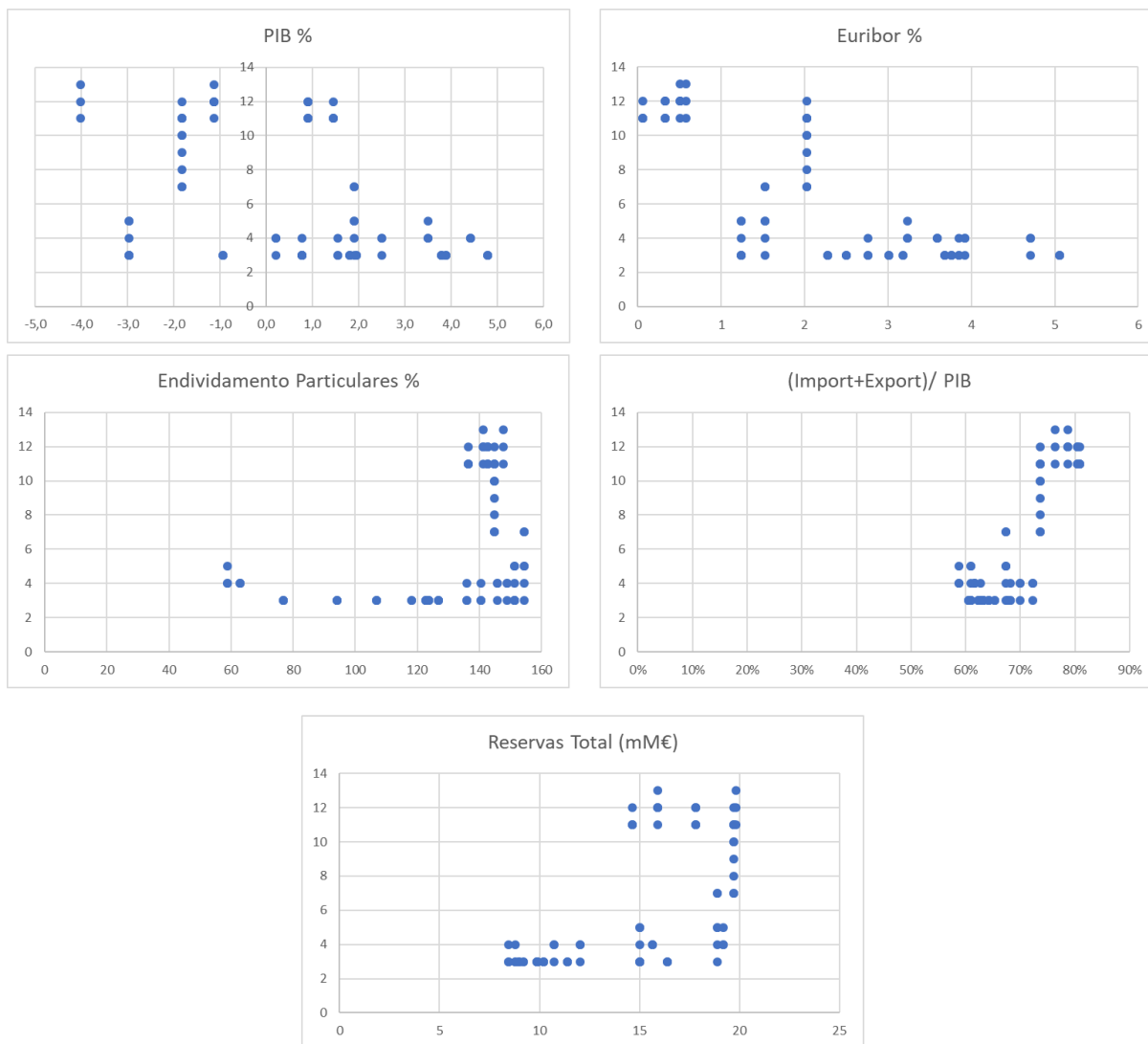


Figura 8 - Variáveis independentes contra o *rating*.

4.3.3.3 Análise dos resíduos

Num modelo de regressão linear múltipla espera-se que os resíduos associados às diferentes variáveis independentes apresentem uma dispersão sem nenhum padrão em particular, isto é, que indiquem um determinado tipo de comportamento. Ao analisar a dispersão dos resíduos das cinco variáveis, apresentados na Figura 9, observa-se que não existe um comportamento específico em qualquer uma delas.

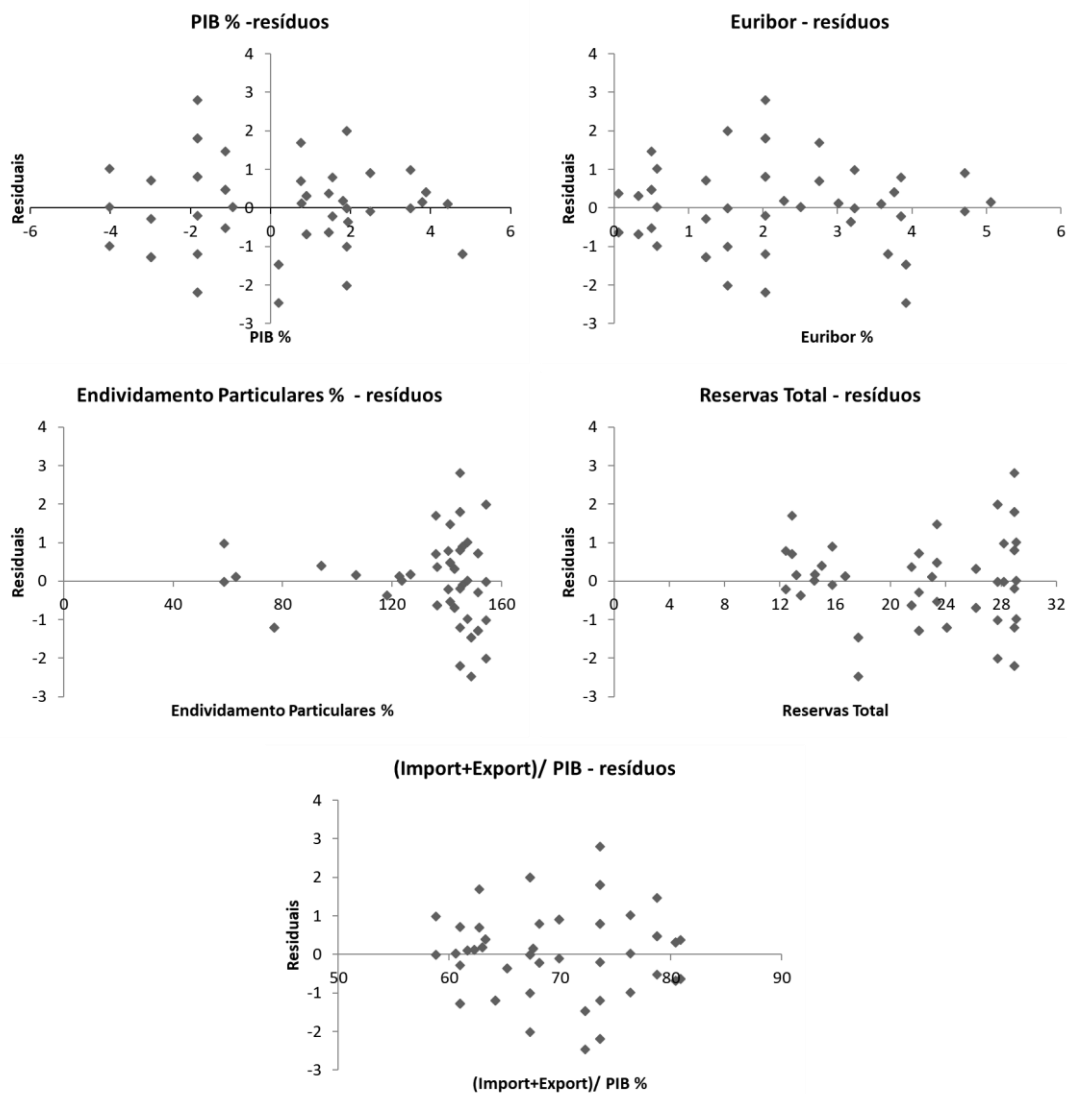


Figura 9 - Representação dos resíduos vs variáveis independentes.

Uma outra análise à qualidade do ajustamento do modelo consiste em verificar se os resíduos seguem uma distribuição normal, $N(0, \sigma^2)$. Esta análise pode passar pela realização do teste do Qui-Quadrado, após a aplicação da regra de Sturges para dividir em classes a amostra dos resíduos. A Tabela 26 apresenta as principais características amostrais dos resíduos.

Mínimo	Máximo	Média	σ^2	Amplitude	Regra Sturges
-2,47	2,80	0,00	1,017	5,27	0,753

Tabela 26 - Principais características amostrais dos resíduos e valores necessários para construção do histograma

A Figura 10 mostra a cinzento o histograma dos resíduos e a preto a curva normal que se lhes ajusta, tendo em consideração as características amostrais presentes na Tabela 26.

De acordo com a regra de Sturges, o histograma dos resíduos (ver Figura 10) deve conter 7 classes. Assim, por questões de arredondamento e de forma a respeitar este número de classes, fixou-se a amplitude das classes em 0,8.

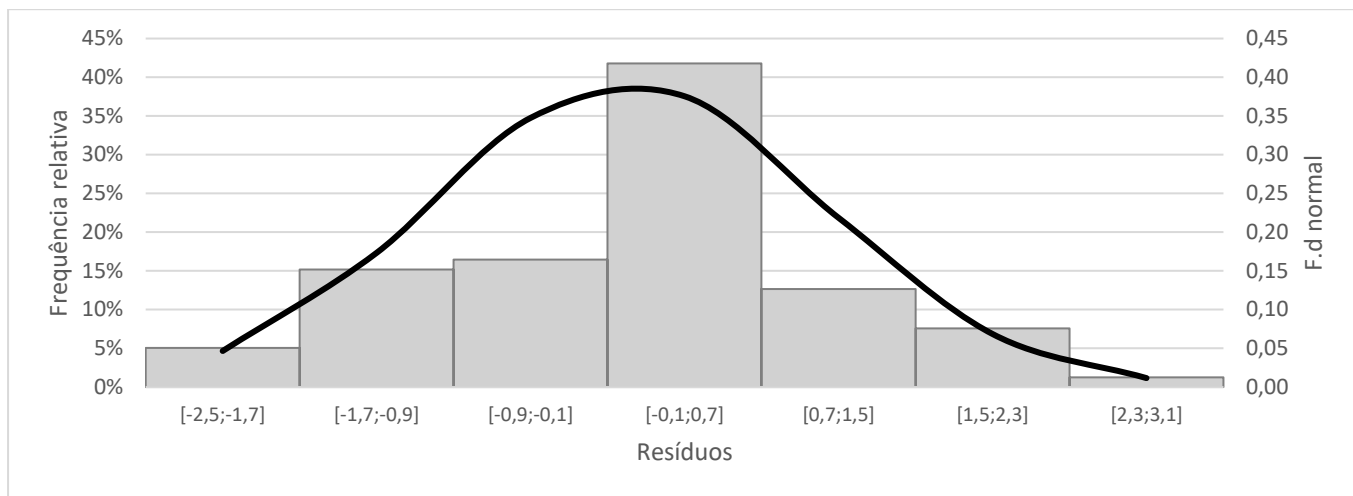


Figura 10 - Histograma dos resíduos.

Para testar a normalidade dos resíduos, procedeu-se ao teste do Qui-Quadrado (Tabela 27). Arredondando o valor mínimo desta amostra a uma casa decimal, tomou-se esse valor como o extremo esquerdo do primeiro intervalo. Respeitando esta condição e a amplitude de 0,8, construiu-se o teste do Qui-Quadrado a partir dos valores apresentados na Tabela 27.

Classes	Frequência esperada	Frequência observada	$(N_j - np_j)^2/np_j$
[-2,5;-1,7]	3,74	4	0,02
[-1,7;-0,9]	11,12	12	0,07
[-0,9;-0,1]	21,54	13	3,39
[-0,1;0,7]	23,18	33	4,16
[0,7;1,5]	13,87	10	1,08
[1,5;2,3]	4,61	6	0,42
[2,3;3,1]	0,85	1	0,03
Graus de liberdade	5	Total Q₀	9,16
		p-value	0,103

Tabela 27 - Resultados do teste do Qui-Quadrado.

A estatística de teste, sob a validade de H_0 , tem distribuição Qui-quadrado com 5 graus de liberdade (i.e., 7 classes - 1 parâmetro amostral - 1), obtendo-se assim um *p-value* de 0,103. Deste modo, não existe evidência para rejeitar a hipótese nula. Considerando os níveis de significância usuais e a amostra referida, constata-se que os resíduos seguem uma distribuição normal.

4.3.3.4 Intervalo de confiança e análise ao conservadorismo do modelo

Por último, procedeu-se à análise da precisão do modelo através da construção de um intervalo de previsão para os diferentes anos em estudo. Para a construção do gráfico que se apresenta na Figura 11, foram usados os resultados apresentados na subsecção 3.1.1.3 e ainda um nível de significância de 10%.

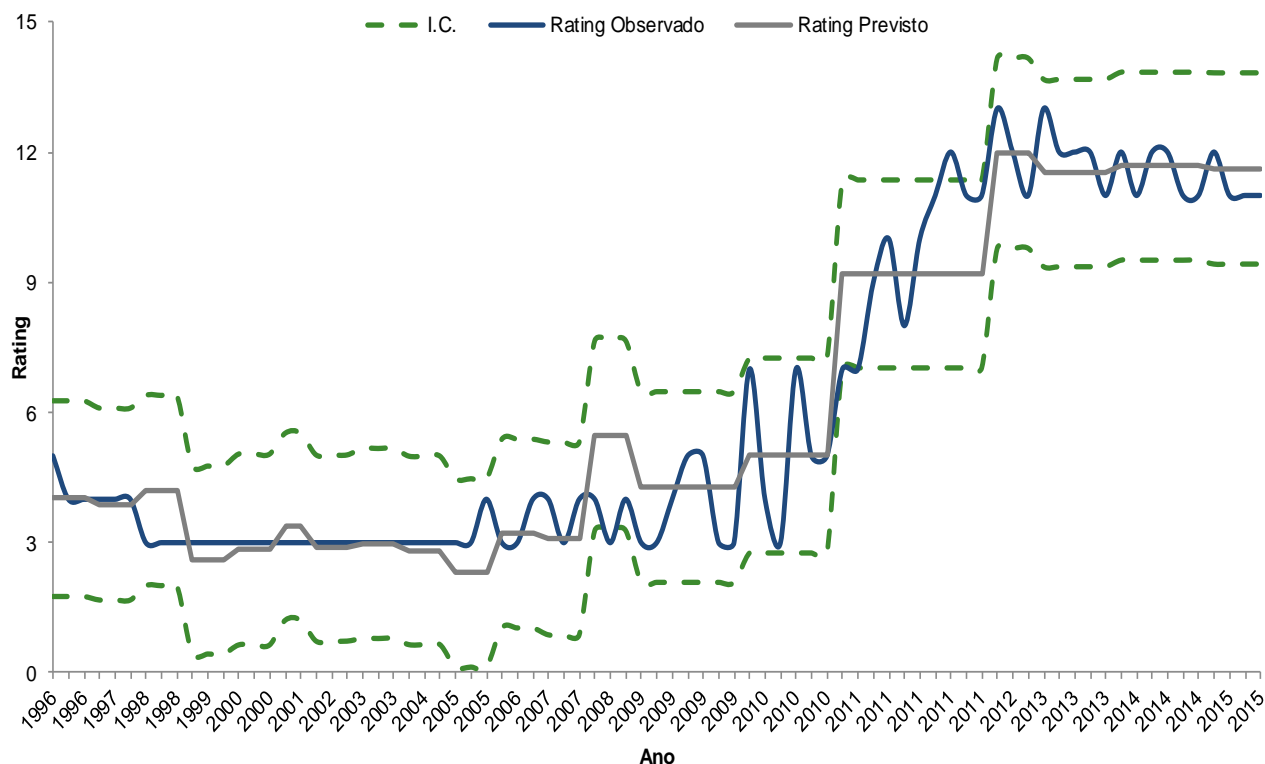


Figura 11 - *Rating* observado, *rating* previsto e intervalo de previsão.

Observa-se que o modelo apresenta um bom nível de precisão, dado que com um intervalo de confiança (I.C.) de 90%, garante-se a cobertura de quase todos os *ratings* efetivamente observados. O intervalo estima uma distância máxima ao *rating* observado de 2 *notches* em sentidos opostos, isto é, apresenta uma amplitude aproximada de 4 *notches*. Ao analisar o comportamento do modelo ao longo do período de observação, verifica-se que este previa um *downgrade* no *rating* nos primeiros sinais da crise de 2008, o qual não se observou. No ano de 2011, o modelo prevê ligeiramente abaixo do *rating* observado (onde se encontram os pontos superiores ao I.C.), sugerindo que outras fontes de informação qualitativas foram relevantes para a atribuição de *rating* (por exemplo, este período coincide com a demissão do governo e a consequente instabilidade política). A partir de 2012, manteve-se relativamente constante e próximo do *rating* efetivamente atribuído pelas agências de notação.

Em suma, os resultados do modelo evidenciam que a variação no *rating* pode ser explicada, em larga medida, pelas variações observadas nas variáveis macroeconómicas incluídas. No entanto, no período de crise económica a variabilidade observada no *rating* atribuído pelas agências indica que outras fontes de informação qualitativas foram utilizadas.

4.4 Análise aos fatores determinantes do risco soberano nos dois períodos de análise

Esta análise visa avaliar o poder preditivo de cada fator e a sua significância no período em questão, no sentido de se concluir quais as variáveis, nos dois períodos distintos, se mantêm significativas. Salienta-se que esta análise apresenta menor significância estatística devido à redução da dimensão da amostra.

Primeiramente, analisou-se a correlação entre as três agências nos dois períodos, tal como mostra a Tabela 28.

Período	Moody's vs Fitch	Moody's vs S&P	Fitch vs S&P
1996-2008	60,1%	43,3%	72,0%
2009-2015	96,4%	94,5%	95,5%

Tabela 28 – Correlações entre as diferentes agências para dois períodos de análise.

Conclui-se que a correlação entre os três pares de agências aumentou no período de crise, essencialmente, onde se registou maior variação no *rating* numérico. Estatisticamente, esta situação deve-se ao facto da diferença de um ou dois *notches* ter menos peso no período pós-crise, dada a maior amplitude entre os *ratings*, relativamente ao período de pré-crise. Adicionalmente, conclui-se que a correlação entre a Moody's e a S&P é a mais baixa nos dois períodos em análise. Já a Moody's e a Fitch têm um comportamento muito semelhante no segundo período em análise. Estas diferenças no valor da correlação também se devem à pequena diferença entre o total de observações nas duas amostras. A Tabela 29 obtém-se por aplicação da metodologia referida na secção 4.1.

Período	Pré-crise				Pós/durante a crise				
	Agências	Contagem	Soma	Média	Variância	Contagem	Soma	Média	Variância
Moody's		13	42	3,23	0,36	17	153	9,0	16,8
Fitch		13	43	3,31	0,23	17	147	8,6	10,7
S&P		13	45	3,46	0,27	17	164	9,6	7,4

Tabela 29 – Características dos *ratings* atribuídos pelas 3 CRA nos dois períodos de análise.

Tal como na amostra global, o resultado da tabela ANOVA para ambos os períodos permite concluir que as publicações de *rating* pertencem todas à mesma população, uma vez que não há evidência para rejeitar a hipótese nula dentro dos níveis de significância usuais, como se pode verificar na Tabela 30.

Período	Estatística F	p-value
1996-2008	0,63	0,54
2009-2015	0,38	0,69

Tabela 30 – Resumo relativo à estatística F e respetivo *p-value* para ambos os períodos em análise.

Na análise à significância das variáveis independentes, aplicou-se o método dos Mínimos Quadrados para cada período em análise, tendo em conta dois aspetos:

- A significância estatística desta avaliação não é robusta, uma vez que cada amostra contém menos de 20 *ratings* observados;

- A volatilidade dos *ratings* em tempos de crise é bastante superior, assim como o comportamento das variáveis independentes, pelo que se espera um melhor ajustamento no período de crise, relativamente ao período pré-crise, onde a variação do *rating* é reduzida.

Nas tabelas abaixo (Tabela 31 e Tabela 32) apresentam-se os resultados do ajustamento de um modelo de regressão, incluindo os EMQ dos coeficientes e respetivos desvios-padrões e ainda a estatística do teste T para avaliar a significância de cada variável.

1996-2008	b ₀	PIB%	Euribor	Endividamento Particulares	(Import+Export)/ PIB	Reservas Totais
EMQ	2,31	-0,06	0,26	0,01	-3,23	0,14
Desvio-Padrão	2,21	0,10	0,20	0,01	5,74	0,04
R² e S	0,37	0,46				
Estatística F e G.L.	3,76	32,00				
Somas Quadrados	3,90	6,65				
Estatística T	1,05	-0,55	1,28	0,66	-0,56	3,15
p-value	0,30	0,58	0,21	0,52	0,58	0,00

Tabela 31 – Ajuste pelo mínimo dos métodos quadrados e significância estatística de cada variável (período 1996-2008)

Na Tabela 31, tal como referido, observa-se um ajustamento no período pré-crise fraco, com um R² de 37%, embora com um erro-padrão relativamente reduzido, devido à limitada variação no *rating* durante este período. Adicionalmente, apenas a variável “reservas totais” é significativa, com um EMQ positivo. No entanto, esta relação deve-se ao facto da maior variação no *rating*, isto é, da sua melhoria entre 1996 e 2005 ter coincidido com a mudança, em 2001, para a moeda única, da qual resultou a diminuição significativa do montante de reservas. Assim, face a este acontecimento, um maior montante de Reservas Totais, implica um *rating* mais elevado, ou seja, pior.

Já na Tabela 32, apresenta-se o resultado para o período de crise, onde se verifica o aumento no R² para 89%, com um desvio-padrão dos termos de erro de 1,23, o que se deve ao aumento da volatilidade nas observações de *rating*. Neste período verifica-se que apenas as seguintes variáveis se mantêm significativas: a “variação do PIB” e a “(Importações+Exportações)/PIB”. Ambas refletem um EMQ com o mesmo comportamento comparativamente ao modelo final, ou seja, uma variação positiva no PIB melhora o *rating* e um aumento das transações comerciais face ao PIB contribui para uma degradação do *rating*.

2009-2015	b ₀	PIB%	Euribor	Endividamento Particulares	(Import+Export) / PIB	Reservas Totais
EMQ	-65,71	-0,51	0,84	0,20	68,20	-0,33
Desvio-Padrão	42,35	0,14	1,09	0,22	22,97	0,47
R² e SRM	0,89	1,23				
Estatística F e GL	54,83	35,00				
Soma Regressão	416,01	53,11				
Estatística T	-1,55	-3,56	0,77	0,90	2,97	-0,70
p-value	0,13	0,00	0,45	0,38	0,01	0,49

Tabela 32 – Ajuste pelo mínimo dos métodos quadrados e significância estatística de cada variável (período 2009-2015)

4.5 Impacto do *rating* e dos fatores determinantes do risco português na taxa de juro da dívida portuguesa e no prémio de mercado

O âmbito desta análise engloba três vertentes distintas, tal como descrito anteriormente na Figura 5:

- A comparação entre o juro da dívida soberana portuguesa (a 10 anos) e a variação do juro face ao mês n-1 com o *rating*. Também se efetuou esta comparação apenas considerando os momentos em que foram atribuídos uma nova notação de *rating*;
- A comparação entre o prémio de mercado e a variação deste face ao mês n-1 com o *rating*, considerando também, uma amostra com os registos do momento da publicação do *rating*;
- A avaliação da influência dos fatores determinantes do *rating* no prémio de risco de mercado.

Analisando a primeira vertente, conclui-se, a partir da Tabela 33, que não será possível ajustar um modelo de regressão linear com as variáveis juro e *rating*:

Método	Observações	R ² (%)	S ²	p-value (<i>rating</i>)
Rating vs Juro	275	2,6	3,66	0,008
Rating vs variação juro	274	0,2	0,06	0,494
Publicação <i>rating</i> vs juro (apenas no mês de publicação)	45	9,9	2,82	0,035

Tabela 33 – Análise *rating* vs juro da dívida soberana portuguesa.

Sumariamente, verifica-se que a variável *rating* não é significativa, para um nível significância de 5 e 10%, para explicar a variação na taxa de juro observada, caso se considere a amostra com o histórico mensal durante os 21 anos. No entanto, se se considerar o valor da taxa de juro, o *rating* passa a ser significativo, embora o valor de R² seja reduzido. Se se optasse por utilizar uma amostra apenas com os registos na data de emissão do *rating*, o valor de R² aumentaria para 9,9%, ainda assim, bastante abaixo do que é considerado aceitável numa regressão linear.

Estes resultados sugerem que, adicionalmente ao *rating*, outros fatores qualitativos e quantitativos foram utilizados pelo mercado, como por exemplo, o risco associado ao facto de Portugal ser um país PIIGS e a instabilidade política e financeira vivida no ano de 2011. Adicionalmente, os efeitos do *quantitative easing* acionado pelo BCE em 2015, permitiram um alívio da taxa de juro da dívida soberana portuguesa num período em que o *rating* se mantinha abaixo de “*Investment Grade*”.

Em suma, os resultados sugerem que o *rating* terá alguma influência no juro da dívida, se considerarmos o mês em que é publicada uma nova notação. Contudo, no caso português, a sua influência é mais reduzida do que seria expectável.

Na Figura 12 procurou-se analisar o comportamento da taxa de juro, através da comparação entre a sua variação mensal ao longo do período de 21 anos e o *rating* observado. Por observação do gráfico, poder-se-á concluir que nos anos 90, a variação do juro foi, em média, negativa, com a maioria dos pontos verdes abaixo do eixo das abcissas. A partir de 2005, inverte-se esta tendência, com o aumento paulatino da taxa de juro. Já em 2011, no mesmo momento em que o *rating* começa a subir exponencialmente,

regista-se uma volatilidade acentuada nas diferenças entre os juros mensais. Isto deve-se, essencialmente, ao aumento do comportamento especulativo que tipicamente se gera quando o *rating* se encontra abaixo do limiar para uma notação *investment grade* [32]. De notar, ainda, o facto de apesar da crise financeira se ter iniciado em 2008, apenas a partir de 2010 se verifica a subida da taxa de juro, em linha com a primeira deterioração significativa do *rating* efetuada pela S&P (de A+ em 2009 para A- em 2010).

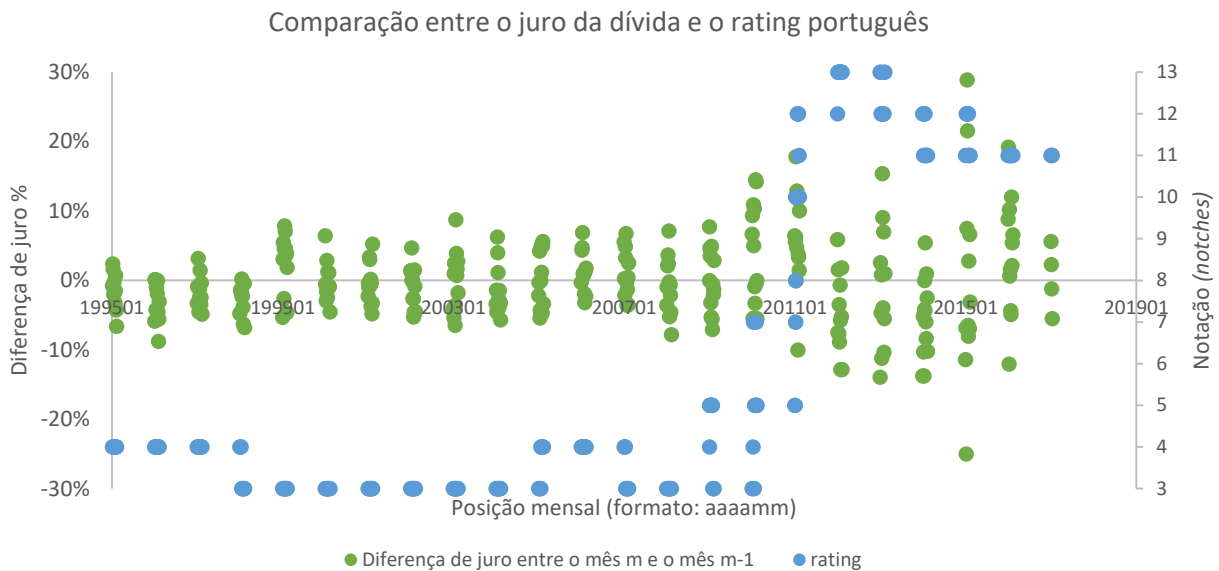


Figura 12 – Diferença dos juros entre o mês m e o mês m-1 e o *rating* (análise: 01/1995-04/2017).¹³

Após esta análise, procurou-se retirar os efeitos da variação originada por fatores de mercado. Assim, efetuou-se uma análise semelhante, mas considerando o prémio de mercado da taxa de juro da dívida, dado pela diferença entre o juro da dívida portuguesa e o juro da dívida alemã. A consideração desta última para cálculo do prémio de risco deve-se ao facto da Alemanha se tratar de uma das entidades soberanas europeias com menor risco de incumprimento, isto é, com um *rating* de AAA desde 1983, e nesse sentido, mais próxima de um investimento com taxa de juro sem risco.

Método	Observações	R ² (%)	S ²	p-value (rating)
Rating vs prémio de risco do juro	275	54,0	2,51	6,5x10 ⁻⁴⁸
Publicação rating vs prémio de risco do juro (apenas no mês de publicação)	45	40,1	2,47	3,1x10 ⁻⁶

Tabela 34 – Análise *rating* vs prémio de mercado.

A utilização do prémio de risco do juro da dívida portuguesa permite melhorar significativamente os resultados da regressão linear simples nas duas amostras. Adicionalmente, é de notar um valor do erro-padrão de cerca de 2,5, relativamente a uma variável com amplitude de 12%. Em termos de correlação com o *rating*, tem-se que nas duas amostras as correlações são fortes (basta tomar o valor de R² e aplicar a raiz quadrada para observar que esses valores são de 0,73 e 0,63), sendo que, contrariamente à análise anterior que utiliza a taxa de juro, os resultados melhoram quando se utiliza o histórico mensal dos últimos

¹³ Data de referência: junho de 2017.

21 anos. Em suma, as correlações são significativamente mais elevadas com a supressão do efeito da oferta e da procura, dado que o prémio de mercado é mais baixo no início do período, onde as taxas de juro se igualavam às taxas do período pós-crise, embora o *rating* numérico fosse muito inferior (melhor), tal como se pode verificar na Figura 13.

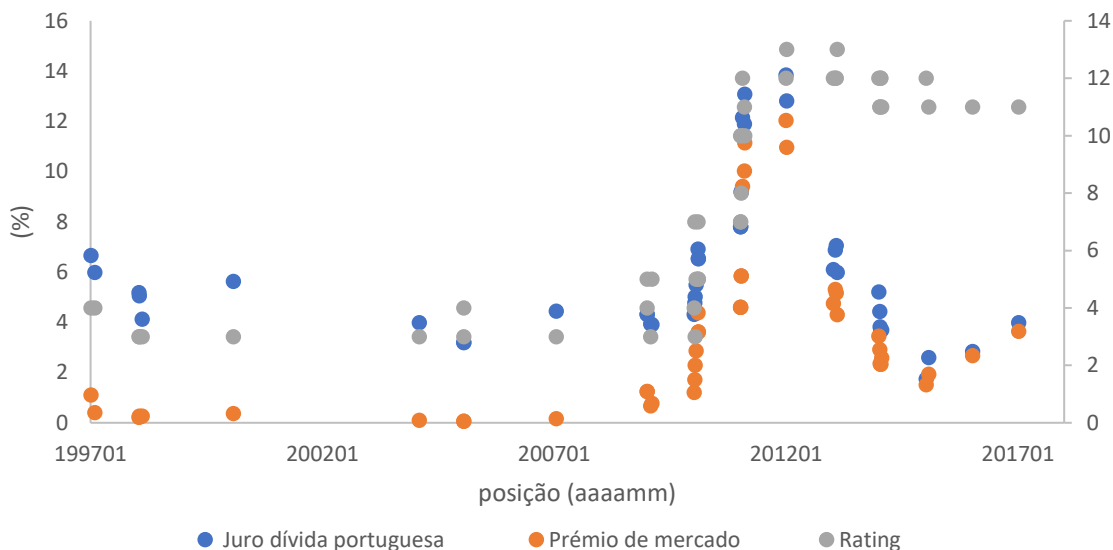


Figura 13 - Juro da dívida portuguesa, prémio de mercado e o *rating* para o período de análise (45 observações)

Por fim, avaliou-se a significância dos fatores determinantes do *rating* no prémio de mercado. Na Tabela 35, apresentam-se o R^2 , o S^2 e ainda o *p-value* do teste ANOVA. Esta análise foi efetuada com base na amostra que contém apenas os registos à data de publicação do *rating*. Optou-se por testar o impacto do *rating* e dos fatores determinantes nas variáveis taxa de juro e prémio de mercado considerando esta amostra, dado que permitiu obter melhores resultados na primeira análise (correlação mais forte com o *rating*).

Método	Observações	R^2 (%)	S^2	<i>p-value</i> (ANOVA)
Variáveis + <i>rating</i> vs juro	43	79,9	1,43	$3,5 \times 10^{-11}$
Variáveis + <i>rating</i> vs prémio de mercado		81,4	1,50	$8,6 \times 10^{-12}$
Variáveis vs juro		68,1	1,77	$2,6 \times 10^{-8}$
Variáveis vs prémio de mercado		70,7	1,86	$5,4 \times 10^{-9}$

Tabela 35 – Análise variáveis vs juro ou prémio de mercado.

Em todas as análises observa-se um elevado valor de R^2 e um valor do desvio-padrão aceitável. Adicionalmente, verifica-se que, quer na variável taxa de juro quer no prémio de mercado, o *rating* permite aumentar o R^2 dos modelos em, aproximadamente, 10pp, ainda assim, os fatores determinantes explicam, a maior parte, cerca de 70%, da variação do prémio de mercado ou da taxa de juro. Contudo, dever-se-á avaliar a significância de cada fator, uma vez que um valor de R^2 muito elevado não garante, por si só, a sua robustez.

A análise será efetuada par a par, onde o *driver* será a inclusão ou não do *rating* na explicação do juro ou do prémio de mercado. O primeiro conjunto contempla a inclusão do *rating* como variável explicativa:

	Juro da dívida portuguesa				Prémio de mercado			
	EMQ	Desvio-padrão	Estatística T	p-value	EMQ	Desvio-padrão	Estatística T	p-value
Termo Constante	4,753	4,855	0,979	0,334	-2,004	5,111	-0,392	0,697
Rating	0,828	0,180	4,607	0,000	0,864	0,189	4,567	0,000
PIB %	-0,505	0,140	-3,601	0,001	-0,483	0,148	-3,274	0,002
Euribor	1,967	0,304	6,467	0,000	1,411	0,320	4,407	0,000
Endividamento Particulares %	0,005	0,014	0,362	0,719	0,016	0,015	1,034	0,308
(Import+Export)/ PIB	-0,223	0,081	-2,745	0,009	-0,165	0,085	-1,928	0,062
Reservas Total	0,426	0,079	5,391	0,000	0,372	0,083	4,467	0,000

Tabela 36 – Análise das variáveis para cada uma das taxas em estudo (com *rating*)

Nas duas variáveis taxa de juro e prémio de mercado, observa-se que o endividamento dos particulares é a variável não significativa, ou seja, é a menos explicativa da variação na taxa de juro ou no prémio de mercado. Adicionalmente, o rácio de Importações e Exportações sobre o PIB também deixa de ser significativo no prémio de mercado, para níveis de significância de 1 e 5%. Deste modo, conclui-se que os fatores variação do PIB, Euribor, Reservas totais e *rating* permitem explicar aproximadamente 80% da variação observada no juro da dívida portuguesa ou no prémio de mercado. De notar que, embora a utilização do prémio de mercado apresente resultados bastante semelhantes relativamente à modelação direta da taxa de juro, a sua utilização possibilita um aumento no R² (+1,5pp) com menos uma variável (o rácio de Importações e Exportações sobre o PIB).

	Juro da dívida portuguesa				Prémio de mercado			
	EMQ	Desvio-padrão	Estatística T	p-value	EMQ	Desvio-padrão	Estatística T	p-value
Termo Constante	-7,854	4,988	-1,574	0,124	-15,158	5,234	-2,896	0,006
PIB %	-0,816	0,153	-5,337	0,000	-0,808	0,160	-5,036	0,000
Euribor	1,339	0,338	3,960	0,000	0,755	0,355	2,130	0,040
Endividamento Particulares %	-0,021	0,016	-1,291	0,205	-0,012	0,017	-0,691	0,494
(Import+Export)/ PIB	0,100	0,051	1,948	0,059	0,172	0,054	3,201	0,003
Reservas Total	0,474	0,097	4,866	0,000	0,422	0,102	4,125	0,000

Tabela 37 – Análise das variáveis para cada uma das taxas em estudo (sem *rating*).

A Tabela 37 permite verificar a capacidade explicativa dos fatores determinantes, excluindo-se deste modo o *rating* como variável explicativa. A exclusão do *rating* resulta numa diminuição do R² nas duas taxas. Ainda assim, a variável “Endividamento dos particulares em %” mantém-se não significativa na explicação da taxa de juro e do prémio de mercado. Assim, revela-se importante avaliar o impacto da exclusão desta variável no valor do R².

Neste contexto, a análise seguinte inclui todas as variáveis, à exceção da referida anteriormente, de forma a explicar o prémio de mercado. A Tabela 38 apresenta os resultados obtidos utilizando o método de seleção de variáveis *backwards*.

	Iteração #1				Iteração #2			
	EMQ	Desvio-padrão	Estatística T	p-value	EMQ	Desvio-padrão	Estatística T	p-value
Termo Constante	-15,96	4,179	-3,820	0,000	-11,231	1,291	-8,701	0,000
Rating	0,34	0,165	2,087	0,040	0,511	0,088	5,796	0,000
PIB %	-0,72	0,105	-6,844	0,000	-0,699	0,104	-6,721	0,000
Euribor	1,58	0,239	6,617	0,000	1,612	0,238	6,778	0,000
(Import+Export)/ PIB	0,08	0,066	1,190	0,238	-	-	-	-
Reservas Total	0,54	0,071	7,627	0,000	0,514	0,067	7,692	0,000

Tabela 38 – Análise das variáveis para cada o prémio de mercado (com *rating* e sem endividamento de particulares).

Observa-se que, na primeira iteração (sem o “Endividamento de particulares em %”), que as variáveis *rating* e “(Importações+Exportações) / PIB” registam um *p-value* superior a 0,01, embora para esta última seja ainda superior a 0,10. Assim, rejeitando-se a significância deste rácio, procedeu-se a uma segunda iteração sem o mesmo. Neste modelo todas as variáveis são estatisticamente significativas. Como se pode observar na Tabela 39, a remoção desta variável provoca uma ligeira redução no R² (5 pb) e um aumento pouco significativo do desvio-padrão (0,005), concluindo-se que a mesma não contribuía para a qualidade da regressão linear.

Em suma, aliado ao *rating*, concluiu-se que a variação do PIB, a Euribor e as Reservas Totais explicam cerca de 76% da variação do prémio de mercado. Com efeito, a variação do PIB é uma das variáveis prementes na explicação da variação do *spread* da dívida, segundo a bibliografia. [33]

Iteração	R ²	Erro-padrão
1	76,2%	1,811
2	75,7%	1,816

Tabela 39 – Qualidade da regressão linear através da seleção de variáveis pelo método *backwards*.

5. Conclusões

O risco soberano da dívida portuguesa é um tema premente na conjuntura económica atual. De facto, o risco associado à dívida portuguesa mantém-se, desde 2011, num nível de risco superior em aproximadamente 6 a 8 *notches* relativamente ao nível de risco do início da primeira década do século XXI. Aliado a uma conjuntura económica desfavorável e um fraco desempenho macroeconómico, outros fatores qualitativos levaram a que os investidores reduzissem a confiança no investimento em títulos de dívida pública portuguesa. Com efeito, a atribuição de *ratings* pertencentes à categoria de “lixo” por parte das três principais CRA contribuiu significativamente para a alteração da perceção do perfil de risco da dívida soberana, conduzindo ao aumento acentuado, num período muito reduzido, do prémio de risco da taxa de juro da dívida portuguesa. Contudo, a política monetária expansionista adotada pelo BCE em 2015, através de medidas de *quantitative easing*, tem permitido a Portugal, como membro da moeda única, aliviar a escalada da taxa de juro da dívida, embora para este efeito também tenha contribuído uma ligeira melhoria das perspetivas económicas.

Neste contexto, o principal objetivo deste trabalho visou a seleção das variáveis determinantes na evolução do risco soberano da dívida portuguesa, isto é, que explicam a variação observada no *rating*. Deste modo, não se propõe a construção de um modelo de estimação do risco soberano, mas uma abordagem que tem por base o uso do *rating* como *proxy* do nível de risco da dívida soberana.

Após a revisão da literatura, que possibilitou a compreensão das variáveis mais utilizadas para avaliar o risco da dívida soberana, decidiu-se selecionar um conjunto de vinte e duas variáveis com um racional económico que coincidissem com a realidade portuguesa e que apresentasse uma magnitude de correlação aceitável com a variável resposta, o *rating*. Visando a maximização da qualidade da regressão linear múltipla, previamente procedeu-se à identificação e mitigação das situações de multicolinearidade, através das análises VIF e das Componentes da Variância. Este tratamento reduziu a amostra inicial a um conjunto de apenas seis variáveis com magnitudes de correlação aceitáveis. Atendendo a que este trabalho incorpora fundamentalmente variáveis macroeconómicas, verificaram-se magnitudes de multicolinearidade elevadas que levaram à eliminação da maioria das variáveis com maior correlação com o *rating*. Posteriormente, aplicando os três métodos de seleção de variáveis mais usuais na literatura, *stepwise*, *backwards* e *forward linear*, concluiu-se que as cinco variáveis determinantes do risco da dívida soberana portuguesa são “Variação do PIB (%)”, “Euribor”, “Endividamento dos Particulares (%)”, o rácio “(Importações+Exportações) / PIB” e ainda as “Reservas Totais”, sendo a variável “Inflação (%)” rejeitada nos três métodos com um *p-value* superior para todos os níveis usuais de significância. Aliado ao facto de os três métodos selecionarem as mesmas cinco variáveis, sintomático da inexistência de multicolinearidade, concluiu-se que a qualidade da regressão é elevada, permitindo explicar 92,4% da variação observada na atribuição do *rating*, com um valor relativamente reduzido de erro-padrão e ainda uma distribuição normal para os resíduos obtidos. Por outro lado, verificou-se que a regressão é mais estável do que a os *ratings* publicados pelas agências, o que indica que estas utilizam outras informações qualitativas na decisão do nível de risco, as quais o modelo não contemplou. Denotou-se também uma

ligeira tendência a prever um *rating* pior que o observado, ou seja, um maior nível de conservadorismo, tendo em conta um intervalo de 2 *notches*, em módulo, face ao valor observado.

Analisando os períodos pré e durante/pós-crise, é importante realçar que os resultados não são estatisticamente significativos, uma vez que se dispõe de um reduzido número de observações. No entanto, poder-se-á destacar um R^2 mais elevado no segundo período, resultado de uma maior volatilidade no *rating* e nas variáveis explicativas. Por outro lado, destaca-se que poucas variáveis são significativas em cada um dos períodos analisados: para o período pré-crise, apenas a variável “Reservas Totais” é significativa para explicar o *rating*; no período durante/pós-crise, apenas as variáveis “variação do PIB (%)” e o rácio “(Importações+Exportações) / PIB” são significativas, sugerindo que a variação do PIB é fundamental para explicar a deterioração do *rating* por parte das agências. A alteração do comportamento deve-se também ao facto de ambas se encontrarem em fases do ciclo económico diferentes. Nesta análise, não foi testada a existência de multicolinearidade entre elas.

Por fim, concluiu-se que, no caso português, a relação entre o nível do *rating* e a taxa de juro da dívida portuguesa é fraca. Este resultado, prende-se com o facto de os níveis da taxa de juro no período pré-crise serem muito semelhantes aos registados durante o período de crise, para um nível de risco muito mais baixo no início da série. Em suma, esta magnitude de correlação reduzida deve-se essencialmente a dois fatores: alteração do paradigma da lei da procura e a oferta durante o período analisado e ações de *quantitative easing*. No sentido de reduzir o efeito destes fatores, calculou-se o prémio de mercado com base na diferença entre a taxa de juro da dívida portuguesa e a da dívida alemã. Nesta hipótese, verificou-se que além do *rating*, a “variação do PIB (%)”, a “Euribor” e as “Reservas Totais” permitem explicar 76% da variação observada no prémio de mercado da taxa de juro da dívida portuguesa.

6. Sugestões de trabalho futuro

Este trabalho recorreu a uma regressão linear múltipla para encontrar os fatores determinantes do risco soberano português, com base num histórico de 21 anos. Uma das limitações que poder-se-á identificar neste período histórico é a ausência de observações que pertençam a uma fase do ciclo económico mais favorável (alargar o histórico para a década de 80), i.e., que possam contemplar um ciclo económico completo. Uma opção possível para esta limitação seria a atualização da informação até ao ano de 2017.

No que respeita à seleção de variáveis e tratamento de multicolinearidade, seria possível a utilização de outras abordagens, como por exemplo:

- Remover a variável menos correlacionada com o *rating* e não remover a variável com maior multicolinearidade;
- *Expert* económico – Definir previamente variáveis relevantes que deveriam ser consideradas, procurando promover essas em detrimento de outras;
- Inclusão de outras variáveis com racional económico no caso português, como por exemplo, “Índice de preço médio da habitação”.

Por outro lado, do ponto de vista da metodologia estatística utilizada, o facto de existirem variáveis que parecem influenciar o *rating* de uma forma diferente consoante a fase do ciclo económico (expansão e recessão) leva a crer que a utilização de um modelo estrutural, isto é, um modelo de regressão com coeficientes variáveis no tempo, poderá conduzir a melhores previsões.

Referências Bibliográficas

- [1] Jiao, Y, LI, S. (2015). "Modelling sovereign risks: from a hybrid model to the generalized density approach".
- [2] Canuto, O., Santos, P., Porto, P. (2004). "Macroeconomics and Sovereign Risk Ratings", Washington
- [3] Kraay, A., Nehru, V. (2006). "When is External Debt Sustainable?" World Bank Economic Review, 20 (pp. 341-365)
- [4] Villemot, S. (2012). "Essays on Modelling the Sovereign Default Risk"
- [5] Diniz, T., Amaral, H., Ferreira, B. (2012). "Avaliação do risco soberano - O impacto de variáveis económicas, políticas e sociais."
- [6] Phelps, L. (2015). "Government Debt Chapter 15"
- [7] Bresser-Pereira, L. (2009). "A crise financeira de 2008." Em Revista de Economia Política, vol. 29, nº 1 (113) (pp. 133-149).
- [8] Elliott, L. (2011). "Global financial crisis: five key stages 2007-2011". The Guardian.
- [9] Gray, D., Jobst, A. (2011). "Modelling systemic financial sector and sovereign risk". Sveriges Riskbank Economic Review.
- [10] Lane, P. (2012). "The European Sovereign Debt Crisis. Em Journal of Economic Perspectives", Volume 26, Nº3 (pp. 49-68).
- [11] Beirne, J., Fratzscher, M. (2013). "The pricing of sovereign risk and contagion during the European sovereign debt crisis". Alemanha. Banco Central Europeu.
- [12] PORDATA. (2017). "PIB e PIB per capita a preços constantes (base=2011)" [http://www.pordata.pt/Portugal/PIB+e+PIB+per+capita+a+pre%C3%A7os+constantes+\(base+2011\)-2953](http://www.pordata.pt/Portugal/PIB+e+PIB+per+capita+a+pre%C3%A7os+constantes+(base+2011)-2953) (consultado última vez a 24 de fevereiro de 2017)
- [13] Diário de Notícias. (2011). "Sócrates pede demissão e abre caminho a pedido de ajuda." (consultado última vez a 30 de março de 2017)
- [14] Jornal de Negócios. (2015). "DBRS mantém "rating" de Portugal com perspectiva "estável"
- [15] PORDATA. (2017). "Taxa de desemprego: total e por sexo (%)". [Www.pordata.com](http://www.pordata.com) (consultado última vez a 24 de fevereiro de 2017)
- [16] World Bank. (2017). "GDP annual growth (%)". <http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG>, 2017 (consultado última vez a 26 de fevereiro de 2017)
- [17] McKinsey. (2011). "Portugal: o imperativo do crescimento."
- [18] De Sentis, A. (2016). "Working paper: impact of the asset purchase programme on euro area government bond yields using market news."
- [19] Arefjevs, I. (2013). "Determinants of sovereign credit ratings - Example of Latvia". New Challenges of Economic and Business Development.
- [20] Cantor, R., Packer, F. (1996). "Determinants and Impact of Sovereign Credit Ratings."
- [21] NIST. (2012). "NIST/SEMATECH e-Handbook of Statistical Methods"
- [22] Scott, D. (2009). "Sturges' Rule"

- [23] Akinwande, M., Dikko, H., Samson, A. (2015). "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis". Em Open Journal Of Statistics.
- [24] Faisca, L. (2010). "Seminários de métodos e análise de dados"
- [25] Jeeshim. (2002). "Multicollinearity in Regression Models."
- [26] Bank, C. N. (2017). "The Credit Rating of the Czech Republic". https://www.cnb.cz/en/monetary_policy/inflation_reports/2011/2011_IV/boxes_and_annexes/zoi_2011_IV_box_2.html (consultado última vez a 15 de março de 2017)
- [27] Afonso, A., Gomes, P., Rother, P. (2007). "What "hides" behind sovereign debt ratings?". ECB.
- [28] Morgan, D. (2002). "Rating Banks: Risk and Uncertainty in an Opaque Industry". American Economic Review, 92(4):874-888.
- [29] Livingston, M., Wei, J., & Zhou, L. (2010). "Moody's and S&P Ratings: Are They Equivalent? Conservative Ratings and Split Rated Bond Yields". Journal of Money, Credit and Banking.
- [30] Roy, P. V. (2005). "Credit Ratings and the Standardised Approach to Credit Risk in Basel II"
- [31] Carneira, M., Ganda, R., Romano, P. (2017). "A economia portuguesa no divã".
- [32] Gray, D., Gross, M., Paredes, J., Sydow, M. (2013). "Modeling Banking, Sovereign, and Macro Risk in a CCA Global VAR". IMF Working Paper.
- [33] Hilscher, J., Nosbusch, Y. (2007). "Determinants of Sovereign Risk: Macroeconomics Fundamentals and the Pricing of Sovereign Debt."
- [34] Lowenstein, R. (2015). "Economic Historic Repeating". The Wall Street Journal.

Anexos

A. Descrição do set inicial de variáveis

Variáveis	Cálculo	Fonte
<i>Banks</i> NPL/Empréstimo (%)	Total NPL/Valor total de empréstimo do portfólio	http://databank.worldbank.org/data/reports.aspx?source=2&country=PRT
Dívida bruta em % do PIB	Dívida Bruta / PIB	http://www.pordata.pt/Portugal/Administra%C3%A7%C3%B5es+P%C3%BAblicas+d%C3%ADvida+bruta+em+percentagem+do+PIB-2786
Desemprego %	Desempregados / Activos	http://www.pordata.pt/Portugal/Taxa+de+desemprego+total+e+por+sexo+(percentagem)-550
Formação bruta de capital fixo (% PIB)	Investimento / PIB	http://www.pordata.pt/Portugal/Investimento+Forma%C3%A7%C3%A3o+bruta+de+capital+fixo+em+percentagem+do+PIB-2833
Balança Comercial	Exportações-Importações	http://www.pordata.pt/Portugal/Balan%C3%A7a+comercial-2594
(Importações+Exportações)/ PIB	(Importações + Exportações)/PIB	http://www.pordata.pt/Portugal/Exporta%C3%A7%C3%B5es++importa%C3%A7%C3%B5es+e+saldo+da+balan%C3%A7a+de+bens+total+e+por+principais+grupos+de+pa%C3%ADses+parceiros+comerciais-2341 http://www.pordata.pt/Portugal/Produto+Interno+Bruto+na+%C3%B3ptica+da+despesa+(base+2011)-2283
Balança Corrente/PIB	(Balança Bens e Serviços + Balança de Rendimentos + Transferências Correntes) / PIB	http://www.pordata.pt/Europa/Balan%C3%A7a+de+pagamentos+salDOS+(Euro)-1550 http://www.pordata.pt/Portugal/Produto+Interno+Bruto+na+%C3%B3ptica+da+despesa+(base+2011)-2283
Despesas Estado Social/PIB	Gastos em Ação e Segurança Social / PIB	https://www.pordata.pt/Portugal/Despesas+do+Estado+em+percentagem+do+PIB+por+algumas+fun%C3%A7%C3%B5es-2778-236920
PIB/Hora trabalhada	PIB/ Horas de trabalho	https://data.oecd.org/lprdy/gdp-per-hour-worked.htm#indicator-chart
Poupanças % Famílias	(Rendimento + Fundo de Pensões) -Gastos	https://data.oecd.org/natincome/saving-rate.htm
Impostos sobre o rendimento e património	Imposto Rendimentos + Imposto Patrimónios	http://www.pordata.pt/Portugal/Receitas+das+Administra%C3%A7%C3%B5es+P%C3%BAblicas+total+e+por+classifica%C3%A7%C3%A3o+econ%C3%B3mica+(base+2011)-2796
Euribor	Média das taxas dos juros interbancários para os bancos proeminentes europeus à excepção das 15% mais baixas e mais altas	https://www.dnb.nl/binaries/se2000q2_tcm46-147362.pdf + http://pt.global-rates.com/taxa-de-juros/euribor/euribor-informacao-de-referencia.aspx http://data.okfn.org/data/core/euribor#readme
Divida Externa/PIB	Dívida acumulada ao estrangeiro / PIB	http://www.pordata.pt/Portugal/D%C3%ADvida+externa+l%C3%ADquida+em+percentagem+do+PIB-2950
Unit Labour Costs	Custo de uma hora de trabalho / Valor produzido numa hora	https://data.oecd.org/lprdy/unit-labour-costs.htm#indicator-chart
PIB per Capita	PIB / População	https://data.oecd.org/gdp/gross-domestic-product-gdp.htm
Reservas Total	Reservas + Reservas de ouro	http://databank.worldbank.org/data/reports.aspx?source=2&country=PRT#

Leverage banking %	Ativos financeiros / Capital próprio	https://data.oecd.org/corporate/banking-sector-leverage.htm
Divida externa / Exportações	Dívida acumulada ao estrangeiro / Total de exportações	http://www.pordata.pt/Portugal/D%C3%ADvida+externa+!%C3%ADquida+em+percentagem+do+PIB-2950 http://www.pordata.pt/Portugal/Exporta%C3%A7%C3%B5es++importa%C3%A7%C3%B5es+e+saldo+da+balan%C3%A7a+de+bens+total+e+por+principais+grupos+de+pa%C3%ADses+parceiros+comerciais-2341
PIB %	(PIB no ano n - PIB no ano n-1)/PIB no ano n	http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG
Envidamento Particulares %	Total de Créditos/Rendimento líquido disponível	https://data.oecd.org/hha/household-debt.htm
Inflação	(Preço n - Preço n-1)/Preço n	http://www.pordata.pt/Portugal/Taxa+de+Infla%C3%A7%C3%A3o+(Taxa+de+Varia%C3%A7%C3%A3o+do+%C3%8Dndice+de+Pre%C3%A7os+no+Consumidor)+total+e+por+consumo+individual+por+objectivo-2315
Défice público(M€)	Receitas Públicas - Despesas Públicas	http://www.pordata.pt/Portugal/Administra%C3%A7%C3%B5es+P%C3%BAblicas+despesas++receitas+e+d%C3%A9fice+excedente+(base+2011)-2784

Tabela 40 - Variáveis, descrição e fonte de dados.

B. Análise VIF preliminar ao set inicial de variáveis

Inflação	Balança Comercial	PIB %	Desemprego %	Leverage banking %	PIB/ Hora trabalhada	Unit Labour Costs	Euribor	Poupanças % Famílias	Endividamento Particulares %	Formação bruta de capital fixo (%)	Défice/excedente público (ME)	Dívida bruta em % do PIB	Impostos	(Importações + Exportações) / PIB	Divida externa/ PIB	Divida externa / Exportações	Reservas Total	Bank NPL / total gross loans (%)	Despesas Estado Social/PIB	Balança Corrente/PIB	PIB per Capita
2E+14	-7E+14	-1E+14	-2E+14	-1E+14	2E+13	-5E+13	2E+14	2E+14	-1E+14	-7E+13	-9E+13	1E+14	2E+13	-2E+14	-2E+14	-9E+13	-1E+14	6E+14	3E+14	4E+14	3E+14
-7E+14	3E+15	6E+14	1E+15	6E+14	-1E+14	2E+14	-8E+14	-8E+14	7E+14	2E+14	4E+14	-6E+14	-5E+13	8E+14	9E+14	3E+14	5E+14	-3E+15	-2E+15	-2E+15	-1E+15
-1E+14	6E+14	1E+14	2E+14	1E+14	-3E+13	4E+13	-1E+14	-1E+14	1E+14	3E+13	8E+13	-1E+14	-6E+12	2E+14	1E+14	6E+13	9E+13	-5E+14	-3E+14	-4E+14	-2E+14
-2E+14	1E+15	2E+14	4E+14	2E+14	-6E+13	8E+13	-3E+14	-3E+14	2E+14	1E+14	1E+14	-2E+14	-3E+13	3E+14	2E+14	2E+14	2E+14	-9E+14	-5E+14	-7E+14	-4E+14
-1E+14	6E+14	1E+14	2E+14	1E+14	-2E+13	5E+13	-2E+14	-2E+14	1E+14	6E+13	8E+13	-1E+14	-2E+13	2E+14	2E+14	8E+13	9E+13	-5E+14	-3E+14	-4E+14	-2E+14
8E+12	-3E+13	-2E+13	-4E+13	-4E+12	1E+14	5E+12	3E+13	-1E+13	-1E+14	7E+13	-4E+13	-5E+13	-3E+13	-4E+13	2E+14	-1E+14	-2E+12	-5E+13	4E+13	9E+13	1E+13
-5E+13	3E+14	5E+13	9E+13	5E+13	-6E+11	2E+13	-6E+13	-6E+13	5E+13	2E+13	3E+13	-5E+13	-6E+12	6E+13	8E+13	2E+13	4E+13	-2E+14	-1E+14	-2E+14	-1E+14
2E+14	-8E+14	-1E+14	-3E+14	-2E+14	5E+13	-6E+13	2E+14	2E+14	-2E+14	-4E+13	-1E+14	1E+14	5E+12	-2E+14	-2E+14	-9E+13	-1E+14	6E+14	4E+14	5E+14	3E+14
2E+14	-8E+14	-1E+14	-3E+14	-2E+14	4E+12	-6E+13	2E+14	2E+14	-1E+14	-9E+13	-9E+13	2E+14	3E+13	-2E+14	-2E+14	-1E+14	-1E+14	7E+14	4E+14	5E+14	3E+14
-1E+14	7E+14	1E+14	2E+14	1E+14	-1E+14	4E+13	-2E+14	-1E+14	4E+14	-1E+14	1E+14	-6E+13	7E+13	1E+14	-4E+13	8E+12	9E+13	-4E+14	-3E+14	-6E+14	-3E+14
-6E+13	2E+14	3E+13	1E+14	6E+13	6E+13	2E+13	-4E+13	-9E+13	-1E+14	1E+14	-9E+12	-8E+13	-7E+13	1E+14	2E+14	1E+14	4E+13	-3E+14	-1E+14	2E+12	-2E+13
-8E+13	4E+14	8E+13	1E+14	8E+13	-5E+13	3E+13	-1E+14	-9E+13	1E+14	-7E+12	6E+13	-6E+13	1E+13	1E+14	3E+13	5E+13	6E+13	-3E+14	-2E+14	-3E+14	-2E+14
1E+14	-6E+14	-1E+14	-2E+14	-1E+14	-3E+13	-5E+13	1E+14	2E+14	-7E+13	-9E+13	-6E+13	1E+14	3E+13	-2E+14	-3E+14	-3E+13	-1E+14	6E+14	3E+14	4E+14	2E+14
2E+13	-4E+13	-4E+12	-3E+13	-2E+13	-3E+13	-5E+12	2E+12	3E+13	7E+13	-7E+13	1E+13	3E+13	4E+13	-4E+13	-6E+13	-6E+13	-1E+13	1E+14	4E+13	-5E+13	-2E+13
-2E+14	7E+14	1E+14	3E+14	2E+14	-5E+13	6E+13	-2E+14	-2E+14	1E+14	1E+14	1E+14	-1E+14	-4E+13	2E+14	9E+13	2E+14	1E+14	-7E+14	-4E+14	-4E+14	-2E+14
-2E+14	1E+15	2E+14	3E+14	2E+14	2E+14	1E+14	-2E+14	-3E+14	-3E+11	2E+14	5E+13	-3E+14	-7E+13	2E+14	9E+14	-2E+14	2E+14	-1E+15	-4E+14	-6E+14	-4E+14
-6E+13	1E+14	3E+13	1E+14	4E+13	-1E+14	3E+12	-4E+13	-5E+13	-3E+13	1E+14	3E+13	8E+12	-6E+13	2E+14	-3E+14	4E+14	3E+13	-2E+14	-2E+14	4E+13	7E+13
-1E+14	5E+14	9E+13	2E+14	9E+13	-1E+13	4E+13	-1E+14	-1E+14	9E+13	4E+13	6E+13	-9E+13	-1E+13	1E+14	1E+14	5E+13	7E+13	-4E+14	-2E+14	-3E+14	-2E+14
6E+14	-3E+15	-5E+14	-9E+14	-5E+14	2E+12	-2E+14	6E+14	7E+14	-4E+14	-3E+14	-3E+14	5E+14	1E+14	-7E+14	-8E+14	-4E+14	-4E+14	2E+15	1E+15	2E+15	9E+14
3E+14	-1E+15	-3E+14	-5E+14	-3E+14	7E+13	-1E+14	4E+14	4E+14	-3E+14	-1E+14	-2E+14	3E+14	4E+13	-4E+14	-3E+14	-2E+14	-2E+14	1E+15	7E+14	1E+15	5E+14
5E+14	-2E+15	-4E+14	-7E+14	-4E+14	1E+14	-2E+14	6E+14	5E+14	-6E+14	-1E+13	-3E+14	4E+14	-4E+13	-5E+14	-5E+14	-7E+13	-3E+14	2E+15	1E+15	2E+15	9E+14
3E+14	-1E+15	-2E+14	-4E+14	-2E+14	4E+13	-1E+14	3E+14	3E+14	-3E+14	-3E+13	-2E+14	2E+14	-1E+13	-3E+14	-4E+14	-2E+12	-2E+14	1E+15	6E+14	9E+14	5E+14

Tabela 41 – Análise VIF (1996-2015).