

Escola de Ciências Sociais e Humanas

Departamento de Economia Política

O Google como medida de sentimento nos mercados financeiros

Ana Inês Martins Mourão

Dissertação submetida como requisito parcial para a obtenção do grau de
Mestre em Economia Monetária e Financeira

Orientador:

Doutor Luís Filipe Farias de Sousa Martins, Professor Auxiliar com Agregação,

ISCTE – Instituto Universitário de Lisboa

Maio, 2018

“An investment in knowledge always pays the best interest.”

Benjamin Franklin

Agradecimentos

Às duas grandes mulheres da minha vida, mãe e avó Clementina. Por me terem ensinado o que é ser uma verdadeira mulher. Por nunca me terem deixado desistir de nada, mostrando sempre que a luta só é vencida por quem se dá de corpo e alma.

Aos meus pais por tudo o que me proporcionaram, por terem apostado em mim e na minha educação sem receio do que poderia correr mal. Obrigada pelo voto de confiança. Sem vocês nada teria sido possível.

Ao Miguel por estar sempre lá. Por me amar incondicionalmente. Por ser sempre o meu maior apoio. Por me fazer sempre ver que sou mais do que aquilo que penso. Amo-te.

Ao Pedro, para quem sempre quis ser um exemplo. Nada é impossível.

Ao Matias, por ter sido o meu companheiro de estudo durante estes últimos anos.

A todos os que de alguma forma contribuíram para que isto fosse possível. Família, amigos, colegas de faculdade, professores e todos os que não me tocando diretamente me inspiraram para prosseguir em frente.

Ao professor Luís por todo o apoio e confiança. Sem a sua preciosa ajuda não teria sido possível concluir este trabalho. Foi para mim uma honra tê-lo como orientador.

Obrigada.

Resumo

Esta dissertação insere-se no âmbito das finanças comportamentais e tem como objetivo inicial construir uma *proxy* de sentimento baseada nos dados de pesquisa online do Google, através da sua plataforma *Google Trends* e além disso compará-la com outras previamente utilizadas, no que toca a precisão e disponibilidade.

Posteriormente centra-se no estudo da relação existente entre este índice de sentimento com os retornos do mercado acionista português e dos Estados Unidos, no período compreendido entre dezembro 2007 a junho 2017, usando para isso regressões lineares em que a variável Google é usada como uma das variáveis independentes para explicar os retornos de cada um dos mercados.

Dado que se conclui que os dados do Google são realmente capazes de medir adequadamente a influência do sentimento em ambos os mercados, e que este mostra resultados mais robustos do que outras *proxies* previamente utilizadas noutros trabalhos, de seguida, optou-se por estudar a sua capacidade preditiva quanto aos movimentos dos mercados acionistas em estudo. Verificou-se que existe, de facto, um poder de previsão deste índice de sentimento contemporâneo, no entanto este deve ser usado em simultâneo com indicadores mais teóricos, como o CAPE, especialmente em mercados financeiros mais desenvolvidos como nos EUA.

Palavras-chave: Finanças comportamentais, Sentimento de mercado, Google Trends, Previsão mercados financeiros

JEL Classification: G41, C53

Abstract

This dissertation aims in the first place to add to the behavioural finance literature a proxy for sentiment based on Google's online search data, through its Google Trends platform and, in addition, by comparing it with other previously used proxies, regarding precision and availability.

Subsequently, it focuses on the relationship between this sentiment index and the Portuguese and American stock markets returns, between december 2007 and june 2017, using linear least squares regressions in which the Google variable is used as one of the independent variables to explain the returns of each market.

By the fact that it was concluded that Google data are indeed capable of adequately measuring the influence of sentiment in both markets, and that it displays more thorough results than other proxies used in previous work, this study proceeds to examining its predictive capacity in regard to the stock markets movements under investigation. I conclude that this contemporary sentiment index displays predictive power but it should be used with other theoretical measures like CAPE, especially in more developed financial markets such as EUA.

Keywords: Behavioral finance, Market sentiment, Google Trends, Forecasting financial markets.

Índice Geral

1.	Introdução.....	1
2.	Enquadramento Teórico e Revisão de Literatura	3
2.1.	Economia Comportamental	6
2.2.	Sentimento.....	7
2.3.	Medidas de sentimento	8
2.4.	Internet e Google	11
2.5.	Google Trends	12
3.	Comparação da realidade financeira e tecnológica dos EUA e de Portugal.....	15
4.	Dados e Metodologia.....	21
4.1.	Índice de sentimento Google	21
4.2.	Índice CAPE.....	24
4.3.	Retornos e variáveis de controlo	26
4.4.	Tratamento dos dados.....	27
4.5.	Hipóteses a testar, metodologia e período amostral	28
5.	Resultados	29
5.1.	Matriz de Correlações	29
5.2.	Causalidade à Granger.....	31
5.3.	Estimações OLS	32
5.4.	Previsão	33
6.	Conclusões	39
7.	Bibliografia.....	41
8.	Fontes	46
9.	Anexos.....	46

Índice de Tabelas

Tabela 3.1: Americanos que investem no mercado de ações- por idade e rendimento	17
Tabela 3.2: Nível de penetração da internet	19
Tabela 5.1: Matriz de Correlações - EUA.....	30
Tabela 5.2: Matriz de Correlações - Portugal.....	30
Tabela 5.3: Matriz de Correlações - retornos dos mercados acionistas em Portugal e nos EUA..	31
Tabela 5.12: Conteúdo informativo das variáveis	35
Tabela 5.13: Desempenho absoluto fora da amostra.....	35
Tabela 5.14: Desempenho relativo fora da amostra	37
Tabela 5.15: Teste Diebold-Mariano.....	38
Anexo A - Tabela 4.1: Conjunto de termos de pesquisa	48
Tabela 4.2: Definição das variáveis.....	49
Anexo B - Tabela 5.4: Regressões OLS - EUA	51
Tabela 5.5: Regressões OLS - Portugal.....	52

Índice de Gráficos

Gráfico 2.1: Partilha mundial de mercado dos principais motores de pesquisa (2010-2017)	12
Gráfico 3.1: Capitalização de Mercado das empresas domesticas listadas em bolsa.....	15
Gráfico 3.2: Percentagem de americanos adultos que investem no mercado de ações	16
Gráfico 3.3: Portugueses com investimentos em valores mobiliários.....	18
Gráfico 4.1: Sazonalidade	23
Gráfico 4.2: CAPE- EUA.....	25

Índice de Figuras

Anexo A - Figura 4.1: Gráfico do Google Trends	47
---	----

1. Introdução

A história do mercado de ações está repleta de eventos catastróficos, como o Crash de 1929, que não têm explicação aos olhos do modelo padrão das finanças em que os preços do mercado de ações tendem para o valor presente dos fluxos de caixa esperados no futuro. Já em 1936, John Maynard Keynes argumentava que os mercados flutuavam sob a influência dos “*animal spirits*” dos investidores, no entanto, o estudo da economia comportamental tem vindo apenas a desenvolver-se nos últimos tempos. Para este desenvolvimento contribuiu, mais recentemente, a atribuição do prémio nobel da economia de 2017 a Richard Thaler, cujos estudos ajudam a compreender como o comportamento humano por vezes se desvia dos modelos económicos teóricos. Ao explorar as consequências da racionalidade limitada, as preferências sociais e a falta de autocontrolo, Thaler mostrou como esses traços humanos afetam, de forma sistemática, as decisões individuais e, por outro lado, os resultados económicos. As suas contribuições construíram assim uma ponte entre a análise económica e a análise psicológica no processo de tomada de decisão. Além de Thaler, vários economistas se têm dedicado à investigação deste tema, e a grande conclusão a que chegaram é que existem, de facto, influências cognitivas, sociais e emocionais no comportamento económico dos agentes.

Daqui surgiu o conceito de sentimento económico o qual se designa como o consenso geral sobre a evolução futura dos mercados financeiros, entre os investidores, que, geralmente, padece de enviesamentos como o excesso de confiança e se traduz no risco que os investidores estão dispostos a tomar (Baker e Wurgler, 2007).

Existem, atualmente, imensos estudos empíricos que provam que o sentimento de mercado influencia, de facto, os mercados financeiros e, portanto, agora a questão passa a ser como medir esse sentimento (Baker e Wurgler, 2007).

O objetivo central desta dissertação é perceber de que forma as medidas de sentimento estão relacionadas com os movimentos do mercado acionista, quer em Portugal quer nos EUA, tal como sugerido por Ferreira (2016), fazendo, além disso, uma comparação entre os mercados acionistas dos dois países. De acordo com Simon (1955), os investidores começam os seus processos de tomada de decisão tentando recolher informações, sendo que no mundo tecnológico de hoje parte desta coleta consiste em pesquisas on-line. Assume-se que as atitudes dos investidores reveladas pelas suas pesquisas na internet podem ser consideradas uma medida de sentimento (Da et al. 2011, 2015) e por isso usa-se como *proxy* o índice de pesquisas no

Google tal como em Ferreira (2016). Brochado (2015) usando também este índice conclui que, de facto, este é uma boa reprodução do sentimento económico e que, além disso, se correlaciona bem com as medidas de sentimento alternativas mais conhecidas. Além de ser uma base de dados das intenções dos investidores revelando o comportamento espontâneo dos mesmos, o índice de sentimento Google oferece vantagens adicionais já que está disponível com maior frequência e menor custo. Na presente dissertação, e com o objetivo de perceber se as medidas mais tradicionais como o índice CAPE¹ de Robert Shiller são mais eficazes a explicar os movimentos do mercado de ações do que as medidas de sentimento mais atuais como o índice Google, comparam-se o poder explicativo destas duas *proxies* para ambos os países, no que toca aos retornos dos mercados bolsistas.

Posteriormente, é feita uma avaliação do poder preditivo do CAPE e do índice de sentimento Google no que toca aos movimentos da bolsa quer em Portugal quer nos EUA, seguindo a conclusão do estudo de Preis et al. (2013) que, ao analisarem as mudanças nos volumes de pesquisa no Google para termos financeiros, encontraram padrões que podem ser interpretados como "sinais de alerta precoce" dos movimentos do mercado de ações. Isto é bastante relevante uma vez que pode ajudar os vários agentes da economia a tomar decisões financeiras mais proveitosas.

Esta dissertação torna-se assim importante para a literatura uma vez que faz um estudo comparativo sobre o impacto do sentimento em Portugal e nos EUA, países tão diferentes quer em termos de mercados acionistas, quer de dimensões, e além disso compara o índice de sentimento negativo Google com o CAPE de Robert Shiller em termos de previsão *out-of-sample* como, tanto quanto me foi permitido apurar, nunca antes feito.

Aplicando regressões lineares OLS e modelos de previsão *in-sample* e *out-of-sample*, no período compreendido entre dezembro 2007 a junho 2017, conclui-se que, de facto, existe um poder explicativo da variável de sentimento google criada na dissertação e que além disso, esta apresenta capacidade preditiva no que toca aos retornos do mercado acionista de ambos os países, no entanto deve ser conjugada com indicadores mais teóricos como o CAPE.

O estudo subdivide-se em cinco grandes secções. Começa-se com uma introdução à economia comportamental e aos conceitos e definições a ela subjacentes, incluindo uma visão geral sobre as medidas de sentimento e o seu papel nas finanças comportamentais, com ênfase particular

¹ Cyclically adjusted price-to-earnings ratio

nos dados de pesquisa na Internet. Na secção dois é feita uma comparação da realidade tecnológica e financeira dos dois países em estudo. A secção três apresenta os dados e a metodologia usados para chegar ao índice de sentimento baseado nas pesquisas no Google e todo o tratamento aplicado às restantes variáveis usadas no estudo. A secção seguinte apresenta os resultados em resposta às questões centrais da tese, ou seja, averigua se o índice Google tem uma relação significativa com os retornos do mercado acionista dos dois países e se, além disso, é possível afirmar que esta relação se reflete em termos de previsão. E, por último, na secção cinco, expõem-se as conclusões do estudo.

2. Enquadramento Teórico e Revisão de Literatura

Antes de mais é necessário definir o quadro teórico em que a investigação se insere e por isso procede-se à revisão da literatura e também à descrição do estado atual da investigação neste tema.

A forma de funcionamento dos mercados financeiros é alvo de intenso debate há muitas décadas, talvez mesmo séculos (Sewell, 2011), sendo ano após ano uma área altamente dinâmica em termos de produção científica. Um importante marco neste domínio é a hipótese dos mercados financeiros eficientes (HME) avançada por Eugene Fama (Fama, 1969,1995). A HME diz-nos que os preços dos ativos financeiros refletem toda a informação referente quer a factos passados, quer a factos que o mercado espera, fundamentadamente, que venham a acontecer.

Roberts (1967) divide a HME em três formas: Fraca, Semi-Forte e Forte. A forma forte afirma que toda a informação relevante está completamente refletida no preço dos ativos. De acordo com esta hipótese, mesmo pessoas com informações privilegiadas não conseguiriam “vencer” o mercado. A forma semi-forte da HME diz-nos que os preços dos ativos refletem toda a informação publicamente disponível não só acerca das empresas em causa como no que diz respeito a fatores macroeconómicos. Por fim, a forma fraca afiança que todos os dados relativamente a preços anteriores estão incorporados nos preços dos ativos.

Ao contrário daquilo que é muitas vezes afirmado, a HME não obriga a que todos os agentes envolvidos na compra e venda de ativos sejam perfeitamente racionais. Os proponentes da HME acreditam que uma grande parte dos agentes será racional, sim, mas também aceita que uma

parte importante dos investidores financeiros não o seja. Mesmo assim, a HME verificar-se-ia porque por um lado os agentes irracionais efetuariam compras e vendas de forma relativamente aleatória e que, portanto, se cancelariam mutuamente, e, por outro lado, quando as ações destes agentes irracionais não se cancelam, elas geram possibilidades de lucros extraordinários que serão aproveitadas pelos investidores racionais, nulificando assim as ações daqueles.

Existe muita evidência empírica favorável à HME. Fama constatou que não existia correlação suficiente entre os preços passados e futuros de ativos individuais para originar lucros. Por outro lado, vários estudos evidenciam que fundos geridos ativamente não conseguem, consistentemente, obter melhor retorno que simples índices que sigam a tendência geral do mercado (Kahn e Rudd, 1995; Malkiel, 1995). Existem até, estudos que parecem oferecer algum suporte à forma forte da HME (Maloney e Mulherin, 1998). No entanto, a HME padece de um grave problema que é o facto de não considerar a possível existência de retornos anormais, o que vai contra o que se verifica nos mercados financeiros onde os retornos nem sempre apresentam valores ditos “normais” como esta avança. Isto contrasta com o pensamento dos investigadores que acreditam na previsibilidade dos retornos do mercado de ações e nos seus valores anormais, afirmando que é de facto possível prever os retornos com base no seu histórico.

Diversos autores se têm dedicado ao estudo da previsibilidade dos retornos financeiros e, apesar de ainda não existir consenso sobre quais as variáveis que possuem esse poder preditivo, grande parte dos trabalhos empíricos neste tema sugere que a *dividend-yield* é a variável mais eficaz para prever os retornos dos mercados de ações (Shiller, 1980), (Campbell e Shiller, 1988), (Fama e French, 1988), (Cochrane, 2008), (Golez e Koudijs, 2017). No entanto, noutros estudos ela deixa de possuir poder preditivo estatisticamente significativo (Ang e Bekaert, 2006). Já no que toca aos rácios de lucro (*Price-Earnings ratios*) a literatura é mais consensual dado que praticamente a totalidade dos investigadores encontram neles poder preditivo. Também existem artigos que mencionam o rácio CAPE de Robert Shiller, o qual será desenvolvido mais à frente, como sendo bastante confiável no que toca à previsão dos retornos do SP500 (Keimling, 2016). Além destas variáveis, uma das causas apontadas para a variação dos retornos da bolsa é o ciclo em que a economia se encontra, sendo que os retornos tendem a ser superiores quando a economia está em recessão (Fama e French, 1988) daí que o ciclo económico seja muitas vezes visto como tendo algum poder preditivo no que toca aos retornos das ações.

Apesar de todos os estudos já existentes nesta temática da previsibilidade dos retornos dos mercados financeiros, desde cedo, foi sendo chamada a atenção para o facto de os agentes não tomarem decisões da forma que a HME sugere. Tversky e Kahneman (1974) mostraram que, de facto, existiam diversas lacunas no processo de decisão dos seres humanos e propuseram modelos que põem seriamente em causa os fundamentos da HME (Kahneman e Tversky, 1979). Tal como foi referido anteriormente, já em 1936 Keynes argumentava que existia algo, o qual ele designou por “*animal spirits*” que fazia com que os preços se afastassem dos seus fundamentais, mas a evidência empírica contrária à HME foi-se acumulando mais nos últimos anos sendo hoje abundante. Para dar um exemplo, Robert Shiller, vencedor do Prémio de Ciências Económicas em Memória de Alfred Nobel em conjunto com Eugene Fama em 2013, mostrou que os preços das ações exibem demasiada volatilidade que não pode ser explicada por novas informações relativas às mesmas (Shiller, 1980).

Diversos autores se têm dedicado ao estudo do sentimento dos investidores no mercado de ações, sendo um bom exemplo, Baker e Wurgler (2007) que concluíram que de facto, olhando para a história do mercado de ações é notória a quantidade de eventos que se afastam do comportamento teórico referido nos modelos económicos. Na verdade, o pressuposto do modelo padrão de que os investidores irracionais forçam sempre os preços do mercado de capitais a igualar o valor presente dos fluxos de caixa esperados não se conseguiu demonstrar empiricamente.

Isto deve-se em parte ao facto dos seres humanos se afastarem dos agentes económicos típicos dos modelos teóricos que regem a economia. E tal como afirma Thaler (2015): “Não há nada de errado nas pessoas, todos somos apenas seres humanos- homo sapiens”. O grande problema consiste na não utilização destes homo-sapiens nos modelos económicos, o que leva a erros de previsão que podem ser bastante nefastos para a economia.

O que se quer dizer aqui não é que a teoria económica está errada, mas sim que “é necessária uma abordagem mais rica da investigação económica, uma abordagem que reconheça a existência e a relevância dos seres humanos.” - Thaler (2015). É essencial compreender os mecanismos através dos quais a irracionalidade financeira contagia a economia real e é isso que Bes (2010) tenta fazer analisando as cinco bolhas mais importantes da história: "A Tulipomania" holandesa do século XVII, a bolha da "Companhia dos Mares do Sul" em Inglaterra, no século XVIII, o crash de 1929, nos EUA, o Japão dos anos 90 e a crise do subprime. O autor conclui que importam mais as condições presentes do que os fundamentais

dos mercados. Dessas condições, destaca-se a facilidade de acesso ao "jogo", normalmente porque, muitas vezes com facilidades de crédito, as pessoas se endividam para poder negociar, e a inveja que caracteriza os investidores e serve como guia das suas atitudes.

2.1. Economia Comportamental

Durante anos, os economistas resistiram ao apelo para basearem os seus modelos numa caracterização mais exata do comportamento humano, mas tornou-se premente a criação de uma base teórica alternativa à HME. Tendo como ponto de partida os modelos económicos já existentes surge a chamada economia comportamental a qual, nas palavras de Thaler (2015): “continua a ser economia, mas é uma economia com fortes injeções de boa psicologia e de outras ciências sociais”. Ainda o mesmo autor reclama que a principal razão para adicionar os seres humanos à teoria económica, é melhorar a exatidão das previsões económicas que antes padeciam de enviesamentos inexplicáveis pela teoria económica tradicional.

As finanças comportamentais partem de dois pilares fundamentais. Por um lado, utilizam descobertas ligadas à psicologia para identificar vários desvios, em relação ao comportamento previsto pelo agente racional usado na teoria económica clássica, no processo de tomada de decisão por parte dos investidores. Sabemos hoje que os agentes sofrem de excesso de confiança, otimismo infundado, conclusões retiradas de pequenas amostras, conservadorismo entre muitos outros afastamentos em relação ao ideal de racionalidade (Barberis e Thaler, 2002). No entanto, já vimos que a HME não necessita que todos os investidores sejam perfeitamente racionais. No caso de uma minoria não o ser, existirão investidores profissionais prontos a lucrar através de processos de arbitragem, ou seja, às custas dos investidores irracionais, e acabarão por fazer voltar os preços ao valor intrínseco dos ativos. É então aqui que entra o segundo pilar das finanças comportamentais: a identificação dos limites da arbitragem. Os defensores das finanças comportamentais começam por documentar empiricamente os falhanços da arbitragem, ou seja, instâncias em que esta não foi capaz de trazer os preços para o seu valor fundamental (Mitchell et al., 2002). Para além disso é fácil identificar, mesmo em termos teóricos, limites à arbitragem como seja o facto de nem sempre existirem ativos substitutos, ou o risco que representa para um investidor profissional ir contra a opinião do mercado por grandes períodos de tempo (Barberis e Thaler, 2002).

A economia comportamental é agora um ramo em crescimento da economia e, recentemente, tornou-se importante também ao nível das decisões de políticas públicas do governo de alguns

países, como o Reino Unido que desde 2010 formou uma equipa de perspetivas comportamentais.

2.2. Sentimento

Investigações no campo da psicologia e da economia comportamental têm-nos mostrado que não só os agentes económicos estão sujeitos a desvios, na forma como pensam e decidem, em relação aquilo que a HME prevê, como também esses desvios estão correlacionados na medida em que são idênticos na maioria dos agentes “irracionais” (Shleifer e Summers, 1990). De facto, estes tendem a sofrer de excesso de confiança, extrapolam dados passados como se estes servissem de guia fidedigno para o futuro, seguindo assim as tendências e aumentando a correlação nos desvios mencionados (Andreassen e Kraus, 1990), e dão demasiada importância a informações sobre eventos recentes acabando por reagir exageradamente (Tversky e Kahneman, 1974). Assim, os preços nos mercados financeiros são em grande medida influenciados por estes e outros modos de pensamento enviesados.

Passados 50 anos do nascimento da expressão “*animal spirits*” de Keynes, De Long et al. (1990) formalizaram o que passou a ser definido como o sentimento dos investidores nos mercados financeiros, o qual designam como o consenso geral sobre os fluxos de caixa futuros e os riscos a eles associados que não são justificados pelos fundamentais. Este sentimento, tal como afirmado por Hui et al. (2013) é muitas vezes aceite como o motor dos preços das ações.

Na verdade, nem todos os investidores são racionais uma vez que sendo influenciados pelas suas emoções e tendo acesso a diferentes informações sobre os mercados, alguns agentes, irracionais, têm comportamentos que fazem com que os preços se afastem dos valores defendidos pela teoria económica. Assim e devido aos limites à arbitragem defendidos por Shleifer e Vishny (1997) os agentes racionais não têm capacidade de forçar os preços para os seus valores fundamentais, como sugerem os modelos económicos. Existem inúmeros exemplos que comprovam que nem sempre as informações financeiras são importantes nas tomadas de decisão dos agentes. Cutler et al. (1989) concluíram que as notícias com informação quantitativa macroeconómica explicam apenas um quinto da variação do preço de uma ação. A verdade é que os agentes padecem de racionalidade limitada e o ambiente e as emoções que os rodeiam são elementos chave na sua tomada de decisão. Hirshleifer e Shumway (2003), por exemplo, concluíram que em dias em que o céu está nublado os retornos diários de ações são 9% inferiores aos dias de sol. Também Edmans et al. (2007) concluíram que as partidas

internacionais de futebol afetam o retorno diário de ações para o país perdedor. Tais conclusões não são justificadas pela teoria económica.

Podemos assim concluir que fenómenos como heurísticas, preconceitos, emoções, humor, reações a notícias, todos causados pela incerteza e a informação assimétrica contribuem para uma perceção de mercado que se distancia da realidade financeira existente. Tudo isto leva aos já mencionados desvios dos fundamentais que estão na base das performances imprevisíveis e aleatórias que caracterizam a história dos mercados de ações.

2.3. Medidas de sentimento

Tentar perceber se o sentimento do investidor afeta os mercados de ações é um dos principais temas de debate em finanças comportamentais. Os estudos realizados por Black (1986), De Long et al. (1990) e Barberis et al. (1997) modelaram o papel do sentimento dos investidores nos mercados financeiros, analisando o comportamento dos *noise traders* (dois primeiros) e a reação às notícias dos media (último) e concluíram que os investidores são, de facto, influenciados pelo sentimento.

Existem três resultados essenciais que se retiram dos estudos de Baker e Wurgler (2007) e Da et al. (2015). Primeiro, o sentimento do investidor explica o retorno das ações. Em segundo lugar, o sentimento tem uma influência maior nas ações cujos preços são mais subjetivos e difíceis de arbitrar (por exemplo, aquelas ações com maior β^2 , maior risco de queda e maior volatilidade). Por último, o sentimento do investidor está sujeito a reversões, isto é, sentimentos mais positivos correspondem a retornos baixos e nos dias que se seguem às transações, essa relação inverte-se.

De acordo com Baker e Wurgler (2007), neste momento "a atenção académica passou de estudar" se o sentimento do investidor afeta os preços das ações "para avaliar" como medir o sentimento do investidor e quantificar os seus efeitos".

Um número crescente de estudos empíricos procurou medir esse sentimento e tradicionalmente, os investigadores tomaram duas abordagens. A maioria desses estudos identifica medidas de sentimento diretas e indiretas (Qiu e Welch, 2004). As medidas de sentimento diretas derivam

² O β é uma medida do grau de risco de uma ação. Este mede a sua correlação com o resto do mercado, ou seja, de que forma o preço da ação se movimenta em relação ao mercado. Por exemplo, uma ação que não tem qualquer relação com o mercado tem um β de zero.

de pesquisas que perguntam diretamente aos indivíduos como se sentem sobre as condições do mercado acionista e as condições económicas atuais ou futuras. Estas medidas baseadas em inquéritos são auto-avaliações dos investidores e por isso estão sujeitas a erros de medição, isto porque nem sempre os inquiridos respondem de forma verdadeira às questões, distorcendo, por vezes, as suas respostas em prol do que acham que deveria ser a resposta adequada apesar de não ser o que efetivamente fazem ou acham. Além disso, exigem um painel representativo de populações-alvo, tendo a desvantagem de serem custosas de produzir e, ademais, a sua periodicidade é reduzida. A maioria dos resultados da pesquisa são gradualmente apresentados durante uma semana, um mês ou um trimestre, e por isso os resultados não correspondem ao sentimento do investidor num determinado momento, mas a uma mistura de opiniões recentes e antigas. Todas estas desvantagens aumentam a popularidade crescente das medidas de sentimento indiretas que representam as variáveis económicas e financeiras que servem para capturar o estado de espírito dos investidores.

Nesta altura, abundam já os estudos empíricos que encontram em várias *proxies* usadas para medir o sentimento, capacidade explicativa do sentido dos mercados [Baker e Wurgler (2013); Kumar e Lee (2006); Lee et al. (1991); Tetlock (2007)]. Posteriormente surge a questão sobre qual o indicador de sentimento que melhor explica as variações nos mercados financeiros. Já foram utilizados, com diferentes graus de sucesso, vários indicadores tais como índices de confiança do consumidor e da indústria e inquéritos aos investidores, no entanto não existe ainda consenso sobre qual será a melhor *proxy* do sentimento.

Um outro tipo de medida inovadora e que sai fora das medidas-padrão são as chamadas meta-medidas. Algumas dessas medidas coletam dados de fontes de informação tradicionais, maioritariamente, boletins informativos e jornais. Manela e Moreira (2017), por exemplo, construíram uma medida de incerteza que se baseia no conteúdo dos artigos da primeira página do Wall Street Journal (WSJ) com o objetivo de medir a perceção das pessoas sobre a incerteza quanto ao futuro nos Estados Unidos. Os autores concluíram que o índice NVIX³ criado no estudo aumentava significativamente durante as quedas no mercado de ações, os tempos de incerteza política, as guerras mundiais e as crises financeiras.

Baker et al. (2016) desenvolveram um novo índice de incerteza de política económica (EPU) nos Estados Unidos com base também na frequência de artigos de jornais, mas neste caso dos dez mais prestigiados dos EUA, que continham um certo trio de termos relacionados com

³ *News Implied Volatility*

economia⁴ e concluíram que este índice aumenta perto de eleições presidenciais, dos ataques do 11 de setembro, do fracasso do Lehman Brothers, da disputa do tecto da dívida de 2011 e outras grandes batalhas sobre a política fiscal.

Tetlock (2007) relaciona também os media com o mercado de ações nos Estados Unidos, mas baseia-se no conteúdo de uma das colunas do interior do WSJ a qual tem como título “*Abreast of the Market*”. Este autor usa para analisar essa coluna um programa de análise de conteúdo quantitativo que se designa *General Inquirer*, o qual define uma regra de medição para converter o conteúdo das colunas em valores numéricos. A conclusão é que o alto pessimismo dos media prevê uma pressão descendente sobre os preços de mercado seguido de uma reversão em direção aos fundamentais. Também usando este programa de análise de conteúdo, Kaminsky e Schmukler (1999) estudaram o contexto financeiro caótico da crise Asiática (1997-1998), analisando o tipo de notícias que movem os mercados nos dias de mais “nervosismo” de mercado usando como base as notícias da *Bloomberg*. Este estudo conclui que os movimentos mais bruscos são desencadeados por notícias locais ou de países vizinhos, sendo que as que têm mais peso são as relacionadas com acordos com organizações internacionais e agências de notação de crédito.

Brown e Cliff (2004) investigaram a formação de perspectivas sobre o mercado, através de questionários, e tentaram relacioná-las com o sentimento dos investidores, avaliar a sua relação com outras medidas de sentimento e a sua capacidade de prever retornos e descobriram que estes questionários tinham pouco poder em prever retornos a curto-prazo. No entanto, no seu artigo seguinte, Brown e Cliff (2005), mantendo como objetivo o estudo da relação entre o sentimento do investidor e os retornos do mercado, mas agora numa perspectiva de longo-prazo, concluíram que o mercado, regra geral, está sobrevalorizado em períodos de otimismo e que os questionários realizados aos investidores têm uma capacidade preditiva no que toca aos retornos de mercado e são capazes de explicar os desvios dos seus valores fundamentais, querendo isso dizer que, de facto, o sentimento afeta o nível de retornos dos mercados de ações.

Seria de enorme interesse utilizar todo este tipo de medidas de incerteza e de sentimento na dissertação, mas devido à dificuldade de construção da maioria, a qual não está disponível para Portugal, não foi possível.

⁴ “economic” or “economy”; “uncertain” or “uncertainty”; e um ou mais dos seguintes termos: “Congress”, “deficit”, “Federal Reserve”, “legislation”, “regulation” or “White House”.

2.4. Internet e Google

Tal como afirmam Barber e Odean (2001) as atitudes e decisões dos investidores são possíveis de ser afetadas pelos progressos tecnológicos contínuos associados à internet. Assim, nos últimos anos, começaram a surgir medidas de sentimento inovadoras que pretendem lidar com os desenvolvimentos tecnológicos mais recentes e com o uso cada vez mais acentuado das redes sociais e dos motores de pesquisa da internet por parte dos agentes económicos. Este tipo de medidas tornar-se-ão indispensáveis para as ciências sociais e irão complementar as fontes de dados mais tradicionais.

Alguns estudos reúnem dados sobre o consenso público a partir do conteúdo das redes sociais, ou seja, conteúdo gerado por utilizadores, como mensagens em chats da Internet sobre ações (Antweiler e Frank, 2004), *feeds* do Twitter (Bollen et al., 2011) e atividade do Facebook (Mao et al., 2015), ou, até mesmo, mensagens no Yahoo (Das e Chen, 2007).

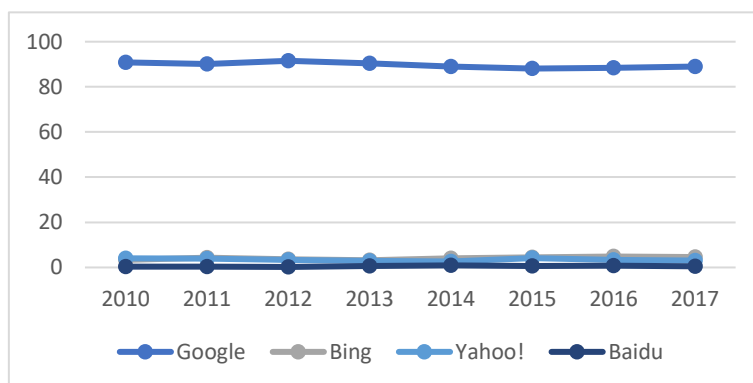
Outra dessas medidas contemporâneas baseia-se em dados sobre a frequência de pesquisas na Internet, conforme proposto por Da et al. (2015), Preis et al. (2013) e Beer et al. (2013). Tal como já foi referido, parte da recolha de informação que sustenta as decisões dos investidores é hoje obtida através dos motores de pesquisa da internet, daí que o comportamento on-line seja presumivelmente representativo das tendências da população em geral. Além disso, dado que deriva da sua espontaneidade apresenta propriedades de sinalização interessantes que permitem produzir um conjunto de dados que possibilita aos investigadores fazer progressos significativos na exploração e compreensão do comportamento humano. Tal como afirma Singer (2002) esta espontaneidade é uma das principais vantagens quando comparado com os dados dos inquiridos à população uma vez que muitas vezes há pouco incentivo para responder às perguntas da pesquisa com sinceridade, especialmente quando as questões são sensíveis.

O Google define como missão: *“to organize the world’s information and make it universally accessible and useful”* e, comparado com outros motores de pesquisa, constitui a fonte privilegiada de informação na referida coleta de informações por parte dos investidores. Segundo os dados da *Statista*⁵, desde o seu “aparecimento” em 1997 que o Google domina o mercado de motores de pesquisa, mantendo, em outubro de 2017, uma quota de aproximadamente, 87%. Este valor contrasta com a percentagem de 5,1% dos utilizadores do

⁵ *Statista* - base de dados de estatísticas online e pesquisas de mercado. Fornece acesso a dados de instituições de pesquisa de mercado e de opinião, bem como de organizações empresariais e instituições governamentais.

Bing, o segundo motor de pesquisa mais utilizado em todo o mundo, tal como podemos verificar no gráfico abaixo.

Gráfico 2.1.: Partilha mundial de mercado dos principais motores de pesquisa (2010-2017)



Fonte: Statista

Dado que satisfaz de forma rápida e eficiente as questões dos investidores, o motor de pesquisa Google tornou-se numa poderosa fonte de procura de informações. Prova disso mesmo é o aumento substancial do volume de pesquisas no Google da palavra “Lehman” após, em setembro de 2008, o banco Lehman Brothers ter declarado falência.

Battelle (2005) classifica assim o motor de pesquisas Google como uma “base de dados de desejos, necessidades, vontades e preferências que podem ser descobertos, intimados, arquivados, seguidos e explorados para toda a espécie de fins”. Por isso mesmo, começa a ser também utilizada, com bons resultados, a base de dados do motor de buscas Google como medida de sentimento económico.

2.5. Google Trends

O Google permite aceder, gratuitamente, aos dados referentes à evolução do número de pesquisas sobre uma determinada palavra ou conjunto de palavras, em termos relativos, através da sua plataforma *Google Trends* disponível desde 2008. Esta disponibilidade de dados e a conclusão de Schmidt e Vosen (2009) de que, em quase todas as suas experiências de previsão, um indicador de atividade de pesquisa no Google supera os conhecidos indicadores baseados em inquéritos, aumentou ainda mais a popularidade das medidas de sentimento que têm como base as pesquisas na internet tanto nas ciências sociais como nas experimentais.

Ao nível da saúde pública, por exemplo, tem sido desenvolvida uma ferramenta, com base nas consultas de termos de pesquisa relacionados com determinadas doenças, como os seus sintomas e tratamentos, que ajuda a prever a incidência dessas doenças. No seu estudo, Ginsberg et al. (2009) conseguiram através destes dados construir uma medida que permite acompanhar a progressão da gripe nos EUA de uma forma muito mais rápida do que o habitual.

No campo da economia, Goel et al. (2010) observaram o comportamento de pesquisas na internet de forma a conseguir prever as vendas de bilhetes de cinema ao fim de semana. Os autores realçam que a utilidade deste indicador se prende com a sua disponibilidade e acessibilidade quase imediata, e não com a sua superioridade em relação aos outros indicadores. McLaren e Shanbhogue (2011) utilizam este método para prever o comportamento dos mercados de trabalho e de habitação, concluindo que existe de facto uma importância cada vez maior na utilização deste tipo de informações para conseguir ter perspetivas sobre a economia do Reino Unido.

São diversos os estudos que empregam estes dados para prever variáveis macroeconómicas como por exemplo o desemprego ou o mercado trabalho. Mas além disso, investigar os ciclos económicos também se tornou uma área com bastante investigação. Chen et al. (2015), por exemplo, concluíram que os dados do volume de pesquisas no Google permitem identificar com maior precisão o ponto de viragem do ciclo económico durante a recessão dos EUA em 2007-2008. Outro exemplo é o comportamento agregado do consumidor, como mostram por exemplo Choi e Varian (2012) que usam dados da atividade do Google para determinar as vendas de automóveis no EUA, os seus destinos de viagem e até mesmo o seu grau de confiança. Schmidt e Vosen (2009) utilizam a mesma fonte de dados para prever o consumo privado dos Estados Unidos.

Nas finanças, Da et al. (2011) utilizam os dados relativos ao volume de pesquisas no Google para avaliar o grau de atenção dos investidores individuais baseado na ideia de que, dadas as restrições de tempo e processamento de informações, os investidores individuais são propensos a adquirir as ações que chamam mais à sua atenção. Noutro estudo, os autores constroem, uma medida de sentimento do investidor com base nas pesquisas da internet realizadas pelas famílias americanas. Uma das conclusões a que chegam é que este índice FEARS⁶ tem uma correlação

⁶ *Financial and Economic Attitudes Revealed by Search*

negativa com os retornos do mercado, mas mostra reversões no futuro, devido ao sentimento dos investidores.

Para Portugal, os estudos de Ferreira (2016) e Brochado (2015) pressupõem também que as atitudes dos investidores reveladas pelo comportamento de pesquisa na internet podem ser consideradas uma medida de sentimento. Combinando a lista final de termos de pesquisa criados por Da et al. (2015) e os resultados de 90 entrevistas pessoais, chegam a uma lista final de palavras que estão positiva e negativamente correlacionadas com os retornos do PSI20 e usam-na como medida de sentimento do investidor.

Noutros estudos, tem vindo a ganhar destaque o poder preditivo destes dados. Bank et al. (2011), por exemplo, mostraram que um aumento no volume de pesquisas está associado a um aumento na atividade de negociação e liquidez das ações. Smith (2012) evidencia que o número de pesquisas no Google para palavras-chave específicas ajuda a prever a volatilidade no mercado da moeda estrangeira. D'Amuri e Marcucci (2017) avaliam o desempenho do índice de volume de pesquisas de emprego no Google como indicador preditivo para a taxa de desemprego mensal dos EUA, concluindo que ele é de facto um bom preditor. Preis et al. (2013) focam-se nos mercados financeiros, tal como nesta dissertação, e afirmam que as fontes de dados resultantes da interação humana com a internet além de refletirem aspetos do estado atual da economia podem oferecer uma visão das tendências futuras dos agentes económicos e sinais de alerta precoce dos movimentos dos mercados acionistas. Joseph et al. (2011) tendo como objetivo perceber se as pesquisas online podem prever retornos anormais do mercado de ações, especialmente de ações mais voláteis e mais difíceis de arbitrar, constroem uma carteira de ações que apresentam maior intensidade de pesquisas através da classificação das ações do SP500 em quantiles e confirmam que a intensidade das pesquisas, de facto, prevê retornos anormais.

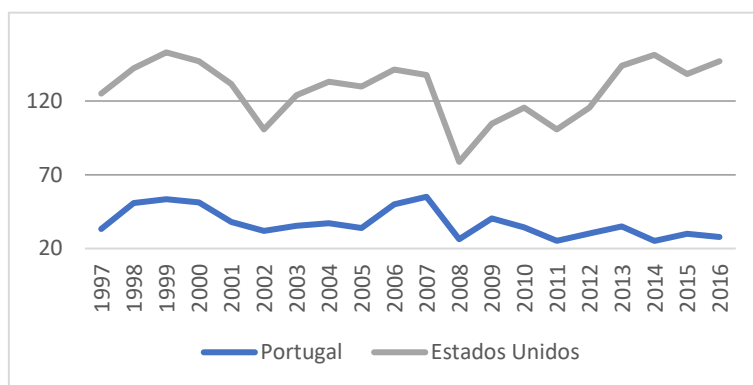
Nesta dissertação, tendo como base estas conclusões achou-se premente a necessidade de comparar este poder preditivo das pesquisas online no Google no que toca aos retornos dos mercados financeiros, em países que em tanto diferem um do outro, daí que se estude Portugal e os Estados Unidos.

3. Comparação da realidade financeira e tecnológica dos EUA e de Portugal

Mais do que determinar de que forma as medidas de sentimento se relacionam com o mercado acionista de cada país, é importante perceber como é que esse efeito se discerne em diferentes países. Na verdade, a maioria das pesquisas anteriores sobre os efeitos do sentimento quase exclusivamente se concentra nos resultados para o mercado de ações dos EUA.

Dado as suas diferentes dimensões, quer em termos populacionais, quer em termos de mercado de ações, mas também dado o nível de penetração da internet em Portugal e nos EUA, nesta dissertação irá fazer-se a comparação entre os dois países.

Gráfico 3.1: *Capitalização de Mercado das empresas domesticas listadas em bolsa (% do PIB)*



Fonte: Banco Mundial

Olhando para os dados do Banco Mundial⁷ referentes à capitalização de mercado, em percentagem do PIB, das empresas nacionais cotadas em bolsa, é notória a diferença entre Portugal e os Estados Unidos. O nível de capitalização das empresas portuguesas nunca ultrapassa sequer o mínimo dos Estados Unidos e a verdade é que existe uma tendência crescente para que os dois se distanciem ainda mais, tal como é visível no gráfico acima.

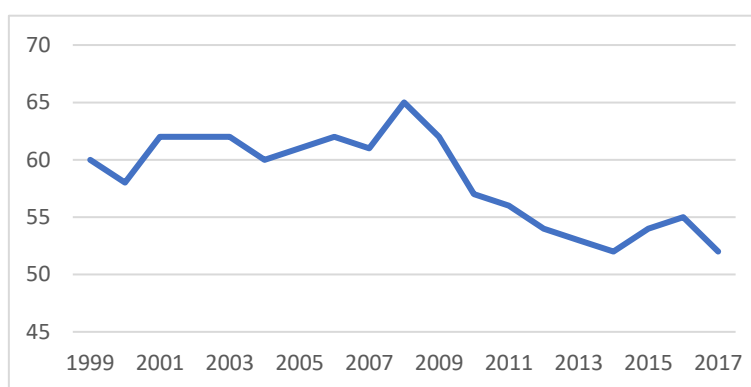
Quanto à percentagem de indivíduos que investem no mercado de ações, a diferença também é notória. Segundo dados da pesquisa anual de economia e finanças da Gallup⁸ cerca de metade

⁷ Banco Mundial (The World Bank) - fonte de assistência financeira e técnica para países em desenvolvimento em todo o mundo. Tem como principal objetivo reduzir a pobreza e apoiar o desenvolvimento.

⁸ <http://news.gallup.com/poll/190883/half-americans-own-stocks-matching-record-low.aspx>

dos americanos (52%) investem em ações, sendo que este valor é o mais baixo desde 1999. Em 2007, cerca de 65% dos americanos investiam no mercado de ações, no entanto, essa percentagem tem vindo a diminuir devido à grande recessão que contribuiu para o aumento da incerteza e insegurança quanto ao futuro da economia e aos meios financeiros sobre o qual se deveria investir. A verdade é que, segundo esta empresa de pesquisa de opinião dos Estados Unidos, o número de ações detido pelos americanos ainda não recuperou para o nível reportado antes da recessão, apesar dos níveis elevados apresentados pelo índice Dow Jones.

Gráfico 3.2: *Percentagem de americanos adultos que investem no mercado de ações*



Fonte: pesquisa anual de economia e finanças da Gallup

Ainda seguindo a mesma fonte, é de realçar que, embora, no geral, haja uma menor propensão para deter investimentos em ações do que antes da recessão, os americanos ditos de classe média foram os que mais “fugiram” deste mercado, ficando apenas no mercado cerca de metade dos adultos desta classe. Esta queda de 22 pontos percentuais face a 2007 é mais do dobro das mudanças observadas no investimento entre os grupos de rendimentos mais altos e mais baixos.

Tabela 3.1: *Americanos que investem no mercado de ações- por idade e rendimento*

	2007	2016	Variação
	%	%	(pontos percentuais)
Adultos	65	52	-13
Menos de \$30.000	28	23	-5
\$30.000-\$74.999	72	50	-22
mais de \$75.000	90	79	-11
18-34 anos	52	38	-14
35-54 anos	73	62	-11
mais de 55 anos	65	56	-9

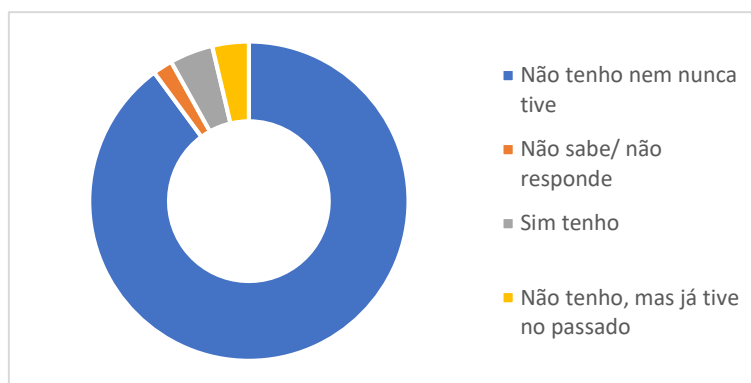
Fonte: pesquisa anual de economia e finanças da Gallup

Os adultos com menos de 35 anos também apresentam uma menor propensão para investir desde a recessão, tal como é visível no quadro acima. Comparativamente com 2007, em que cerca de metade (52%) investiam na bolsa, houve uma diminuição de 14 pontos percentuais no que toca à detenção de ações para esta faixa etária. Embora haja uma convicção crescente de que os jovens devem alocar grande parte do seu rendimento em ações para maximizar os seus retornos de longo-prazo, a maioria dos jovens (18 a 34 anos), não está atenta a este conselho dos consultores financeiros. Eles preferem economizar o seu dinheiro ou investir em imóveis, tal como é afirmado pelo estudo que tem vindo a ser mencionado até aqui.

Concluindo, apesar de mais de metade dos americanos continuar a investir no mercado de ações, esta percentagem está longe dos níveis de pré-recessão. A confiança no mercado e os níveis de literacia financeira sofreram nos últimos anos e isso levou a que as taxas de investimento não fossem capazes de acompanhar a recuperação geral do mercado.

Mesmo este valor historicamente baixo dos investidores americanos contrasta com a percentagem de portugueses que detêm investimentos em valores mobiliários, a qual, segundo o inquérito à literacia financeira realizado pelo Conselho Nacional de Supervisores Financeiros em todo o território nacional entre maio e junho de 2015 era apenas de 4,4%. Destes 4,4%, cerca de 80% investiam em ações e 41,7 % em unidades de participação de fundos de investimento. Este conjunto de investidores aponta como principais razões para investir em valores mobiliários, o rendimento esperado superior ao dos depósitos bancários e o facto de estes serem recomendados nos balcões dos seus bancos. Cerca de 40 % dos investidores afetam a valores mobiliários menos de 10% do seu património e nenhum investidor tem mais de 50% do seu património investido nestes produtos. A generalidade dos entrevistados financia os seus investimentos através de poupanças próprias.

Gráfico 3.3: *portugueses com investimentos em valores mobiliários*



Fonte: inquérito à literacia financeira, ASF

Entre os entrevistados que não têm investimentos em valores mobiliários, 65,3% indicam não ter rendimentos que justifiquem este tipo de investimento e 29,2% têm como principal receio a incerteza inerente ao mercado, especialmente no que toca à ocorrência de *crashes*.

Quanto à “sabedoria” dos investidores relacionada com os produtos financeiros, a perceção dos portugueses sobre os próprios conhecimentos financeiros encontra-se algo desalinhada do seu conhecimento efetivo. A generalidade não compreende o conceito de capital garantido num valor mobiliário, no entanto os entrevistados revelam algum conhecimento sobre o mercado de ações, com 72,4% a identificarem corretamente a relação entre risco de investimento e diversificação da carteira de ações. Assim, Portugal surge em 10º lugar no indicador global de literacia financeira da OCDE sendo este liderado pela França.

Já em 2009, um inquérito conduzido pela CMVM sobre o perfil do investidor particular português afirmava que menos de 10% das famílias portuguesas possuíam, além dos tradicionais depósitos a prazo e certificados de aforro, outros ativos financeiros tais como ações e obrigações. Esta reduzida participação das famílias portuguesas no mercado de valores mobiliários materializou-se numa também reduzida percentagem da capitalização acionista das nossas empresas que é detida pelas famílias portuguesas.

Esse estudo concluiu também que o escalão etário é importante na probabilidade de ser investidor, o seu gráfico assume assim a forma de U invertido, pois a probabilidade aumenta até ao escalão que termina nos 64 anos, diminuindo para as idades mais avançadas. A diferenciação por género também se mostra relevante, com os homens a participarem mais do que as mulheres nos mercados financeiros. Tal facto é também defendido por Barber e Odean (2001) que usam uma das conclusões da psicologia para justificar este facto. A verdade é que

os homens são mais propensos a um excesso de confiança do que as mulheres, em particular nos domínios controlados pelos homens, como as finanças e por isso tendem a comercializar mais do que as mulheres.

Em simultâneo, aumentos do rendimento mensal do agregado familiar, assim como do nível de escolaridade do indivíduo, têm um impacto positivo na propensão para ser investidor. Profissões menos qualificadas estão associadas a uma menor probabilidade de ser investidor e profissões mais qualificadas permitem aumentar a participação nos mercados financeiros.

No que toca ao nível de penetração da internet é importante perceber que a adoção das novas tecnologias difere e continuará a diferir tanto em países como em indivíduos (Askitas e Zimmermann, 2015). Olhando os dados da *Internet World Stats*⁹ é notório o maior nível de penetração da internet nos Estados Unidos, chegando quase aos 90% em 2017, comparando com os pouco mais de 70% em Portugal. A diferença entre os dois países é ligeiramente maior no que toca à percentagem de indivíduos que utilizam o Facebook, uma das maiores redes sociais do mundo. É também visível o que foi mencionado acima relativamente ao facto de nem todos os indivíduos com acesso à internet terem uma vida social em rede ativa.

Tabela 3.2: *Nível de penetração da internet*

	População	Utilizadores de internet	Penetração (%população)	Facebook
Portugal	10264797	7430762	72,40%	57%
EUA	326474013	286942362	87,90%	74%

Fonte: *internet world stats*

Esta análise comparativa é de extrema importância porque apesar de todos termos uma ideia de que Portugal se diferencia dos EUA é importante olhar para valores que nos permitam ficar com uma ideia mais clara de como é que são efetivamente essas diferenças. Nesta tese, e dado que normalmente os artigos se centram no estudo de apenas um país, decidiu-se analisar estes dois países em simultâneo, de forma a perceber como é que estas desigualdades, quer em termos de mercados acionistas, quer em termos de penetração de internet, quer também a nível de literacia financeira, se materializam no estudo da previsão dos retornos que usam medidas de

⁹ *Internet World Stats* - site internacional que apresenta estatísticas relativas ao atual uso da Internet mundial, à população, à media social e aos dados de pesquisa de mercado na Internet, para mais de 243 países e regiões do mundo.

sentimento que têm por base pesquisas na internet. Será que numa economia com maior acesso à internet estas medidas são mais adequadas para prever retornos? Será que o maior nível de literacia financeira tem impacto na relação entre estas medidas de sentimento contemporâneas e os retornos das ações? Será que as medidas mais teóricas, como o CAPE, tem vantagem preditiva sobre um índice de sentimento Google, num mercado acionista mais desenvolvido? São tudo perguntas ao qual se procura dar uma resposta fazendo a análise dos dois mercados acionistas, durante o mesmo período temporal, usando as mesmas variáveis e a mesma metodologia para que tudo seja comparável.

4. Dados e Metodologia

4.1. Índice de sentimento Google

Este estudo irá, primeiramente, debruçar-se sobre a construção de uma variável através dos dados de pesquisa on-line do Google, que servirá como *proxy* do sentimento do mercado dado que, tal como visto anteriormente, se concluiu em alguns artigos que este é um importante preditor dos retornos do mercado. Através do *Google Trends* é possível ter acesso aos dados referentes às consultas na internet de um termo de pesquisa específico e restringir a pesquisa em termos de localização geográfica, bem como em termos temporais. O gráfico em anexo (A - Figura 4.1) retirado do *Google Trends* reflete a quantidade de pesquisas do termo “recessão” em Portugal entre 2004 e o presente. É de facto notório o aumento deste termo de pesquisa entre 2008-2009, plena recessão no nosso país após a grande crise financeira mundial 2007-2008.

É necessário salientar que os dados fornecidos por esta ferramenta não são um indicador do volume de pesquisa absoluto, sendo sim medidos relativamente à maior quantidade total de pesquisas, ou seja, os dados fornecidos são normalizados, sendo apresentados em referência ao número máximo de *clicks* que o termo de pesquisa alcançou durante o período de tempo especificado. Os valores variam, assim, de 0 a 100 sendo que o ponto mais alto do gráfico terá sempre o valor de 100, traduzindo o período em que a quantidade de pesquisas foi mais elevada.

A definição dos termos de pesquisa relevantes para a investigação é considerada um passo crucial para um estudo deste tipo (Da, et al., 2015). É importante usar termos precisos e apropriados à pesquisa que se realiza. Existem algumas palavras que não são objetivas. Um desses exemplos é a palavra depressão que tanto tem uma conotação financeira como pode ser pesquisada, por exemplo, para descobrir os sintomas do distúrbio mental com o mesmo nome. Assim, e para eliminar alguma subjetividade, usar-se-á o *Harvard IV Dictionary*, tal como Da, et al. (2015), Tetlock (2007) e Beer, et. al (2013) que classifica um conjunto de termos em económicos e financeiros e em positivos e negativos. Além disso, e ao contrário do estudo qualitativo usado pela Brochado (2015) que teve como principal objetivo captar um conjunto de termos mais técnicos usados por profissionais da área económica aquando de pesquisas online, aqui, sendo o objetivo captar o sentimento da população em geral, usar-se-ão termos pouco técnicos tal como em Da, et al. (2015).

Seguindo a metodologia dos mesmos autores aquando da construção do seu índice de sentimento FEARS¹⁰ chegamos a uma lista final de cerca de 120 termos de pesquisas que revelam o sentimento dos investidores em relação às condições económicas do seu país (Ver Tabela 4.1 do anexo A). Esta lista obtém-se depois de se identificar, através da plataforma Google Trends, os 10 principais termos de pesquisa relacionados com cada um dos termos e se eliminarem os termos não económicos. Deste conjunto de 120 termos eliminam-se, tal como referido anteriormente, termos muito técnicos e não diretamente económicos e além disso termos com o mesmo significado, ou seja, ter inflação e taxa de inflação é praticamente a mesma coisa, tal como referido por Beer, et al. (2013) e por isso não há necessidade de ter os dois termos na amostra. Ficamos assim com uma lista final de cerca de 60 termos com significado económico geral e familiar.

Neste caso, e uma vez que existe uma preferência para o investidor comum realizar pesquisas usando a sua língua-mãe em vez de usar uma linguagem mais global onde poderia obter informações mais fidedignas e diversificadas, utiliza-se os termos, referentes à lista mencionada acima, em português para a construção do índice de sentimento para Portugal e em inglês para o índice americano, restringindo a pesquisa a cada um dos países e para o período compreendido entre 2004-2017, devido à disponibilidade dos dados do *Google Trends*.

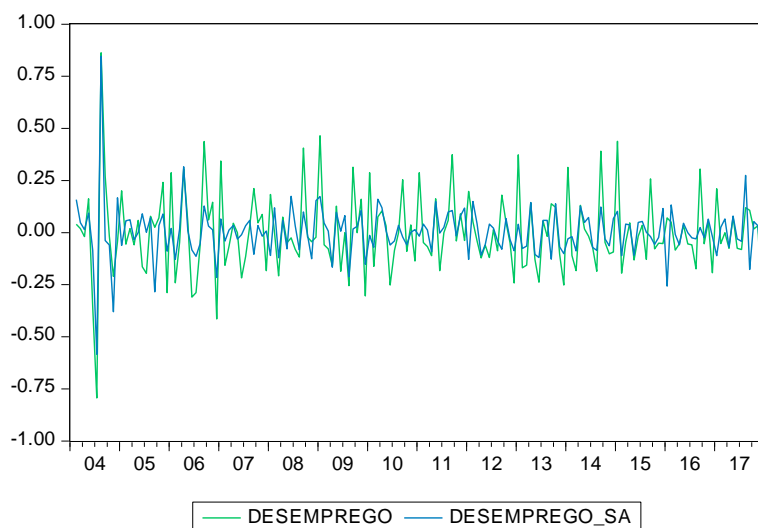
Depois de fazer o download das séries temporais com frequência mensal referentes a todos os termos, construiu-se o logaritmo da série, para que esta tivesse uma melhor e mais fácil interpretação. Para isso tiveram de se substituir por 0,1 as series temporais associadas aos termos que possuíam um índice do volume de pesquisas igual a zero num ou mais períodos. Em seguida, os dados foram transformados em variações de forma a captar as mudanças que se verificam entre os meses.

Para tornar a lista final de termos comparável e perante a possível presença de sazonalidade e heteroscedasticidade nos dados, realizaram-se outras transformações. No que toca à sazonalidade, e dado que os dados do Google têm, segundo D'Amuri e Marcucci (2017) uma sazonalidade peculiar uma vez que, geralmente, em novembro e dezembro, o denominador do índice é inflacionado por pesquisas relacionadas com o Natal, começou-se por fazer um teste de Holt-winters a cada uma das séries para verificar a existência ou não de sazonalidade. Quando existia, usava-se a série desazonalizada que se obtém com a realização do teste census X12. Por exemplo, analisando o output referente ao Holt-Winters da série do termo desemprego

¹⁰ Financial and Economic Attitudes Revealed by Search

é notória a existência de sazonalidade daí que se passe a usar a série desazonalizada. Tal como se pode verificar no gráfico abaixo existe uma grande diferença entre a varável desazonalizada (desemprego_sa) e a série primitiva (desemprego), provando assim o que já se exibia no output do Holt-winters.

Gráfico 4.1: sazonalidade



Para abordar a heterocedasticidade nos dados e para tornar as séries temporais comparáveis, padronizou-se cada série temporal pelo desvio padrão. Para isso analisou-se a heterocedasticidade condicional dos resíduos verificando se existiam efeitos ARCH. Nos casos em que não existiam dividiu-se a série pelo desvio-padrão amostral e nos casos em que existia, estimou-se o modelo com efeitos ARCH, e dividiu-se a série pelo desvio-padrão estimado, ele próprio uma série temporal.

Depois do breve tratamento dos dados referido acima, analisou-se quais os termos mais negativamente correlacionados respetivamente com o mercado dos EUA e com o mercado português. Optou-se pela análise apenas dos termos negativamente correlacionados uma vez que esses termos são mais adequados para identificar o sentimento (Tetlock, 2007). Da, et al. (2015) tendo isto como base construíram o seu índice de sentimento negativo Google (FEARS) usando apenas termos que estão negativamente correlacionados com os retornos. Mesmo outros estudos que optaram pela construção do índice positivo e negativo acabaram por concluir que o índice Google é de facto mais eficaz a explicar o sentimento negativo dos investidores (Ferreira, 2016). Isto prova que o sentimento negativo tem um impacto maior nas forças que impulsionam o mercado daí esta opção.

Deste conjunto de termos, selecionam-se, opcionalmente, os 15 termos com maior correlação negativa com os retornos de cada mercado usando o total da amostra e para cada um desses termos aplicou-se a metodologia de Da, et al. (2015), ou seja, realizaram-se regressões *rolling backward-looking* com o objetivo de identificar os termos com maiores correlações históricas com os retornos do mercado. Os primeiros 4 anos da amostra (48 observações mensais) foram deixados de fora, devido à necessidade de uma janela inicial de pelo menos seis meses, e cada uma destas regressões é feita de 6 em 6 meses. Optou-se por ordenar os 10 termos com maior significância estatística em cada um dos subperíodos de estimação e construiu-se o índice de sentimento negativo Google para esse período semestral sendo ele a soma desses termos a dividir por 10. Como era de esperar estes termos vão mudando ao longo do tempo daí que nem sempre o índice seja constituído pelos mesmos 10 termos. Só para se ficar com uma ideia das palavras que constituem este índice mencionam-se, em seguida, os 10 termos constituintes do índice de cada país no período compreendido entre dezembro de 2007 e junho 2008.

EUA									
Deficit	Debt	Bankruptcy	Government	Depression	Jobless	Corruption	Capitalize	Recession	Unemployment
Portugal									
Produção	Recessão	Inflação	Lucrativo	Emprego	Tarifa	Empreendedorismo	Desemprego	Défice	Produtividade

Chega-se assim ao fim da construção do índice de sentimento negativo Google português e americano, o qual será usado posteriormente para averiguação da sua capacidade explicativa e de previsibilidade em termos dos retornos da bolsa dos respetivos países.

4.2.Índice CAPE

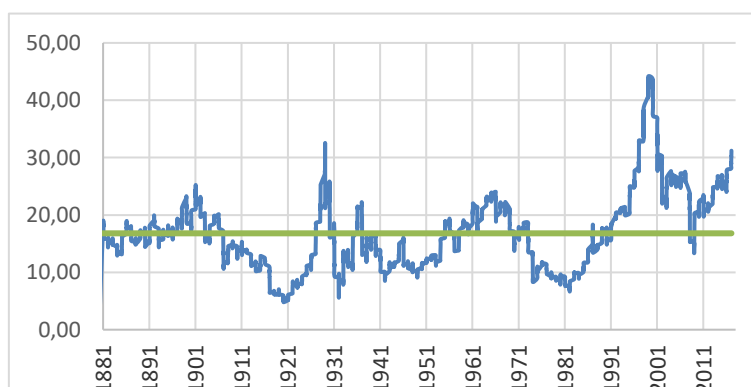
Os ganhos anuais das empresas podem ser voláteis, especialmente para empresas com altos custos fixos, dependendo da fase do ciclo económico. Como resultado, um índice de preço/lucro, com base nos ganhos do ano que se segue, varia significativamente de ano para ano, tornando difícil a sua utilização como uma medida de previsibilidade dos ganhos do mercado de ações.

Depois de alguns trabalhos nesta área, Campbell e Shiller (2001) concluíram que os retornos do mercado de ações de longo prazo não eram um passeio aleatório, mas, em vez disso,

poderiam ser previstos por uma medida de avaliação do mercado de ações a que chamaram rácio CAPE. Este rácio calcula-se dividindo o *real price* do SP500 (para o mercado americano) pela média dos retornos reais dos últimos 10 anos. Os economistas preferem ter esta longa visão retrospectiva dos retornos uma vez que, desta forma, alisam as flutuações temporárias que se dão ao longo do ciclo económico, e além disso conseguem perceber mais facilmente o que um investidor gostaria de ter feito em determinada altura do ciclo. Quando o mercado diverge das suas tendências históricas acaba sempre por regressar à média, por exemplo quando as ações parecem baratas acabaram por recuperar. De modo que deste rácio preço/lucros a longo-prazo parece resultar alguma capacidade de previsão.

O CAPE tem sido recentemente tema de debate no seio financeiro, uma vez que, tal como referido por Boushelle (2014), o índice se encontra bastante acima da média (Gráfico 4.2). Este facto preocupou os especialistas, receando que os mercados estivessem a atingir um “*bubble territory*”. No entanto, o índice CAPE não deve ser usado como "um mecanismo de cronometragem", mas sim como uma forma de avaliar a ponderação relativa dos ativos que um investidor possui no seu portfólio para a tomada de decisões de alocação de ativos de longo prazo. O autor sugere, assim, que os investidores devem tentar reduzir o seu risco nas carteiras, mas devem continuar a investir em ações.

Gráfico 4.2: CAPE- EUA



Fonte: página pessoal de Robert Shiller

Um dos motivos sugeridos por Siegel (2016) para os elevados valores do CAPE é que os investidores são excessivamente otimistas sobre o crescimento dos ganhos futuros e, quando

esse crescimento não se materializa, os investidores vendem fazendo com que os preços das ações diminuam.

Dado que para os EUA este índice se encontra disponível na página pessoal de Shiller¹¹, para o caso português retiraram-se os dados trimestrais dos *earnings* e mensais do *last price* do PSI20 na Bloomberg e do IPC¹² mensal no Banco de Portugal e replicou-se a metodologia para o nosso país, ficando assim com o índice CAPE para Portugal. Tal índice é usado também nos relatórios da CMVM sobre a sua atividade e sobre os mercados de valores mobiliários. Nesta dissertação servirá como uma importante variável de controlo dado que estudos como o de Campbell e Shiller (1998) mostraram que existe uma forte correlação negativa entre o CAPE e os retornos futuros do mercado de ações de longo-prazo, em média e segundo Keimling (2016) e Gray e Vogel (2014) este é um dos fatores explicativos dos retornos do mercado de ações.

4.3. Retornos e variáveis de controlo

Antes de mais é necessário definir os índices de mercado de ambos os países que servirão de base para o estudo. Os retornos destes índices constituem a variável dependente dos modelos, ou seja, a que se pretende explicar e prever através da conjugação das outras variáveis do estudo e calculam-se através dos dados mensais referentes aos *opening prices* (P) de cada um dos índices fornecidos pela Bloomberg.

$$\text{Retornos} = \ln(P_t) - \ln(P_{t-1})$$

O PSI20 é o índice mais importante da bolsa portuguesa agregando a capitalização bolsista de cerca de vinte grandes empresas que cumprem certos critérios definidos pela Euronext. O SP500 trata-se de um índice de mercado americano baseado nas capitalizações de mercado de 500 grandes empresas com ações listadas na NYSE ou na NASDAQ. Os constituintes do índice SP500 e as suas ponderações são determinados pelos índices SP Dow Jones. Difere de outros índices do mercado de ações dos EUA, como a *Dow Jones Industrial Average* ou o índice *Nasdaq Composite* por causa da sua diversa metodologia de constituição e ponderação. É um dos índices de ações mais comumente seguidos e muitos consideram que é uma das melhores representações do mercado de ações dos Estados Unidos e uma base para a economia do país. Com o objetivo de estudar de que forma o índice de sentimento Google influencia os mercados

¹¹ <http://www.econ.yale.edu/~shiller/data.htm>

¹² Índice de preços ao consumidor

financeiros mencionados acima, além de o comparar com o CAPE descrito anteriormente, achou-se necessário introduzir no modelo algumas variáveis de controlo: taxa de crescimento do PIB, taxa de Desemprego e Inflação (calculada através da variação homóloga do IHPC) que servem para incorporar informações fundamentais de carácter macroeconómico nos mercados financeiros. Como a literatura considera a *dividend yield*¹³ (Golez e Koudijs, 2017) de enorme importância para a previsão dos retornos usa-se também esta variável. Além disso, é usado o *spread* da dívida pública¹⁴ uma vez que existe, de facto, uma forte relação entre os bilhetes do tesouro e as ações. Ao longo dos últimos 15 anos, as ações tiveram um bom desempenho quando o *spread* estava relativamente elevado.

Servindo como termo de comparação para o índice Google adicionaram-se alguns indicadores de sentimento económico. O indicador de confiança do consumidor que é obtido através de inquéritos à população em geral, ou seja, a indivíduos que não detêm, necessariamente, formação económica, que incidem sobre as suas intenções de consumo atuais e futuras. O índice de sentimento económico que diz respeito a um índice composto obtido através da conjugação de cinco indicadores de confiança de vários sectores da economia, ou seja, além da confiança dos consumidores, é um indicador de confiança de outros setores da economia. E, por fim, o *turnover* que se refere ao total de ações negociado durante um determinado período de tempo, ou seja que mede o volume de negociação dos mercados de ações também é aqui usado como indicador de sentimento económico tal como em Da, et al. (2015) e espera-se que este aumente à medida que as pessoas se tornam mais otimistas.

Estas variáveis de controlo encontram-se descritas na Tabela 4.2, do anexo A, no qual se identifica a sua denominação, a sua frequência e a sua fonte.

4.4. Tratamento dos dados

A validade do trabalho empírico, nomeadamente no que toca às regressões lineares, pressupõe a estacionariedade das séries. Assim, começou-se por realizar testes Augmented Dickey-Fuller (ADF) com vista a aferir se as séries usadas na dissertação eram estacionárias concluindo-se que eram todas estacionárias. Além disso, foram submetidas a uma análise de sazonalidade e

¹³ Dividendo de 12 meses por ação a dividir pelo seu preço.

¹⁴ Diferença entre a taxa mensal de rendibilidade de obrigações do tesouro emitidas a taxa fixa, por prazo residual - 10 anos e a taxa mensal de rendibilidade de obrigações do tesouro emitidas a taxa fixa, por prazo residual - 2 anos.

heterocedasticidade seguindo a metodologia aplicada aos termos de pesquisa constituintes do índice de sentimento Google.

4.5. Hipóteses a testar, metodologia e período amostral

O objetivo da próxima secção é responder à pergunta de investigação desta dissertação: de que forma a medida de sentimento Google está relacionada com os movimentos do mercado acionista, quer em Portugal quer nos EUA e se existem ganhos preditivos na sua utilização relativamente aos fundamentais, no período compreendido entre dezembro 2007 e junho 2017. Assim, a metodologia que se propõe de forma a responder a esta questão consiste essencialmente em três fases.

Para perceber melhor como é que as várias medidas de sentimento se relacionam entre si, começa-se por fazer uma matriz de correlações e por estudar a relação de causalidade à Granger existente entre elas. Além disso, dado que um dos objetivos secundários da dissertação é fazer uma comparação entre os mercados acionistas de Portugal e dos Estados Unidos, procede-se também ao estudo das relações existentes entre os índices e algumas das variáveis macroeconómicas que caracterizam a economia dos dois países.

Numa segunda fase, o estudo centra-se na relação do sentimento de mercado medido pelos dados fornecidos pelo Google, com os retornos do PSI20 e do SP500. Para realizar este teste, aplicar-se-ão regressões lineares OLS (Hamilton, 1992; Wooldridge, 2009), usando os dados do Google como variável independente, enquanto que ao mesmo tempo se usam outras variáveis de controlo e alguns indicadores de sentimento económico tal como referido anteriormente. Além disso, a *proxy* de sentimento criada será comparada com o índice CAPE de Robert Shiller no que toca à sua capacidade para explicar o comportamento dos mercados em análise.

Posteriormente, faz-se uma análise do poder preditivo do índice Google e do índice CAPE no que toca aos retornos dos mercados acionistas usando uma série de modelos como em D'Amuri e Marcucci (2017) em que o modelo vai sendo aumentado com diferentes indicadores como o índice de sentimento Google negativo, para cada um dos mercados, conforme se esteja a estudar Portugal ou os EUA. É necessário salientar que foram deixadas de parte 24 observações, o que corresponde a dois anos, para se proceder à avaliação da previsão *out-of-sample*. Ou seja, começa-se por averiguar de que forma o índice de sentimento google e o CAPE melhoram o

modelo dito de base *in-sample* (dezembro 2007 a junho 2015) e posteriormente, usando-se esse modelo estimado *in-sample*, procede-se à previsão *out-of-sample*. Começa-se por estimar cada modelo para o período dezembro 2007 a julho 2015, faz-se a previsão a h passos (meses) para esse período e depois vai-se fazendo o *rolling* da amostra, adicionando sempre mais uma observação à amostra observada e reestimando o modelo, isto é, a amostra passa a ser dezembro 2007 a agosto 2015, dezembro 2007 a setembro 2015 e assim sucessivamente.

5. Resultados

5.1. Matriz de Correlações

Com o objetivo de obter uma imagem geral sobre a relação existente entre as várias medidas de sentimento usadas na dissertação, fez-se uma matriz de correlações entre elas para cada um dos países em estudo. Nos Estados Unidos existe uma correlação bastante significativa entre o índice CAPE e o índice de confiança dos consumidores e entre o índice de sentimento Google e o *turnover* do SP500. Quanto ao índice de sentimento Google criado no estudo este é negativamente correlacionado, embora não estatisticamente significativo, com o índice de confiança dos consumidores como já se esperava uma vez que o índice de sentimento Google se foca nas palavras negativamente correlacionadas com os retornos do mercado os quais dependem positivamente da confiança dos consumidores.

Olhando para a matriz de correlações referente às *proxies* de sentimento para Portugal podemos afirmar que as correlações existentes entre as variáveis não são estatisticamente significativas. No entanto, a correlação mais elevada, apesar de negativa, dá-se entre o índice de sentimento económico e o CAPE. Quanto ao índice de sentimento Google este é negativamente correlacionado com o CAPE e o índice de confiança dos consumidores tal como acontece nos Estados Unidos, não sendo esta correlação significativa, tal como afirmado anteriormente.

No geral, a variável de sentimento Google criada não tem grande correlação com as outras *proxies* de sentimento mais conhecidas o que pode ser um indício de que os dados do Google não englobam totalmente os efeitos do sentimento no mercado. Os valores estimados das correlações apresentam-se em seguida nas tabelas 5.1 e 5.2.

O google como medida de sentimento nos mercados financeiros

Tabela 5.1: *Matriz de Correlações - EUA.*

Correlação valor-p	CAPE_EUA	CONFIANCA_EUA	GOOGLE_EUA	TURNOVER_EUA	SENTIMENTO_EUA
CAPE_EUA	1,000 -----				
CONFIANCA_EUA	0,759*** 0,000	1,000 -----			
GOOGLE_EUA	-0,009 0,924	-0,063 0,501	1,000 -----		
TURNOVER_EUA	0,085 0,366	0,112 0,234	0,332*** 0,000	1,000 -----	
SENTIMENTO_EUA	0,099 0,292	0,073 0,438	0,064 0,494	0,073 0,437	1,000 -----

Nota: *, **, *** denotam níveis de significância estatística de 10%, 5% e 1%, respetivamente.

Tabela 5.2: *Matriz de Correlações - Portugal.*

Correlação valor-p	CAPE_PT	CONFIANCA_PT	GOOGLE_PT	TURNOVER_PT	SENTIMENTO_PT
CAPE_PT	1,000 -----				
CONFIANCA_PT	0,008 0,931	1,000 -----			
GOOGLE_PT	-0,029 0,756	-0,069 0,464	1,000 -----		
TURNOVER_PT	-0,025 0,791	-0,058 0,536	0,039 0,679	1,000 -----	
SENTIMENTO_PT	-0,107 0,258	0,091 0,332	0,017 0,858	0,014 0,885	1,000 -----

Nota: *, **, *** denotam níveis de significância estatística de 10%, 5% e 1%, respetivamente.

No que toca à relação existente entre os mercados acionistas dos dois países, olhando para os retornos pode-se caracteriza-la como tendo uma correlação bastante significativa, tal como se pode verificar no quadro abaixo.

Tabela 5.3: *Matriz de Correlações - Portugal e EUA,*

Correlação valor-p	RETORNOS_EUA	RETORNOS_PT
RETORNOS_EUA	1,000 -----	
RETORNOS_PT	0,545*** 0,000	1,000 -----

Nota: *, **, *** denotam níveis de significância estatística de 10%, 5% e 1%, respetivamente.

5.2. Causalidade à Granger

Nesta secção analisam-se a eventual existência e, a existir, o sentido das causalidades entre as diversas variáveis em estudo, usando para isso testes de causalidade com 2 e 3 lags. O teste de causalidade previsional mais conhecido e utilizado na literatura e no seio da econometria deve o nome a Granger e resumidamente ajuda a perceber de que forma duas séries A e B se relacionam entre si, ou seja, se A precede a B ou B precede a A ou se A e B ocorrem simultaneamente. Aqui o objetivo é avaliar se as variáveis de sentimento causam, no sentido de Granger, os retornos da bolsa de cada um dos países e não o contrário para que o estudo seja relevante.

Depois de efetuados os testes *pairwise*, ou seja, em que se estuda a relação de causalidade entre os retornos da bolsa de cada país com cada uma das *proxies* de sentimento, individualmente, confirma-se que as variáveis de sentimento causam, no sentido de Granger, os retornos da bolsa nos dois países e não no sentido contrário, o que ajuda a explicar a escolha da regressão linear dos retornos em função das medidas de sentimento. É também de realçar a relação de causalidade entre os retornos da bolsa dos Estados Unidos e os do PSI20 sendo estes últimos causados pela performance do SP500.

Fazendo um estudo mais abrangente com o mesmo objetivo e apesar de não achar necessário expor os outputs referentes, é de interesse realçar algumas das relações existentes entre certas variáveis do estudo. No caso americano destaca-se o facto do desemprego causar, no sentido de Granger, o índice de sentimento Google, quer com 2 quer com 3 lags, o que faz sentido dado que havendo mais desemprego, as medidas de sentimento irão sofrer alterações devido ao estado de espírito mais deprimido dos investidores daí esse desemprego se refletir no índice de sentimento desenvolvido no estudo. Pode também ser olhado na perspetiva de que havendo desemprego, sendo este um dos termos constituinte do índice Google ao longo da amostra é natural que se pesquise mais esta palavra daí esta relação. O desemprego aparece também como causa dos retornos, os quais causam, no sentido de Granger, as yields, o que já era de esperar tendo em conta a sua relação descrita no capítulo das variáveis de controlo. Olhando para as relações de causalidade existentes entre as variáveis portuguesas destaca-se o facto de o índice de confiança do consumidor causar, no sentido de Granger, os retornos do PSI20, os quais por sua vez causam o CAPE. Existe uma causalidade bilateral entre o índice de sentimento económico e o PIB português o qual é também afetado pelos retornos da bolsa portuguesa.

5.3. Estimações OLS

Focando o objetivo primordial desta dissertação, nesta secção averiguar-se-á de que forma o índice Google transmite o sentimento dos investidores e como influencia os mercados financeiros quantificando o poder explicativo desta nova *proxy* no que toca aos retornos dos mercados acionistas dos dois países em estudo. Para realizar este teste, aplicaram-se regressões lineares OLS, usando os dados do Google como variável independente.

Começa-se por uma regressão simples usando como variável dependente os retornos do mercado acionista de cada um dos países e o índice de sentimento Google como variável explicativa, tal como mencionado anteriormente. Depois vai-se acrescentando individualmente as restantes variáveis de controlo do estudo, ficando assim com um total de 10 regressões para cada país. À medida que o modelo se torna mais complexo, capta de maneira mais completa os diferentes aspetos que influenciam os retornos do mercado.

Olhando aos outputs do anexo B (Tabela 5.4 e 5.5) conclui-se que, quer em Portugal quer nos Estados Unidos, o índice de sentimento Google apresenta uma grande significância estatística e correlaciona-se negativamente com os retornos do mercado acionista de cada um dos países, tal como já era de esperar. Nos EUA um dos indicadores de sentimento mais significativos,

além do índice Google, é o *turnover* o qual apresenta uma correlação positiva com os retornos do SP500. No entanto, a sua significância estatística diminui à medida que se adicionam variáveis de controlo na regressão. Comparando com o índice de sentimento negativo Google, o *turnover* dos EUA apresenta menores coeficientes e menor significância estatística daí que se conclua que, tal como mencionado em artigos anteriores, este índice Google criado no estudo se relacione melhor com os retornos do mercado acionista do país. Tal é também visível em Portugal sendo o índice Google o indicador de sentimento com maiores coeficientes e maior significância estatística, os quais perduram, não obstante a introdução das variáveis de controlo nas regressões.

5.4.Previsão

Dados os resultados empíricos da última secção de que o índice de sentimento Google apresenta uma correlação contemporânea significativa com o mercado acionista quer de Portugal quer dos EUA e dado que este índice causa à Granger os retornos dos mercados acionistas mencionados, aqui o objetivo é testar se este índice Google tem também capacidade preditiva no que toca aos retornos do mercado de ações dos dois países, e verificar se existem ganhos de previsão quando se adicionam aos modelos este índice de sentimento Google ou o CAPE. As previsões dentro da amostra testam o poder preditivo do respetivo indicador ao longo de todo o período da amostra, variando de dezembro de 2007 a junho de 2015, enquanto os testes fora da amostra investigam a estabilidade e qualidade desse poder preditivo ao longo de vários subperíodos, desde julho 2015 a junho 2017.

Começa-se com um modelo base *in-sample* onde se inclui apenas as variáveis consideradas pela literatura como sendo fundamentais para prever os retornos do mercado de ações desfasadas um período e as variáveis de sentimento mais tradicionais também desfasadas um período.

Ou seja, começa-se por usar como variáveis explicativas a *dividend-yield*, o *spread* e as variáveis macroeconómicas: desemprego, IHPC e PIB. E para Portugal os próprios retornos desfasados 4 períodos devido à sua FACP (função autocorrelação parcial) indicar para a presença de alguma relação entre ambas e este lag se mostrar estatisticamente significativo aquando da realização de um AR(4). Pela mesma razão, para os EUA usam-se também os retornos com lag de 1 mês. Além disso, adicionam-se ao modelo as variáveis de sentimento também desfasadas 1 período (índice de confiança ao consumidor, índice de sentimento económico e *turnover*):

$$\begin{aligned} \text{retornos}_{pt} = & c + \beta_1 * \text{retornos}_{pt}(-4) + \beta_2 * \text{yield}_{pt}(-1) + \beta_3 * \text{spread}_{pt}(-1) \\ & + \beta_4 * \text{desemprego}_{pt}(-1) + \beta_5 * \text{ihpc}_{pt}(-1) + \beta_6 * \text{pib}_{pt}(-1) + \beta_7 \\ & * \text{confiança}_{pt}(-1) + \beta_8 * \text{sentimento}_{pt}(-1) + \beta_9 * \text{turnover}_{pt}(-1) \\ & + \mu_{pt} \end{aligned}$$

$$\begin{aligned} \text{retornos}_{eua} = & c + \beta_1 * \text{retornos}_{eua}(-1) + \beta_2 * \text{yield}_{eua}(-1) + \beta_3 \\ & * \text{spread}_{eua}(-1) + \beta_4 * \text{desemprego}_{eua}(-1) + \beta_5 * \text{ihpc}_{eua}(-1) \\ & + \beta_6 * \text{pib}_{eua}(-1) + \beta_7 * \text{confiança}_{eua}(-1) + \beta_8 \\ & * \text{sentimento}_{eua}(-1) + \beta_9 * \text{turnover}_{eua}(-1) + \mu_{eua} \end{aligned}$$

Posteriormente, analisa-se o contributo, a existir, da adição do índice de sentimento Google e do CAPE desfasados 1 período no que toca à previsibilidade dos retornos do mercado de ações de cada um dos países em estudo. Para testar a melhoria em relação ao modelo base da adição destas variáveis, calculou-se a redução relativa na variância inexplicada (R^2 incremental) da respetiva equação aumentada pelo indicador comparada com a do modelo base.

A tabela 5.12 apresenta os resultados da avaliação *in-sample* para os modelos aumentados pelos indicadores CAPE e Google em relação ao modelo de base para o período dezembro 2007 a junho 2015, reportando o R^2 ajustado que resulta do aumento do modelo base com o respetivo índice. Para Portugal, apesar dos reduzidos R^2 incrementais, quer o CAPE quer o Google melhoram significativamente o modelo de base, sendo que é de realçar que o modelo sofre uma melhoria mais considerável aquando da adição de ambos os indicadores ao modelo de base. Nos Estados Unidos, o CAPE não mostrou ser muito significativo (R^2 incremental perto de zero) quando adicionado ao modelo base, no entanto quando adicionado em conjunto com o índice de sentimento Google o R^2 incremental já é de 1 ponto percentual. Isto mostra que os modelos de previsão dos retornos do mercado acionista não devem assentar apenas em indicadores de sentimento económico baseados em pesquisas e em variáveis macroeconómicas. Se a estes modelos base se acrescentar o índice de sentimento Google ou o CAPE obter-se-á um desempenho muito melhor.

Tabela 5.12: Conteúdo informativo das variáveis,

	Portugal	EUA
	increm, R²	
Apenas Google vs Base	0,009	0,0109
Apenas CAPE vs Base	0,019	0,0011
Google e CAPE vs Base	0,030	0,0114

Além disso, avalia-se o desempenho dos modelos *out-of-sample* em relação ao modelo base comparando o RMSFE (raiz quadrada do erro de previsão quadrático médio) de cada modelo com o do modelo base, com o propósito de analisar o valor do RMSFE de cada um dos modelos individualmente, para que seja possível concluir se o erro de previsão cometido pelos modelos per si é ou não elevado. Olhando para a Tabela 5.13 verifica-se que para Portugal, o modelo base supera os modelos de previsão que usam quer o índice de sentimento Google quer o CAPE, quer ambos. No entanto, nos Estados Unidos há vantagem em usar ambos os indicadores em estudo (Google e CAPE) no modelo de previsão, dado que o seu RMSFE é inferior ao do modelo base e aos modelos que usam cada um dos indicadores individualmente. A grande conclusão a que se chega é que apesar do índice de sentimento Google e do CAPE não ser muito vantajoso na previsão dos retornos do PSI20, quando usados simultaneamente no mercado americano estes mostram ser bastante benéficos para a previsão, olhando aos valores absolutos dos RMSFE referentes aos modelos apresentados.

Tabela 5.13: Desempenho absoluto fora da amostra

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
RMSFE	Portugal											
Base	0,9053	0,9155	0,9381	0,9352	0,8830	0,7994	0,7979	0,8255	0,8015	0,7236	0,6950	0,7034
Google	0,9496	0,9558	0,9803	0,9773	0,9309	0,8219	0,8297	0,8581	0,8495	0,7666	0,7462	0,7499
Cape	0,9170	0,9259	0,9493	0,9459	0,8836	0,8030	0,8011	0,8298	0,8004	0,7219	0,6947	0,7026
Google+Cape	0,9656	0,9705	0,9956	0,9928	0,9336	0,8257	0,8332	0,8632	0,8498	0,7664	0,7482	0,7517
RMSFE	EUA											
Base	0,8857	0,8921	0,8726	0,7602	0,7524	0,7941	0,8347	0,7566	0,7699	0,7492	0,6886	0,7080
Google	0,9282	0,9282	0,9141	0,8209	0,7967	0,8487	0,8804	0,7820	0,7845	0,7421	0,6911	0,7029
Cape	0,8422	0,8712	0,8335	0,7011	0,6983	0,7142	0,7631	0,7076	0,7494	0,7271	0,6980	0,6723
Google+Cape	0,8211	0,8289	0,8202	0,6573	0,6707	0,7395	0,7808	0,7095	0,6948	0,6787	0,6623	0,6185

A Tabela 5.14 compara a precisão dos modelos aumentados pelos indicadores (CAPE e Google) com os modelos base de cada um dos países dentro do período junho 2015 a junho 2017. A primeira linha de cada país diz respeito à relação entre o RMSFE obtido para o modelo usando o indicador de sentimento Google e o do respectivo modelo base (rácio entre os dois RMSFEs). A segunda linha documenta a proporção entre o RMSFE do modelo usando o CAPE e o do modelo de base, e assim por diante. Os modelos aumentados pelos indicadores também são comparados entre si. Valores menores que um indicam que o primeiro modelo supera o segundo, isto é, o primeiro modelo na linha tem um RMSFE menor que o segundo e, portanto, supera-o.

Para todos os horizontes de previsão (de $h=1$ a $h=12$), o modelo com Google é inferior ao modelo base para Portugal. Nos Estados Unidos, os modelos que usam o indicador de sentimento Google têm maior capacidade preditiva do que o modelo base em horizontes mais longos, no entanto não têm uma performance muito boa nos horizontes de previsão mais reduzidos. O mesmo acontece com o modelo CAPE, quando comparado com o modelo base, em Portugal. No que toca aos modelos aumentados com os dois indicadores em estudo, seja qual for o horizonte de previsão, em Portugal a sua performance é inferior ao modelo base, mas nos EUA é superior.

Deixando um pouco de lado o modelo de base, compara-se a performance dos vários modelos aumentados, quer pelo CAPE, pelo Google ou por ambos. Em Portugal, conclui-se que o modelo aumentado por ambos os indicadores, em simultâneo, é mais eficaz na previsão dos retornos do mercado acionista do que os modelos aumentados apenas por um dos indicadores. Comparando o modelo Google com o modelo CAPE, conclui-se que o modelo usando apenas o CAPE apresenta uma melhor performance do que o modelo Google. Nos Estados Unidos da América, as conclusões, no geral, mantêm-se, mas é importante realçar que quando se compara o modelo aumentado pelos dois indicadores com o modelo usando apenas o índice CAPE, a performance passa a ser diferente a partir do horizonte 5, isto é, o CAPE passa a ter uma maior capacidade explicativa do que o modelo aumentado pelo Google e pelo CAPE, a qual se volta a reverter para os h mais elevados.

Tabela 5.14: Desempenho relativo fora da amostra

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
RMSFE rel.	Portugal											
Google/Base	1,0490	1,0441	1,0450	1,0450	1,0542	1,0281	1,0399	1,0395	1,0599	1,0595	1,0737	1,0661
Cape/Base	1,0129	1,0114	1,0118	1,0115	1,0006	1,0046	1,0040	1,0052	0,9986	0,9978	0,9995	0,9989
Google+Cape/Base	1,0666	1,0601	1,0613	1,0616	1,0573	1,0328	1,0443	1,0456	1,0602	1,0592	1,0765	1,0686
Cape/Google	0,9657	0,9687	0,9683	0,9679	0,9491	0,9771	0,9654	0,9670	0,9422	0,9417	0,9309	0,9369
Google+Cape/Google	1,0168	1,0153	1,0156	1,0159	1,0029	1,0046	1,0042	1,0059	1,0003	0,9998	1,0026	1,0023
Google+Cape/Cape	1,0530	1,0481	1,0489	1,0495	1,0567	1,0282	1,0402	1,0402	1,0617	1,0616	1,0770	1,0698
RMSFE rel.	EUA											
Google/Base	1,0480	1,0404	1,0475	1,0797	1,0588	1,0688	1,0547	1,0336	1,0189	0,9905	1,0037	0,9929
Cape/Base	0,9509	0,9765	0,9551	0,9222	0,9281	0,8995	0,9142	0,9352	0,9734	0,9705	1,0137	0,9495
Google+Cape/Base	0,9271	0,9291	0,9399	0,8646	0,8914	0,9312	0,9355	0,9378	0,9024	0,9060	0,9618	0,8736
Cape/Google	0,9074	0,9386	0,9118	0,8541	0,8765	0,8416	0,8668	0,9048	0,9553	0,9798	1,0099	0,9563
Google+Cape/Google	0,8847	0,8930	0,8973	0,8007	0,8418	0,8713	0,8869	0,9073	0,8856	0,9146	0,9583	0,8799
Google+Cape/Cape	0,9750	0,9514	0,9841	0,9376	0,9605	1,0353	1,0233	1,0027	0,9271	0,9335	0,9489	0,9200

Com o objetivo de estudar a significância das diferenças entre as previsões anteriores, utiliza-se o teste Diebold-Mariano (1955), o qual tem uma distribuição normal *t-student* e mostra se as diferenças nos RMSFE's são estatisticamente significativas. Um resultado menor que -1,96 indica que o primeiro modelo é preferível ao segundo, dado que o seu RMSFE é inferior, pelo contrário, um valor superior a 1,96 indica que o RMSFE do primeiro modelo é superior ao do segundo daí que o segundo modelo seja preferível. Valores entre -1,96 e 1,96, indicam que os dois modelos têm a mesma capacidade de previsão, ou seja, não se rejeita a hipótese nula.

Olhando para a tabela 5.15, em baixo, verificamos que, em Portugal o modelo base apresenta uma capacidade de previsão superior quer ao modelo com Google, quer ao modelo com Google e CAPE. O modelo usando apenas o CAPE é superior ao modelo com os dois indicadores do estudo. Para os EUA, o modelo aumentado pelo Google e pelo CAPE é superior quer ao modelo base, quer ao modelo usando apenas o indicador de sentimento negativo Google.

Tabela 5.15: Teste Diebold-Mariano

	Portugal
Base vs Google	-5,203507
Base vs CAPE	-0,853299
Base vs Google+CAPE	-4,809851
Google vs Google+CAPE	-1,063310
CAPE vs Google+CAPE	-4,949105
	EUA
Base vs Google	-1,332811
Base vs CAPE	1,706044
Base vs Google+CAPE	3,304050
Google vs Google+CAPE	2,988079
CAPE vs Google+CAPE	0,939535

Fazendo a sumula deste ponto 5.4 pode-se afirmar que não existe uma conclusão direta e indiscutível no que toca ao benefício do uso destes dois indicadores em estudo. É verdade que olhando para os R^2 incrementais parece existir vantagem em acrescentar aos modelos ditos de base quer o índice de sentimento Google quer o CAPE, no entanto quando se procede à previsão *out-of-sample* e olhando para os valores dos RMSFE e para os resultados do teste Diebold-Mariano verificamos que em Portugal nem sempre essa vantagem se materializa sendo o modelo base por vezes visto como o mais vantajoso e o CAPE superior ao índice de sentimento Google. Nos Estados Unidos, já é mais notória esta vantagem dado que os modelos usando os dois indicadores, ou apenas o indicador de sentimento Google superam os modelos de base.

6. Conclusões

O presente estudo assume que o comportamento online dos investidores aquando de pesquisas na internet pode ser considerado uma boa medida de sentimento. Construiu-se um índice de sentimento negativo Google tendo como base uma lista de termos de pesquisa económicos e financeiros que se relacionam negativamente com os retornos de cada um dos mercados acionistas, dado que Tectlock (2007) afirma que os termos de pesquisa negativos são os mais adequados para se medir o sentimento. Concluiu-se que esse índice possui validade de conteúdo no que toca à mensuração do sentimento do investidor individual quer em Portugal quer nos Estados Unidos da América e que, além disso, este indicador de sentimento oferece vantagens adicionais quando comparado com as medidas de sentimento económico usadas no estudo, já que os primeiros estão disponíveis com maior frequência e menor custo, algo absolutamente fundamental tendo em conta o contexto económico-financeiro dos diversos países do mundo.

Por meio de uma série de regressões de mínimos quadrados, a hipótese de que essa nova *proxy* para o sentimento de mercado possui poder explicativo sobre os retornos do índice PSI20 e do SP500, conforme o país em estudo, foi comprovada. Além disso, no que toca à previsão *in-sample*, os resultados mostram que, quando comparados com outros *proxies* de sentimento, os dados do Google têm a vantagem de mostrar poder de previsibilidade sobre os retornos, justificando a crença de que as consultas na Internet podem prever o desempenho do mercado, principalmente em mercados menos desenvolvidos como Portugal onde o nível de literacia financeira é inferior aos Estados Unidos daí que o comportamento dos investidores sejam mais bem representado por este tipo de medidas mais contemporâneas.

No que concerne à previsão *out-of-sample*, não é possível afirmar que existe de facto uma vantagem em usar este índice de sentimento Google para prever os movimentos do mercado acionista de ambos os países, dado que as conclusões são um pouco ambíguas quando se compara os modelos com os modelos de base. No entanto, pode-se realçar que, no geral, existe uma superioridade preditiva aquando da utilização deste índice de sentimento Google em conjunto com o CAPE, uma medida mais teórica e fundamentada na literatura, em vez da utilização de cada um dos indicadores individualmente, em ambos os países. Por exemplo, nos EUA, o modelo aumentado pelos dois indicadores em estudo, simultaneamente, apresenta vantagem preditiva dos retornos do SP500 comparativamente aos modelos aumentados por cada um dos indicadores em separado. Estes resultados podem advir da falta de especificidade dos termos que constituem o índice Google, isto porque noutros estudos como Brochado (2015)

são usados termos mais específicos e mais relacionados com os mercados financeiros que advém das entrevistas pessoais a pessoas qualificadas nestes tópicos e que seguem mais de perto os mercados.

Em suma, esta tese mostra que é possível usar os dados do Google Trends para criar variáveis de sentimento, abrangendo o pessimismo do investidor em relação ao mercado, no entanto convém complementar a análise com medidas mais teóricas com o índice CAPE. A principal vantagem de usar pesquisas na Internet é que elas são geradas por meio do comportamento espontâneo dos investidores, oferecendo, assim, propriedades de sinalização que se tornam bastante aliciantes para os investigadores. A combinação deste índice de sentimento com dados de negociação financeira pode abrir novas percepções, desenvolvendo uma melhor compreensão do comportamento coletivo complexo dos indivíduos que nem sempre se encontra descrito nas medidas de sentimento mais utilizadas na literatura. Isto pode ser bastante importante para o estudo das bolhas especulativas e de todos os eventos catastróficos que os mercados financeiros vão sofrendo ao longo dos anos, dado que a espontaneidade dos investidores associada a este índice de sentimento Google consegue captar de uma forma bastante eficaz o seu comportamento e isso é, sem dúvida, crucial para uma economia que se faz de seres humanos que tantas vezes são esquecidos pela teoria económica.

Voltando de novo às sugestões de Ferreira (2016) e tendo por base as descobertas existentes neste tópico penso que seria interessante enriquecer o estudo introduzindo outros produtos financeiros, como títulos e derivados, ou até mesmo usando a análise dos principais componentes como em Beer (2013) em que se estuda o sentimento usando como base também os dados de pesquisa na internet dos investidores, neste caso, franceses. Além disso, fugindo um pouco ao Google, seria também de enorme interesse desenvolver uma base de dados de notícias em Portugal uma vez que, dado a sua inexistência, não é possível desenvolver uma medida de sentimento baseada em notícias para o nosso país. O Banco Central do Brasil desenvolveu recentemente uma medida de sentimento com base na cobertura jornalística de cada empresa como medida de atenção dos investidores, divulgadas no jornal de maior circulação e no jornal de negócios líder no Brasil (Souza, et al., 2017) e concluiu que o volume de negociação no mercado de ações brasileiro é elevado no dia de publicação das notícias nos jornais (ou seja, no dia seguinte ao evento). Esse resultado sugere que investidores menos especializados são influenciados pelos jornais de grande circulação o que contribui ainda mais para a necessidade de criação desta base de dados para Portugal.

Seria também de enorme interesse para o desenvolvimento das finanças comportamentais no nosso país a construção de um índice de incerteza económica como o desenvolvido por Baker, et al. (2016), o qual é usado como medida de sentimento alternativa em alguns estudos dos Estados Unidos e não está disponível para Portugal.

7. Bibliografia

- Andreassen, P. B. and Kraus, S. J. (1990) ‘Judgmental extrapolation and the salience of change.’ *J. Forecast.* 9, 347–372.
- Ang, A., and Bekaert, G. (2006). ‘Stock Return Predictability: Is it There?’. *Review of Financial Studies*, 20(3), 651-707.
- Antweiler, W. and Frank, M. Z. (2004) ‘Is all that talk just noise? The information content of Internet stock message boards’, *Journal of Finance*, 59(3), pp. 1259–1294.
- Askitas, N. and Zimmermann, K. F. (2015) ‘In Social Sciences The Internet as a Data Source for Advancement in Social Sciences’, (8899).
- Baker, M. and Wurgler, J. (2007) ‘Investor Sentiment in the Stock Market’, *Journal of Economic Perspectives*, 21(2), pp. 129–151.
- Baker, M. and Wurgler, J. (2013) ‘Behavioral Corporate Finance: An Updated Survey’, *Handbook of the Economics of Finance*. Elsevier Inc.
- Baker, S. R., Davis, S. J. and Bloom, N. (2016) ‘Measuring Economic Policy Uncertainty’, *The Quarterly Journal of Economics*, 131(4), pp. 1593–1636.
- Bank, M., Larch, M. and Peter, G. (2011) ‘Google search volume and its influence on liquidity and returns of German stocks’, *Financial Markets and Portfolio Management*, 25(3), pp. 239–264.
- Barber, B. M. and Odean, T. (2001) ‘Boys will be Boys: Gender, Overconfidence, and Common Stock Investment’, *The Quarterly Journal of Economics*, 116(1), pp. 261–292.
- Barber, B. M. and Odean, T. (2001) ‘The Internet and the Investor’, *Journal of Economic Perspectives*, 15(1), pp. 41–54.
- Barberis, N., Shleifer, A. and Vishny, R. (1997) ‘A Model of Investor Sentiment’, 49.
- Barberis, N. and Thaler, R. H. (2002) ‘A Survey of Behavioral Finance’, *SSRN Electronic Journal*.
- Battelle, John (2005), ‘The Search – Como o Google mudou as regras do negócio e revolucionou a cultura’, Cruz Quebrada, Casa das Letras.
- Beer, F., Herve, F. and Zouaoui, M. (2013) ‘Is big brother watching us? Google, investor

- sentiment and the stock market’, *Economics Bulletin*, 33(1), pp. 454–466.
- Bes, F. T. (2010) ‘O homem que trocou a casa por uma tulipa’, *Editorial Presença*.
- Black, F. (1986) ‘Noise’, *The Journal of Finance*, 41(3), pp. 529–543.
- Bollen, J., Mao, H. and Zeng, X. (2011) ‘Twitter mood predicts the stock market’, *Journal of Computational Science*, 2(1), pp. 1–8.
- Boushelle, B. (2014) ‘An Introduction to the Shiller CAPE Ratio An Introduction to the Shiller CAPE Ratio (continued)’, (May), pp. 2–5.
- Brochado, A. (2015) ‘Paper CMVM retail investor sentiment: can we google it?’ *Working Paper CMVM*.
- Brown, G. and Cliff, M. (2004) ‘Investor Sentiment and the Near-term Stock Market’, *Journal of Empirical Finance*, Vol. 11, Issue 1, pp 1-27
- Brown, G. and Cliff, M. (2005) ‘Investor Sentiment and Asset Valuation’, *The Journal of Business*, Vol. 78, Issue 2, pp. 405-440
- Campbell, John Y, and Shiller, R. J. (1988) ‘Stock prices, Earnings and Expected Dividends. *Science*’.
- Campbell, J. Y. and Shiller, R. J. (2001) ‘Valuation ratios and the long-run stock market outlook: an update.’
- Chen, T., So, E. P. K., Wu, L. and Yan, K. M. Y. (2015) ‘The 2007-2008 US Recession: What Did the Real-Time Google Trends Data Tell The United States?, *Contemporary Economic Policy*, vol. 33, no. 2, pp. 395-403.
- Choi, H. and Varian, H. (2012) ‘Predicting the Present with Google Trends’, *Economic Record*, 88(SUPPL.1), pp. 2–9.
- Cochrane, J. H. (2008) ‘The dog that did not bark: a defense of return predictability. ’ *Rev. Financ. Stud.* 21, 1533-1575.
- Cutler, D. M., Poterba, J. M. and Summers, L. H. (1989) ‘What moves stock prices?’, *The Journal of Portfolio Management*, pp. 4–12.
- D’Amuri, F. and Marcucci, J. (2017) ‘The predictive power of Google searches in forecasting US unemployment’, *International Journal of Forecasting*. Elsevier B.V., 33(4), pp. 801–816.
- Da, Z., Engelberg, J. and Gao, P. (2011) ‘In Search of Attention’, *Journal of Finance*, 66(5), pp. 1461–1499.
- Da, Z., Engelberg, J. and Gao, P. (2015) ‘The sum of all FEARS investor sentiment and asset prices’, *Review of Financial Studies*, 28(1), pp. 1–32.
- Das, S. R. and Chen, M. Y. (2007) ‘Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web’, *Management Science*, 53(9), pp. 1375–1388.

- De Long, J. B., Shleifer A., Summers L. H. and Waldmann R. J. (1990) ‘Noise Trader Risk in Financial Markets’, *Journal of Political Economy*, 98(4), pp. 703–738.
- Diebold, F. X. and R. S. Mariano, 1995, ‘Comparing Predictive Accuracy’, *Journal of Business and Economic Statistics* 13, 253-263.
- Edmans, A., García, D. and Norli, O. (2007) ‘Sports sentiment and stock returns’, *Journal of Finance*, 62(4), pp. 1967–1998.
- Fama, E. F. (1969) ‘American Finance Association Efficient Capital Markets : A Review of Theory and Empirical Work’, *The Journal of Finance*, 25(2), pp. 28–30.
- Fama, E. F. (1995) ‘Random Walks in Stock Market Prices’, *Financial Analysts Journal*, 51(1), pp. 75–80.
- Fama, E. F., and French, K. R. (1988) ‘Dividend yields and expected stock returns’, 22(March), 325.
- Fama, E. F. and French, K. R. (1988) ‘Permanent and Temporary Components of Stock Prices’, 96(2), 246-273.
- Ferreira, B. (2016) ‘Can Google data measure market sentiment’, Dissertação de Mestrado em Monetary and Financial Economics, Lisboa, ISEG.
- Ginsberg, J., Mohebbi M. H., Patel R. S., Brammer L., Smolinski M. S. and Brilliant L. (2009) ‘Detecting influenza epidemics using search engine query data’, *Nature*. Nature Publishing Group, 457(7232), pp. 1012–1014.
- Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M. and Watts, D. J. (2010) ‘Predicting consumer behavior with Web search’, *Proceedings of the National Academy of Sciences*, 107(41), pp. 17486–17490.
- Golez, B. and Koudijs, P. (2017) ‘Four centuries of return predictability’, *Journal of Financial Economics*. Elsevier B.V., 127(2), pp. 248–263.
- Gray, W. R. and Vogel, J. (2014) ‘The Cross-Section Predictability of Cyclically-Adjusted Valuation Measures’.
- Hamilton, L. C. (1992) ‘Regression with Graphics: A Second Course in Applied Statistics.’ Pacific Grove.
- Hirshleifer, D. and Shumway, T. (2003) ‘Good Day Sunshine: Stock Returns and the Weather’, *Journal of Finance*, 58(3), pp. 1009–1032.
- Hui, E., Zheng, X. and Wang, H. (2013) ‘Investor Sentiment and Risk Appetite of Real Estate Security Market’, *Applied Economics*, 45 (19), 2801-2807.
- Joseph, K., Wintoki, M. and Zhang, Z. (2011) ‘Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search’, *International Journal of Forecasting*, Vol.27, Issue 4, pp.1116-112.

- Kahn, R. N. and Rudd, A. (1995) 'Does Historical Performance Predict Future Performance?' *Financial Analysts Journal* 51, 43–52
- Kahneman, D. and Tversky, A. (1979) 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica*, 47(2), pp. 263–292.
- Kaminsky, G. L. and Schmukler, S. L. (1999) 'What triggers market jitters?', *Journal of International Money and Finance*, 18(4), pp. 537–560.
- Keimling, N. (2016) 'Predicting Stock Market Returns Using the Shiller CAPE An Improvement Towards Traditional Value Indicators?', *SSRN Electronic Journal*, (January), pp. 1–32.
- Kumar, A. and Lee, C. (2006) 'Retail Investor Sentiment and Return Comovements', LXI(5).
- Lee, C., Shleifer, A. and Thaler, R. H. (1991) 'Investor sentiment and the closed-end fund puzzle'.
- Malkiel, B.G. (1995) 'Returns from Investing in Equity Mutual Funds 1971 to 1991.' *The Journal of Finance* 50, 549–572.
- Maloney, M. T. and Mulherin, J. H. H. (1998) 'The Stock Price Reaction to the Challenger Crash: Information Disclosure in an Efficient Market', *SSRN Electronic Journal*, (July 1992), pp. 1–20.
- Manela, A. and Moreira, A. (2017) 'News implied volatility and disaster concerns', *Journal of Financial Economics*. Elsevier B.V., 123(1), pp. 137–162.
- Mao, H., Counts, S. and Bollen, J. (2015) 'Quantifying the effects of online bullishness on international financial markets', *ECB Statistics Paper Series*, 9(9), pp. 1–22.
- McLaren, N. and Shanbhogue, R. (2011) 'Using internet search data as economic indicators', *Bank of England Quarterly Bulletin*, Q2(1), pp. 134–140.
- Mitchell, M., Pulvino, T. and Stafford, E. (2002) 'Limited arbitrage in equity markets', *Journal of Finance*.
- Preis, T., Moat, H. S. and Stanley, H. E. (2013) 'Quantifying Trading Behavior in Financial Markets Using Google Trends', *Scientific Reports*, 3(1), p. 1684.
- Qiu, L. and Welch, I. (2004) 'Investor sentiment Measures', *Working paper, Brown University*.
- Roberts (1967) 'Statistical versus Clinical Prediction in the Stock Market'
- Schmidt, T. and Vosen, S. (2009) 'Forecasting Private Consumption', *Economic Papers*, 155, p. 23.
- Sewell, M. (2011) 'History of the Efficient Market Hypothesis', *Rn*, 11(4), p. 4.
- Smith, G.P. (2012) 'Google Internet Search Activity and Volatility Prediction in the Market for Foreign Currency.' *Finance Research Letters* 9, 103-110

- Shiller, R. J. (1980) 'Do stock prices move too much to be justified by subsequent changes in dividends?', *National Bureau of Economic Research*, 71(456), pp. 421–436.
- Shleifer, A. and Summers, L. H. (1990) 'The Noise Trader Approach to Finance', *Journal of Economic Perspectives*, 4(2), pp. 19–33.
- Shleifer, A. and Vishny, R. W. (1997) 'The limits of arbitrage', *Review of Finance*, 14(1), pp. 157–187.
- Siegel, J. J. (2016) 'The shiller CAPE ratio: A new look', *Financial Analysts Journal*, 72(3), pp. 41–50.
- Simon, H. A. (1955) 'A Behavioral Model of Rational Choice', *The Quarterly Journal of Economics*, 69(1), pp. 99–118.
- Singer, E. (2002) 'The use of incentives to reduce nonresponse in household surveys.' *Working Paper*, University of Michigan
- Souza, H. E., Barbedo, C. H. and Araújo, G. S. (2017) 'Does investor attention affect trading volume in the Brazilian stock market?', *Research in International Business and Finance*.
- Thaler, R. H. (2015) 'Comportamento inadequado', Actual Editora.
- Tetlock, P. C. (2007) 'Giving Content to Investor Sentiment: The Role of Media in the Stock Market' Published by: Wiley for the American Finance Association Stable, 62(3), pp. 1139–1168.
- Tversky, A. and Kahneman, D. (1974) 'Judgment under Uncertainty: Heuristics and Biases', *Science*, 185(4157), pp. 1124–1131.
- Wooldridge, J. M. (2009) 'Introductory Econometrics: A Modern Approach.' South Western College Publications, fourth ed.

8. Fontes

Foram usadas as seguintes fontes/noticias (durante o mês de janeiro) para recolher a informação usada como base para o capítulo 3 da dissertação:

<http://www.asf.com.pt/NR/exeres/2E260022-2D02-4891-991C-F1C0AC522767.htm>

<http://news.gallup.com/poll/190883/half-americans-own-stocks-matching-record-low.aspx>

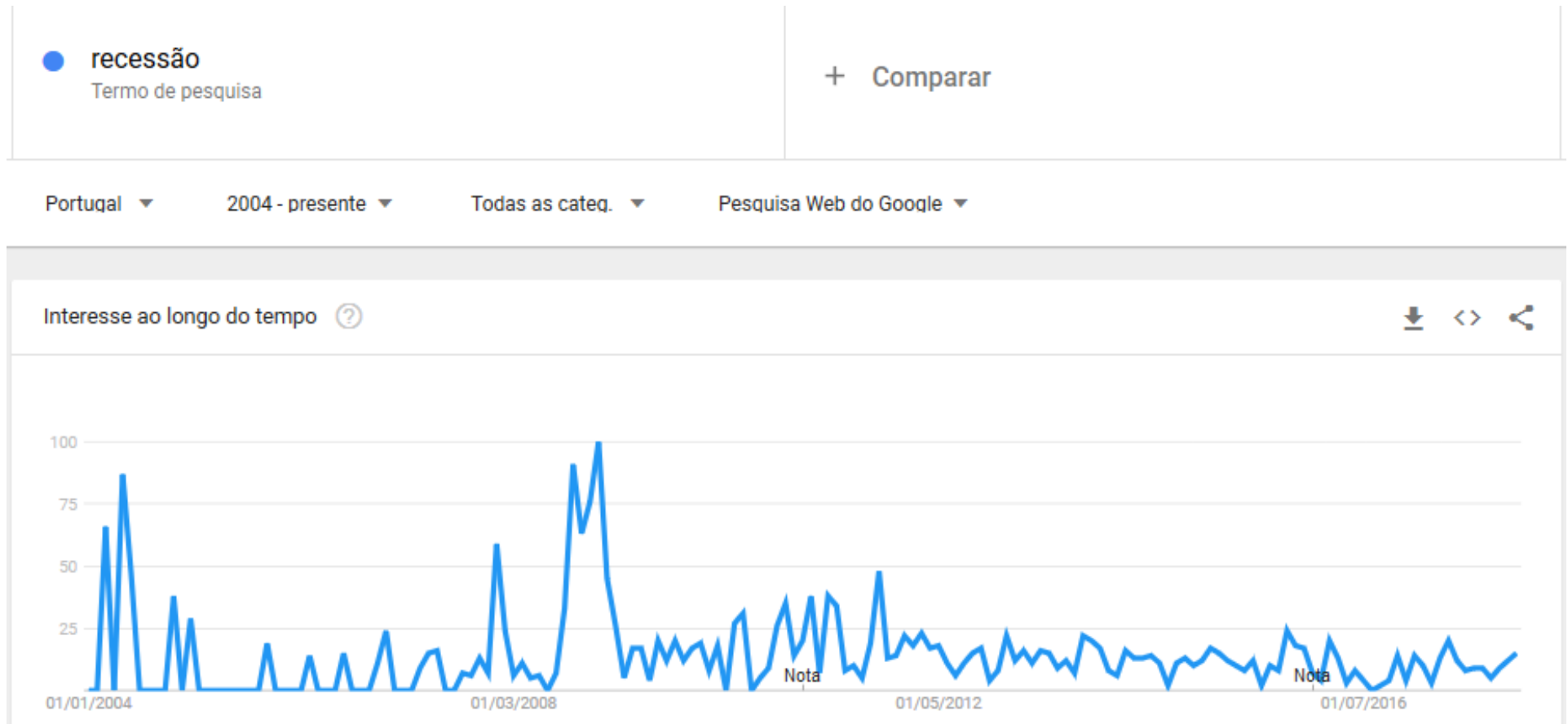
<http://www.cmvm.pt/pt/EstatisticasEstudosEPublicacoes/Estudos/Pages/O%20Perfil%20do%20Investidor%20Particular%20Portugu%C3%AAs.aspx?v=>

<https://data.worldbank.org/>

<https://www.internetworldstats.com/>

9. Anexos

Anexo A- Figura 4.1: Gráfico do Google Trends para o termo de pesquisa “recessão” (print screen)



Fonte: Google Trends

Anexo A - Tabela 4.1: *Conjunto de termos de pesquisas que revelam o sentimento dos investidores em relação às condições económicas do seu país.*

2008 recession, Abundance, Accounting, Affluence, Allowance, Bank beneficiary, Bankrupt, Bankruptcy, Bargain, Benefit, Bequest, Bonus, Bribery, Budget, Budget deficit, Business, Business success, Buy, Capital gain, Capitalize, Catalonia, Cheap, Compensation, Competitive advantage, Contribute, Contribution, Corruption, Cost, Cost benefit, Cost of living, Costly, Creditor, Crisis, Crisis intervention, Debt, Debt equity, Debtor, Deficit, Depreciation, Depression, Economic backwardness, Economic crisis, Economic prosperity, Economic recession, Endowment, Entrepreneurial, Entrepreneurship, Equity ratio, Estado benefactor, Expense, Expensive, Financial crisis, Fiscal deficit, Fixed cost, Frugal, Frugal living, Gdp, Gold, Gold price, Gold rate, Government, Great depression, Great recession, Gross profit, Inexpensive, Inflation, Inflation rate, Inheritance, Insurance beneficiary, Job, Jobless, Jobless rate, Liquidation, Low cost, Lucrative, Luxury, Mastercard, National savings, Net worth, Opportunity cost, Owe, Owe money, Owe taxes, Partner, Partnership, Patronage, Poverty, Poverty level, Poverty rate, Precious metal, Private equity, Production, Productivity, Profit, Profit margin, Profitable, Prosperity, Recession, Recompense, Reward, Rich, Richness, Savings, Savings rates, Social security, Squander, Subsidy, Target, Tariff, Tax allowance, Tax benefit, Thrift, Thrift savings, Unemployed benefits, Unemployment, Unemployment rate, Unprofitable, Valuable, Value.

Nota: *as palavras em português foram obtidas por tradução.*

Anexo A - Tabela 4.2: Definição das variáveis.

	Denominação	Variável	Frequência	Fonte
Portugal	Mercado de capitais			
	RETORNOS_PT	Retornos do PSI20	Mensal	calculados através dos open values do PSI20 retirados da Euronext Lisboa
	Medidas de Sentimento			
	GOOGLE_PT	Índice de sentimento negativo Google	Mensal	Google Trends
	TURNOVER_PT	<i>Turnover</i> PSI-20	Diário	Euronext Lisboa
	CONFIANÇA_PT	Índice de confiança dos consumidores	Mensal	Eurostat
	SENTIMENTO_PT	Índice de sentimento económico	Mensal	Eurostat
	Variáveis macroeconómicas			
	DESEMPREGO_PT	Desemprego	Trimestral	INE
	PIB_PT	Produto interno bruto	Trimestral	INE
	IHPC_PT	Índice harmonizado de preços ao consumidor	Mensal	Eurostat
	YIELD_PT	Dividend yield	Mensal	Bloomberg
	SPREAD_PT	<i>Yield Spread</i> PT-treasury	Mensal	Banco de Portugal
	CAPE_PT	Cyclically adjusted price-to-earnings ratio	Mensal	Replicou-se a metodologia de Robert Shiller usando para isso: o Last Price e os Earnings do PSI20 retirados da Bloomberg e o CPI do banco de Portugal

	Denominação	Variável	Frequência	Fonte
Estados Unidos	Mercado de capitais			
	RETORNOS_EUA	Retornos do SP500	Mensal	calculados através dos open values do SP500 retirados da Bloomberg
	Medidas de Sentimento			
	GOOGLE_EUA	Índice de sentimento negativo Google	Mensal	Google Trends
	TURNOVER_EUA	<i>Turnover</i> SP500	Mensal	Yahoo Finance
	CONFIANÇA_EUA	Índice de confiança dos consumidores	Mensal	Reuters
	SENTIMENTO_EUA	Índice de sentimento económico	Mensal	OCDE
	Variáveis macroecómicas			
	DESEMPREGO_EUA	Desemprego	Trimestral	US Bureau of Labor Statistics
	PIB_EUA	Produto interno bruto	Trimestral	Eurostat
	IHPC_EUA	Índice harmonizado de preços ao consumidor	Mensal	FRED
	YIELD_EUA	Dividend yield	Mensal	Quandl - Fontes: Standard & Poor's for current S&P 500 Dividend Yield, Robert Shiller and his book Irrational Exuberance for historic S&P 500 Dividend Yields
	SPREAD_EUA	Yield <i>Spread</i> EUA-treasury	Mensal	FRED
	CAPE_EUA	Cyclically adjusted price-to-earnings ratio	Mensal	Página pessoal de Robert Shiller

Nota: os dados trimestrais transformados em mensais usando a técnica de Shiller [exemplo: Jan=2/3(Dez)+1/3(Mar)] e dados diários transformados em mensais usando o último dia de cada mês.

Anexo B - Tabela 5.4: Regressões OLS-EUA

	OLS 1	OLS 2	OLS 3	OLS 4	OLS 5	OLS 6	OLS 7	OLS 8	OLS 9	OLS 10
GOOGLE_EUA	-0,546993***	-0,5489**	-0,7941***	-0,79227***	-0,7832***	-0,78453***	-0,82407***	-0,81832***	-0,82197***	-0,82504***
CAPE_EUA		0,002128	0,001937	-0,00372	-0,008	-0,00889	-0,0084	-0,00854	-0,00774	-0,00789
TURNOVER_EUA			0,397413***	0,374692***	0,376625***	0,376632***	0,338327**	0,336703**	0,318729*	0,319938*
SENTIMENTO_EUA				0,00086	0,00078	0,000766	0,000797	0,000795	0,000788	0,000778
CONFIANCA_EUA					0,014937	0,012713	0,013993	0,014765	0,016791	0,016888
YIELD_EUA						0,000999	0,001011	0,00057	0,000451	0,00044
IHPC_EUA							0,000274	0,000279	0,000335	0,000329
PIB_EUA								0,006421	0,004728	0,003711
DESEMPREGO_EUA									-0,00194	-0,00195
SPREAD_EUA										-0,00193
C	0,220398	0,199179	-1,21831	-1,28399	-1,3237	-1,32489	-1,22962	-1,23777	-1,16706	-1,1504
R-squared	0,05959	0,059941	0,134562	0,156118	0,157477	0,157713	0,160536	0,160864	0,16266	0,163202
Adjusted R-squared	0,049023	0,038576	0,104719	0,116867	0,107917	0,09755	0,089738	0,078997	0,069622	0,058602
Prob (F-statistic)	0,019708	0,065891	0,005473	0,005216	0,011175	0,022267	0,036625	0,061226	0,091351	0,133399
Durbin-Watson stat	1,649027	1,649396	1,57677	1,61871	1,626883	1,625964	1,598841	1,604318	1,622575	1,6216

Nota: o modelo mais simples utiliza apenas a variáveis Google_EUA como variável explicativa da variável dependente, Retornos_EUA. Cada coluna representa uma nova regressão na qual uma nova variável de controle é adicionada ao modelo individualmente. À medida que o modelo se torna mais complexo, ele capta de maneira mais completa os diferentes aspectos que influenciam os retornos do mercado e fornece uma base para comparação com outras proxies de sentimento conhecidos. *, **, *** correspondem a 90%, 95%, 99% de significância estatística, respetivamente.

Anexo B - Tabela 5.5: Regressões OLS-PT

	OLS 1	OLS 2	OLS 3	OLS 4	OLS 5	OLS 6	OLS 7	OLS 8	OLS 9	OLS 10
GOOGLE_PT	-1,12589***	-1,11533***	-1,12382***	-1,10478***	-1,11032***	-1,05928***	-1,057422***	-1,018588***	-1,026851***	-0,961696***
CAPE_PT		0,0016	0,001656	0,001513	0,001522	0,00162	0,001615	0,001671	0,001628	0,001809
TURNOVER_PT			0,173627	0,190387	0,189796	0,195398	0,194419	0,206411	0,220826	0,247501
SENTIMENTO_PT				-0,00676	-0,00677	-0,01195	-0,011965	-0,016740**	-0,017062**	-0,016674**
CONFIANCA_PT					-0,00468	-0,00409	-0,003954	-0,005065	-0,006215	-0,011062
YIELD_PT						0,043348	0,042867	0,042514	0,041168	0,045586
IHPC_PT							-0,000111	-0,000140	-0,000192	0,000739
PIB_PT								0,041115*	0,042850*	0,040224*
DESEMPREGO_PT									-0,002070	-0,002811
SPREAD_PT										0,074723*
C	-0,20523	-0,22787	-0,50748	-0,41498	-0,43537	-0,55993	-0,549408	-0,548069	-0,536043	-0,894473
R-squared	0,152148	0,163221	0,173824	0,185112	0,185759	0,200287	0,200347	0,232861	0,235194	0,265835
Adjusted R-squared	0,142621	0,144203	0,145335	0,14721	0,137862	0,143164	0,132906	0,158018	0,150216	0,174064
Prob(F-statistic)	0,000132	0,000393	0,000812	0,001347	0,003251	0,003824	0,007878	0,004058	0,007093	0,003861
Durbin-Watson stat	1,673953	1,692382	1,700009	1,731361	1,718197	1,73541	1,735477	1,773881	1,786313	1,891289

Nota: o modelo mais simples utiliza apenas a variáveis Google_PT como variável explicativa da variável dependente, Retornos_PT. Cada coluna representa uma nova regressão na qual uma nova variável de controle é adicionada ao modelo individualmente. À medida que o modelo se torna mais complexo, ele capta de maneira mais completa os diferentes aspectos que influenciam os retornos do mercado e fornece uma base para comparação com outras proxies de sentimento conhecidos. *, **, *** correspondem a 90%, 95%, 99% de significância estatística, respetivamente.