

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

BASES FUNDAMENTAIS DA SOBREVIVÊNCIA DAS EMPRESAS: EVIDÊNCIA EMPÍRICA DAS NOVAS EMPRESAS CRIADAS EM PORTUGAL

Ana Isabel Rolo de Paiva

Mestrado em Gestão

Orientador:

Prof. Doutor Paulo Viegas de Carvalho, Professor Auxiliar Convidado
ISCTE Business School

Novembro, 2020



**BUSINESS
SCHOOL**

Departamento de Marketing, Operações e Gestão Geral

**BASES FUNDAMENTAIS DA SOBREVIVÊNCIA DAS EMPRESAS: EVIDÊNCIA
EMPÍRICA DAS NOVAS EMPRESAS CRIADAS EM PORTUGAL**

Ana Isabel Rolo de Paiva

Mestrado em Gestão

Orientador:

Prof. Doutor Paulo Viegas de Carvalho, Professor Auxiliar Convidado
ISCTE Business School

Novembro, 2020

Agradecimentos

A realização da dissertação não seria possível sem o apoio, direto ou indireto, de várias pessoas.

Ao meu orientador Paulo Carvalho, pela prontidão no esclarecimento de qualquer dúvida, conhecimento transmitido, eficácia das sugestões dadas e toda a ajuda e disponibilidade prestada ao longo deste ano.

Ao Dr. Filipe Andrade, da Informa D&B, que teve a gentileza de despende duas horas do seu tempo, para me auxiliar na compreensão do funcionamento da base de dados SABI.

À Dr^a. Cristina Roque, pela atenção e flexibilidade oferecida no horário de trabalho.

Aos meus amigos, pelo apoio virtual ao longo de um ano que não permitiu grandes encontros presenciais.

À minha família, em particular, aos meus avós por acreditarem sempre em mim, à minha irmã, à minha mãe por todas as palavras encorajadoras e preciosa ajuda na revisão final e ao meu pai, pela força e incentivo nos tempos mais difíceis.

Por último, ao meu namorado Gonçalo, pela ajuda nas formatações, paciência e cedências durante todo o processo, compreensão pelas horas que não estive presente e, acima de tudo, pelo apoio e motivação diários.

Resumo

A presente dissertação consiste no estudo da sobrevivência empresarial em Portugal, tendo por base o efeito das condições iniciais de constituição das empresas no desempenho futuro. As hipóteses colocadas incidem sobre a importância dos efeitos intrínsecos das empresas, setor de atividade económica, região e ambiente económico na sobrevivência a 5 anos.

Os dados relativos a 178.437 novas empresas portuguesas, criadas entre 2005 e 2014, foram obtidos da base de dados SABI, do Bureau Van Dijk. Os modelos estatísticos desenvolvidos resultaram da aplicação da regressão logística, sendo validados posteriormente.

Verificou-se que os fatores externos às empresas apresentam uma influência superior às características inerentes às mesmas, particularmente: ao nível da fase do ciclo económico, um período de recessão económica providencia condições ao desenvolvimento de empresas mais resilientes, contribuindo para uma sobrevivência a longo prazo, ao contrário do que é geralmente observado; ao nível do setor, o dos outros serviços apresenta um contributo maioritariamente negativo, face aos principais setores; e, ao nível da localização, não existe evidência que permita a atribuição de um papel fundamental ao fator região na capacidade de atividade permanente das empresas.

Palavras-chave: novas empresas, sobrevivência, Portugal, condições iniciais

Classificação JEL: L25, M13

Abstract

The present dissertation consists on the study of firm survival in Portugal, based on the effect of the initial conditions of firms' constitution in their future performance. The hypothesis set focus on the importance of firms' intrinsic effects, economic activity sector, region and economic environment on a 5 years survival.

The data related to 178.437 new Portuguese firms, created between 2005 and 2014, were obtained from the data basis SABI, of Bureau Van Dijk. The developed statistic models resulted from the application of logistic regression, being validated later.

It was found that external factors to firms present a superior influence to their own inherent characteristics, particularly: at the level of the economic cycle's phase, a period of economic recession provides conditions to the development of more resilient firms, contributing to a long-term survival, unlike what is generally observed; at the sector level, the other services present a mostly negative contribution, facing the main sectors; and, at the location level, there is no evidence that allows the attribution of a fundamental part to the region factor in the capacity of permanent activity of firms.

Key-words: new firms, survival, Portugal, initial conditions

JEL Classification: L25, M13

Índice

Agradecimentos.....	v
Resumo.....	vii
Abstract	ix
Índice de Quadros.....	xiii
Índice de Figuras	xiii
Glossário de siglas.....	xv
Introdução.....	1
CAPÍTULO 1. Revisão de Literatura.....	3
1.1. Conceito de empresa.....	3
1.2. Novas empresas	4
1.3. Sobrevivência	4
1.4. Fatores com influência na sobrevivência das empresas	5
1.4.1. Fatores relacionados com a empresa.....	6
1.4.1.1. Tamanho	6
1.4.1.2. Capital.....	7
1.4.1.3. Forma Jurídica	8
1.4.1.4. Pertença a um grupo.....	9
1.4.2. Fatores externos.....	9
1.4.2.1. Setores de atividade económica.....	10
1.4.2.2. Região	11
1.4.2.3. Ambiente Macroeconómico.....	13
1.4.2.3.1. Ciclos Económicos.....	14
1.5. Breve descrição da economia portuguesa entre 2005 e 2015.....	15
1.6. Formulação de Hipóteses.....	17
CAPÍTULO 2. Metodologia.....	19
2.1. Regressão Logística (Binária)	19
2.2. Dados.....	20
CAPÍTULO 3. Análise de dados	23
3.1. Variável dependente	23
3.2. Análise Univariada	23
3.2.1. Tamanho.....	25
3.2.2. Capital Social	25

3.2.3.	Forma jurídica	25
3.2.4.	Pertença a um grupo de empresas	26
3.2.5.	Setor de atividade económica.....	26
3.2.6.	Região.....	28
3.2.7.	Ano de constituição	29
3.3.	Seleção de variáveis.....	29
CAPÍTULO 4. Resultados.....		33
4.1.	Modelo de variáveis associadas à empresa.....	33
4.2.	Modelo com inclusão de todas as variáveis.....	34
4.3.	Modelo com exclusão da variável setorial	36
4.4.	Modelo com exclusão da variável regional	37
4.5.	Modelo com exclusão da variável temporal.....	38
CAPÍTULO 5. Robustez		41
5.1.	Log likelihood ratio test (LRT)	41
5.2.	AIC vs BIC	42
5.3.	Tabela de Classificação	44
5.4.	Curva ROC	45
5.5.	Calibração.....	47
CAPÍTULO 6. Validação de Hipóteses		49
Conclusão		51
Referências Bibliográficas		53
Anexos.....		57
Anexo A – Taxa de sobrevivência a 5 anos de empresas constituídas entre 2005 e 2014... 57		57
Anexo B – Frequências das variáveis contínuas		58
Anexo C – Tabulação Cruzada e Medidas de Associação		59
Anexo D - Setor vs Região		62
Anexo F – Correlações		64
Anexo G - Ausência de Multicolinearidade		65
Anexo H - Modelo com interações.....		66
Anexo I – Tabelas de Classificação.....		67
Anexo J - Curvas ROC		68
Anexo L - Comparação AUC		69
Anexo M - Calibração		70

Índice de Quadros

Quadro 3.1. - Potenciais variáveis independentes e respetivo efeito conforme os vários autores	24
Quadro 3.2. - Regressão logística univariada (inclusão de cada variável individual)	31
Quadro 3.3. - Amostra para estimação	32
Quadro 3.4. - Amostra para validação.....	32
Quadro 4.1. - Modelo A (apenas variáveis relativas à empresa).....	34
Quadro 4.2. - Modelo B (com todas as variáveis selecionadas)	36
Quadro 4.3. - Modelo C (com exclusão da variável setor)	37
Quadro 4.4. - Modelo D (com exclusão da variável região).....	38
Quadro 4.5. - Modelo E (com exclusão da variável anos)	39
Quadro 5.1. - Teste da razão de verosimilhança	42
Quadro 5.2. - AIC vs BIC	43
Quadro 5.3. - Coeficiente de correlação de Matthews	45
Quadro 5.4. - AUC dos modelos	46
Quadro 5.5. - Comparação de AUCs.....	47

Índice de Figuras

Figura 1.1. - Taxa crescimento do PIB em Portugal, entre 2005 e 2019	15
Figura 1.2. - Desempregados inscritos no centro de emprego (média anual, em milhares)	16
Figura 2.1. - Percentagem de empresas encerradas por ano.....	21
Figura 2.2. - Período de sobrevivência (5 anos) por ano de constituição (2005 a 2014).....	22
Figura 3.1. - N.º de empresas encerradas e ativas por tipo de empresa	25
Figura 3.2. - N.º de empresas encerradas e ativas, conforme a (não) pertença a um grupo.....	26
Figura 3.3. - N.º de empresas encerradas e ativas, de acordo com o SAE	27
Figura 3.4. - N.º de empresas constituídas entre 2005 e 2014, ativas após 5 anos, por SAE...	27
Figura 3.5. - N.º de empresas encerradas e ativas, por região de Portugal	28
Figura 3.6. - N.º de empresas encerradas e ativas, constituídas entre 2005 e 2014	29
Figura 5.1. – Curva ROC do Modelo B	47

Glossário de siglas

- AIC - Akaike Information Criterion
- AUC - Area Under the Curve
- BIC - Bayesian Information Criterion
- CE – Comissão Europeia
- CSC – Código das Sociedades Comerciais
- GEM - General Entrepreneurship Monitor
- OE – Organizational Ecology
- I&D – Investigação e Desenvolvimento
- INE – Instituto Nacional de Estatística
- LDA – Limitada
- LRT – Likelihood Ratio Test
- MAR – Marshall-Arrow-Romer
- MCC - Matthews Correlation Coefficient
- MES – Minimum Efficient Scale
- OECD - Organisation for Economic Co-operation and Development
- IO - Industrial Organization
- PIB – Produto Interno Bruto
- RBV – Resource Based View
- ROC – Receiver Operating Curve
- SA – Sociedade Anónima
- SAE – Setor de Atividade Económica
- SNC – Sociedade em Nome Coletivo
- SPSS - Statistical Package for the Social Sciences
- TIC – Tecnologias de Informação e Comunicação
- TOL - Tolerance
- VIF – Variance Inflation Factor

Introdução

A sobrevivência empresarial é um tema amplamente discutido, com várias vertentes de estudo e em diferentes países. Desde meados dos anos 50 que se investiga este fenómeno e até aos dias de hoje que se verificam limitações na compreensão detalhada do mesmo, em consequência da falta de integração entre os resultados alcançados, pela existência de diversas abordagens adotadas (Josefy et al, 2017).

Segundo Schutjens e Wever (2000), a criação de empresas contribui para a diminuição do desemprego, pelo que o sucesso empresarial deve ser assegurado, por exemplo, com o desenvolvimento de estratégias e orientações fundamentadas e planeadas, que melhorem a probabilidade de sobrevivência das novas empresas. Em Portugal, a formação de empresas representa 35% da criação de emprego do país e a sua saída apresenta um impacto de 40% na destruição do emprego (Centeno et al., 2007). As novas empresas são consideradas essenciais para o crescimento local e para a mudança estrutural (Brixy e Grotz, 2007) e a manutenção da atividade futura deverá ser garantida.

A presente dissertação tem como objetivos a identificação das condições iniciais que propiciam a sobrevivência empresarial a longo prazo e impacto das mesmas. Embora seja virtualmente impossível traçar um perfil da empresa ideal a constituir, em virtude das especificidades de cada empreendedor, da própria empresa e do contexto espacial, temporal e económico do negócio em causa é, ainda assim, relevante compreender as características distintivas da fundação das empresas que são comuns às que sobrevivem e apresentam capacidade de manter a sua atividade empresarial a longo prazo. De acordo com o observado, poder-se-ão retirar conclusões relevantes e extrapolar evidência útil para as empresas que ainda não foram criadas.

A maioria dos estudos elaborados sobre esta matéria versam sobre um determinado país. O facto de esta dissertação incidir apenas sobre as empresas portuguesas deve-se ao ambiente macroeconómico, regulamentação, cultura, entre outras características, que são específicas e subjacentes a Portugal. Empreender numa economia pequena e exposta à concorrência exterior como acontece em Portugal é desafiante, razão pela qual este trabalho pode trazer contribuições úteis para os empreendedores em economias com características idênticas à portuguesa.

Por forma a concluir sobre as hipóteses formuladas referentes à relevância das características inerentes à empresa, setor, região e ambiente económico na sobrevivência a 5 anos, foram recolhidas informações relativas a 178.437 novas empresas portuguesas,

constituídas entre 2005 e 2014, através da base de dados SABI (empresa analítica da Moody's) e aplicados modelos estatísticos baseados na regressão logística.

Tendo por base os modelos obtidos, é possível afirmar que os fatores extrínsecos às empresas apresentam maior relevância face às características específicas da mesma, na explicação da sobrevivência. A um nível particular, observa-se que o setor de atividade apresenta influência significativa, nomeadamente empresas criadas no setor da agricultura e pescas têm uma possibilidade de sobreviver cerca de 2 vezes superior à do setor dos outros serviços, seguido do setor da indústria transformativa; que, além de não se verificar o domínio do fator urbanização, a localização da empresa não é fundamental; e, que as condições subjacentes a um período de recessão económica são favoráveis ao desenvolvimento de empresas com uma esperança de vida superior a 5 anos, verificando-se que as empresas constituídas em 2008 têm cerca do dobro da possibilidade de sobreviver a longo prazo face a empresas constituídas em 2014.

A constatação dos fatores que são a base para a sobrevivência das empresas, interessam aos investidores, pela escolha antecipada das empresas com maior rendimento futuro; interessam aos gestores, pela adoção de estratégias a adotar numa fase inicial com impacto posterior; e aos políticos, importam pela criação de apoios tendo em conta o perfil de cada empresa.

A dissertação está organizada de acordo com a seguinte estrutura. O primeiro capítulo, respeita à revisão de literatura sobre o tema e formulação de hipóteses. O segundo capítulo é relativo à metodologia, com a apresentação dos dados e modelo a utilizar. O terceiro capítulo trata a análise dos dados de forma individual e da seleção de variáveis. O quarto capítulo está relacionado com a construção dos modelos. O quinto capítulo refere-se à análise do poder discriminatório de cada modelo, comparações entre eles, validação e calibração. No sexto capítulo é realizada a discussão das hipóteses formuladas. Por fim, apresentam-se os principais resultados, limitações e sugestões relativas a estudos futuros.

CAPÍTULO 1.

Revisão de Literatura

Existem diversos estudos relacionados com a sobrevivência das empresas, com o objetivo de compreender quais os determinantes que a sustentam (Schutjens e Wever, 2000; Mata e Portugal, 2002; Fritsch et al., 2006; Acs et al., 2007; Kato e Honjo, 2015). No entanto, os resultados alcançados registam divergências, por vezes bastante pronunciadas, o que dificulta a obtenção de conclusões claras. Esta contrariedade de evidências deve-se à natureza das diferentes variáveis utilizadas, fase do ciclo de vida das empresas em estudo e teorias consideradas. As análises empreendidas tanto se podem centrar nas novas empresas como nos novos empreendedores. Além disso, há autores que se focam apenas no contexto industrial, ignorando o contexto regional/local. É ainda essencial, considerar que os resultados observados em dado país não são diretamente aplicáveis a outros países, pelas especificidades económicas, sociais e ambientais de cada um. Por fim, até o próprio conceito de “sobrevivência” levanta algumas dúvidas.

1.1. Conceito de empresa

Durante o século XX, designavam-se as empresas apenas como unidades de produção, responsáveis pelo fornecimento de bens e/ou serviços à população. O trabalho desenvolvido por Coase (1937) suscitou novas perspetivas sobre o conceito. Passa a reconhecer-se a existência de fronteiras entre mercados e empresas, ora com base no mecanismo do preço, através do qual poderão ocorrer transações no mercado, isto é, entre empresas ou em alternativa ao mesmo, a contratação de fatores de produção por parte do empreendedor (O’Kelley, 2012). Esta nova perspetiva gera os desafios relacionados com a criação de oferta e procura. De acordo com Machlup (1967:27-28), existem pelo menos 21 conceitos de empresa, extensíveis a vários outros, que definem a empresa enquanto “reator abstrato às mudanças do ambiente” nos termos da concorrência de preços, “conjunto de ativos e passivos” no âmbito da contabilidade ou ainda, “pessoa jurídica, com propriedade, direitos e obrigações”.

Uma das definições de empresa - e a mais antiga -, assenta na função de produção, desenvolvida pela teoria económica neoclássica, com base na qual a empresa cria um conjunto de produtos/serviços (*output*) através dos fatores de produção (*input*), tendo por objetivo a maximização do lucro. No entanto, com a evolução tecnológica e a inovação, é requerida uma adaptação constante por parte das empresas, alterando também a perspetiva da sua essência. De

acordo com os designados “contractarianos”, que consideram a empresa como um “nexus-de-contratos”, a essência da empresa é denotada pela autoridade com poder para alocar recursos (O’Kelley, 2012). Não obstante, a empresa definida como um contrato, ou seja, um acordo entre partes interessadas e que regula as relações entre elas (Georgiev, 2015), tem sido alvo de controvérsias, ao nível da separação entre controlo e propriedade.

Pode ainda ser definida pelas suas competências principais (*core*). Esta designação, socorre-se da Teoria Baseada nos Recursos (*Resource Based View* ou RBV), assumindo que a empresa detém vários recursos, tais como competências dos trabalhadores e tecnologia, e que consegue explorá-los de forma eficiente, por forma a alcançar vantagens competitivas sustentáveis, ou seja, que os seus concorrentes não conseguem imitar (Javidan, 1998).

1.2. Novas empresas

Existem algumas dúvidas quanto à denominação de uma empresa como “nova”, nomeadamente se esta definição está relacionada com a propriedade ou com o quadro contabilístico (Johnson, 2018), uma vez que se confundem com as empresas “adquiridas”. Ainda que a mudança de acionista/gestor tenha implicações ao nível da organização de recursos, alterações de estatutos, entre outras mudanças profundas, a empresa continua a existir. De igual forma, a troca de nome, criação de novas linhas de produtos ou expansão para novos mercados, não resulta na formação de uma nova empresa. No entanto, a elaboração de demonstrações financeiras e relatório e contas que apontam para a constituição de uma entidade totalmente diferente, permitem admitir a constituição de uma “nova” empresa.

1.3. Sobrevivência

Brüderl e Preisendörfer (1998) consideraram a sobrevivência como o indicador mínimo do sucesso empresarial. De facto, este conceito tem sido aplicado como medida do desempenho das empresas (Kato e Honjo, 2015), nomeadamente em termos da continuidade de operações (Josefy et al., 2017). A maioria dos investigadores avalia a sobrevivência empresarial, em termos da taxa de saída ou “taxa de morte das novas empresas”, isto é, a “percentagem de empresas criadas em certo ano, que após um determinado período deixam de existir ou de estar listadas no registo de empresas” (Schutjens e Wever, 2000:137), ou da taxa de risco definida como a “probabilidade de saída de uma empresa durante um período específico, dado que sobreviveu até ao início desse período” (Mata e Portugal, 2002:332). De acordo com Weterings e Marsili (2015), esta medida, assumida como o insucesso da empresa é inconsistente, visto que há casos em que a saída pode ser considerada um resultado positivo e uma estratégia de sucesso,

como no caso das fusões e aquisições. Adicionalmente, não tem em atenção a decisão voluntária de sair por parte do empreendedor tendo em vista o investimento num outro negócio, perspectivas de maiores rendimentos ou dada a proximidade da idade da reforma (Kato e Honjo, 2015). Por outro lado, coloca em causa as empresas que são vistas como bem-sucedidas apenas por se encontrarem registadas (Schutjens e Wever, 2000). Para Pajunen e Järvinen (2018:757) “fracasso, sobrevivência e sucesso são aspetos distintos do desempenho empresarial”. Mais recentemente, Josefy et al. (2017:772) defenderam que a sobrevivência, enquanto medida do desempenho empresarial, tem sido utilizada como uma “variável dicotómica, (...) não tendo em consideração o desempenho financeiro ou a alteração de propriedade”.

Por forma a resolver esta visão da saída como um evento homogéneo, há estudos que excluem os dados relativos à mudança de propriedade (Mata e Portugal, 2002; Geroski et al., 2010) ou que consideram as fusões e aquisições como estratégias bem-sucedidas, considerando-as *right-censored* (Pajunen e Järvinen, 2018). No estudo sobre novas empresas na Holanda, Weterings e Marsili (2015) analisaram as saídas por fusões e aquisições face às saídas por insolvência. Kato e Honjo (2015) na análise da sobrevivência de novas empresas japonesas, na indústria transformativa, consideraram três tipos de saídas, comparando falência como indicador de fracasso, com liquidação voluntária e fusões, como indicadores de sucesso. Em Espanha, Tong e Saladrigues (2018) utilizaram como critério o não reporte de receitas operacionais das empresas na indústria transformativa e distribuição durante 2 anos consecutivos. Com base na análise em 222 artigos dos jornais científicos mais relevantes, Josefy et al. (2017) sugerem a solvência enquanto medida do desempenho das novas empresas.

De acordo com os dados recolhidos de empresas americanas pelo U.S. Bureau of Labor Statistics entre 1994 e 2009, apenas 50% das novas empresas sobrevive ao fim de 5 anos e um terço encerra nos primeiros dois anos de atividade (Clayton et al., 2013). De forma semelhante, a OECD (1996) observa que menos de metade das start-ups é considerada sobrevivente por mais de 5 anos. Ainda é de referir que uma das mais baixas taxas de sobrevivência das empresas europeias é a das empresas portuguesas, com dois terços das novas empresas a terminar a atividade empresarial ao fim de 5 anos (European Commission, 2019).

1.4. Fatores com influência na sobrevivência das empresas

Nas últimas três décadas tem-se assistido a um aumento no número de pesquisas sobre a sobrevivência das empresas (Josefy et al., 2017). A literatura relativa aos fatores críticos para o desempenho das novas empresas é bastante vasta e há uma diversidade de evidências contraditórias. A Organização Industrial (*Industrial Organization* ou IO), Ecologia

Organizacional (*Organization Ecology* ou OE), bem como a RBV, entre outras, são exemplos de literaturas que têm sido abordadas e utilizadas como suporte ao desenvolvimento da análise da sobrevivência (Geroski et al., 2010). Entre os anos 50 e 60, a maioria dos estudos centrava-se na relação entre o crescimento da empresa e o seu tamanho (Acs et al., 2007). Seguindo este trajeto, procedeu-se à avaliação de outras variáveis intrínsecas à empresa e, posteriormente, ao próprio mercado e impacto do capital humano. Tendo por base estudos anteriores, pode-se assumir a existência de três vetores de análise, nomeadamente o de fatores relacionados com o empreendedor, características da nova empresa e fatores externos, no qual se incluem determinantes relacionados com indústria, região e ambiente macroeconómico (Schutjens e Wever, 2000; Strotmann, 2007).

As características do empreendedor que são maioritariamente objeto de análise para a sobrevivência das empresas centram-se na experiência de gestão/indústria, educação, idade e género (Cooper, 1981, Brüderl e Preisendörfer, 1998, Schutjens e Wever, 2000). Devido a limitações na recolha de dados relativos às características do empreendedor, o estudo vai-se centrar nas especificidades das empresas e em determinantes extrínsecas às mesmas.

1.4.1. Fatores relacionados com a empresa

1.4.1.1. Tamanho

Na literatura da sobrevivência das novas empresas, o tamanho inicial é um dos fatores mais explorados e cujo consenso relativo à sua influência é, de uma forma geral, unânime. Brixy e Grotz (2007) sugeriram que as novas empresas são geralmente mais pequenas e não possuem experiência relevante, sujeitando-se a uma competição superior. A adoção de uma estrutura pequena na entrada no mercado pode estar relacionada com um conjunto de condicionantes, que se traduzem em maiores problemas na obtenção de financiamento pelo desconhecimento da marca por parte dos investidores (Honjo e Kato, 2019), desvantagens em termos de eficiência de escala (Audretsch e Mahmood, 1994), na escolha de minimizar o risco do investimento derivado da inexperiência inicial ou a fraca qualidade do capital humano (Strotmann, 2007). A incerteza relativa à eficiência nos primeiros anos de atividade limita a entrada em grande escala das novas empresas (Jovanovic, 1982). Esta desvantagem das novas empresas denominada *liability of smallness*, poderá ainda resultar dos impostos rigorosos, burocracia excessiva e desvantagens superiores na competição pelo custo do trabalho (Falck, 2007).

Pelo contrário, as empresas de maior dimensão são geralmente diversificadas, apresentando alternativas de produtos consoante o mercado, e reconhecidas pela gestão eficiente permitindo-

lhes optarem por uma dimensão superior a custos inferiores (Geroski et al., 2010). Adicionalmente, a entrada em larga escala está relacionada com expectativas otimistas quanto ao futuro sucesso, pelo que a ocorrência de gastos excessivos e inesperados durante mais tempo não se apresenta como um problema maior.

Esta desigualdade em termos de escala, obriga a que as empresas recentes cresçam mais rapidamente, tendo como única alternativa a saída do mercado (Weterings e Marsili, 2015). De facto, há várias evidências de que a probabilidade de sobrevivência está positivamente relacionada com o tamanho (Schutjens e Wever, 2000; Mata e Portugal, 2002; Falck, 2007; Strotmann, 2007; Weterings e Marsili, 2015; Simón-Moya et al., 2016; Sarmiento e Forte, 2019). Mata e Portugal (2002), com base na análise em 124.249 novas empresas portuguesas e filiais multinacionais constituídas em Portugal, entre 1983 e 1989, concluíram que um aumento do tamanho em 10%, diminuía a probabilidade de não sobreviver em 3,7%. Geroski et al. (2010), observaram uma correlação negativa e significativa entre o tamanho da empresa e o risco de saída do mercado. Na verdade, Simon Moya et al. (2016), verificaram que o risco de insucesso diminui 22,6% por cada colaborador extra. Ainda assim, Sarmiento e Forte (2019) verificaram um efeito positivo do tamanho na probabilidade de sair, sugerindo que as empresas de maior dimensão enfrentam maiores obstáculos aquando a realização de ajustamentos necessários em períodos de crise, derivados da complexidade de procedimentos e coordenação.

Falck (2007) constatou que o tamanho é o melhor indicador, ao nível da empresa, para prever o risco de fracasso das novas empresas. Este pode ser mensurado em função do número de trabalhadores que constituem a empresa¹, valor do capital inicial ou ainda rotação do número de colaboradores ou do ativo (Strotmann, 2007). Os efeitos do tamanho poderão depender do tipo de indústria (Hannan e Carroll, 1992) e existem contradições no que concerne à fase que apresenta maior impacto na probabilidade de sobrevivência, se o tamanho inicial ou o corrente (Falck, 2007; Geroski et al., 2010). Tendo em conta que o estudo presente tem por base as condições iniciais, o indicador a considerar é o tamanho inicial.

1.4.1.2. Capital

A disponibilidade de recursos financeiros adequados ao estabelecimento de novas empresas é um determinante importante na criação de novas empresas (Acs et al., 2007). Existem várias

¹ De acordo com o indicador da OECD – *Enterprises by business size* -as empresas são geralmente categorizadas de acordo com o número de colaboradores: micro empresa (menos de 10 colaboradores), pequena empresa (entre 10 a 49 colaboradores), empresa média (entre 50 a 249 colaboradores) e grande empresa (igual ou superior a 250 colaboradores).

medidas financeiras que são utilizadas para o estudo da sobrevivência, como sejam a solvência, liquidez e lucro obtido ao longo dos anos (Tong e Saladrighes, 2018). Pese embora sejam fatores a considerar, a análise centra-se nas componentes iniciais com possível influência a longo prazo, no momento de criação das empresas.

O capital social de uma empresa diz respeito aos montantes investidos pelos sócios ou acionistas da empresa, numa fase inicial da atividade empresarial, condicionando o direito aos lucros e o direito de voto do sócio (Caixa Geral de Depósitos, 2015). Nos termos do n.º 5 do artigo 276º do Código das Sociedades Comerciais (CSC), o montante mínimo do capital social para a constituição de uma sociedade anónima é de 50.000 euros.

Segundo Fotopoulos e Louri (2000), as empresas que possuem um montante elevado de capital inicial são menos suscetíveis ao risco de sair, tendo uma maior capacidade para suportar as adversidades a longo prazo. De facto, por cada 1000 euros adicionais, a empresa diminui o risco de saída por insolvência em 0,4% (Simón-Moya et al., 2016). No entanto, Schutjens e Wever (2000) num estudo realizado a cerca de 2000 *start-ups* holandesas entre 1994 e 1997, não encontraram qualquer efeito significativo do capital inicial na sobrevivência das empresas. Barbosa (2016) argumenta ainda que a alavancagem financeira poderá limitar a sobrevivência a longo prazo das empresas, devido ao aumento dos serviços de dívida, caso os grandes investimentos realizados resultem num risco real ou pela concorrência com disponibilidade financeira. O aumento do rácio da dívida gera um incremento de 33% na probabilidade de saída por parte das empresas (Sarmiento e Forte, 2019).

1.4.1.3. Forma Jurídica

A estrutura legal é um dos fatores que tem sido incluído no estudo da sobrevivência empresarial (Mata e Portugal, 2002, Kato e Honjo, 2015, Huggins et al., 2017). Em Portugal, as diversas formas jurídicas dividem-se em duas categorias – singulares (como o empresário em nome individual, sociedade unipessoal por quotas e estabelecimento individual de responsabilidade limitada) e coletivas (como a Sociedade por Quotas (Limitada -LDA), a Sociedade Anónima (S.A.), Sociedade em Nome Coletivo (SNC), Sociedade em Comandita e Cooperativa) que poderão variar entre sociedades de responsabilidade limitada e ilimitada.

Nas empresas de responsabilidade limitada, é o capital social que responde pelas dívidas da empresa, pelo que a atividade se prolonga até à insolvência. Por outro lado, nas empresas de responsabilidade ilimitada são, de uma forma geral, os proprietários que respondem pelas dívidas contraídas pela empresa através da entrega dos bens que compõem o seu património pessoal e empresarial, pelo que a existência da empresa depende da capacidade dos seus

proprietários (Kato e Honjo, 2015). Num estudo relativo a 1.425 start-ups do País de Gales, entre 1995 e 1999, Huggins et al. (2017) verificaram que o facto de uma empresa se apresentar sob a forma legal de “responsabilidade limitada” ou como uma associação tem influência positiva para a sobrevivência, ao contrário do comerciante individual. De igual modo, Mata e Portugal (2002) verificaram que as empresas cuja estrutura legal seja de responsabilidade limitada sobrevivem mais tempo.

1.4.1.4. Pertença a um grupo

A criação de uma nova empresa implica ainda a definição do seu tipo, como seja uma empresa singular e independente, que poderá (ou não) resultar da intenção por parte de um ex-trabalhador de uma empresa estabelecida, ou subsidiária de uma empresa já existente. Strotmann (2007) sugere que um estabelecimento, como parte constituinte de uma empresa, beneficia de recursos financeiros e experiência. No entanto, o mesmo autor acrescenta que a reestruturação e criação de um novo negócio resultante da decisão voluntária de empreender, providencia ganhos de eficiência e proporciona o aumento da probabilidade de sobrevivência das empresas singulares, evidenciando esse efeito positivo.

Neste sentido, estudos recentes têm avaliado as relações entre empresas, nomeadamente ao nível *parent-spinoff*, i.e., novas empresas criadas por outras já estabelecidas, geralmente na mesma região, que podem ficar detentoras de uma participação financeira na recente empresa e, até, providenciar recursos ou ainda empresas criadas por indivíduos que trabalharam noutras empresas (Bagley, 2019). As *spinoffs* poderão ainda beneficiar da experiência, reputação e credibilidade da “empresa-mãe” e do seu conhecimento de novos mercados, bem como explorar novas tecnologias. Num estudo relativo a empresas de Estocolmo, Bagley (2019) concluiu que as empresas com conhecimento hereditário e que estão centralmente posicionadas na cadeia de canais de conhecimento da empresa-mãe têm maiores probabilidades de sobreviver. Weterings e Marsili (2014) não constataram qualquer efeito significativo. Ainda assim, Brixy e Grotz (2007) sugeriram que as longas relações entre grandes empresas podem limitar a formação de novas, comprometendo o ambiente propício à inovação.

1.4.2. Fatores externos

O desempenho das empresas poderá ainda ser influenciado pelo setor no qual opera, pela dimensão local ou por componentes macroeconómicas (Mata e Portugal, 2002; Acs et al., 2007; Simón-Moya et al., 2016).

1.4.2.1. Setores de atividade económica

Encontram-se diversos estudos empíricos que verificam a debilidade das novas empresas nos anos iniciais de existência, essencialmente derivados da intensidade competitiva do mercado no qual atuam (Fritsch et al., 2006). Para Pajunen e Järvinen (2018), o setor apresenta efeitos diretos no desempenho das empresas, derivados das limitações existentes e modelos de negócio.

É observado por vários autores que as indústrias cujo tamanho eficiente mínimo (*Minimum Efficient Size* ou MES) é relativamente mais elevado, ou seja, na qual subsistem empresas de maior dimensão, apresentam valores inferiores de taxas de sobrevivência (Audretsch, 1995; Fritsch et al., 2006; Mata e Portugal, 2002; Strotmann, 2007). No caso das indústrias de capital intensivo é requerido que as novas empresas possuam um certo montante de recursos, por forma a atingir o tamanho necessário à sobrevivência. Deste modo, tanto o superior MES, como a intensidade de capital na indústria, atuam não só como barreiras à entrada, mas também propiciam o processo de seleção autónomo das novas empresas, contribuindo para a alta qualidade das remanescentes e, conseqüentemente, futura sobrevivência (Fritsch et al., 2006). O setor da indústria transformativa é caracterizado por um MES superior e empresas especializadas em capital intensivo, ao contrário de setores relacionados com os serviços.

Adicionalmente, o impacto positivo do crescimento da indústria reflete de igual forma a sobrevivência das empresas (Audretsch e Mahmood, 1994; Mata e Portugal, 2002; Strotmann, 2007). Quanto maiores os lucros, menor necessidade de competir pelos clientes existentes (Mata e Portugal, 2002). Ainda que a necessidade de realização de ajustamentos contínuos ao longo do ciclo de vida da indústria obrigue as empresas a realizar gastos, numa primeira fase do ciclo, as pequenas empresas poderão beneficiar da criação de produtos inovadores pela sua estrutura organizacional achatada e menos rígida (Falck, 2007).

A inovação é um dos potenciais determinantes do sucesso das empresas e uma indústria promotora da inovação tanto poderá promover a sustentabilidade das novas empresas pela localização próxima às empresas especializadas no conhecimento tecnológico, como induzi-las a um alto risco de insucesso (Falck, 2007). Geralmente, além do estudo da inovação da indústria pela existência de departamentos de Investigação e Desenvolvimento (I&D), é relevante a análise do seu “regime tecnológico”, considerado mais assertivo. De acordo com Audretsch e Mahmood (1994), num regime rotinizado, caracterizado pela presença de empresas estabelecidas, a necessidade de mais tecnologia obriga de igual forma a uma maior capacidade de financiamento. Já no caso de um regime empreendedor, as novas empresas representam a grande parte da atividade inovadora desenvolvida, pela criação de produtos originais e, competem pelas novas ideias no âmbito da destruição criativa, saindo do mercado em caso de

fraco desempenho (Kato e Honjo, 2015). O setor das atividades de informação e comunicação compreende as atividades relativas aos media, telecomunicações e tecnologias de informação. Segundo Audretsch (1995), as empresas pertencentes aos setores intensivos em tecnologias têm maiores probabilidades de encerrar a sua atividade empresarial.

Existe uma forte evidência de que o comportamento das taxas de entrada da indústria na qual a nova empresa se estabelecerá é bastante similar ao comportamento das taxas de saída (Mata e Portugal, 2002; Strotmann, 2007; Audretsch e Fritsch, 2002; Weterings e Marsili, 2014). Audretsch (1995:590) afirma que a significância dos efeitos de escala, resulta na existência de um regime *revolving-door*, desenvolvido por Jovanovic (1982), associado a um regime rotinizado no qual as empresas estabelecidas beneficiam de escala superior, eliminando as novas empresas do mercado. No entanto, Brixy e Grotz (2007) constataram a formulação de um paradoxo, pois tendo em conta que o objetivo das políticas económicas é a criação e o crescimento de novas empresas, a sua permanência no mercado poderá, simultaneamente, limitar a formação de outras e estabelecer limitações e implicações ao nível do emprego.

Dado que os fatores associados à indústria, variam de acordo com o setor em estudo, é relevante analisar o efeito do mesmo na sobrevivência. Os custos irrecuperáveis são relativamente mais baixos no setor dos serviços, o que potencia de forma positiva uma maior criação de empresas, e de forma negativa igual saída, tendo em conta o supramencionado relativo aos fluxos nas indústrias (Brixy e Grotz, 2007). Há estudos que confirmam a maior probabilidade de não sobrevivência por parte de novas empresas no setor dos serviços (Brixy e Grotz, 2007, Weterings e Marsili, 2014, Simón-Moya et al., 2016). Já no setor da indústria, pela necessidade de elevados investimentos iniciais e da formulação de um planeamento cuidadoso, a situação inverte, e a probabilidade de sobrevivência é maior. Ainda assim, Pajunen e Järvinen (2018) encontraram evidência empírica de que empresas no setor dos serviços apresentam menores taxas de insucesso, contrariamente às novas empresas no setor tecnológico. Estas indústrias caracterizadas pela maior assimetria de informação, rápida alteração tecnológica e ciclos de produto mais curtos, requerem uma competição constante por parte das empresas pelo desenvolvimento de produtos inovadores (Kato e Honjo, 2015).

1.4.2.2. Região

Para além da dimensão da indústria, que tem sido amplamente estudada, é igualmente relevante ter em conta o efeito do espaço, ou seja, a dimensão regional. Geralmente, tem-se em consideração o tipo de região e o seu crescimento, como medida do desenvolvimento

económico local (Falck, 2007). Nos estudos recentes, têm-se relacionado ambas as dimensões, de modo a realizar análises mais completas.

As grandes aglomerações tendem a providenciar bases mais favoráveis à constituição das empresas, em termos de recursos locais e criação de um ambiente propício ao desenvolvimento da atividade empresarial (Brixy e Grotz, 2007). As vantagens traduzem-se numa proximidade a fornecedores e clientes, a centros de investigação e desenvolvimento, bem como no acesso a um conjunto alargado de serviços especializados ou ainda, maior diversificação do mercado de trabalho. De facto, a teoria da aglomeração argumenta que regiões com maior número de pessoal qualificado observam uma elevada formação de empresas. No entanto, a taxas de sobrevivência registam um comportamento oposto. Embora a literatura da geografia económica acrescente outras externalidades positivas como os baixos custos de transporte ou a dimensão do mercado, a teoria do sistema urbano neoclássico, defende que a aglomeração nem sempre acarreta externalidades positivas (Krugman, 1996). Assiste-se a um maior congestionamento, necessidade de oferecer salários mais elevados e maior competição (Acs et al., 2007), assim como a uma forte variabilidade da procura, pela maior disponibilidade de fornecedores substitutos (Brixy e Grotz, 2007). Estas diferenças de rendas e salários entre áreas urbanas e rurais, normalmente avaliadas pelo fator da urbanização, influenciam a estrutura de custos das novas empresas, apresentando implicações para a sua sobrevivência. Contrariamente a Fotopoulos e Louri (2000), que comprovaram que as grandes áreas urbanas registam taxas de risco reduzidas, Strotmann (2007), encontrou evidências de que há 30% maior risco de saída de empresas em áreas urbanas. Não obstante, outros estudos não detetaram qualquer efeito da aglomeração (Schutjens e Wever, 2000, Simón-Moya et al., 2016).

Para além disto, são discutidas questões ao nível dos benefícios proporcionados pela localização geográfica entre empresas da mesma indústria, com base no modelo Marshall-Arrow-Romer (MAR), por oposição ao conhecimento partilhado por empresas de outras indústrias, defendido por Jacobs (1969). De acordo com o modelo MAR, as cidades permitem um maior contacto entre entidades, o que facilita o desenvolvimento de *knowledge spillovers* entre indústrias similares (Acs et al., 2007). De forma semelhante, He e Yang (2016) sugeriram que empresas localizadas em indústrias chave e que se relacionam entre elas poderão, conseqüentemente, beneficiar das economias de localização. A proximidade geográfica assume assim, um papel importante na transmissão de conhecimentos às novas empresas após a sua entrada no mercado. Bagley (2019) constatou que o efeito da distância geográfica entre empresas do mesmo cluster assume a forma de um U invertido, devendo-se à substituição dos benefícios das economias de aglomeração pelos efeitos da competição, que se diluem

posteriormente. Ainda assim, He e Yang (2016) constataram a existência de uma competição intensa intra-indústrias-chave, contribuindo para o fraco desempenho empresarial. De uma outra perspectiva, Weterings e Marsili (2015) sugerem que o conhecimento partilhado é limitado ao número de empresas presentes, seguindo a teoria da ecologia organizacional. No lado oposto, Jacobs (1969) argumenta que a variedade de indústrias e a sua proximidade promove a inovação e o crescimento, pela complementaridade de ideias/conhecimento – *inter-industry spillover*. Acs et al. (2007) observaram, na análise a mais de 11 milhões de estabelecimentos nos Estados Unidos da América, que a intensidade de indústrias diversificadas proporcionava maiores probabilidades de sobrevivência. Por oposição, Weterings e Marsili (2014) encontraram evidência de que a probabilidade de insolvência diminui 8,6% quanto maior a percentagem de empresas ativas em dada indústria pelo total de empresas ativas em todas as indústrias. Strotmman (2006) não constatou qualquer efeito significativo da heterogeneidade.

A localização escolhida pelas empresas é de igual forma inerente à proximidade do capital humano (Brixy e Grotz, 2007). Quanto maior o número de pessoas qualificadas, maior o capital humano regional enraizado nas competências fundamentais à criação de novas ideias, desenvolvendo-se um ambiente ideal à propagação do conhecimento - *knowledge spillovers*.

Outros dos aspetos da localização das empresas inclui o nível de acesso a fontes de financiamento, quer públicas como subsídios do governo ou fundos comunitários, quer privadas, como capitais próprios, *crowdfunding*, *leasing*, *business angels*, empréstimos bancários, entre outros. Os apoios locais permitem mitigar os efeitos da competição e são atribuídos frequentemente no setor das novas tecnologias pela maior incerteza da procura (He e Yang, 2016). Não obstante, promovem a permanência no mercado das empresas menos produtivas e eficientes e reduzem a sobrevivência daquelas que não usufruem das mesmas condições de financiamento, aumentando os custos sociais com a intervenção governamental. Pajunen e Järvinen (2018) constaram que o financiamento público proporcionava uma redução na taxa de insucesso, por oposição ao financiamento privado.

1.4.2.3. Ambiente Macroeconómico

Outra das componentes com potencial influenciador na atividade futuras das empresas é o próprio ambiente macroeconómico (Acs et al., 2007; Geroski et al., 2010; Barbosa, 2016; Simón-Moya et al., 2016; Sarmiento e Forte, 2019), que vigora no ano de constituição das empresas. Este indicador traduz-se no mercado de trabalho, procura, taxas de juro ou crescimento do Produto Interno Bruto (PIB).

1.4.2.3.1. Ciclos Económicos

Os ciclos económicos dizem respeito a períodos alternados de contração e expansão da economia de determinado país. Uma das principais razões da sua importância traduz-se na mudança que provocam nas expectativas futuras das empresas e no seu desempenho a longo prazo (Geroski et al., 2010).

Em períodos de recessão, as novas empresas têm maior tendência a sair pela falta de liquidez e tempo para se darem a conhecer ao mercado (Gersoki et al., 2010). De um ponto de vista divergente, acredita-se que os indivíduos têm maior propensão para a criação de novos negócios, uma vez que existe a necessidade de arranjar emprego face à falta de opções de trabalho e devido aos menores custos de oportunidade (Simón-Moya et al., 2016), os designados “empreendedores por necessidade”, na terminologia do *General Entrepreneurship Monitor (GEM)* ou *push*, por oposição aos “empreendedores por oportunidade” ou *pull* (Simón-Moya et al., 2016), motivados pelas próprias ideias e pela possibilidade de exploração de diferentes negócios. De facto, Acs et al. (2007) observaram que altas taxas de desemprego contribuem para a manutenção de elevadas taxas de sobrevivência, tanto em períodos de recessão como de crescimento económico, sobretudo na área dos serviços, visto que não requerem elevado capital inicial. Ainda assim, Simón-Moya et al. (2016:160) evidenciaram que a probabilidade de sobrevivência é maior durante crises. O resultado alcançado é consequência da falta de alternativas de emprego em períodos de contração, que além de forçar o início da atividade empresarial, obriga à permanência no negócio, ou seja, o efeito dos baixos custos de oportunidade para qualquer tipo de empreendedores é superior ao “efeito da diminuição da procura que leva ao insucesso das novas empresas”. Este facto suporta novamente as teorias que afirmam que a sobrevivência não é similar ao sucesso, visto que a empresa poderá continuar no mercado e simultaneamente não gerar lucros. Por outro lado, Barbosa (2016) defende que a contração da procura agregada influencia de forma negativa as vendas e o lucro das empresas, promovendo a sua saída do mercado, e que a competição sobre as novas empresas é superior em períodos de recessão.

Geroski et al. (2010) concluíram que empresas criadas durante períodos de crescimento económico têm maiores probabilidades de sobreviver. A expansão da economia e subjacente aumento da procura, cria oportunidades para a criação de empresas e aumenta as oportunidades de trabalho (Simón-Moya et al., 2016). Tendo em conta que algumas das condições iniciais de criação das empresas têm um efeito a longo prazo na sobrevivência das empresas (Geroski et al. 2010; Moreira, 2018), um choque positivo da procura agregada permite que as novas empresas acumulem reputação para o futuro. Numa outra perspetiva, Acs et al. (2007)

propuseram que a alteração de emprego poderá estimular o desenvolvimento de novas empresas. Com a melhoria das perspetivas económicas, há o alargamento a várias alternativas de nichos de mercado com poucas barreiras, aumentando a propensão à criação de mais empresas e melhores perspetivas de sobrevivência. No entanto, as empresas mais pequenas e pouco produtivas formadas em períodos de expansão económica, terão maior tendência para sair do mercado, assim que as condições comecem a piorar (Moreira, 2018).

Tendo em conta a potencial influência do ambiente macroeconómico, é necessário efetuar a sua contextualização para um melhor entendimento das suas repercussões na análise posterior.

1.5. Breve descrição da economia portuguesa entre 2005 e 2015

Durante a década de 90, especialmente a partir de 1995, registou-se um crescimento significativo da economia e, posteriormente, entre 2006-2009 (Observatório sobre Crises e Alternativas, 2013). O primeiro período ficou marcado pela abundância de créditos bancários com baixos juros, maioritariamente concedidos ao setor da construção e do imobiliário, resultando num endividamento exponencial promotor de taxas de investimento elevadas. Em meados de 2006, assiste-se a nova atribuição de créditos para colmatar as necessidades de liquidez, com o aproveitamento dos fluxos financeiros europeus. Esta estratégia incentivou ao contínuo refinanciamento dos empréstimos e aumento do défice externo.

A recessão que se seguiu resultou em taxas de crescimento do PIB na ordem dos -3,12% em 2009 e -4,06% em 2012 (Figura 1.1.), as mais baixas desde que há registo.

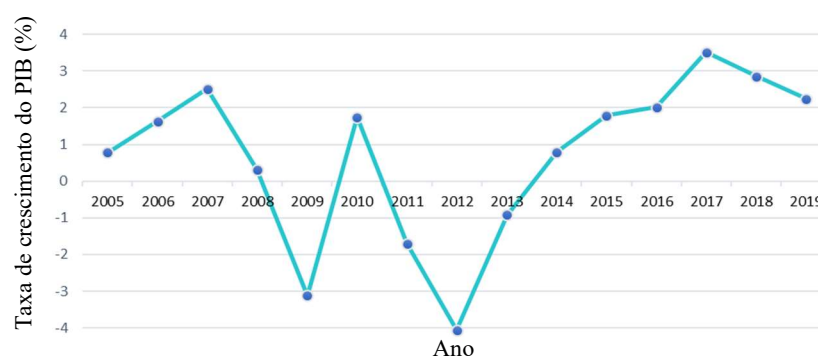


Figura 1.1. - Taxa crescimento do PIB em Portugal, entre 2005 e 2019

Fonte: Pordata, 2020

Como se pode observar na figura 1.2., em 2013, registaram-se 707.800 (16,2%) desempregados e a partir desse ano, observa-se uma diminuição nesse número.

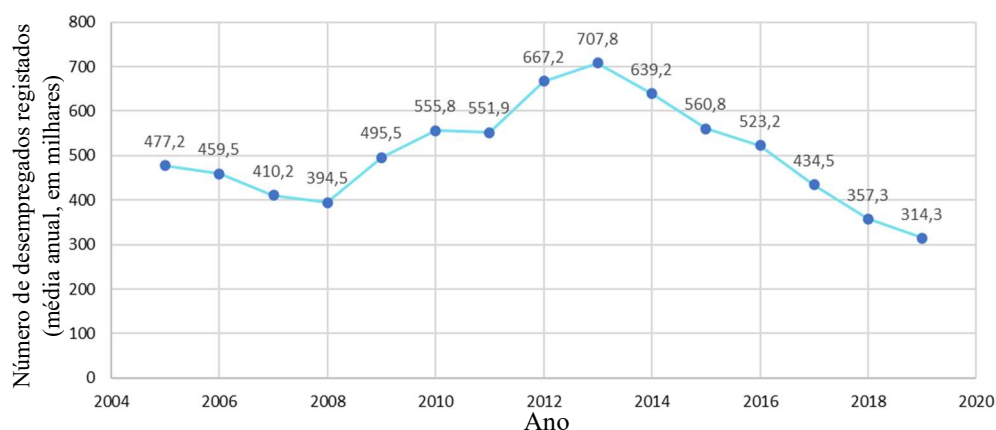


Figura 1.2. - Desempregados inscritos no centro de emprego (média anual, em milhares)
Fonte: Pordata, 2020

Por ano são constituídas cerca de 30.000 empresas. Em 2019, registou-se a criação de 31.054 empresas desde o início do ano até final de julho, assistindo a uma variação homóloga de 10,5% (Informa D&B, 2019). Ainda assim, no mesmo período, mais de 10.000 empresas encerraram. De acordo com dados do INE (2020), foi precisamente em 2013 que se registou a maior taxa de criação de empresas (10,83%), ou seja, 38.191 empresas, localizadas sobretudo na região Norte, Centro e Alentejo, sendo de igual modo o ano em que a taxa de sobrevivência a 5 anos foi superior (52,88%), desde 2008.

A região Norte foi a que mais beneficiou com a atribuição de fundos comunitários, entre 2007 e 2013, com uma percentagem na ordem dos 39,6%, correspondente a 8.873.206 euros. A região Centro registou participações de 6.132.161 euros (27,3% do total nacional), seguida da região do Alentejo, que beneficiou de 3.024.019 euros (13,5%) para desenvolvimento regional/empresarial/económico (CCDR Alentejo, 2014).

A contrariedade de evidências relacionadas com os mais diversos fatores torna relevante a realização de uma análise empírica relativa a Portugal. Dos estudos abrangidos, que versam sobre as empresas portuguesas, Geroski et al. (2010), focaram-se na criação e sobrevivência das mesmas em anos de relativa estabilidade e Sarmento e Forte (2019) estudaram o efeito das empresas estrangeiras na probabilidade de saída das empresas portuguesas. Sarmento e Forte, (2019) consideraram o período de 2007-2013. No entanto, analisaram as empresas já estabelecidas e não tiveram em consideração os fatores iniciais de constituição, assim como não analisaram o período de recuperação e estabilidade económica. Neste sentido, esta dissertação contribui para o desenvolvimento do estudo da sobrevivência das empresas portuguesas, colmatando pontos essenciais, i.e., fatores iniciais com impacto nos futuros resultados.

1.6. Formulação de Hipóteses

Considerando que as empresas têm características inerentes às mesmas com potencial influência na sua capacidade de sobrevivência a longo prazo, torna-se necessário compreender se são mais ou menos significativas que as propriedades extrínsecas, como sejam a região, setor de atividade económica ou o ambiente de constituição, em virtude da adoção de estratégias para o futuro. Diferentes contextos económicos proporcionam maiores probabilidades de sobrevivência para empresas com perfis distintos (Simón-Moya et al., 2016). No entanto, serão esses fatores externos que têm maior impacto? De acordo com Falck (2007), à exceção do tamanho da empresa, todas as restantes variáveis relacionadas com as suas características não apresentam qualquer significância no risco de insucesso.

H1: As variáveis intrínsecas às empresas têm menor influência na probabilidade de sobrevivência que as variáveis extrínsecas.

O setor no qual as empresas atuam apresenta limitações no seu ambiente (Pajunen e Järvinen, 2018), como, por exemplo, as indústrias caracterizadas pela rápida mudança, derivada de ciclos de produto curtos, que geram uma forte competitividade entre empresas, privilegiando as que têm uma estrutura mais flexível (Kato e Honjo, 2015) ou, ainda, a existência de uma escala de eficiência mínima da indústria superior que gera desvantagens em termos de custo para as pequenas empresas, nomeadamente para as mais recentes no mercado (Audretsch, 1995). Deste modo, o desempenho das empresas é influenciado pelos efeitos derivados do setor.

H2: O setor de atividade económica é fator fundamental para a sobrevivência das empresas.

O local no qual a empresa se estabelece não apresentou qualquer impacto para Schutjens e Wever (2000). Não obstante, para Fritsch et al. (2006) a inclusão do fator localização que tem subjacente condições inerentes ao mesmo (tais como, o número de empresas presentes, emprego ou atividade inovadora local) é essencial à compreensão da sobrevivência.

H3: A região é fator fundamental para a sobrevivência das empresas.

As diferentes fases dos ciclos económicos apresentam diversos efeitos na capacidade das empresas para manter a sua atividade empresarial. De forma contrária aos períodos de alto crescimento, promotores da criação de empresas e manutenção do emprego (Simón-Moya et al., 2016), as recessões costumam afetar de forma negativa a atividade das empresas, pela contração da procura (Barbosa, 2016). Ainda assim, esta fase depressiva poder-se-á traduzir em oportunidades para a exploração de ineficiências ou ainda, num alívio da competição sobre as novas empresas, com a saída de mais empresas do mercado (Simón-Moya et al., 2016).

H4: O ambiente macroeconómico é fator fundamental para a sobrevivência das empresas.

CAPÍTULO 2.

Metodologia**2.1. Regressão Logística (Binária)**

O modelo de regressão logística é um modelo estatístico associado aos modelos lineares generalizados (*generalized linear models*), utilizado para análise da relação de um conjunto de variáveis explicativas numéricas (nominais, ordinais ou contínuas) com uma variável dependente (Y). A principal distinção entre o modelo de regressão logística e o de regressão linear traduz-se na natureza da variável dependente, que, no caso da regressão logística é de natureza binária (ou dicotómica).

Além da utilização de uma variável dependente dicotómica, os pressupostos deste modelo são os seguintes:

- i) observações independentes
- ii) utilização de uma grande amostra ²
- iii) ausência de multicolinearidade
- iv) linearidade do *logit* (*log odds*) e as variáveis independentes contínuas

Os problemas de regressão relacionam-se com a estimação do valor esperado da variável dependente (Y), dado o valor das variáveis independentes X, o que se traduz na designada “média condicionada” $E[Y | x]$. Tendo em conta que a variável dependente assume um carácter dicotómico, a média apenas poderá tomar valores entre 0 e 1. Apresentando-se como um modelo não-linear, devido ao não cumprimento dos pressupostos de normalidade e variância constante dos resíduos (Sarkar e Midi, 2010), a regressão logística de carácter univariado, descreve esta relação entre variáveis com base na seguinte expressão:

$$\pi(x) = E[Y|x] = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

A curva resultante assume assim a “forma em S”, com implicações na média esperada. Por forma a adequar o modelo, a regressão logística aplica o método da verossimilhança máxima (*Maximum Likelihood*), que estima os valores dos parâmetros do modelo que maximizam a probabilidade de obter o conjunto de dados observados, de modo iterativo. Com

² A amostra mínima (N) deverá ser igual ao rácio de 10 a multiplicar pelo n.º de variáveis independentes (k) pelo menor dos casos positivos ou negativos (p), ou seja, $N=10k/p$ (Peduzzi et al., 1996).

base nos coeficientes gerados é calculada a “transformação *logit*” ou o logaritmo do *odds ratio* (razão de possibilidades) da probabilidade em estudo, de caráter linear:

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) \quad (2)$$

em que $-\infty \leq g(x) \leq +\infty$, $-\infty \leq x \leq +\infty$.

O *odds ratio* ou razão de possibilidades é descrita da seguinte forma:

$$\text{odds} = \frac{\text{probabilidade de um evento ocorrer}}{\text{probabilidade de um evento não ocorrer}} \quad (3)$$

No caso em que:

- OR=1, o *odds* do evento é equivalente para os dois grupos;
- OR>1, o *odds* do evento é superior para o 1º grupo;
- OR<1, o *odds* do evento é inferior para o 1º grupo.

Nas situações com mais de uma variável independente é aplicada a regressão logística multivariada. Tendo em consideração *as n* observações independentes referentes ao par (x_i, y_i) , no qual o vetor nas *m* variáveis explicativas é dado por x_i e a variável dependente representada por y_i , obtém-se a função logística:

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}} \quad (4)$$

2.2. Dados

Neste capítulo será efetuada a análise do conjunto de dados, respeitantes às empresas criadas entre 2005 e 2014 em Portugal, recolhidos da base de dados SABI do Bureau van Dijk. Os dados são de natureza secundária e o seu tratamento e estudo teve por base a utilização do software *SPSS (Statistical Package for the Social Sciences)*. O objetivo do estudo centra-se na análise dos fatores iniciais de constituição das empresas que contribuem para a sobrevivência das mesmas e compreender a influência dos fatores indústria, espaço e tempo na sua capacidade de continuidade da atividade empresarial.

Como referido no capítulo 2, é afirmado que mais de metade das empresas não sobrevive além de 5 anos. Com base na amostra de 178.437 empresas, das quais 96.579 encerraram, cerca de 57% não sobreviveram mais de 5 anos.

Com base na figura 2.1., verifica-se que a taxa de encerramento mais expressiva, cerca de 13,5%, se registou no segundo ano de atividade das empresas em estudo. Ainda que seja tendencioso a escolha de uma análise com base nos 2 anos de sobrevivência, pela evidência de uma maior taxa de encerramentos nesse período, é de considerar todo o procedimento moroso a que estão sujeitas as empresas de maior dimensão. Devido à necessidade de investimentos significativos e processos burocráticos, espera-se que o desenvolvimento da sua atividade empresarial em pleno seja mais tardia, pelo que opção por um período de análise superior favorece a implementação das novas empresas (Brixy e Grotz, 2007).

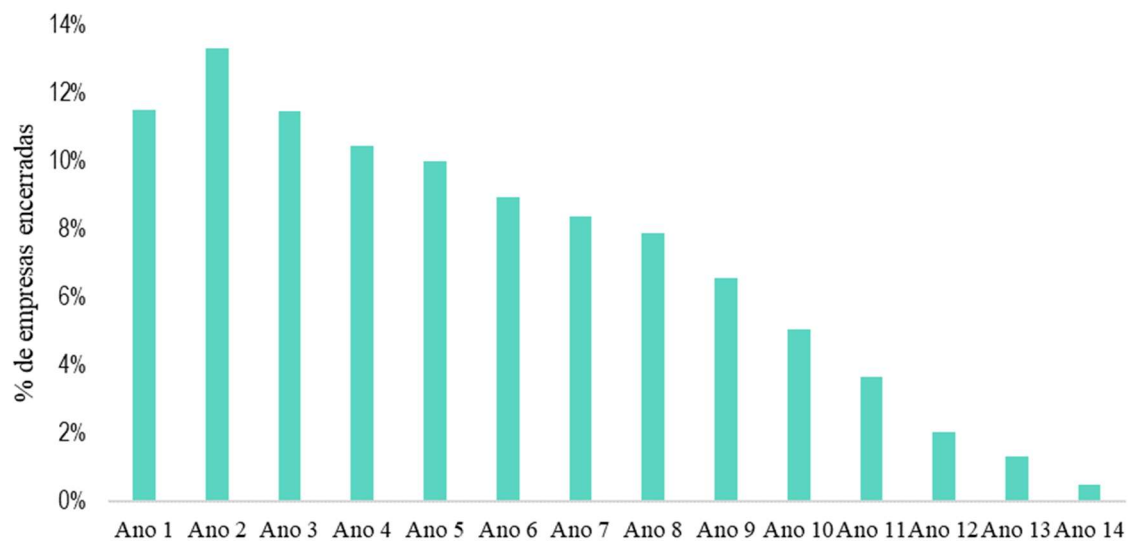


Figura 2.1. - Percentagem de empresas encerradas por ano

A variável dependente é a sobrevivência a 5 anos, que tem por base o estado (ativa ou encerrada) das empresas ao fim de 5 anos, recolhidas para amostra. A título de exemplo foram consideradas como sobreviventes as empresas que constituídas em 2005, tiveram a capacidade de manter a sua atividade até 31 de dezembro de 2010. Por outro lado, foram consideradas como não sobreviventes, as empresas que constituídas em 2005, não mantiveram a atividade empresarial após 31 de dezembro de 2010, e sucessivamente, até às empresas constituídas em 2014 (Figura 2.2.).

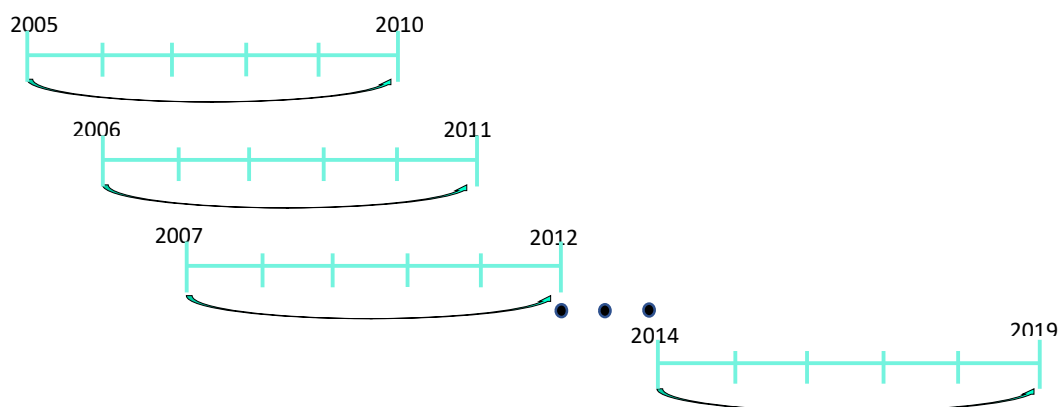


Figura 2.2. - Período de sobrevivência (5 anos) por ano de constituição (2005 a 2014)

O estado “encerrada” engloba os estados “insolvência”, “liquidação” e “dissolução” e exclui as empresas “temporariamente inativas” e “adquiridas”. As empresas que estão registadas de acordo com a primeira, encontram-se inativas há mais de 2 anos. Ainda que não se encontrem operacionais, mantêm-se no mercado, não se enquadrando enquanto empresas ativas ou encerradas. No que respeita às empresas adquiridas, apesar da mudança de propriedade e diferente designação não resultar na formação de uma nova empresa, a sua atividade também não encerrou, pelo que a designação de “encerrada” também não seria o mais apropriado.

De acordo com o Anexo A, a taxa de sobrevivência a 5 anos para as empresas em estudo é mais baixa em 2011, alcançando os 72% face aos 77% registados em 2013.

Em linha com a evidência reportada na literatura prévia sobre este tema, são seleccionados como potenciais fatores explicativos da sobrevivência: a forma jurídica (singular ou coletiva), grupo (pertença ou não a um grupo de empresas), número de empregados, capital social (mil euros), setor de atividade económica³ (Agricultura e pescas, Indústria transformadora, Energia e água, Construção e atividades imobiliárias, Comércio, Transportes e armazenagem, Alojamento e restauração, Informação e comunicação, Outros serviços), região (Lisboa e Vale do Tejo, Algarve, Alentejo, Centro, Norte e Regiões Autónomas) e o ano de criação das empresas (2005 a 2014).

Foram excluídas da análise empresas pertencentes ao setor de atividade económica da Administração Pública e Defesa, bem como da Atividade Financeira e Seguros, pela complexidade do processo de criação que as distingue.

³ Inicialmente, os setores obtidos da base de dados foram organizados de acordo com as secções da CAE-Rev.3 (Classificação Portuguesa de Atividades Económicas, Revisão 3) e, posteriormente, agrupados conforme a agregação realizada pelo INE nos anuários estatísticos disponibilizados.

CAPÍTULO 3.

Análise de dados**3.1. Variável dependente**

Esta variável tem por base a continuidade da atividade empresarial de um conjunto de empresas portuguesas. Tratando-se de uma variável binária, é categorizada em 0 e 1, em que o insucesso é a não sobrevivência da empresa nos primeiros 5 anos de existência (0) e o sucesso é a sobrevivência da empresa após os primeiros 5 anos de existência (1). Totalizaram-se 178.437 empresas válidas, das quais 130.255 (73%) sobreviveram e 48.182 (27%) encerraram ao fim de 5 anos.

3.2. Análise Univariada

Como referido anteriormente, o objetivo do estudo centra-se no conhecimento prévio da sobrevivência das empresas tendo por base os fatores presentes no momento da sua constituição. O quadro 3.1. reflete as potenciais variáveis explicativas com base nos artigos previamente analisados.

Procedeu-se então à análise das variáveis face à variável dependente (ativa ou encerrada), com o apoio do coeficiente de contingência V de Cramer e Phi , no caso da associação entre variáveis binárias, introdução do *t-test* à igualdade de duas médias para as variáveis contínuas, bem como com o desenvolvimento de um modelo de regressão simples para análise da influência de cada variável. Pese embora se opte pela utilização do teste do qui-quadrado, os valores apresentam-se bastantes expressivos e estatisticamente significantes no caso de amostras de grande dimensão, não acrescentando informação útil.

O coeficiente Phi permite mensurar a intensidade da associação entre duas variáveis nominais, com base numa tabela de contingência 2x2 e que varia entre 0 e 1. Embora o V de Cramer seja uma medida de associação semelhante⁴, aplica-se a tabelas superiores a 2x2 e apresenta a seguinte fórmula:

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}} \quad (5)$$

⁴ O coeficiente Phi obtém-se pela divisão do qui-quadrado com a amostra e aplicando a raiz quadrada ao valor resultante (Field, 2017).

em que χ^2 corresponde ao qui-quadrado de Pearson, N representa o número de observações e k o menor número de categorias de entre as variáveis consideradas. Quanto mais próximo de 0, significa que existe fraca ou nenhuma associação; superior a 0,25 considera-se uma associação forte (Akoglu, 2018).

Quadro 3.1. - Potenciais variáveis independentes e respetivo efeito conforme os vários autores

Variável independente	Designação	Autores/Ano	Efeito na probabilidade da empresa sobreviver
tamanho	N.º de empregados	Autdretsch e Mahmood, 1994	Positivo
		Schutjens e Wever, 2000	Positivo
		Mata e Portugal, 2002	Positivo
		Strotmann, 2007	Positivo/Negativo
		Geroski et al., 2010	Positivo
		Weterings e Marsili, 2014	Positivo/Negativo
		Huggins et al., 2015	Positivo
		Simón-Moya et al., 2016	Positivo
		Sarmiento e Forte, 2019	Negativo
capsocial	Capital social (em mil euros)	Schutjens e Wever, 2000	-
		Fotopoulos e Louri, 2000	Positivo
		Harada, 2003	Positivo
		Simón-Moya et al., 2016	Positivo
formajuridica	Forma jurídica (ref. Responsabilidade Limitada)	Mata e Portugal, 2002	Positivo
		Huggins et al, 2015	Positivo
grupo	Pertença a um grupo (ref. Única)	Schutjens e Wever, 2000	Negativo
		Strotmann, 2007	Positivo
		Weterings e Marsili, 2014	-
		Bagley, 2019	Negativo
setor	Setor de atividade económica (ref. Serviços)	Harada, 2003	Positivo
		Fritsch et al., 2006	Negativo
		Brixy e Grotz, 2007	Negativo
		Weterings e Marsili, 2014	Negativo
		Simón-Moya et al., 2016	Negativo
		Pajunen e Järvinen, 2018	Positivo/Negativo
regiao	Região (ref. Urbano)	Schutjens e Wever, 2000	-
		Strotmann, 2007	Negativo
		Weterings e Marsili, 2015	-
		Simón-Moya et al., 2016	-
anos	Ambiente macroeconómico (ref. Crescimento)	Gersoki et al., 2010	Positivo
		Simón-Moya et al., 2016	Negativo
		Barbosa, 2016	Positivo
		Acs et al., 2007	Negativo

Nota: Nem todos os autores listados apresentam como referência a variável indicada, no caso das variáveis categóricas. No entanto, pela necessidade de uniformização dos efeitos foi considerada a categoria referida entre parêntesis, sem alterar as conclusões obtidas.

3.2.1. Tamanho

Cerca de 117.800 empresas não possuem nenhum colaborador na sua constituição. Na verdade, 99% das empresas, têm no seu momento de início de atividade até 15 empregados, pelo que a maioria é categorizada como micro e pequena empresa. Isto deve-se, entre outros fatores, à minimização do risco de investimento inicial (Strotmann, 2007). Adicionalmente, as empresas com um número maior relativo de empregados são as que sobreviveram por um período superior a 5 anos (Anexo B).

3.2.2. Capital Social

A assimetria positiva está presente de igual forma, no capital social, ainda que mais acentuada em comparação ao número de colaboradores (ver Anexo B). Em média, as empresas em estudo investem aproximadamente 22 mil euros (média = 22.277,5 euros), na sua constituição. As empresas que encerram ao fim de 5 anos investem, em média, 9 mil euros.

3.2.3. Forma jurídica

Se compararmos as empresas constituídas, verifica-se que não existe uma divergência evidente ao nível da sobrevivência a 5 anos entre as empresas singulares e as coletivas (Figura 3.1.). Não obstante, as denominadas “unipessoal” apresentam um número ligeiramente maior de empresas que não sobrevivem (cerca de 27,3%) face às empresas coletivas encerradas ao fim de 5 anos (cerca de 26,8%) e, as sob a forma coletiva registam um número relativamente superior de sobreviventes (cerca de 73,2% face a 72,7% de empresas singulares), tendo por base o Anexo C. Com base na informação da amostra, admite-se a existência de uma fraca associação entre o estado das empresas e a sua forma jurídica ($\Phi = 0,006$, com $p\text{-value}=0,006$).

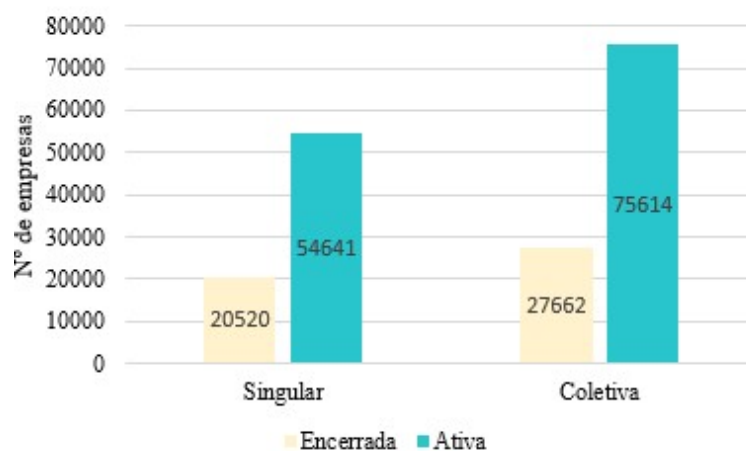


Figura 3.1. - N.º de empresas encerradas e ativas por tipo de empresa

3.2.4. Pertença a um grupo de empresas

Cerca de 66.126 empresas (37,1% do total) é parte constituinte de um grupo. Do conjunto de empresas que encerraram, uma maioria de 47.803 (99,2%) empresas não pertence a um grupo, conforme a figura 3.2.. Com base na análise da distribuição das empresas (ver Anexo C), observa-se que há uma associação forte-moderada entre o seu estado e a pertença a um grupo ($\Phi=0,457$).

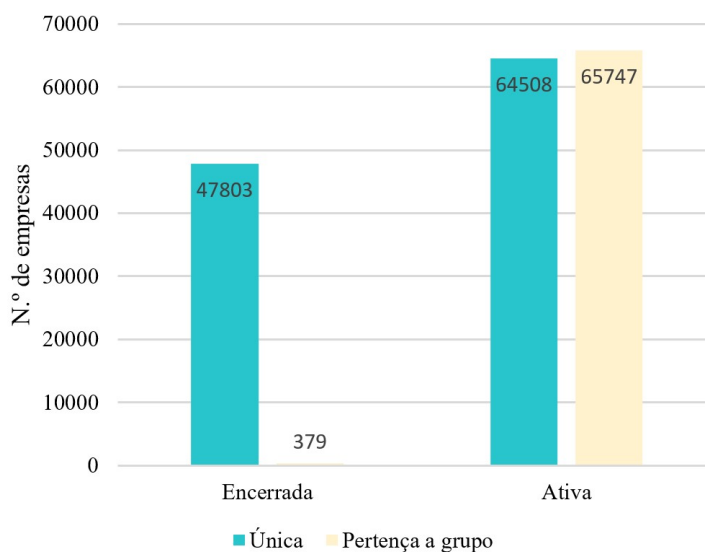


Figura 3.2. - N.º de empresas encerradas e ativas, conforme a (não) pertença a um grupo

3.2.5. Setor de atividade económica

O setor dos serviços alberga diversas atividades financeiras, educação, saúde, bem como todas as atividades de consultoria, científicas e técnicas, entre outras

. Cerca de 35.010 (70,4%) empresas tiveram a capacidade de manter atividade neste setor para além dos 5 anos definidos, superando o dobro das empresas encerradas (Figura 3.3.). Ainda assim, é o setor da agricultura e pescas que apresenta o maior número de empresas ativas ao fim de 5 anos, 84,8%, face às empresas encerradas (ver Anexo C). Analisando o inverso, são os setores da Informação e comunicação e Alojamento e restauração, que registam uma superioridade de empresas encerradas, com 31,6% e 31%, respetivamente. O setor da Energia e água encontra-se em inferioridade, com cerca de 786 empresas constituídas entre 2005 e 2014. Com base na amostra recolhida, existe alguma associação entre as variáveis relacionadas (V de Cramer = 0,103).

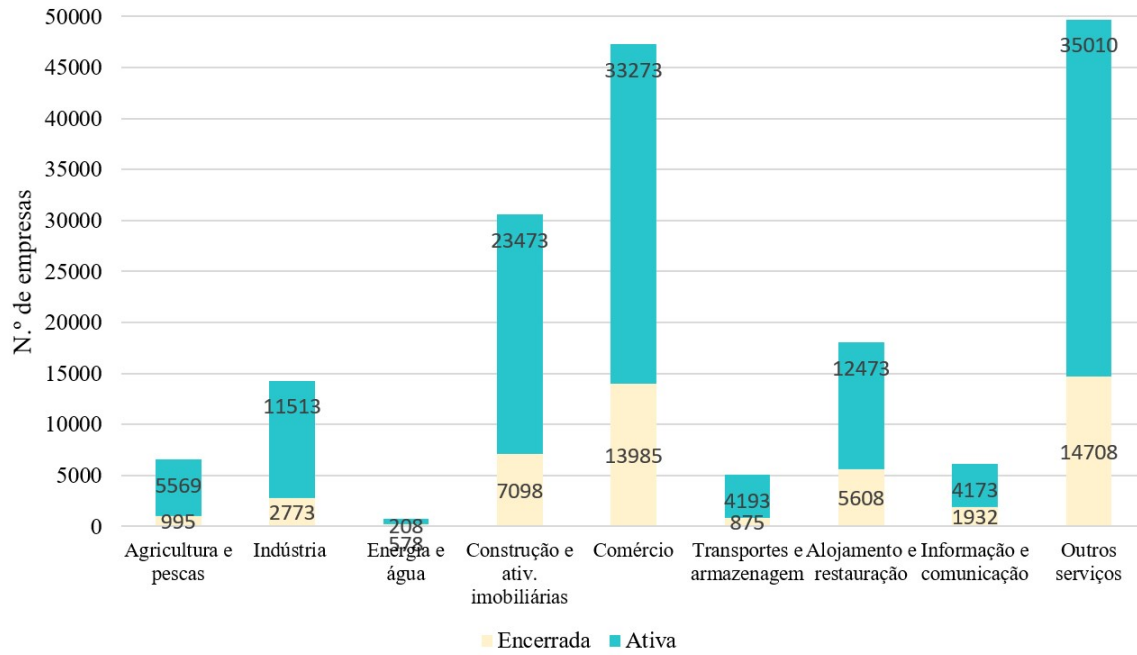


Figura 3.3. - N.º de empresas encerradas e ativas, de acordo com o SAE

As empresas pertencentes ao setor de Outros serviços que sobreviveram em número superior foram constituídas em 2014 (4.443 empresas), conforme observado na figura 3.4.. O setor do Comércio registou o maior número de empresas constituídas em 2013 (4.225 empresas) e em 2006 (3.293 empresas) e que sobreviveram após 5 anos de atividade face aos restantes setores. Em relação ao setor da Construção e dos Transportes e armazenagem, verificou-se que grande parte das empresas sobreviventes foram constituídas em 2007, período de concessão intensiva de créditos (3.074 empresas e 540 empresas, respetivamente).

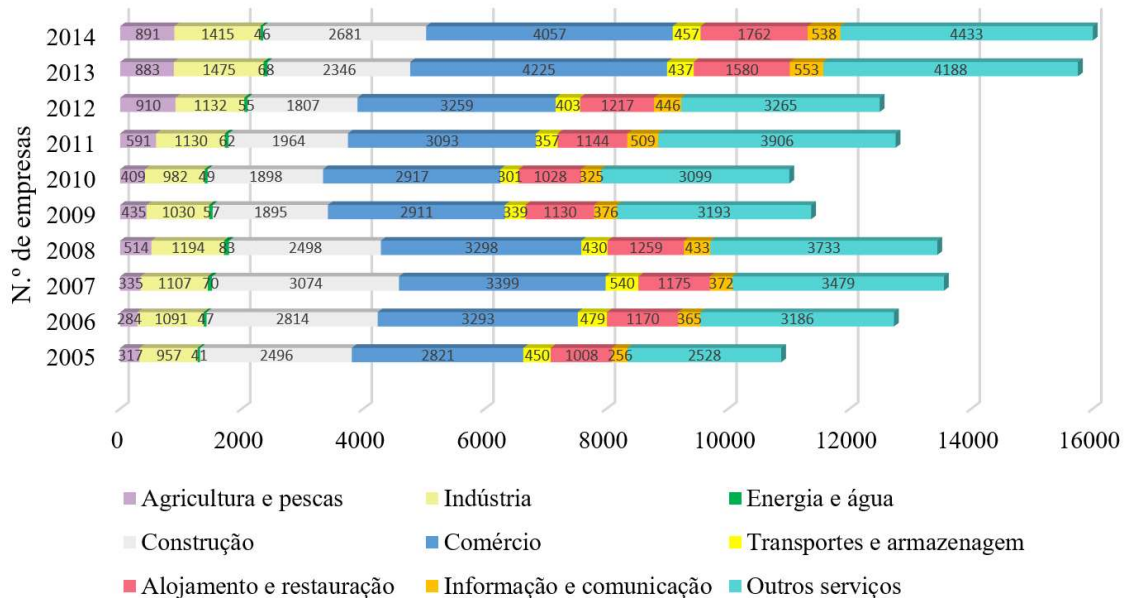


Figura 3.4. - N.º de empresas constituídas entre 2005 e 2014, ativas após 5 anos, por SAE

3.2.6. Região

A região de Lisboa e Vale do Tejo (LVT) apresenta os números mais expressivos, registando cerca de 68.209 (38,2%) empresas constituídas, entre 2005 e 2014, seguida da região Norte, que contabiliza um total de 63.400 (35,5%) empresas (ver Anexo C). De igual forma, é a região que regista mais encerramentos, cerca de 19.446 empresas, alcançado os 40,4% do total de encerradas ao fim de 5 anos, conforme observado na figura 3.5.. Por outro lado, a região do Alentejo, é aquela que apresenta um número inferior de não sobreviventes, atingindo pouco mais de 1.440 (21,7%) empresas do total da região. Verifica-se uma associação entre as variáveis em estudo, ainda que pouco significativa (V de Cramer = 0,037).

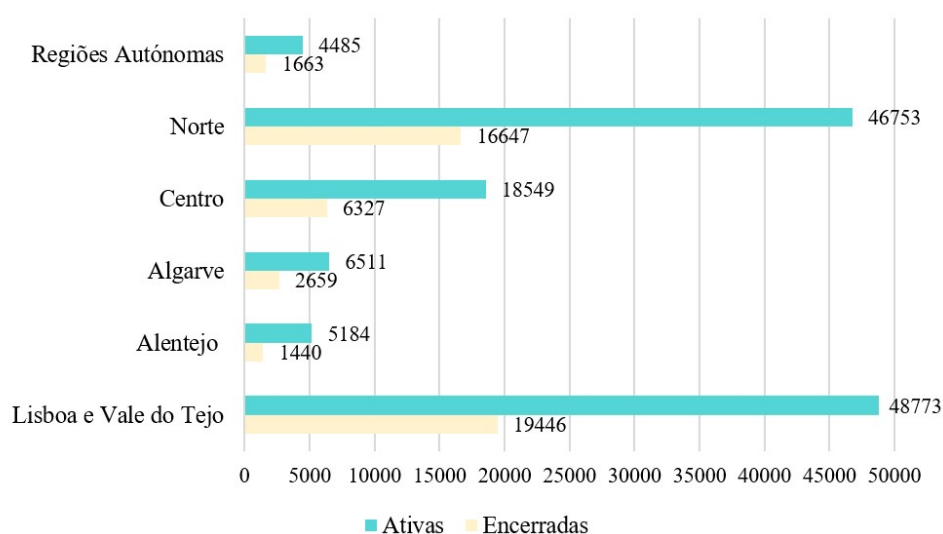


Figura 3.5. - N.º de empresas encerradas e ativas, por região de Portugal

Conforme observado no Anexo D, acresce ainda que cerca de 13.000 empresas localizadas na região Norte e pertencentes ao setor do Comércio e cerca de 16.000 empresas localizadas na região LVT em outros serviços sobreviveram ao fim de 5 anos. Não obstante, é nessas mesmas regiões e nos setores respetivos que também se verifica um número superior de empresas encerradas, 5.294 e 6.898 empresas respetivamente. Por semelhança à região Norte, as regiões do Centro e Regiões Autónomas registam um número superior de empresas em atividade no setor do Comércio, 5.194 e 1.203, respetivamente. O número de empresas sobreviventes no setor da Agricultura e pescas é predominante na região do Norte (1.514 empresas), seguida da região LVT (1.329 empresas) e da região do Alentejo (1.222 empresas). Esta última tem um maior número de empresas ativas neste setor. Apesar das 749 empresas encerradas no setor dos Outros Serviços na região do Algarve, esta região aparenta ser favorável à permanência de empresas no setor da Construção (1.701 empresas). Já o setor Informação e comunicação regista 2.393 empresas ativas ao fim de 5 anos na região LVT e cerca de metade desse número de empresas encerradas no mesmo período.

3.2.7. Ano de constituição

Assume-se que a variável temporal reflete as condições macroeconómicas, uma vez que os anos em que se registou uma grande diminuição das empresas ativas corresponderam também aos anos de recessão em Portugal.

Das empresas constituídas em 2005, cerca de 11.000 mantiveram atividade ao fim de 5 anos, enquanto que cerca de 3.000 deixaram de existir (Figura 3.6.). De salientar, a queda de 2008 para 2009, no número de empresas constituídas e, sobretudo, na redução da quantidade de empresas que constituídas em 2011 mantiveram atividade ao fim de 5 anos (passaram de 78,8% de empresas sobreviventes face às encerradas em 2005, para as 69,7% em 2011, conforme Anexo C). Não obstante, observa-se uma associação significativamente fraca entre o estado das empresas e o ciclo da economia (V de Cramer = 0,058).

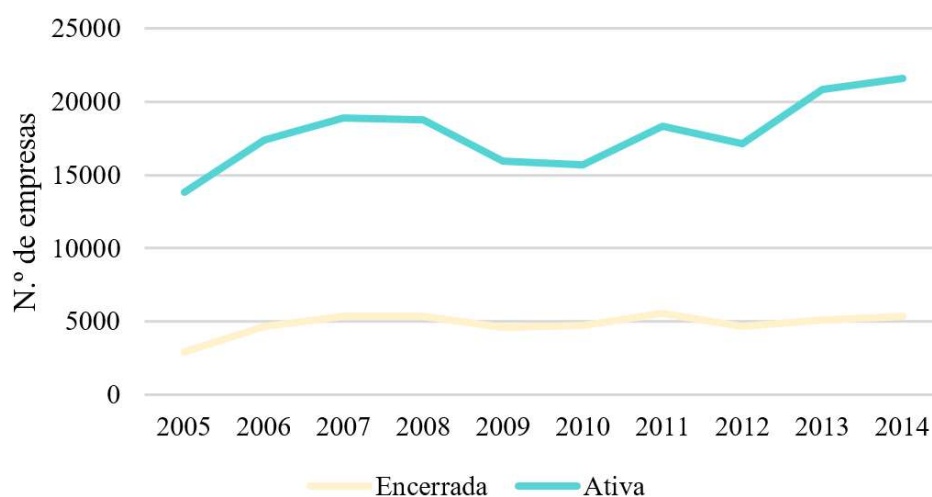


Figura 3.6. - N.º de empresas encerradas e ativas, constituídas entre 2005 e 2014

3.3. Seleção de variáveis

No sentido de eliminar influências significativas de eventuais problemas de heterocedasticidade e dos *outliers* das variáveis *tamanho* e *capsocial* nos modelos a desenvolver, procedeu-se à transformação da variável para a sua forma “log”.

Adicionalmente, a introdução do *t-test* à igualdade de duas médias permite a realização de uma análise prévia à potencial divergência dos indicadores supramencionados entre as empresas sobreviventes e não sobreviventes. Aplicando o teste de Levene, verifica-se que o *p-value* é inferior a 0,001, pelo que a decisão é rejeitar H_0 , podendo admitir-se a diferença de variâncias nos dois grupos (ver Anexo E). Tendo em conta o resultado previamente obtido e com base no teste-t de igualdade de duas médias, a média populacional do tamanho e do capital das empresas é divergente entre empresas encerradas e as ativas ao fim de 5 anos.

Observando a não correlação entre as variáveis em estudo (ver Anexo F), é ainda necessário verificar o pressuposto de ausência de multicolinearidade.

A medida que permite aferir a proporção de variância da variável independente que não é explicada pelas restantes é designada de *Tolerance (TOL)*, sendo dada por $1 - R_i^2$, em que R_i^2 é o coeficiente de determinação de X_i sobre as remanescentes variáveis independentes. Admite-se que se $TOL < 0,1$, então existe uma forte multicolinearidade. Não obstante, a medida mais relevante, é a *Variance Inflation Factor (VIF)*, que quantifica a inflação da variância do estimador do coeficiente em função da existência de multicolinearidade. Admite-se que uma $VIF > 10$ implica a existência de uma forte multicolinearidade. Observando os resultados obtidos (ver Anexo G), admite-se a ausência de um problema de multicolinearidade.

Por fim, a importância de cada variável explicativa é dada pelo *p-value* referente ao teste de Wald, que deverá apresentar um valor inferior a 0,25 (pe), de modo a que esta seja considerada para a construção do modelo. De acordo com Bendel e Afifi (1977), um *p-value* de 0,05 poderá subestimar o poder de variáveis que se poderão tornar fulcrais para o modelo. Ainda assim, existe o risco de se incluírem variáveis não essenciais.

Observando os resultados individuais no quadro 3.2., é estimado que os *odds* de sobrevivência de uma empresa que é constituída por um elevado número de colaboradores é relativamente superior quando comparado com empresas com um número de colaboradores inferior. Por sua vez, empresas cujo investimento inicial é mais significativo, apresentam um *odds* de sobrevivência de cerca de 1,5 vezes superior às restantes.

Verificou-se ainda, que do conjunto de variáveis intrínsecas à empresa, a forma jurídica é a que tem menos influência para a sua continuidade futura e que o facto de ser parte constituinte de um grupo, aumenta consideravelmente a probabilidade de prevalência das empresas no mercado.

Relativamente às variáveis externas, constatou-se que os *odds* de sobrevivência de empresas constituídas em setores como a Agricultura e pescas são 2,35 vezes superiores face às empresas constituídas no setor Outros serviços, com destaque semelhante para os setores Transporte e armazenagem e Indústria, como propícios à sobrevivência das empresas.

Ao nível da região, prevê-se que as empresas localizadas no Alentejo, seguidas das empresas do Centro e Regiões Autónomas apresentem maiores probabilidades de sobreviver face às estabelecidas na grande metrópole nacional.

Adicionalmente, as empresas constituídas em anos de pré-crise e recessão tem maior tendência para apresentar menor capacidade de sobrevivência.

Quadro 3.2. - Regressão logística univariada (inclusão de cada variável individual)

Variável independente	Regressão univariada			
tamanho	β	OR	IC	<i>p-value</i>
Constante	0,868			<,0001
N.º de colaboradores	0,2	1,221	(1,201;1,242)	<,0001
capsocial	β	OR	IC	<i>p-value</i>
Constante	0,229	1,257	(1,492;1,527)	<,0001
Capital social	0,441	1,508		<,0001
formajuridica	β	OR	IC	<i>p-value</i>
Singular (Ref)	0,979			<,0001
Coletiva	0,026	1,027	(1,005;1,048)	0,015
grupo	β	OR	IC	<i>p-value</i>
Única (Ref)	0,3			<,0001
Pertença a um grupo	4,856	128,552	(116,126;142,307)	<,0001
regiao	β	OR	IC	<i>p-value</i>
Lisboa e Vale do Tejo (Ref)	0,92			<,0001
Alentejo	0,361	1,435	(1,351;1,525)	<,0001
Algarve	-0,024	0,976	(0,93;1,024)	0,328
Centro	0,156	1,169	(1,131;1,208)	<,0001
Norte	0,113	1,12	(1,093;1,147)	0,015
Regiões Autónomas	0,073	1,075	(1,014;1,140)	<,0001
setor	β	OR	IC	<i>p-value</i>
Outros serviços (Ref)	0,867			<,0001
Agricultura e pescas	0,855	2,351	(2,192;2,522)	<,0001
Indústria	0,556	1,744	(1,083;1,150)	<,0001
Energia e água	0,155	1,167	(0,913;1,144)	0,057
Construção e atividades imobiliárias	0,329	1,389	(1,110;1,160)	<,0001
Comércio	0	1	(0,786;0,818)	0,973
Transportes e armazenagem	0,7	2,013	(1,276;1,399)	<,0001
Alojamento e restauração	-0,068	0,934	(0,853;0,898)	<,0001
Informação e comunicação	-0,097	0,907	(0,897;0,972)	0,001
anos	β	OR	IC	<i>p-value</i>
2014 (Ref)	1,115			<,0001
2005	0,195	1,215	(1,155;1,279)	<,0001
2006	-0,111	0,895	(0,855;0,937)	<,0001
2007	-0,183	0,833	(0,797;0,871)	<,0001
2008	-0,192	0,826	(0,790;0,863)	<,0001
2009	-0,204	0,815	(0,778;0,854)	<,0001
2010	-0,265	0,767	(0,733;0,803)	<,0001
2011	-0,284	0,753	(0,720;0,786)	<,0001
2012	-0,127	0,881	(0,841;0,922)	<,0001
2013	0,015	1,015	(0,971;1,061)	0,512

Uma vez que todas as variáveis têm um *p-value* <0,25, as mesmas são consideradas aptas para inclusão na análise.

Tendo em conta a necessidade de validação do modelo a elaborar, a amostra foi separada em dois grupos: amostra para estimação, cerca de 70% das observações (Quadro 3.3.) e amostra para validação, cerca de 30% das observações (Quadro 3.4.). Como não existem valores omissos, o número de empresas ativas e encerradas é proporcional às percentagens criadas.

Quadro 3.3. - Amostra para estimação

Sobrevivência a 5 anos (Amostra para estimação)		
	Frequência	Porcentagem
Encerrada	33 702	27
Ativa	91 209	73
Total	124 911	100

Quadro 3.4. - Amostra para validação

Sobrevivência a 5 anos (Amostra para validação)		
	Frequência	Porcentagem
Encerrada	14 480	27
Ativa	39 046	72
Total	53 526	100

CAPÍTULO 4.

Resultados**4.1. Modelo de variáveis associadas à empresa**

De modo a analisar as hipóteses colocadas, foram elaborados modelos apenas com variáveis associadas à empresa, adicionando-se posteriormente as restantes variáveis. Foram ainda criados modelos que excluem, individualmente, as variáveis *setor*, *regiao* e *anos*, avaliando-se a influência de cada fator na capacidade preditiva do modelo.

Neste primeiro modelo foram assim introduzidas todas as variáveis relativas à empresa e consideradas previamente aptas (Quadro 4.1.). Todos os coeficientes das variáveis introduzidas aparentam ter significância estatística ($p\text{-value}<0,05$), contribuindo positivamente para a sobrevivência das empresas.

Um destaque para a variável *grupo*, uma vez que é 198 vezes mais provável que uma empresa mantenha a sua atividade empresarial se pertencer a um grupo. Adicionalmente, a constituição sob o tipo “coletiva” aumenta a probabilidade de sobrevivência das novas empresas.

Considera-se o número de colaboradores como indicativo do tamanho (Schutjens e Wever, 2000; Mata e Portugal, 2002; Weterings e Marsili, 2015; Simón-Moya et al., 2016). Apesar da sua significância estatística, o tamanho adotado pela empresa tem uma contribuição inferior para a obtenção de resultados positivos, quando comparado com os restantes coeficientes. Ao contrário do que se esperava, tendo em conta o estudo de Falck (2007), o tamanho não é a variável fundamental de entre as variáveis relativas exclusivamente à empresa.

O logaritmo do capital social (em mil euros) representa o investimento inicial em linha com Schutjens e Wever (2000) e Simón-Moya et al. (2016). Os resultados obtidos indicam que uma empresa que tenha um capital superior na sua constituição tem maiores probabilidades de sobreviver, ou seja, cada 1000 euros adicionais no capital social inicial resultam num aumento de 60,5% na possibilidade de sobrevivência a 5 anos.

Quadro 4.1. - Modelo A (apenas variáveis relativas à empresa)

Modelo A									
95% C.I. para EXP(B)									
	β	S.E.	Wald	df	Sig.	Exp(β)	Inferior	Superior	
formajuridica(1)	,906	,017	3002,803	1	,000	2,475	2,396	2,556	
grupo(1)	5,288	,060	7640,574	1	,000	197,912	175,784	222,826	
tamanho	,134	,012	125,511	1	,000	1,143	1,117	1,170	
capsocial	,473	,008	3138,473	1	,000	1,605	1,579	1,632	
constante	-1,311	,021	3900,814	1	,000	,269			
Verosimilhança do log (-2)								102.189,885	
R ² de Nagelkerke								0,427	

Nota: Modelo estimado com base na regressão logística multivariada. Sig representa o nível de significância. N=124.911

4.2. Modelo com inclusão de todas as variáveis

De modo a complementar a análise, considerou-se a variável espacial, *regiao*. Brixy e Grotz (2007) alegam que os efeitos da região, quer ao nível do emprego, clima ou potencialidade de recursos são fatores relevantes para constituição das empresas e essenciais à sua sobrevivência.

Com o intuito de introduzir dinâmica ao modelo, procedeu-se à introdução da variável temporal, assumindo que as restantes variáveis se mantêm constantes no tempo. Outro dos motivos centra-se no objetivo a alcançar e, tendo em conta que o estudo pretende concluir sobre as variáveis que contribuem para a sobrevivência das empresas numa fase inicial, o uso do ano de constituição potencia a captação das condições macroeconómicas no momento de criação da empresa que poderão ter um papel nesse sentido.

Por fim, ainda se incluiu a variável setorial, de acordo com outros autores que argumentam a importância deste fator. Cada setor tem objetivos de mercado, produtos e empresas de diferentes dimensões. Existem setores mais competitivos do que outros, o desenvolvimento das tecnologias é divergente e a maioria apenas distingue o setor serviços com o setor da indústria transformativa (Harada, 2003; Brixy e Grotz, 2007; Simón-Moya et al., 2016; Sarmento e Forte, 2019). Não obstante, a consideração de outras atividades económicas, permite obter o potencial de cada setor pertencente à economia do país para a sobrevivência das empresas que os constituem.

Verificou-se que com a introdução destas três variáveis, os coeficientes estimados relativos a *grupo*, *formajuridica* e *tamanho* reforçaram a sua significância (Quadro 4.2.). É cerca de 2,55 vezes mais provável que empresas constituídas enquanto unidades coletivas sobrevivam ao fim de 5 anos relativamente a empresas singulares e, por cada colaborador adicional, a possibilidade de sobrevivência aumenta 24,9%. De forma inversa, registou-se uma redução do impacto do capital face ao modelo A, contribuindo para um aumento de 48,2% no *odds* de sobrevivência.

Relativamente à variável *setor*, observou-se que o coeficiente do setor alojamento e restauração não apresenta significância estatística ($p\text{-value}=0,717$). Comparativamente ao setor Outros serviços, é cerca de 2 vezes mais provável que as empresas pertencentes ao setor Agricultura e pescas sobrevivam ao fim de 5 anos. Já a possibilidade de sobreviver para empresas pertencentes ao setor Energia e água, Informação e comunicação e Comércio é 28,5%, 12,6% e 8,6% inferior, respetivamente, face ao setor de referência. Relativamente ao setor da informação e comunicação, os resultados obtidos evidenciam a competitividade criada entre empresas, correspondendo ao observado por Pajunen e Järvinen (2018).

Verifica-se ainda que, a um nível de 5%, a região do Algarve, Centro e Regiões Autónomas não apresentam significância estatística. Se considerarmos a um nível de 1%, a variável *regiao* é excluída do modelo. Ao contrário das empresas localizadas na região do Norte, a chance de sobreviver é cerca de 12% superior para as que foram constituídas na região do Alentejo quando comparadas com as constituídas na região de LVT. Este resultado contradiz o fenómeno da urbanização, que argumenta que empresas localizadas em áreas urbanas têm condições maioritariamente propícias à sua sobrevivência. Não obstante, também verificamos anteriormente que LVT é a região que tem um maior número relativo de empresas encerradas, podendo atribuir-se ao facto de ser uma área bastante concorrida (Krugman, 1996), na qual as empresas acabam por competir pelos mesmos recursos, como sejam o capital, conhecimento ou clientes (Carroll e Hannan, 1995).

Tendo por referência ao ano de 2014, todos os restantes aparentam contribuir para a sobrevivência das empresas, o que não é expectável, uma vez que o período abrangido pelos anos em análise respeita essencialmente à recessão económica atravessada pelo país. Segundo Weterings e Marsili (2015), favoráveis condições macroeconómicas contribuem para a sobrevivência e não o contrário. No entanto, Simón-Moya et al. (2016), que obtiveram igual evidência, argumentaram que empresas constituídas em anos de crise desenvolvem capacidades que lhes permitem manter a sua atividade a longo prazo. É ainda argumentado que empresas de menor dimensão poderão ter maior flexibilidade na adaptação às dificuldades sentidas no período de recessão, como seja a exploração de oportunidades existentes em nichos de mercados ou influência diminuta dos custos *sunk* (Liu et al, 1999).

Quadro 4.2. – Modelo B (com todas as variáveis selecionadas)

Modelo B								
95% C.I. para EXP(B)								
	β	S.E.	Wald	df	Sig.	Exp(β)	Inferior	Superior
formajuridica(1)	,937	,017	3066,630	1	,000	2,552	2,469	2,638
grupo(1)	5,447	,061	7976,312	1	,000	232,011	205,871	261,470
tamanho	,222	,013	298,222	1	,000	1,249	1,218	1,281
capsocial	,393	,009	1988,759	1	,000	1,482	1,456	1,508
2014(Ref)			1279,675	9	,000			
2005	1,069	,038	778,790	1	,000	2,912	2,701	3,139
2006	,626	,034	333,187	1	,000	1,869	1,748	1,999
2007	,529	,034	249,029	1	,000	1,697	1,589	1,813
2008	,459	,034	185,279	1	,000	1,583	1,482	1,691
2009	,351	,035	99,143	1	,000	1,421	1,326	1,523
2010	,157	,036	19,207	1	,000	1,170	1,091	1,255
2011	,108	,035	9,448	1	,002	1,114	1,040	1,193
2012	,159	,036	19,146	1	,000	1,172	1,092	1,259
2013	,164	,036	21,288	1	,000	1,178	1,099	1,263
Outros serviços (Ref)			338,485	9	,000			
Agricultura e pescas	,456	,051	79,636	1	,000	1,578	1,427	1,744
Indústria	,308	,033	84,785	1	,000	1,361	1,274	1,453
Energia e água	-,335	,117	8,127	1	,004	,715	,568	,901
Construção	,174	,024	53,132	1	,000	1,190	1,135	1,247
Comércio	-,090	,020	19,561	1	,000	,914	,878	,951
Transportes e armazenagem	,285	,056	25,934	1	,000	1,330	1,192	1,485
Alojamento e restauração	,010	,027	,132	1	,717	1,010	,958	1,064
Informação e comunicação	-,134	,043	9,759	1	,002	,874	,804	,951
LVT(Ref)			18,421	5	,002			
Alentejo	,113	,045	6,328	1	,012	1,119	1,025	1,222
Algarve	-,047	,035	1,792	1	,181	,954	,891	1,022
Centro	,015	,024	,365	1	,546	1,015	,968	1,064
Norte	-,042	,018	5,386	1	,020	,959	,926	,994
Ilhas	-,058	,043	1,754	1	,185	,944	,867	1,028
constante	-1,643	,034	2304,610	1	,000	,193		
Verosimilhança do log (-2)	100.484,504							
R ² de Nagelkerke	0,441							

Nota: Modelo estimado com base na regressão logística multivariada. Sig representa o nível de significância. N=124.911

4.3. Modelo com exclusão da variável setorial

No sentido de estudar sobre as hipóteses formuladas foram criados modelos nos quais foram excluídas as variáveis extrínsecas à empresa, de forma individual.

Com a exclusão da variável *setor* (Quadro 4.3.), verificamos que a probabilidade de sobrevivência a 5 anos de uma empresa com um capital superior é maior face às restantes empresas, quando comparada com o modelo constituído por todas as variáveis. Os coeficientes das restantes variáveis intrínsecas evidenciam valores semelhantes ao do modelo saturado.

Relativamente à variável região, verifica-se que apenas o coeficiente associado à *dummy* Alentejo apresenta significância estatística, traduzindo-se num *odds ratio* superior para as empresas localizadas nesta região face às empresas constituídas na região LVT.

Por último, é de referir que observando o resultado do teste Wald para a variável *anos*, face ao modelo B, a maioria dos períodos de constituição registaram uma diminuição da sua relevância no modelo em apreço.

Quadro 4.3. Modelo C (com exclusão da variável setor)

Modelo C								
95% C.I. para EXP(B)								
	β	S.E.	Wald	df	Sig.	Exp(β)	Inferior	Superior
formajuridica(1)	,931	,017	3063,404	1	,000	2,536	2,454	2,621
grupo(1)	5,453	,061	8000,317	1	,000	233,445	207,153	263,074
tamanho	,234	,013	338,678	1	,000	1,264	1,232	1,295
capsocial	,409	,009	2245,046	1	,000	1,505	1,480	1,530
2014(Ref)			1301,922	9	,000			
2005	1,070	,038	786,275	1	,000	2,915	2,705	3,141
2006	,615	,034	324,632	1	,000	1,850	1,731	1,979
2007	,523	,033	244,803	1	,000	1,686	1,580	1,800
2008	,448	,034	177,471	1	,000	1,565	1,465	1,672
2009	,334	,035	90,393	1	,000	1,397	1,304	1,497
2010	,136	,036	14,633	1	,000	1,146	1,069	1,229
2011	,098	,035	7,871	1	,005	1,103	1,030	1,181
2012	,158	,036	19,114	1	,000	1,172	1,091	1,258
2013	,163	,035	21,149	1	,000	1,177	1,098	1,262
LVT (Ref)			23,065	5	,000			
Alentejo	,169	,044	14,703	1	,000	1,185	1,086	1,292
Algarve	-,027	,035	,588	1	,443	,974	,909	1,043
Centro	,033	,024	1,913	1	,167	1,034	,986	1,083
Norte	-,021	,018	1,457	1	,227	,979	,945	1,013
Ilhas	-,046	,043	1,139	1	,286	,955	,877	1,039
constante	-1,641	,033	2531,426	1	,000	,194		
Verosimilhança do log (-2)	100.826,899							
R ² de Nagelkerke	0,438							

Nota: Modelo estimado com base na regressão logística multivariada. Sig representa o nível de significância. N=124.911

4.4. Modelo com exclusão da variável regional

A componente região é um dos pilares que sustentam a sobrevivência de uma empresa para Falck (2007). Pelo contrário, Weterings e Marsili (2015) não encontraram evidência da significância associada a este fator. Empresas localizadas em sítios pouco atrativos em termos de acessos e recursos ou fraca cultura empresarial são maioritariamente sentenciadas a curtos períodos de continuidade da atividade económica (Huggins et al., 2017).

Embora considerada fundamental, neste caso a componente de localização não o aparenta ser. Ainda que os coeficientes estimados de todas as variáveis apresentem um poder menor relativamente aos homólogos no modelo que inclui a variável *regiao*, a diferença não é expressiva e os sinais são equivalentes (Quadro 4.4.).

Quadro 4.4. - Modelo D (com exclusão da variável *regiao*)

Modelo D								
95% C.I. para EXP(B)								
	β	S.E.	Wald	df	Sig.	Exp(β)	Inferior	Superior
formajuridica(1)	,938	,017	3079,304	1	,000	2,556	2,472	2,642
grupo(1)	5,447	,061	7977,654	1	,000	232,048	205,905	261,509
tamanho	,220	,013	294,306	1	,000	1,247	1,216	1,278
capsocial	,393	,009	1989,173	1	,000	1,481	1,456	1,507
2014(Ref)			1281,633	9	,000			
2005	1,069	,038	779,207	1	,000	2,912	2,701	3,139
2006	,627	,034	335,199	1	,000	1,873	1,751	2,003
2007	,530	,034	250,131	1	,000	1,699	1,591	1,814
2008	,461	,034	186,892	1	,000	1,586	1,484	1,694
2009	,354	,035	100,527	1	,000	1,424	1,329	1,526
2010	,158	,036	19,480	1	,000	1,171	1,092	1,256
2011	,109	,035	9,608	1	,002	1,115	1,041	1,194
2012	,160	,036	19,355	1	,000	1,173	1,093	1,260
2013	,165	,036	21,462	1	,000	1,179	1,100	1,264
Outros serviços (Ref)			342,786	8	,000			
Agricultura e pescas	,473	,051	87,402	1	,000	1,605	1,453	1,772
Indústria	,301	,033	82,487	1	,000	1,352	1,267	1,443
Energia e água	-,332	,117	7,980	1	,005	,718	,570	,903
Construção	,169	,024	50,831	1	,000	1,185	1,131	1,241
Comércio	-,093	,020	20,719	1	,000	,912	,876	,949
Transportes e armazenagem	,287	,056	26,188	1	,000	1,332	1,194	1,487
Alojamento e restauração	,008	,027	,100	1	,752	1,008	,957	1,062
Informação e comunicação	-,132	,043	9,447	1	,002	,876	,805	,953
constante	-1,655	,033	2453,842	1	,000	,191		
Verosimilhança do log (-2)	100.502,984							
R ² de Nagelkerke	0,441							

Nota: Modelo estimado com base na regressão logística multivariada. Sig representa o nível de significância. N=124.911

4.5. Modelo com exclusão da variável temporal

Com a exclusão da variável temporal (Quadro 4.5.), cada acréscimo de 1000 euros no capital social inicial resulta num aumento de 58,1% da possibilidade de sobrevivência.

Ao nível de 5%, os coeficientes da região Algarve, Centro e Ilhas não apresentam significância estatística neste modelo. Pese embora a localização das empresas na região do Alentejo não seja considerada relevante a um nível de significância de 1%, as empresas situadas

na região do Norte apresentam menor probabilidade de sobreviver face às empresas da região LVT. Os resultados referentes aos diversos setores apresentam semelhanças ao modelo anterior.

Quadro 4.5. - Modelo E (com exclusão da variável anos)

Modelo E								
						95% C.I. para EXP(B)		
	β	S.E.	Wald	df	Sig.	Exp(β)	Inferior	Superior
formajuridica(1)	,911	,017	2992,899	1	,000	2,487	2,407	2,570
grupo(1)	5,280	,061	7610,459	1	,000	196,298	174,342	221,020
tamanho	,123	,012	102,394	1	,000	1,131	1,104	1,158
capsocial	,458	,009	2804,076	1	,000	1,581	1,554	1,608
Outros serviços (Ref)			364,117	8	,000			
Agricultura e pescas	,337	,051	55,552	1	,000	1,457	1,320	1,609
Indústria	,342	,033	106,332	1	,000	1,407	1,319	1,502
Energia e água	-,392	,117	11,223	1	,001	,676	,537	,850
Construção	,215	,024	82,800	1	,000	1,240	1,184	1,298
Comércio	-,078	,020	14,881	1	,000	,925	,889	,962
Transportes e armazenagem	,254	,056	20,840	1	,000	1,289	1,156	1,438
Alojamento e restauração	,013	,026	,260	1	,610	1,014	,963	1,067
Informação e comunicação	-,156	,043	13,502	1	,000	,855	,787	,930
LVT (Ref)			20,178	5	,001			
Alentejo	,109	,044	6,042	1	,014	1,115	1,022	1,217
Algarve	-,022	,035	,399	1	,527	,978	,914	1,047
Centro	,018	,024	,544	1	,461	1,018	,971	1,066
Norte	-,050	,018	7,957	1	,005	,951	,918	,985
Ilhas	-,038	,043	,790	1	,374	,962	,885	1,047
constante	-1,313	,025	2845,329	1	,000	,269		
Verosimilhança do log (-2)	101.800,868							
R ² de Nagelkerke	0,430							

Nota: Modelo estimado com base na regressão logística multivariada. Sig representa o nível de significância. N=124.911

Por último, ainda se procedeu à análise da existência de interações entre a variável *regiao* e a variável *setor* (ver Anexo H). Considerando-se que a colaboração ou competição entre empresas do mesmo setor ou entre setores varia consoante as áreas urbanas, onde existe maior proximidade entre as mesmas, e as áreas rurais, pode verificar-se o suportado pelo modelo MAR ou o defendido por Jacobs (1969), o *inter-industry spillover* (Brixy e Grotz, 2007; Weterings e Marsili, 2015). Segundo Hosmer e Lemeshow (2000), se o *p-value* resultante da interação entre as duas variáveis fôr significativa ao nível de 5%, o modelo deverá expressar essa interação.

Com base no modelo que contém a interação supramencionada, observa-se que, embora a variável *regiao* não registre significância estatística (*p-value*>0,05), a sua interação com a variável *setor* é significativa (*p-value*<0,05). Relativamente a esta matéria, verificamos que a sobrevivência a 5 anos se apresenta mais provável para empresas pertencentes ao setor da

Agricultura e pescas (45%) e Comércio (31,6%), localizadas no Alentejo; cerca de 3 vezes mais provável a empresas do setor da Energia e água, localizadas no Algarve e ao setor da Indústria transformativa e Comércio, localizadas no Centro, sendo superior em 23,7% e 12,1%, respetivamente, face a empresas pertencentes ao setor dos serviços, na região LVT. De forma inversa, quando comparadas a empresas estabelecidas neste mesmo setor e região, a chance de sobrevivência a 5 anos é 26% inferior para as que pertencem ao setor dos transportes e armazenagem, presentes no Norte. Considerando as duas perspetivas relativas ao efeito da localização geográfica entre indústrias, poder-se-á admitir que nas grandes áreas urbanas, como seja a região LVT predominantes no setor dos serviços, as empresas da mesma indústria não beneficiam das economias de localização, ao contrário do pressuposto defendido pelo modelo MAR. Já em regiões de carácter rural existe, como o Alentejo, evidência para suportar a presença de *inter-industry spillover*. Este resultado vai de encontro ao argumentado por Weterings e Marsili (2015), de que o conhecimento partilhado pode ser limitado ao número de empresas presentes em dado setor.

A análise está ainda sujeita à verificação do pressuposto da linearidade no *logit* inerente à regressão logística, isto é, se existe uma relação linear entre a variável independente contínua e o *logit* da variável resultado. Caso isso não se verifique, procede-se à transformação da variável contínua em categórica. Existem pelo menos dois métodos para verificar o pressuposto: 1) teste Box-Tidwell e 2) Método dos Quantis.

O teste de Box-Tidwell implica a transformação da variável explicativa na respetiva forma log (Box e Tidwell, 1962). Posteriormente, é incluída no modelo com todas as variáveis e em interação com a variável no seu estado “normal”. Caso se verifique significância estatística ($p\text{-value} < 0,05$), admite-se que não existe linearidade no *logit*.

O Método dos Quantis é sugerido por Hosmer e Lemeshow (2000) e consiste na verificação dos quartis relativos à distribuição da frequência da variável contínua, seguida da criação de variáveis *dummy*, com base nesses quartis, sendo o primeiro nomeado como a categoria de referência. Posteriormente, procede-se à construção de um gráfico, no qual os pontos médios dos quartis são confrontados com os coeficientes estimados com base nos quartis. Caso a curva obtida aparente ser aproximadamente linear verifica-se o pressuposto de linearidade no *logit*.

Introduzindo-se o termo interativo, a variável relativa ao capital social é linear no *logit*, por oposição ao *tamanho* ($p\text{-value} < 0,05$). Não obstante, com a aplicação do Método dos Quantis, obtém-se uma aproximação linear da variável *tamanho* ao *logit* da variável dependente.

CAPÍTULO 5.

Robustez**5.1. Log likelihood ratio test (LRT)**

A razão de verosimilhança tem por objetivo a comparação de um modelo ajustado, constituído por um subconjunto de variáveis em análise, face ao modelo saturado, composto por todas as variáveis (Hosmer e Lemeshow, 2000), isto é,

$$D = -2\ln \left[\frac{\text{verosimilhança do modelo ajustado}}{\text{verosimilhança do modelo saturado}} \right] \quad (6)$$

A estatística D, por vezes designada de *deviance*, tem uma função semelhante à do Método dos Mínimos Quadrados, medindo a diferença entre os valores estimados (modelo estimado) e os observados (modelo perfeito).

Dado que se pretende testar a significância do modelo ajustado em que as m variáveis são independentes, temos uma hipótese nula, H_0 , que pressupõe a não significância de nenhum dos coeficientes. Assim tem-se que, o LRT, sob a H_0 segue a distribuição do X^2 , com m graus de liberdade, definindo-se como,

$$G = -2\ln \left[\frac{\text{verosimilhança sem a variável } m}{\text{verosimilhança com a variável } m} \right] \quad (7)$$

Tendo por base as hipóteses formuladas, o objetivo passa por testar a verosimilhança do modelo com as variáveis subjacentes à empresa face ao modelo saturado e a relevância das variáveis extrínsecas, como sejam a variável temporal, setorial e de espaço.

Observando os resultados obtidos (Quadro 5.1.), verifica-se que o modelo apresenta um maior ajustamento com todas variáveis (Modelo B) relativamente ao modelo que contém apenas as variáveis subjacentes à empresa (Modelo A), pelo que os fatores exteriores contribuem significativamente para a capacidade preditiva do modelo.

Como esperado, a inclusão das variáveis setorial e temporal aumenta a superioridade do modelo, ainda que de uma forma menos expressiva na primeira. Não obstante, a exclusão da variável *regiao* aparenta não ter qualquer efeito sob a capacidade preditiva do modelo, como previamente indicado.

Quadro 5.1. - Teste da razão de verosimilhança

Modelos	TesteLR χ^2	p-value
Modelo A vs Modelo B	1705.381	<,0000
Modelo C vs Modelo B	342.395	<,0000
Modelo D vs Modelo B	18.48	0,8576
Modelo E vs Modelo B	1316.364	<,0000

5.2. AIC vs BIC

Um dos critérios para a seleção de modelos é o designado de *Akaike Information Criterion* (AIC), que é composto pelo valor da função verosimilhança e um ajustamento assente no número de parâmetros estimados (Akaike, 1974). Assumindo que θ é o conjunto de parâmetros (vetor) do modelo, que $L(\hat{\theta})$ é a função máxima de verosimilhança e k representa o n.º de parâmetros e constante, o AIC é calculado da seguinte forma:

$$AIC = -2\log L(\hat{\theta}) + 2k \quad (8)$$

O melhor modelo é o que apresenta o menor AIC, criando um *trade-off* entre a elaboração de um modelo mais complexo e a penalização pela adição de mais parâmetros. De modo a avaliar cada modelo, com base no nível de suporte empírico e respetiva verosimilhança, existem duas medidas (Burnham e Anderson, 2002):

- ΔAIC
- Média ponderada

O melhor modelo é assumido como aquele que tem o menor AIC (AIC*). A diferença entre os restantes modelos ($m=1 \dots M$) e o modelo AIC* é representada pelo Δm que caso seja,

- <2 , existe evidência para considerar o modelo candidato
- $4 < \Delta < 7$, o modelo candidato apresenta potencial inferior
- ≥ 10 , é improvável que seja o melhor modelo

A média ponderada (*Akaike weights*) que representa a verosimilhança do modelo m é dada por:

$$w_m = \frac{\exp\left(-\frac{\Delta_m}{2}\right)}{\sum_{j=1}^M \exp\left(-\frac{\Delta_j}{2}\right)} \quad (9)$$

em que m = modelos candidatos ($m=1 \dots M$).

Por outro lado, utiliza-se o *Bayesian Information Criterion* (BIC), método para a seleção de modelos (Schwarz, 1978), complementar ao AIC e que é calculado da seguinte forma:

$$BIC = -2\log L(\hat{\theta}) + k\log(n) \quad (10)$$

Onde n é o n.º de observações.

De igual modo, o modelo que se apresenta como melhor candidato é aquele que tem menor BIC (BIC*). Não obstante, contrariamente ao AIC, os modelos que tiverem maior diferença/amplitude de BIC, são os que possuem maior potencial (Fabozzi et al., 2014). Assim, se:

- <2 , é improvável que seja o melhor modelo
- $2 < \Delta < 6$, evidencia contra o modelo candidato
- $6 < \Delta < 10$, forte evidencia contra o modelo candidato
- ≥ 10 , candidato muito fraco a melhor modelo

O método de seleção AIC tende a minimizar o erro quadrático médio das previsões e é maioritariamente utilizado em contextos de modelos múltiplos, ao contrário do BIC que define um “modelo verdadeiro” à medida que a amostra aumenta, podendo ir até ao infinito. Caso o conjunto de variáveis que explicam o fenómeno na sua totalidade não estejam disponíveis nos dados considerados, o AIC seleciona o modelo que é o mais próximo do “verdadeiro”, tendo em conta esta perda de informação. O método BIC vai apenas selecionar um modelo “quase-verdadeiro” do conjunto disponível. O método AIC poderá resultar num modelo sobreajustado, ao contrário do BIC que peca pelo subajustamento.

Verifica-se que o AIC resulta na seleção do modelo mais complexo, por oposição ao BIC, que exclui a variável *regiao* (Quadro 5.2.).

O melhor modelo candidato, resultante da menor variação de AIC, é o Modelo D, que apesar de apresentar um ajustamento potencial inferior face ao modelo base ($\Delta AIC=8,48$), é o que apresenta um valor dentro do que é expectável. Em termos de BIC, o Modelo B seria o candidato melhor colocado, uma vez que apresenta o valor mais baixo dos restantes modelos. No entanto, é improvável que o seja ($\Delta BIC > 10$).

Quadro 5.2. - AIC vs BIC

Modelos	AIC	ΔAIC	BIC	ΔBIC
Modelo A	102 199,89	1661,38	102 248,56	1 487,40
Modelo B	100 538,50	0,00	100 801,36	40,19
Modelo C	100 864,90	326,40	101 049,87	288,70
Modelo D	100 546,98	8,48	100 761,17	0,00
Modelo E	101 836,87	1298,36	102 012,10	1 250,94

O melhor modelo, a um nível geral, é o modelo B, uma vez que apresenta o menor valor de entre as medidas AIC e BIC, e os modelos candidatos não apresentam potencial suficiente. Comparando os resultados, poder-se-á concluir que a variável *regiao* pouco acrescenta à qualidade de ajustamento do modelo, ao contrário das restantes variáveis extrínsecas à empresa.

5.3. Tabela de Classificação

A designação “classificação” advém do propósito de associar a denominação correta a cada resultado estimado (Chicco e Jurman, 2020), ou seja, sobrevivente ou não sobrevivente. A Tabela de Classificação permite obter um sumário dos resultados do modelo comparando os valores observados com valores esperados, derivados das probabilidades estimadas. Se a probabilidade estimada exceder o C – ponto de quebra ou *cutpoint*, então a variável é igual a 1, caso contrário é 0. O valor mais utilizado é 0,5 (Hosmer e Lemeshow, 2000). A probabilidade de o teste fornecer um resultado positivo, dado que a empresa ainda permanece em atividade, é designada de sensibilidade. A especificidade é definida como a probabilidade de se obter um teste com um resultado negativo quando a empresa efetivamente já encerrou. A soma de ambas as probabilidades é designada de taxa de sucesso ou *accuracy*, que indica o rácio dos valores corretamente estimados face a todos os restantes.

Os modelos que apresentam uma maior taxa de sucesso são o Modelo B e o Modelo D, com igual taxa de sucesso de 78,2% (ver Anexo I). Estes resultados suportam o LRT realizado previamente, indicando que a variável *regiao* não acrescenta capacidade ao modelo.

O Modelo A (apenas variáveis relativas à empresa) é o que detém uma taxa de sucesso inferior (77,7%), ainda apenas que ligeiramente aos restantes, verificando-se o observado nas análises anteriores

É de salientar que este método tem a tendência de favorecer os resultados das classes com maior número de observações, isto é, como as empresas que efetivamente sobreviveram são em número superior às empresas encerradas, a taxa de sensibilidade é superior à taxa de especificidade. Pese embora exista este limite, estão de igual forma a ser considerados vários modelos, o que permite realizar comparações e não concluir apenas sobre um.

Ainda existem outras medidas para analisar a eficácia do modelo que é necessário ter ainda em conta: precisão (qual a proporção de empresas consideradas sobreviventes que efetivamente sobreviveram de entre o total de verdadeiros positivos e falsos positivos) e *fallout* (qual a proporção de empresas não sobreviventes que foram identificadas incorretamente como sobreviventes, do total de falsos positivos e verdadeiros negativos) e das quais foram desenvolvidos alguns métodos. O coeficiente de correlação de Matthews (*Matthews correlation*

coefficient ou MCC) é um método com base numa matriz de contingência de modo a comparar valores observados com valores previstos, assente nas medidas supramencionadas (Chicco e Jurman, 2020), calculando-se da seguinte forma:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (11)$$

e cujo intervalo é $[-1,1]$, em que os extremos indicam uma péssima classificação e perfeita classificação, respetivamente.

Este método permite lidar com o facto de existirem mais empresas sobreviventes do que encerradas, centrando-se na previsão correta da maioria dos casos positivos e negativos, independentemente dos rácios.

Os resultados observados no quadro 5.3. são semelhantes aos anteriormente obtidos. Os modelos B e D apresentam-se como os modelos com melhor ajustamento (MCC=0,407).

Quadro 5.3. - Coeficiente de correlação de Matthews

Modelos	MCC
Modelo A	0,397
Modelo B	0,407
Modelo C	0,406
Modelo D	0,407
Modelo E	0,398

5.4. Curva ROC

A curva ROC (*Receiver Operating Characteristic*) é outra ferramenta que permite avaliar o desempenho de um modelo de regressão logística, com base na variação da sensibilidade (taxa de verdadeiros positivos) *versus* 1-especificidade (taxa de falsos positivos), para diferentes pontos de quebra. A taxa de falsos positivos representa a proporção de empresas previstas como sobreviventes, que na verdade encerraram. Já a taxa de falsos negativos representa a proporção de empresas sobreviventes que foram previstas como encerradas. A representação da curva ROC evidencia os valores para os quais existe otimização da taxas de verdadeiros positivos (empresas previstas como sobreviventes, que efetivamente sobrevivem) em função da taxa de falsos positivos (empresas previstas como sobreviventes que na verdade encerraram), a qual está relacionada com o ponto que se encontra mais próximo do canto superior esquerdo do gráfico, ponto (0,1).

Esta otimização evidencia um *trade-off* sujeito a erros do tipo I (classificação de uma empresa como sobrevivente, quando na verdade não exerce mais a sua atividade empresarial) e do tipo II (classificação de uma empresa sobrevivente como encerrada).

Por seu lado, a área abaixo da curva de ROC, designada de *Area Under Curve* (AUC), representa uma medida de discriminação, isto é, a capacidade do modelo para discriminar entre as empresas que sobrevivem e as que não sobrevivem ao fim de 5 anos, que nos dá a possibilidade de uma empresa ativa ter uma probabilidade estimada associada mais elevada do que uma empresa encerrada, podendo corresponder a todos os valores *cut-off* iguais ou superiores a 0,5. Uma das principais vantagens desta medida é relativa à sua aptidão para lidar com situações em que a amostra é bastante enviesada.

Seja R , o valor que corresponde à área abaixo da curva de ROC, segundo Hosmer e Lemeshow (2000) é geralmente aceite:

- Se $R = 0,5$ não há discriminação
- Se $0,7 < R < 0,8$ a discriminação é aceitável
- Se $0,8 < R < 0,9$ a discriminação excelente
- Se $R \geq 0,9$ a discriminação é excepcional

É aconselhável que a discriminação e calibração do modelo sejam efetuadas num conjunto diferente de dados, que não o utilizado para o desenvolvimento dos modelos (Bewick et al., 2005). Deste modo, sendo a AUC uma medida de discriminação, os modelos foram validados na subamostra para validação.

Observando as AUCs dos modelos construídos, existe evidência para suportar que o modelo constituído por todas as variáveis e o modelo que apenas exclui a variável região são os que apresentam melhor capacidade preditiva em relação à sobrevivência das empresas (AUC=0,847), conforme figura 5.1., por oposição ao modelo que contém apenas as variáveis da empresa (AUC=0,837). Não obstante, com base no quadro 5.4. e respetivas curvas ROC (Anexo J), verifica-se que todos os modelos apresentam uma discriminação excelente ($0,8 < AUC < 0,9$).

Quadro 5.4. - AUC dos modelos

Modelos	AUC
Modelo A	0,837
Modelo B	0,847
Modelo C	0,845
Modelo D	0,847
Modelo E	0,841

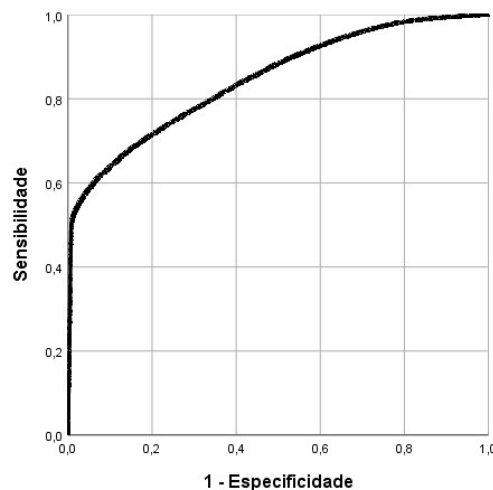


Figura 5.1. – Curva ROC do Modelo B

Por forma a comparar áreas abaixo de duas ou mais curvas ROC derivadas da mesma amostra, neste caso da sub-amostra para validação, é necessário ter em conta a correlação implícita entre essas mesmas curvas (DeLong et al., 1988). Aquando da comparação de duas curvas, utilizou-se a opção *Paired-sample design*, indicado para análises sobre o mesmo conjunto de indivíduos/empresas (ver Anexo L). A H_0 pressupõe que a diferença entre as duas áreas abaixo das respetivas curvas é nula.

Verificou-se então que, com base no *p-value* obtido (Quadro 5.5.), existe evidência estatística para admitir que não há qualquer diferença entre o modelo que não inclui a variável *regiao* face ao modelo completo (*p-value*=0,125). Já a introdução da variável setorial face à variável temporal aparenta contribuir para um menor ajustamento da qualidade do modelo.

Quadro 5.5. - Comparação de AUCs

Modelos	<i>z</i>	IC	<i>p-value</i>
Modelo A vs Modelo B	12,155	(0,013;0,017)	<0,000
Modelo C vs Modelo B	-3,289	(-0,003;-0,001)	0,001
Modelo D vs Modelo B	-1,536	(0,000;0,001)	0,125
Modelo E vs Modelo B	10,203	(0,009;0,013)	<0,000

5.5. Calibração

Segundo Hosmer e Lemeshow (2000:163) “o desempenho de um modelo deve ser avaliado considerando tanto a calibração como a discriminação”. A calibração avalia a aptidão do modelo para retratar a variável de resposta, ou seja, a distância dos valores estimados pelo modelo aos reais/observados (Bewick et al., 2005). Como anteriormente mencionado, apenas é considerada a amostra para validação.

Um dos testes que é frequentemente utilizado para testar a calibração do modelo é o teste de Hosmer e Lemeshow (Huggins et al., 2017; Simón-Moya et al., 2016), através do qual se obtém a estatística \hat{C} cuja distribuição é aproximada ao qui-quadrado de Pearson com $g - 2$ graus de liberdade, onde g é o número de grupos formados, normalmente 10, resultantes das probabilidades estimadas. A estatística \hat{C} é calculada com base no qui-quadrado de Pearson, resultante da tabela $g \times 2$ das frequências estimadas face às observadas. Quanto maior o *p-value* resultante da estatística \hat{C} , melhor a capacidade do modelo para a identificação de divergências entre os valores previstos e os observados. No entanto, este teste não é adequado para amostras com mais de 25.000 observações, uma vez que diminui a capacidade relativa à probabilidade de rejeitar corretamente um modelo pouco calibrado. De acordo com Paul et al. (2013), pese embora um teste à calibração do modelo apropriado devesse ser independente da amostra, este é sensível a amostras desta dimensão ou superiores.

Frequentemente, agrupam-se as probabilidades estimadas em várias categorias (*bins*). De seguida, atribuem-se as probabilidades inferiores ao primeiro *bin* e os restantes por ordem crescente. As observações positivas (casos de empresas sobreviventes) são consequentemente agrupadas com base nas probabilidades de sobrevivência. O objetivo é obter uma relação linear entre as probabilidades estimadas e as observações verificadas.

De forma a validar a calibração dos vários modelos, procedeu-se à elaboração de gráficos, baseados nas probabilidades estimadas face aos valores observados.

Com base nos gráficos obtidos (ver Anexo M), verificou-se que existe uma aproximação das probabilidades estimadas aos dados observados.

CAPÍTULO 6.

Validação de Hipóteses

Com base nos resultados obtidos do LRT e na medida AIC, verifica-se o pressuposto de que a inclusão de variáveis extrínsecas à empresa melhora significativamente o ajuste do modelo face ao que contém apenas as variáveis intrínsecas. Quando comparado o modelo das variáveis subjacentes à empresa com o modelo completo, verifica-se a superioridade do último (H1). Apesar da significância de todas as variáveis relativas à empresa, o modelo apresenta uma maior qualidade de ajustamento quando são integradas as variáveis externas. A evidência obtida não é concordante com o verificado por Falck (2007), designadamente, de que a maioria das componentes relativas à empresa não contribui para a capacidade explicativa do modelo, tendo por base a análise da sobrevivência dos estabelecimentos alemães entre 1993 e 2002. Pese embora se admita a relevância, sobretudo, da forma jurídica da empresa e pertença a um grupo, na verdade, são as restantes variáveis externas que melhoram a capacidade preditiva do modelo.

Relativamente ao SAE, foi previamente mencionado que existem divergências entre autores quanto ao tipo de impacto que cada setor apresenta na atuação futura das empresas (Harada, 2003; Brixy e Grotz, 2007; Weterings e Marsili, 2015). Não obstante, torna-se de igual forma, relevante compreender a importância deste fator. Apesar da não significância do setor Alojamento e restauração (Modelo B), todos os restantes setores apresentam significância estatística. Com base na comparação de AUCs, conclui-se que, apesar da semelhança bastante pronunciada entre modelos com e sem a variável *setor* (Modelo B vs Modelo C), ainda se admite a superioridade de ajustamento do modelo com a variável. Assim, é evidenciado que a inclusão da variável *setor*, contribui, ainda que de uma ligeira forma, para a qualidade do modelo (H2).

No que respeita ao fator região, poder-se-á admitir que o mesmo não é relevante para previsão da sobrevivência das empresas. Numa primeira análise, o LRT indica que não há qualquer diferença na capacidade de ajustamento do modelo com ou sem variável. A medida BIC, desconsidera o Modelo B, privilegiando o modelo sem variável. Por último, observa-se que não existe qualquer divergência entre a AUC do Modelo B e Modelo D ($p\text{-value}>0,05$). Os resultados não permitem corroborar a hipótese estabelecida (H3), pelo que a componente de localização não acrescenta capacidade preditiva ao modelo.

No que concerne ao fator *anos*, admite-se que a sua introdução no modelo contribui para um melhor ajustamento do mesmo, de acordo com o obtido pelo LRT. O modelo E, que não

inclui esta variável, apresenta o segundo valor de MCC mais baixo ($MCC=0,398$) e uma AUC abaixo dos restantes modelos ($AUC=0,841$), com exceção do que apenas inclui as variáveis da empresa. Ainda assim, na comparação de AUCs verifica-se que a divergência entre o modelo sem esta variável face ao modelo completo é significativa ($p\text{-value}<0,001$). Por forma a compreender a importância dos ciclos económicos no desempenho futuro das empresas, isto é, de todo o ambiente macroeconómico subjacente aos anos de constituição das empresas, dever-se-á considerar a inclusão desta variável (H4).

Por último, ainda que a variável *regiao* não seja considerada fundamental conforme mencionado previamente, apresenta significância estatística no Modelo B e este é o melhor modelo, de acordo com a medida AIC. Tendo por base este modelo, por ser mais completo, admite-se que empresas com um maior número de colaboradores apresentam *odds* de sobrevivência mais elevados, indo de encontro aos resultados alcançados por vários autores (Schutjens e Wever, 2000; Geroski et al., 2010; Simón-Moya et al., 2016); assim como as constituídas por um superior capital social, reforçando a importância da disponibilidade financeira no início de atividade (Fotopoulos e Louri, 2000; Acs et al., 2007). As empresas constituídas sob o tipo “coletiva” têm um *odds* de sobrevivência cerca de 2,5 vezes superior face a empresas singulares e existe evidência de que as pertencentes a um grupo de empresas têm maiores chances de sobreviver ($Exp(\beta)=232,01$) ao fim de 5 anos em relação a empresas únicas. A localização fora das grandes cidades, nomeadamente, no Alentejo é aparentemente propícia à sobrevivência futura (11,9% superior face à região LVT), contrariando as externalidades positivas das grandes cidades evidenciadas na literatura da geografia económica. Não obstante, é de salientar que não se apresenta como condição essencial. O *odds* de sobrevivência é de igual forma superior para empresas no setor da Agricultura e pescas (em 57,8%), Indústria transformativa (em 36,1%), Transporte e armazenagem (em 33%) e Construção (em 19%), por oposição ao setor dos Outros serviços, de forma semelhante ao evidenciado por Brixy e Grotz, 2007, entre outros autores. Ainda a acrescentar, que as empresas constituídas em períodos de recessão económica apresentam um *odds* de sobrevivência elevado, contrariamente à generalidade dos estudos considerados. Este resultado poderá ser consequência da superioridade do efeito dos baixos custos de oportunidade face ao efeito da diminuição da procura, saída de outras empresas do mercado pela ausência de flexibilidade em períodos de crise ou exploração de ineficiências (Simón-Moya et al., 2016). Assim, a sobrevivência a 5 anos é mais provável para uma empresa que reúna todas as características mencionadas.

Conclusão

A sobrevivência empresarial continua a ser um tema bastante abordado sob inúmeras perspectivas, com o objetivo de se conhecer as características subjacentes que conduzem a esse fenómeno e, conseqüentemente, as empresas constituídas de acordo com esses atributos, nas quais o investimento é lucrativo a longo prazo. Pese embora se verifique a existência de bases solidamente comprovadas e maioritariamente concordantes entre autores (fatores intrínsecos à empresa, como o tamanho e capital, bem como o setor de atividade económica), observam-se outras que ainda suscitam dúvidas, como é o caso da localização e ambiente macroeconómico.

Esta dissertação permitiu colmatar certos pontos relativos à sobrevivência das empresas em Portugal, nomeadamente no que respeita ao fator localização, com a utilização da variável *regiao* e influência do ambiente macroeconómico nos anos de constituição das empresas, tendo por base um período total de 14 anos. Relativamente à localização, verificou-se que a inclusão da variável não acrescenta qualquer informação significativa. Já a introdução da segunda teve um efeito contrário ao expectável, verificando-se a sobrevivência a longo prazo para empresas constituídas em anos de recessão. Isto poder-se-á dever à maior flexibilidade das empresas de menor dimensão predominantes no país ou à permanência forçada dos novos empreendedores.

Relativamente aos restantes fatores, foi possível corroborar a sua influência e o grau da mesma. A pertença da empresa a um grupo e a sua constituição enquanto empresa individual são as características que mais favorecem a sobrevivência ao fim de 5 anos. Não obstante, o tamanho e capital social superiores contribuem também para a sua permanência no mercado.

É importante referir ainda a existência de limitações. Primeiro, o perfil empresarial de Portugal é caracterizado maioritariamente por PME's (Pequenas e Médias Empresas), o que torna a amostra bastante enviesada em termos de dimensão das empresas. Seria interessante perceber a influência destes fatores ao nível das grandes empresas, com a obtenção de uma amostra mais heterogénea. Em segundo, está em falta a inclusão da variável respeitante ao empreendedor, nomeadamente, motivações para a constituição da empresa, estratégias iniciais, experiência de gestão, experiência de mercado, escolaridade, idade e género. Por serem características maioritariamente subjetivas, seria necessário obter respostas através de inquérito direcionado aos próprios gerentes/proprietários das empresas, o que não permitiria a obtenção de um número tão elevado de empresas. Por último, seria relevante a comparação entre diferentes períodos de sobrevivência, uma vez que a relevância de cada fator poderá ser divergente numa análise a três, cinco ou dez anos.

Referências Bibliográficas

- Acs, Z. J., Armington, C., e Zhang, T. (2007). The determinants of new-firm survival across regional economies: The role of human capital stock and knowledge spillover. *Papers in Regional Science*, 86(3), 367–391. <https://doi.org/10.1111/j.1435-5957.2007.00129.x>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Akoglu, H. (2018). User’s guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3), 91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>
- Audretsch, D. B. (1995). The propensity to exit and innovation. *Review of Industrial Organization*, 10(5), 589–605. <https://doi.org/10.1007/BF01026884>
- Audretsch, D. B., e Mahmood, T. (1994). The rate of hazard confronting new firms and plants in U.S. manufacturing. *Review of Industrial Organization*, 9(1), 41–56. <https://doi.org/10.1007/BF01024218>
- Bagley, M. J. O. (2019). Networks, geography and the survival of the firm. *Journal of Evolutionary Economics*. <https://doi.org/10.1007/s00191-019-00616-z>
- Barbosa, E. G. (2016). *Determinants of Small Business Survival: The Case of Very Small Enterprises of the Traditional Manufacturing Sectors in Brazil*. <http://dx.doi.org/10.2139/ssrn.2802706>
- Bewick, V., Cheek, L., e Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, 9(1), 112–118. <https://doi.org/10.1186/cc3045>
- Box, G.E.P. e Tidwell, P.W. (1962) Transformation of Independent Variables. *Technometrics*, 4, 531-550. <https://doi.org/10.1080/00401706.1962.10490038>
- Brixy, U., e Grotz, R. (2007). Regional patterns and determinants of birth and survival of new firms in Western Germany. *Entrepreneurship and Regional Development*, 19(4), 293–312. <https://doi.org/10.1080/08985620701275510>
- Brüderl, J., e Preisendörfer, P. (1998). Network Support and the Success of Newly Founded Businesses. *Small Business Economics*, 10(3), 213–225. <https://doi.org/10.1023/A:1007997102930>
- Burnham, K. P., e Anderson, D. R. (2002). Model Selection and Inference: A Practical Information-Theoretic Approach. In *The Journal of Wildlife Management* (Vol. 65). <https://doi.org/10.2307/3803117>
- Caixa Geral de Depósitos (2015, Junho 19). *O que é o capital social de uma empresa?*. Disponível em: <https://www.cgd.pt/Site/Saldo-Positivo/negocios/Pages/capital-social-de-uma-empresa.aspx>
- CCDR Alentejo (2014). *Relatório de Aplicação de Fundos Comunitários na Região do Alentejo*. Disponível em: https://www.ccdr-a.gov.pt/docs/ccdra/gestao/Relatorio_Fundos_Comunitarios.pdf
- Centeno, M., Machado C. e Novo, A. A. (2007). A Criação e Destruição de Emprego em Portugal. *Banco de Portugal: Boletim Económico*, 79-108. https://www.bportugal.pt/sites/default/files/anexos/papers/ab200715_p.pdf
- Chicco, D., e Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(6), 1–13. <https://doi.org/10.1186/s12864-019-6413-7>
- Clayton, R. L., Sadeghi A., Talan D. M., e Spletzer J. R. (2013). *High-employment-growth firms: defining and counting them*. Disponível em: <https://www.bls.gov/opub/mlr/2013/article/clayton.htm>
- Coase, R. H. (1937). The Nature of the Firm. *Economica*, 4(16), 386–405. <https://doi.org/10.1111/j.1468-0335.1937.tb00002.x>

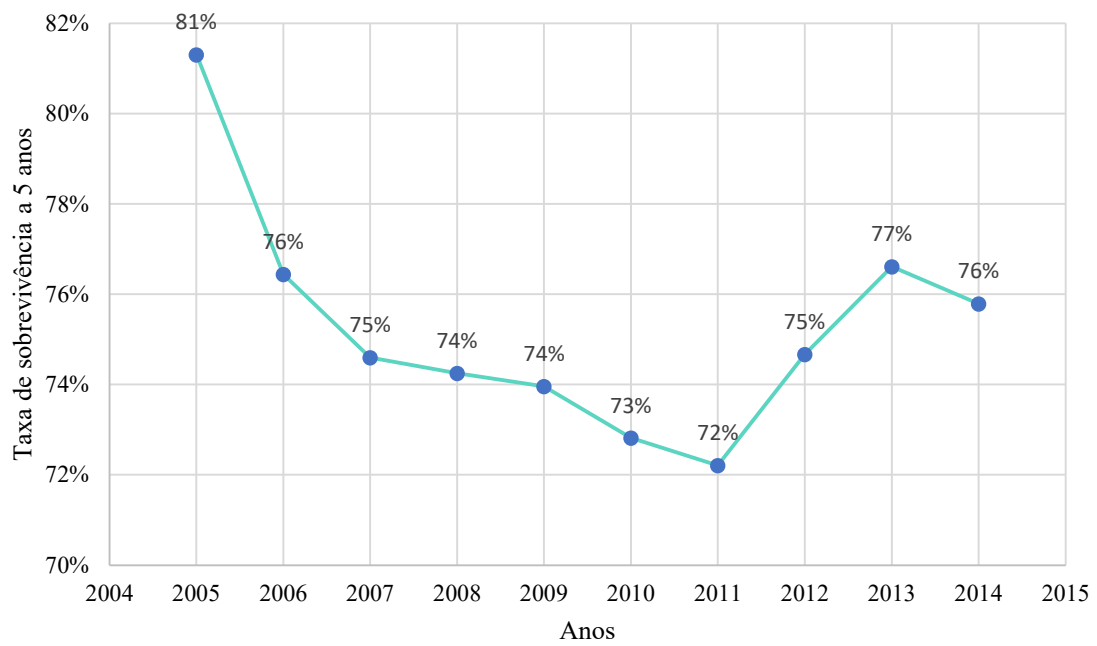
- Cooper, A. C. (1981). Strategic management: New ventures and small business. *Long Range Planning*, 14(5), 39–45. [https://doi.org/10.1016/0024-6301\(81\)90006-6](https://doi.org/10.1016/0024-6301(81)90006-6)
- DeLong, E. R., DeLong, D. M., e Clarke-Pearson, D. L. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44(3), 837–845. <https://doi.org/10.2307/2531595>
- European Commission (2019). *Portugal - SBA Fact Sheet 2019*. Disponível em: <https://ec.europa.eu/docsroom/documents/38662/attachments/23/translations/en/rendition/s/native>
- Fabozzi, F. J., Focardi, S. M., Rachev, S. T., e Arshanapalli, B. G. (2014). Appendix E: Model Selection Criterion: AIC and BIC. *The Basics of Financial Econometrics: Tools, Concepts and Asset Management Applications*. (pp.399–403). <https://doi.org/10.1002/9781118856406.app5>
- Falck, O. (2007). *Emergence and Survival of New Businesses*. Heidelberg: Physica-Verlag. <https://doi.org/10.1007/978-3-7908-1948-9>
- Field, A. (2017). *Discovering statistics using IBM SPSS statistics (5th ed.)*. SAGE Publications.
- Fotopoulos, G., e Louri, H. (2000). Location and Survival of New Entry. *Small Business Economics*, 14(4), 311–321. <https://doi.org/10.1023/A:1008180522759>
- Fritsch, M., Brixy, U., e Falck, O. (2006). The Effect of Industry, Region, and Time on New Business Survival – A Multi-Dimensional Analysis. *Review of Industrial Organization*, 28(3), 285–306. <https://doi.org/10.1007/s11151-006-0018-4>
- Georgiev, M. (2015). The firm as a contract. *Trakia Journal of Science*, 13(Suppl.1), 143–149. <https://doi.org/10.15547/tjs.2015.s.01.024>
- Geroski, P. A., Mata, J., e Portugal, P. (2010). Founding conditions and the survival of new firms. *Strategic Management Journal*, 31, 510–529.
- Hannan, M. T., e Carroll, G. R. (1992). *Dynamics of organizational populations: Density, legitimation, and competition*. Oxford University Press
- Harada, N. (2003). Who succeeds as an entrepreneur? An analysis of the post-entry performance of new firms in Japan. In *Japan and the World Economy* (Vol. 15). [https://doi.org/10.1016/S0922-1425\(02\)00002-6](https://doi.org/10.1016/S0922-1425(02)00002-6)
- He, C., e Yang, R. (2016). Determinants of Firm Failure: Empirical Evidence from China. *Growth and Change*, 47(1), 72–92. <https://doi.org/10.1111/grow.12116>
- Honjo, Y., e Kato, M. (2019). Do initial financial conditions determine the exit routes of start-up firms? *Journal of Evolutionary Economics*, 29(3), 1119–1147. <https://doi.org/10.1007/s00191-019-00623-0>
- Hosmer, D. W., e Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons, Inc. <https://doi.org/10.1002/0471722146>
- Huggins, R., Prokop, D., e Thompson, P. (2017). Entrepreneurship and the determinants of firm survival within regions: human capital, growth motivation and locational conditions. *Entrepreneurship and Regional Development*, 29(3–4), 357–389. <https://doi.org/10.1080/08985626.2016.1271830>
- INE (2020). *Indicadores demográficos das empresas não financeiras em Portugal, 2008-2018*. Disponível em: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=418670737&PUBLICACOESmodo=2
- Informa D&B (2019) *Barómetro*. Disponível em: <https://biblioteca.informadb.pt/read/document.aspx?id=2610#>
- Jacobs, J. (1969). *The Economy of Cities*. New York: Random House.
- Javidan, M. (1998). Core competence: What does it mean in practice? *Long Range Planning*, 31(1), 60–71. [https://doi.org/https://doi.org/10.1016/S0024-6301\(97\)00091-5](https://doi.org/https://doi.org/10.1016/S0024-6301(97)00091-5)

- Johnson, P. (2018). New firms: An economic perspective. In *Routledge*.
<https://doi.org/10.4324/9781351247146>
- Josefy, M. A., Harrison, J. S., Sirmon, D. G., e Matz Carnes, C. (2017). Living and dying: Synthesizing the literature on firm survival and failure across stages of development. *Academy of Management Annals*, 11(2), 770–799.
<https://doi.org/10.5465/annals.2015.0148>
- Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica*, 50(3), 649–670.
<https://www.jstor.org/stable/1912606>
- Kato, M., e Honjo, Y. (2015). Entrepreneurial human capital and the survival of new firms in high- and low-tech sectors. *Journal of Evolutionary Economics*, 25(5), 925–957.
<https://doi.org/10.1007/s00191-015-0427-3>
- Krugman, P. (1996). Urban Concentration: The Role of Increasing Returns and Transport Costs. *International Regional Science Review*, 19(1–2), 5–30.
<https://doi.org/10.1177/016001769601900202>
- Liu, J. T., Tsou, M. W., e Hammitt, J. K. (1999). Do small plants grow faster? Evidence from the Taiwan electronics industry. *Economics Letters*, 65(1), 121–129.
[https://doi.org/10.1016/s0165-1765\(99\)00126-3](https://doi.org/10.1016/s0165-1765(99)00126-3)
- Machlup, F. (1967). Theories of the Firm: Marginalist, Behavioral, Managerial. *The American Economic Review*, 57(1), 1–33. <https://www.jstor.org/stable/1815603>
- Mata, J., e Portugal, P. (2002). The survival of new domestic and foreign-owned firms. *Strategic Management Journal*, 23(4), 323–343. <https://www.jstor.org/stable/20060501>
- Moreira, S. (2018). Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3037178>
- Observatório Sobre Crises e Alternativas (2013). *A Anatomia da Crise: Identificar os problemas para construir as alternativas*. Disponível em:
http://www.qren.pt/np4/np4/?newsId=4087&fileName=Relatorio_Anatomia_Crise_final_.pdf
- OECD (1996). *Small Businesses, job creation and growth: facts, obstacles and best practices*. Disponível em: <https://www.oecd.org/cfe/smes/2090740.pdf>
- O’Kelley, C. C. R. T. (2012). Coase, Knight, and the Nexus-of-Contracts Theory of the Firm: A Reflection on Reification, Reality, and the Corporation as Entrepreneur Surrogate. *Seattle University Law Review*, 35(1999), 1247–1269.
<https://doi.org/10.1099/ijsem.0.002147>
- Pajunen, K., e Järvinen, J. (2018). To survive or succeed? An analysis of biotechnology firms. *Small Business Economics*, 51(3), 757–771. <https://doi.org/10.1007/s11187-017-9963-6>
- Paul, P., Pennell, M. L., e Lemeshow, S. (2013). Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Statistics in Medicine*, 32(1), 67–80.
<https://doi.org/10.1002/sim.5525>
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. e Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49(12), 1373-1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)
- Sarkar, S. K., e Midi, H. (2010). Importance of Assessing the Model Adequacy of Binary Logistic Regression. *Journal of Applied Sciences*, 10(6), 479–486.
<https://doi.org/10.3923/jas.2010.479.486>
- Sarmento, P., e Forte, R. (2019). Does Foreign Presence Induce Host Country Firms’ Exit? The Case of Portugal. *International Advances in Economic Research*, 25(3), 323–337.
<https://doi.org/10.1007/s11294-019-09744-5>
- Schutjens, V. A. J. M., e Wever, E. (2000). Determinants of new firm success. *Papers in Regional Science*, 79(2), 135–153. <https://doi.org/10.1007/s101100050040>

- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464. <https://www.jstor.org/stable/2958889>
- Simón-Moya, V., Revuelto-Taboada, L., e Ribeiro-Soriano, D. (2016). Influence of economic crisis on new SME survival: reality or fiction? *Entrepreneurship and Regional Development*, 28(1–2), 157–176. <https://doi.org/10.1080/08985626.2015.1118560>
- Strotmann, H. (2007). Entrepreneurial survival. *Small Business Economics*, 28(1), 87–104. <https://doi.org/10.1007/s11187-005-8859-z>
- Tong, Y., e Saladríguez, R. (2018). The predictability of financial, accounting-based, and industrial factors on the success of newly incorporated Spanish firms. *Intangible Capital*, 14(1), 127–145. <https://doi.org/10.3926/ic.1106>
- Weterings, A., e Marsili, O. (2015). Spatial Concentration of Industries and New Firm Exits: Does this Relationship Differ between Exits by Closure and by M&A? *Regional Studies*, 49(1), 44–58. <https://doi.org/10.1080/00343404.2012.726708>

Anexos

Anexo A – Taxa de sobrevivência a 5 anos de empresas constituídas entre 2005 e 2014



Anexo B – Frequências das variáveis contínuas

Estatísticas							
	Sobrevivência a 5 anos	N	Média	Mediana	Erro Desvio	Assimetria	Curtose
tamanho	Encerrada	48182	1,22	1,00	1,88	3,901	21,131
	Ativa	130255	1,57	1,00	2,46	3,143	12,065

Estatísticas							
	Sobrevivência a 5 anos	N	Média	Mediana	Erro Desvio	Assimetria	Curtose
capsocial	Encerrada	48182	8,66	5,00	25,55	10,299	132,716
	Ativa	130255	22,28	5,00	57,06	4,854	26,117

Anexo C – Tabulação Cruzada e Medidas de Associação

Tabulação cruzada sobrevivência a 5 anos * formajurídica					
		formajurídica			
			singular	coletiva	Total
Sobrevivência a 5 anos	Encerrada	Contagem	20520	27662	48182
		Contagem Esperada	20295,2	27886,8	48182,0
		% em sobrevivência a 5 anos	42,6%	57,4%	100,0%
		% em formajurídica	27,3%	26,8%	27,0%
		% do Total	11,5%	15,5%	27,0%
	Ativa	Contagem	54641	75614	130255
		Contagem Esperada	54865,8	75389,2	130255,0
		% em Sobrevivência a 5 anos	41,9%	58,1%	100%
		% em formajurídica	72,7%	73,2%	73,0%
		% do Total	30,6%	42,4%	73,0%
Total	Contagem	75161	103276	178437	
	Contagem Esperada	75161,0	103276,0	178437,0	
	% em Sobrevivência a 5 anos	42,1%	57,9%	100,0%	
	% em formajurídica	100,0%	100,0%	100,0%	
	% do Total	42,1%	57,9%	100,0%	

Medidas Simétricas		
	Valor	Significância Aproximada
Fi	,006	,015
N de Casos Válidos	178437	

Tabulação cruzada sobrevivência a 5 anos * grupo					
		Grupo			
			única	pertença a grupo	Total
Sobrevivência a 5 anos	Encerrada	Contagem	47803	379	48182
		Contagem Esperada	30326,5	17855,5	48182,0
		% em Sobrevivência a 5 anos	99,2%	0,8%	100,0%
		% em grupo	42,6%	0,6%	27,0%
		% do Total	26,8%	0,2%	27,0%
	Ativa	Contagem	64508	65747	130255
		Contagem Esperada	81984,5	48270,5	130255,0
		% em Sobrevivência a 5 anos	49,5%	50,5%	100,0%
		% em grupo	57,4%	99,4%	73,0%
		% do Total	36,2%	36,8%	73,0%
Total	Contagem	112311	66126	178437	
	Contagem Esperada	112311,0	66126,0	178437,0	
	% em Sobrevivência a 5 anos	62,9%	37,1%	100,0%	
	% em grupo	100,0%	100,0%	100,0%	
	% do Total	62,9%	37,1%	100,0%	

Medidas Simétricas		
	Valor	Significância Aproximada
Fi	,457	,000
N de Casos Válidos	178437	

BASES FUNDAMENTAIS DA SOBREVIVÊNCIA DAS EMPRESAS

Tabela cruzada sobrevivência a 5 anos * setor												
			Setor									
			Agricultura e pesca	Indústria	Energia e água	Construção	Comercio	Transportes e armazenagem	Alojamento e restauração	Informação e comunicação	Outros serviços	Total
Sobrevivência a 5 anos	Encerrada	Contagem	995	2773	208	7098	13985	875	5608	1932	14708	48182
		Contagem Esperada	1772,4	3857,5	212,2	8254,9	12760,75	1368,5	488,3	1648,5	13425,0	48182,0
		%Sobrevivência a 5 anos	2,1%	5,8%	0,4%	14,7%	29,0%	1,8%	11,6%	4,0%	30,5%	100,0%
		% em setor	15,2%	18,4%	26,5%	23,2%	29,6%	17,3%	31,0%	31,6%	29,6%	27,0%
		% do Total	0,6%	1,6%	0,1%	4,0%	7,8%	0,5%	3,1%	1,1%	8,2%	27,0%
	Ativa	Contagem	5569	11513	578	23473	33273	4193	12473	4173	35010	130255
		Contagem Esperada	4791,6	10428,5	573,8	22316,1	34497,3	3699,5	13198,7	4456,5	36293,0	130255,0
		%Sobrevivência a 5 anos	4,3%	8,8%	0,4%	18,0%	25,5%	3,2%	9,6%	3,2%	26,9%	100,0%
		% em setor	84,8%	80,6%	73,5%	76,8%	70,4%	82,7%	69,0%	68,4%	70,4%	73,0%
		% do Total	3,1%	6,5%	0,3%	13,2%	18,6%	2,3%	7,0%	2,3%	19,6%	73,0%
Total	Contagem	6564	14286	786	30571	47258	5068	18081	6105	49718	178437	
	Contagem Esperada	6564,0	14286,0	786,0	30571,0	47258,0	5068,0	18081,0	6105,0	49718,0	178437,0	
	%Sobrevivência a 5 anos	3,7%	8,0%	0,4%	17,1%	26,5%	2,8%	10,1%	3,4%	27,9%	100,0%	
	% em setor	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	
	% do Total	3,7%	8,0%	0,4%	17,1%	26,5%	2,8%	10,1%	3,4%	27,9%	100,0%	

Medidas Simétricas		
	Valor	Significância Aproximada
V de Cramer	,103	,000
N de Casos Válidos	178437	

Tabela cruzada sobrevivência a 5 anos * regio									
			Região						Total
			LVT	Alentejo	Algarve	Centro	Norte	Ilhas	Total
Sobrevivência a 5 anos	Encerrada	Contagem	19446	1440	2659	6327	16647	1663	48182
		Contagem Esperada	18420,7	1788,6	2476,1	6717,1	17119,4	1660,1	48182,0
		%Sobrevivência a 5 anos	40,4%	3,0%	5,5%	13,1%	34,6%	3,5%	100,0%
		% regio	28,5%	21,7%	29,0%	25,4%	26,3%	27,0%	27,0%
		% do Total	10,9%	0,8%	1,5%	3,5%	9,3%	0,9%	27,0%
	Ativa	Contagem	48773	5184	6511	18549	46753	4485	130255
		Contagem Esperada	49798,3	4835,4	6693,9	18158,9	46280,6	4487,9	130255,0
		%Sobrevivência a 5 anos	37,4%	4,0%	5,0%	14,2%	35,9%	3,4%	100,0%
		% regio	71,5%	78,3%	71,0%	74,6%	73,7%	73,0%	73,0%
		% do Total	27,3%	2,9%	3,6%	10,4%	26,2%	2,5%	73,0%
Total	Contagem	68219	6624	9170	24876	63400	6148	178437	
	Contagem Esperada	68219,0	6624,0	9170,0	2476,0	63400,0	6148,0	178437,0	
	%Sobrevivência a 5 anos	38,2%	3,7%	5,1%	13,9%	35,5%	3,4%	100,0%	
	% regio	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	
	% do Total	38,2%	3,7%	5,1%	13,9%	35,5%	3,4%	100,0%	

Medidas Simétricas		
	Valor	Significância Aproximada
V de Cramer	,037	,000
N de Casos Válidos	178437	

BASES FUNDAMENTAIS DA SOBREVIVÊNCIA DAS EMPRESAS

Tabela cruzada sobrevivência a 5 anos * anos													
		Anos											
		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total	
Sobrevivência a 5 anos	Encerrada	Contagem	2934	4662	5334	5339	4571	4704	5558	4652	5090	5338	48182
		Contagem Esperada	3728,5	4696,0	5099,4	5071,3	4303,3	4242,6	4945,2	4629,8	5628,6	5837,3	48182,0
		%Sobrevivência a 5 anos	6,1%	9,7%	11,1%	11,1%	9,5%	9,8%	11,3%	9,7%	10,6%	11,1%	100,0%
		% em anos	21,2%	26,8%	28,2%	28,4%	28,7%	29,9%	30,3%	27,1%	24,4%	24,7%	27,0%
		% do Total	1,6%	2,6%	3,0%	3,0%	2,6%	2,6%	3,1%	2,6%	2,9%	3,0%	27,0%
	Ativa	Contagem	10874	12729	13551	13442	11366	11008	12756	12494	15755	16280	130255
		Contagem Esperada	10079,5	12695,0	13785,6	13709,7	11633,7	11469,4	13368,8	12516,2	15216,4	15780,7	130255,0
		%Sobrevivência a 5 anos	8,3%	9,8%	10,4%	10,3%	8,7%	8,5%	9,8%	9,6%	12,1%	12,5%	100,0%
		% em anos	78,8%	73,2%	71,8%	71,6%	71,3%	70,1%	69,7%	72,9%	75,6%	75,3%	73,0%
		% do Total	6,1%	7,1%	7,6%	7,5%	6,4%	6,2%	7,1%	7,0%	8,8%	9,1%	73,0%
Total	Contagem	13808	17391	18885	18781	15937	15712	18314	17146	20845	21618	178437	
	Contagem Esperada	13808,0%	17391,0%	18885,0%	18781,0	15937,0	15712,0	18314,0	17146,0	20845,0	21618,0	178437,0	
	%Sobrevivência a 5 anos	7,7%	9,7%	10,6%	10,5%	8,9%	8,8%	10,3%	9,6%	11,7%	12,1%	100,0%	
	% em anos	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	
	% do Total	7,7%	9,7%	10,6%	10,5%	8,9%	8,8%	10,3%	9,6%	11,7%	12,1%	100,0%	

Medidas Simétricas		
	Valor	Significância Aproximada
V de Cramer	,058	,000
N de Casos Válidos	178437	

Anexo D - Setor vs Região

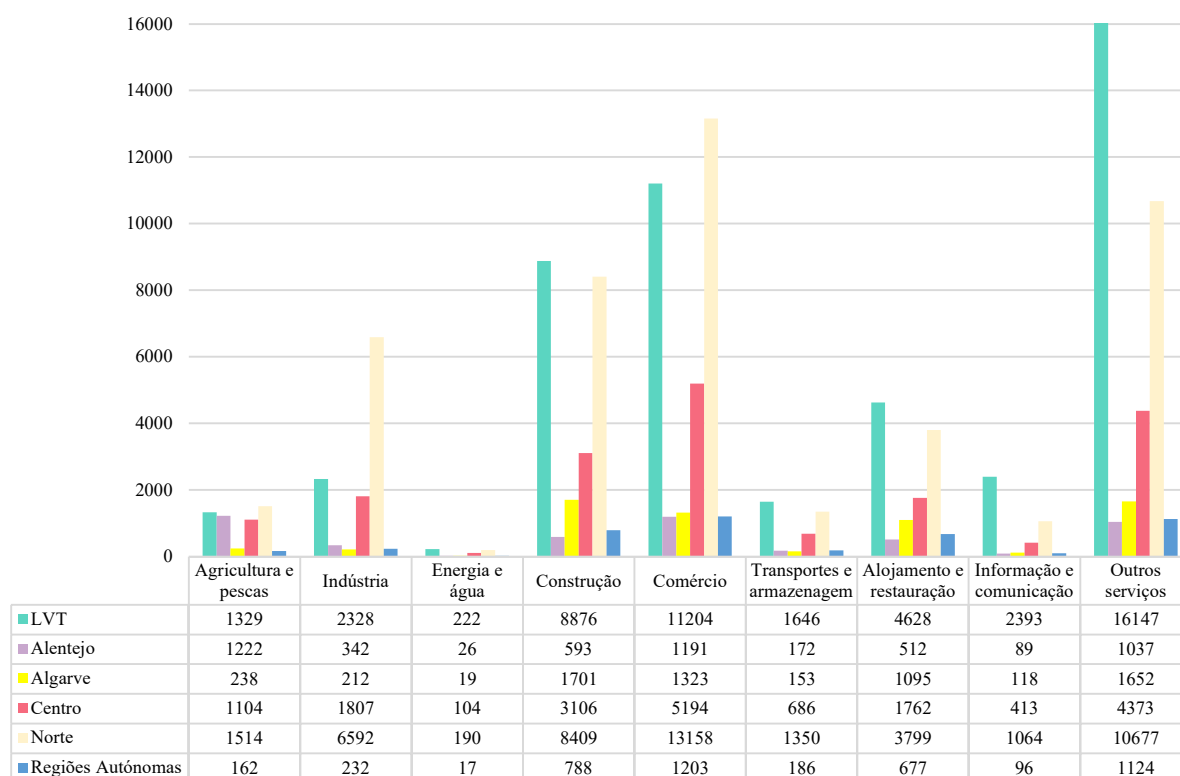


Figura A - Empresas sobreviventes após 5 anos por região e por setor, entre 2005 e 2019



Figura B - Empresas encerradas ao fim de 5 anos por região e por setor, entre 2005 e 2019

Anexo E - Teste – T

Teste de amostras independentes									
tamanho	Teste de Levene para igualdade de variâncias		teste-t para Igualdade de Médias					95% Intervalo de Confiança da Diferença	
	Z.	Sig.	t	df	Sig. (2 extremidades)	Diferença média	Erro padrão de diferença	Inferior	Superior
Variâncias iguais assumidas	611,432	,000	-23,741	178435	,000	-,08232	,00347	-,08912	-,07552
Variâncias iguais não assumidas			-25,312	98181,990	,000	-,08232	,00325	-,08869	-,07594

Teste de amostras independentes									
capsocial	Teste de Levene para igualdade de variâncias		teste-t para Igualdade de Médias					95% Intervalo de Confiança da Diferença	
	Z.	Sig.	t	df	Sig. (2 extremidades)	Diferença média	Erro padrão de diferença	Inferior	Superior
Variâncias iguais assumidas	3890,807	,000	-76,127	178435	,000	-,45809	,00602	-,46989	-,44630
Variâncias iguais não assumidas			-85,924	111848,027	,000	-,45809	,00533	-,46854	-,44764

Anexo F – Correlações

Coeficientes de correlações							
	tamanho	capsocial	forma juridica	grupo	setor	regiao	anos
tamanho	1						
capsocial	,041	1					
formajuridica	-,022	,212	1				
grupo	,049	,031	-,333	1			
setor	-,065	-,128	,030	-,058	1		
regiao	,067	,032	-,030	,038	-,127	1	
anos	,112	-,254	-,137	,250	,010	,018	1

Nota: Os coeficientes de correlação superiores a 0,5 correspondem a uma associação entre variáveis elevada.

Anexo G - Ausência de Multicolinearidade

Colinearidade		
	TOL	VIF
tamanho	,976	1,025
capsocial	,858	1,165
formajuridica	,837	1,194
grupo	,821	1,218
setor	,962	1,039
regiao	,979	1,022
anos	,856	1,168

Anexo H - Modelo com interações

Interações regio*setor								
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
formajuridica(1)	,939	,014	4401,766	1	,000	2,558	2,488	2,630
grupo(1)	5,527	,053	10895,168	1	,000	251,420	226,635	278,915
tamanho	,235	,011	471,697	1	,000	1,265	1,238	1,292
capsocial	,392	,007	2804,784	1	,000	1,480	1,458	1,501
LVT(Ref)			5,177	5	,395			
Alentejo	,009	,072	,016	1	,901	1,009	,877	1,161
Algarve	-,093	,056	2,767	1	,096	,912	,817	1,017
Centro	-,033	,038	,751	1	,386	,968	,899	1,042
Norte	-,039	,028	1,951	1	,162	,962	,911	1,016
Ilhas	-,086	,071	1,491	1	,222	,917	,799	1,054
Outros serviços (Ref)			212,855	8	,000			
Agricultura e pescas	,478	,085	31,859	1	,000	1,612	1,366	1,903
Indústria	,225	,054	17,002	1	,000	1,252	1,125	1,393
Energia e água	-,498	,157	9,996	1	,002	,608	,446	,828
Construção	,195	,031	40,022	1	,000	1,215	1,144	1,291
Comércio	-,138	,026	27,232	1	,000	,871	,827	,917
Transportes e armazenagem	,415	,071	34,073	1	,000	1,515	1,318	1,742
Alojamento e restauração	,055	,035	2,428	1	,119	1,056	,986	1,131
Informação e comunicação	-,129	,048	7,193	1	,007	,879	,800	,966
2014(Ref)			1901,016	9	,000			
2005	1,094	,032	1156,610	1	,000	2,985	2,802	3,179
2006	,640	,029	494,800	1	,000	1,896	1,792	2,006
2007	,531	,028	355,243	1	,000	1,701	1,609	1,797
2008	,470	,028	275,128	1	,000	1,600	1,514	1,692
2009	,377	,030	161,321	1	,000	1,457	1,375	1,545
2010	,168	,030	31,355	1	,000	1,183	1,116	1,255
2011	,099	,029	11,313	1	,001	1,104	1,042	1,170
2012	,172	,031	31,494	1	,000	1,187	1,118	1,260
2013	,159	,030	28,157	1	,000	1,172	1,105	1,243
regiao * setor			69,698	40	,003			
Alentejo *Agricultura e pescas	,372	,152	5,942	1	,015	1,450	1,076	1,955
Alentejo *Indústria	,060	,166	,129	1	,719	1,061	,767	1,470
Alentejo *Energia e água	,197	,504	,154	1	,695	1,218	,454	3,270
Alentejo *Construção	,043	,132	,104	1	,747	1,043	,805	1,352
Alentejo *Comércio	,275	,103	7,105	1	,008	1,316	1,075	1,610
Alentejo *Transportes e armazenagem	-,021	,255	,007	1	,933	,979	,593	1,615
Alentejo *Alojamento e restauração	-,113	,123	,835	1	,361	,893	,702	1,138
Alentejo *Informação e comunicação	-,340	,251	1,837	1	,175	,712	,436	1,615
Algarve *Agricultura e pescas	-,191	,204	,882	1	,348	,826	,554	1,231
Algarve *Indústria	-,161	,174	,858	1	,354	,851	,605	1,197
Algarve *Energia e água	1,046	,531	3,874	1	,049	2,846	1,004	8,063
Algarve *Construção	,037	,082	,198	1	,656	1,037	,883	1,219
Algarve *Comércio	,050	,083	,360	1	,548	1,051	,893	1,238
Algarve *Transportes e armazenagem	-,033	,239	,019	1	,890	,967	,605	1,547
Algarve *Alojamento e restauração	,016	,090	,031	1	,860	1,016	,851	1,213
Algarve *Informação e comunicação	,073	,200	,134	1	,714	1,076	,727	1,592
Centro *Agricultura e pescas	-,066	,127	,270	1	,603	,936	,731	1,200
Centro *Indústria	,213	,092	5,360	1	,021	1,237	1,033	1,481
Centro *Energia e água	,480	,290	2,748	1	,097	1,616	,916	2,852
Centro *Construção	,037	,064	,343	1	,558	1,038	,916	1,175
Centro *Comércio	,115	,053	4,615	1	,032	1,121	1,010	1,245
Centro *Transportes e armazenagem	-,188	,145	1,684	1	,194	,829	,624	1,100
Centro *Alojamento e restauração	-,072	,069	1,065	1	,302	,931	,813	1,066
Centro *Informação e comunicação	-,036	,122	,085	1	,771	,965	,760	1,226
Norte *Agricultura e pescas	-,141	,115	1,503	1	,220	,868	,693	1,088
Norte *Indústria	,089	,068	1,738	1	,187	1,093	,958	1,248
Norte *Energia e água	,077	,238	,105	1	,746	1,080	,677	1,724
Norte *Construção	-,041	,046	,792	1	,373	,960	,877	1,050
Norte *Comércio	,062	,039	2,496	1	,114	1,064	,985	1,150
Norte *Transportes e armazenagem	-,301	,107	7,855	1	,005	,740	,600	,914
Norte *Alojamento e restauração	-,070	,053	1,700	1	,192	,993	,840	1,036
Norte *Informação e comunicação	,053	,085	,386	1	,535	1,054	,892	1,246
Ilhas *Agricultura e pescas	,226	,243	,863	1	,353	1,253	,778	2,018
Ilhas *Indústria	,254	,206	1,521	1	,217	1,289	,861	1,928
Ilhas *Energia e água	-,139	,745	3,230	1	,072	,262	,061	1,129
Ilhas *Construção	-,081	,112	,520	1	,471	,922	,741	1,149
Ilhas *Comércio	,068	,100	,461	1	,497	1,070	,880	1,301
Ilhas *Transportes e armazenagem	-,037	,234	,026	1	,873	,963	,608	1,525
Ilhas *Alojamento e restauração	-,044	,112	,153	1	,696	,957	,768	1,193
Ilhas *Informação e comunicação	,105	,235	,200	1	,655	1,111	,701	1,761
constante	-,1652	,031	2912,839	1	,000	,192		
Verossimilhança do log (-2)								143.192,655
R ² de Nagelkerke								0,443

Nota: Modelo estimado com base na regressão logística multivariada. Sig representa o nível de significância. N=124.911

Anexo I – Tabelas de Classificação

Observado/Previsto	0	1	Total
0	16095	17607	47,8%
1	10250	80959	88,8%
Total			77,7%

Modelo A

Observado/Previsto	0	1	Total
0	15993	17709	47,5%
1	9551	81658	89,5%
Total			78,2%

Modelo B

Observado/Previsto	0	1	Total
0	16102	17600	47,8%
1	9750	81459	89,3%
Total			78,1%

Modelo C

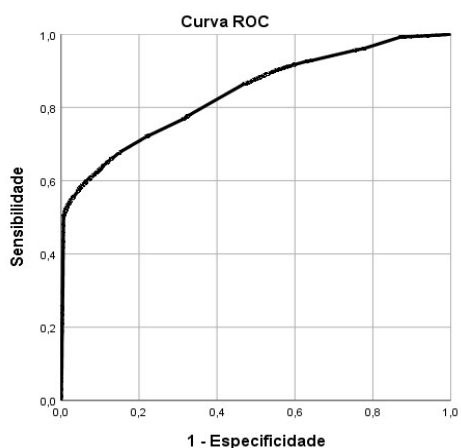
Observado/Previsto	0	1	Total
0	15990	17712	47,4%
1	9518	81691	89,6%
Total			78,2%

Modelo D

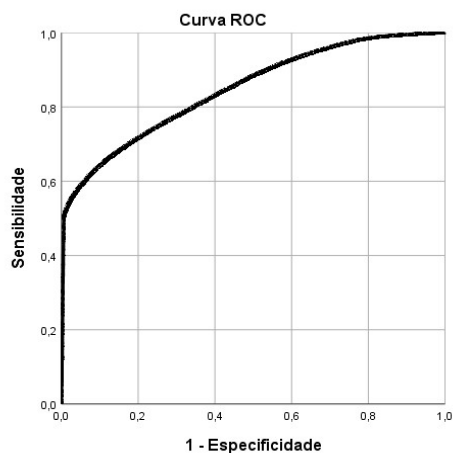
Observado/Previsto	0	1	Total
0	15940	17762	47,3%
1	9967	81242	89,1%
Total			77,8%

Modelo E

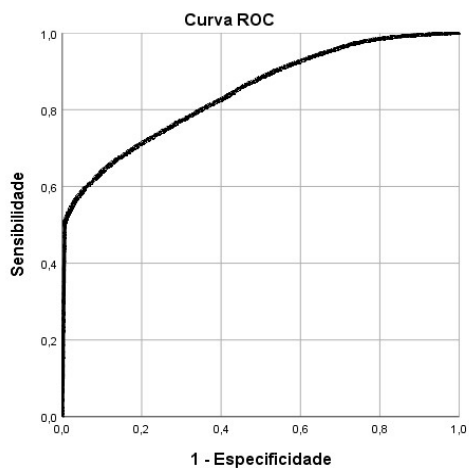
Anexo J - Curvas ROC



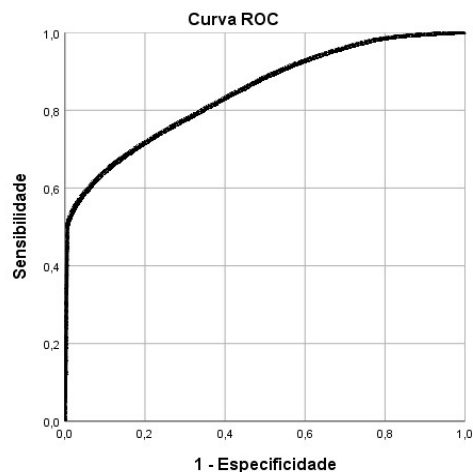
Modelo A



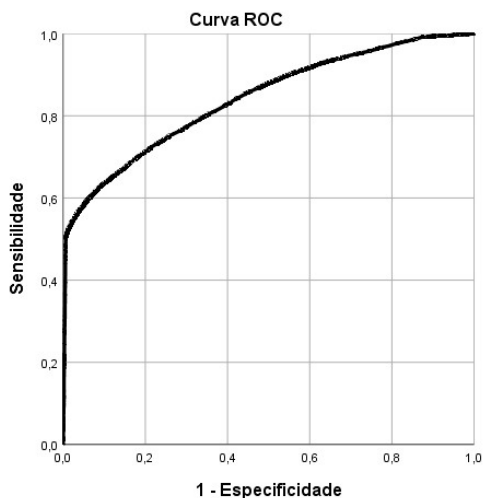
Modelo B



Modelo C

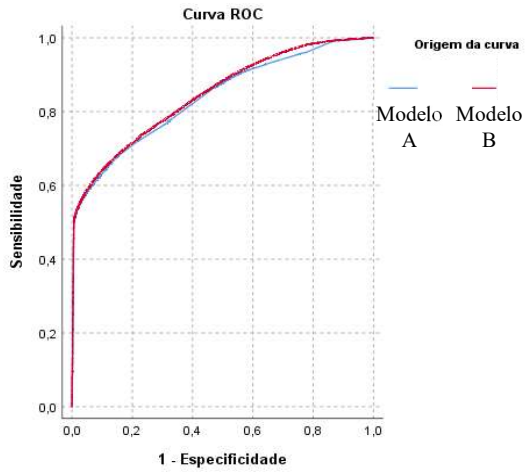


Modelo B

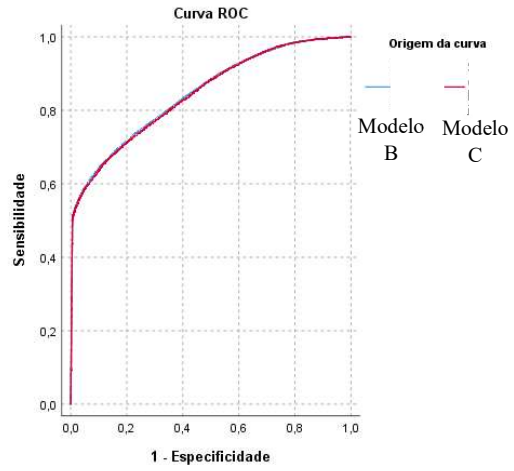


Modelo E

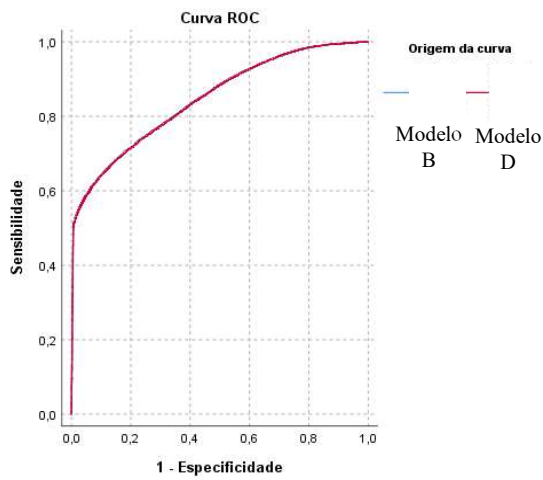
Anexo L - Comparação AUC



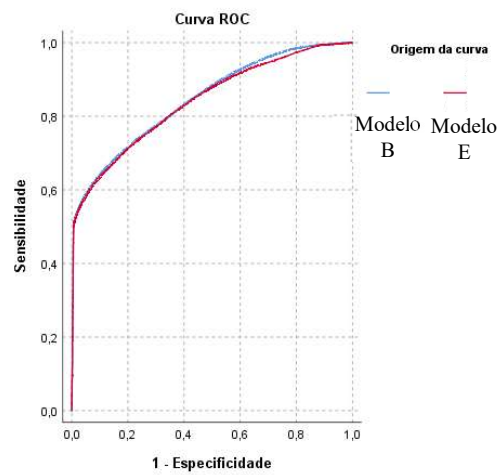
Modelos	AUC
Modelo A	0,837
Modelo B	0,847



Modelos	AUC
Modelo C	0,845
Modelo B	0,847

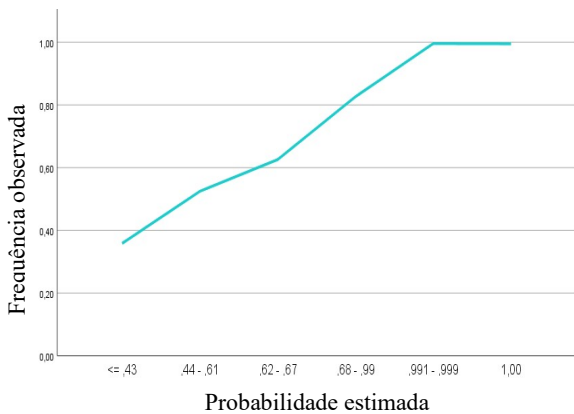


Modelos	AUC
Modelo D	0,847
Modelo B	0,847

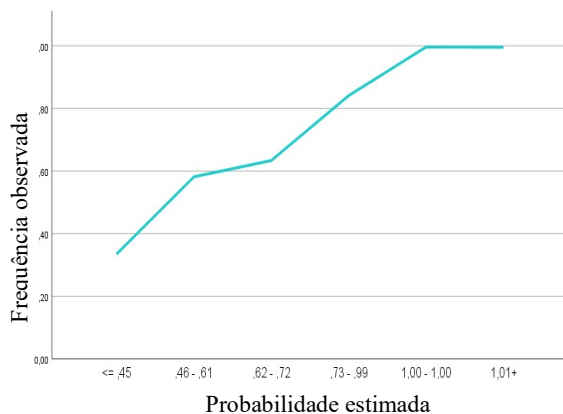


Modelos	AUC
Modelo E	0,841
Modelo B	0,847

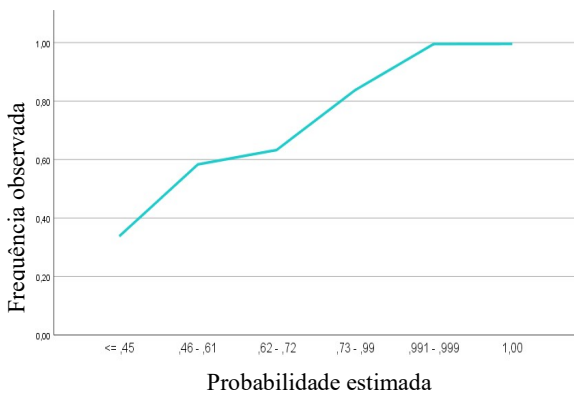
Anexo M - Calibração



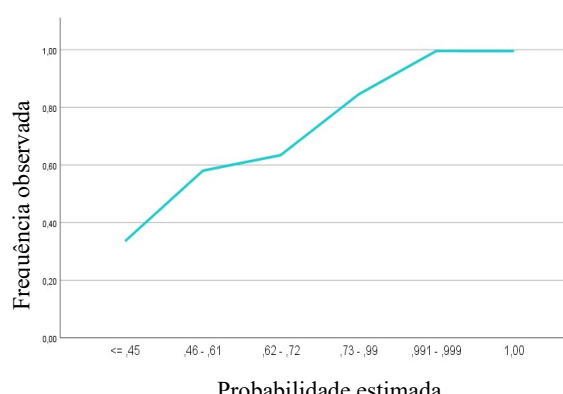
Modelo A



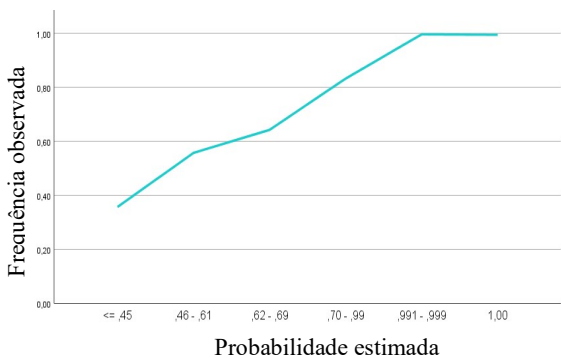
Modelo B



Modelo C



Modelo D



Modelo E