



A influência do espaço na sobrevivência dos estabelecimentos em Portugal

por

Ana Margarida Padrão Cruz Gonçalves Ribeiro

Dissertação para obtenção do grau de Mestre em Economia pela Faculdade
de Economia do Porto

Orientada por:

Maria Isabel Gonçalves da Mota Campos

Pedro José Ramos Moreira de Campos

Julho, 2013

Breve nota biográfica

Ana Margarida Padrão Cruz Gonçalves Ribeiro, nascida a 26 de Setembro de 1990, portuguesa, é natural da Trofa, distrito do Porto. Concluiu o Ensino Secundário no Curso Científico - Humanístico de Ciências Socioeconómicas na Escola Secundária da Trofa em 2008, ano em que ingressou no Ensino Superior. Licenciada em Economia pela Faculdade de Economia da Universidade do Porto (FEP) em 2011, continuou a sua formação académica inscrevendo-se no Mestrado em Economia, área de especialização de Análise Económica, nesse mesmo ano e na mesma instituição (FEP).

Do seu gosto pela Estatística e Econometria em 2013 colaborou com o Observatório de Economia e Gestão de Fraude (OBEGEF). Paralelamente, sempre procurou ter uma participação ativa no meio estudantil ao incorporar diversas organizações de estudantes.

Agradecimentos

Deixo aqui um especial agradecimento aos meus orientadores, Professora Isabel Mota e Professor Pedro Campos, por todo o apoio, disponibilidade, dedicação e motivação que apresentaram ao longo destes meses que se revelaram cruciais para a conclusão desta dissertação.

Agradeço também a todos aqueles que de alguma forma estiveram presentes nesta fase da minha vida, em particular aos meus pais que sempre me incentivaram e me apoiaram na concretização deste objetivo e à minha irmã pelo exemplo que representa para mim.

Resumo

O objetivo desta dissertação é analisar a influência de variáveis espaciais na sobrevivência dos estabelecimentos da Indústria Transformadora em Portugal Continental no período de 1995-2006 (coorte 1995). Para isso, em primeiro lugar procede-se à revisão da literatura que contextualiza a análise de sobrevivência na ciência económica em três diferentes perspetivas (Teoria da Empresa, Ecologia das Organizações e Teoria Evolucionista). Em seguida, identificam-se os principais determinantes e as metodologias utilizadas em análise de sobrevivência. Assim, constata-se que os determinantes da sobrevivência podem ser internos ou externos à empresa entre os quais se encontram as variáveis espaciais definidas: economias de aglomeração e divisão do território em NUTS II. Com recurso ao modelo de Cox e aos *Quadros de Pessoal* (GEP-MTSS, 1995-2006) foram incorporados determinantes internos (dimensão do estabelecimento e número de estabelecimentos) e externos (regime tecnológico, número de entradas, escala mínima eficiente, taxa de crescimento do sector, economias de localização, economias de urbanização e *dummies* por região) como variáveis explicativas da sobrevivência dos estabelecimentos em Portugal. Os resultados da estimação mostram que as variáveis dimensão do estabelecimento, número de estabelecimentos e taxa de crescimento do sector têm uma influência negativa no risco de morte de um estabelecimento. Já o número de entradas e a escala mínima eficiente potenciam este risco. Quanto às variáveis espaciais os resultados são menos conclusivos, sugerindo, contudo: uma influência negativa das economias de urbanização na probabilidade de sobrevivência dos estabelecimentos; e uma influência da localização (por NUTS II) que parece evidenciar melhores níveis de sobrevivência da região Centro face às regiões Norte, Alentejo e Lisboa.

Palavras-Chave: Sobrevivência; Modelo de Cox; Economias de aglomeração; Localização; Estabelecimentos

Códigos JEL: C14; L60; R10

Abstract

This dissertation aims to analyze the influence of spatial variables in the survival of plants operating in the manufacturing sector in Portugal (mainland) between 1995 and 2006 (cohort 1995). For this goal, we start by a literature review that contextualizes the survival analysis in economic sciences in three different perspectives: Theory of the firm, Organizational Ecology and Evolutionary theory. After that, we identify the determinants and methodologies used in survival analysis. Thus, we observed that survival determinants might be internal or external to the firm, where we include the spatial factors we considered: agglomeration economies and division of territory in NUTS II. Using the Cox model and *Quadros de Pessoal* (GEP-MTSS, 1995-2006) some internal (plant size and number of plants) and external determinants (technology, new entries, minimum efficient scale, industrial growth rate, localization economies, urbanization economies and dummies by region) were included as covariates to explain the survival of plants in Portugal. Regarding the variables plant size, number of plants and industrial growth rate, we observed a negative influence in hazard rate; while new entries and minimum efficient scale increase the risk. Focusing on spatial variables the results are less conclusive, while suggesting a negative influence of urbanization economies in the survival likelihood of plants; and an influence of location by region (NUTS II) that seems to exhibit better levels of survival in Centro when comparing to Norte, Alentejo and Lisboa.

Keywords: Survival; Cox model; Agglomeration economies; Location; Plants

JEL Codes: C14; L60; R10

Índice

Breve nota biográfica.....	i
Agradecimentos	ii
Resumo	iii
Abstract.....	iv
Introdução	1
Capítulo 1: A sobrevivência de empresas e a ciência económica: breve contextualização	4
Capítulo 2: Modelização da sobrevivência de empresas: revisão de literatura	8
2.1. Determinantes da sobrevivência.....	8
2.1.1. Determinantes internos à empresa.....	9
2.1.2. Determinantes externos à empresa	12
2.2. Metodologias de análise de sobrevivência	14
Capítulo 3: Espaço e sobrevivência de empresas	20
Capítulo 4: A influência do espaço na sobrevivência dos estabelecimentos em Portugal	34
4.1. Variável dependente	34
4.2. Descrição do modelo	41
4.3. Variáveis explicativas.....	43
4.4. Resultados	50
4.4.1. Pressuposto da <i>Proportional Hazard</i> e Modelo de Cox com variáveis explicativas em interação com o tempo	53
4.4.2. Resultados do Modelo de Cox com efeitos de grupo	55
4.4.3. Testes diagnósticos de qualidade do ajustamento.....	59
4.4.4. Comparação de resultados	60
Conclusões e desenvolvimentos futuros	63
Referências bibliográficas.....	65
Anexo 1 – CAE Rev. 2.1 a dois dígitos	69
Anexo 2 – Nascimentos e sobrevivência (NUTS II e CAE a dois dígitos)	70

Índice de tabelas

Tabela 1 – Espaço e sobrevivência das empresas: estudos empíricos	21
Tabela 2 – Teste de Wilcoxon generalizado para a igualdade das funções sobrevivência por regiões.....	40
Tabela 3 – Variáveis explicativas	44
Tabela 4 – Matriz de correlação entre as variáveis explicativas.....	50
Tabela 5 – Determinantes do risco de morte dos estabelecimentos em Portugal (1995-2006): Modelo de Cox	51
Tabela 6 – Determinantes do risco de morte dos estabelecimentos em Portugal (1995-2006): Modelo de Cox com variáveis em interação com o tempo	54
Tabela 7 – Determinantes do risco de morte dos estabelecimentos em Portugal (1995-2006): Modelo de Cox com efeitos de grupos	57
Tabela 8 – Verificação das hipóteses: resultados obtidos <i>versus</i> sinal esperado.....	61

Índice de figuras

Figura 1 – Sobrevivência dos estabelecimentos (em número e em percentagem dos nascimentos totais) (1995-2006).....	36
Figura 2 – Estrutura dos nascimentos por NUTS II (em número e em percentagem dos nascimentos totais) (1995).....	37
Figura 3 – Estrutura dos nascimentos por CAE a dois dígitos (em número e em percentagem dos nascimentos totais) (1995).....	38
Figura 4 – Sobrevivência de estabelecimentos por NUTS II (em percentagem dos nascimentos por NUTS II) (1995-2006).....	39
Figura 5 – Sobrevivência de estabelecimentos por CAE a dois dígitos (em percentagem dos nascimentos por CAE) (1995-2006)	41
Figura 6 – Teste da proporcionalidade dos riscos por regiões.....	58
Figura 7 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (I)	60
Figura 8 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (II).....	60
Figura 9 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (III)	60
Figura 10 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (IV) ..	60
Figura 11 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (V)....	60
Figura 12 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (VI) ..	60

Introdução

A sobrevivência das empresas (ou estabelecimentos) é abordada na ciência económica em diferentes perspetivas, em particular no âmbito da Teoria da Empresa, Ecologia das Organizações e Teoria Evolucionista. A Teoria da Empresa procura explicar a existência de empresas assim como os seus limites (e.g. Coase, 1937; Williamson, 1985). Assim, chama a atenção para a importância das taxas de entrada e da escala mínima eficiente (entre outros) na sobrevivência das empresas. Por seu lado, e de acordo com Singh e Lumsden (1990), a Ecologia das Organizações procura explicar o processo de nascimento, mudança e morte das empresas, realçando o papel da seleção natural das organizações. São vários os fatores explicativos da sobrevivência identificados por esta escola, de que são exemplo a “responsabilidade da novidade” e a “responsabilidade da pequenez”. Por fim, e segundo Nelson e Winter (1982) (cfr Silva *et al.*, 2004), a Teoria Evolucionista pretende explicar o comportamento das empresas com base na rotina, racionalidade limitada e seleção natural, identificando como determinantes da sobrevivência das empresas variáveis como a idade da empresa.

A literatura que aborda os determinantes da sobrevivência é muito vasta (e.g. Esteve-Pérez e Mañez-Castillejo, 2008; Manjón-Antolín e Arauzo-Carod, 2008). Centrando a atenção no caso português, o estudo precursor de Mata e Portugal (1994) destaca-se ao analisar a sobrevivência das empresas portuguesas com recurso à base de dados *Quadros de Pessoal*, sendo ainda de referir Mata e Portugal (2004), entre outros.

Contudo, quando se incorpora a vertente espacial na análise de sobrevivência, a literatura já não é tão assertiva apresentando conclusões diversas. Neste sentido, a ideia de associar sobrevivência e espaço torna-se o desafio ao qual esta dissertação pretende dar resposta.

Analisando em primeira instância os determinantes que influenciam a sobrevivência das empresas (ou estabelecimentos), verifica-se que estes podem classificar-se como: fatores internos (específicos à empresa ou estabelecimento), de que é exemplo a dimensão e a idade da empresa, as atividades de investigação e desenvolvimento (I&D) e a propriedade; ou fatores externos (específicos da indústria; espaciais; ou relacionados com o ciclo económico), como é o caso da tecnologia, taxas de entrada no sector, escala mínima eficiente e dinâmica do sector de atividade. No

âmbito das variáveis espaciais destacam-se as economias de aglomeração assim como o papel da localização em determinadas áreas geográficas, sugerindo a divisão do espaço em áreas com especificidades distintas.

As economias de aglomeração correspondem a externalidades positivas que resultam da concentração espacial da atividade económica. Significa então que, na sua presença, as empresas possuem vantagens de desempenho devido, entre outros, a fatores como *spillovers* do conhecimento e partilha de capital humano. As economias de aglomeração podem, ainda, ser divididas em economias de localização e economias de urbanização. As primeiras relacionam-se com vantagens conseguidas através da concentração espacial de empresas do mesmo sector ou relacionados, enquanto as segundas prendem-se com os benefícios alcançados através da concentração de atividades de diferentes sectores. As economias de aglomeração enquanto determinantes da sobrevivência são incluídas em alguns estudos (e.g. Shaver e Flyer, 2000; Staber, 2001; Acs *et al.*, 2007; Boschma e Wenting, 2007; Wennberg e Lindqvist, 2010). De acordo com estes estudos, as economias de localização têm um impacto negativo na sobrevivência das empresas (Staber, 2001; Acs *et al.* 2007; Boschma e Wenting, 2007), enquanto as economias de localização que resultam de sectores relacionados possuem influência positiva na sobrevivência (Staber, 2001; Boschma e Wenting, 2007, Wennberg e Lindqvist, 2010). Por seu lado, as economias de urbanização evidenciam uma influência positiva na sobrevivência (Acs *et al.* 2007, Wennberg e Lindqvist, 2010).

A localização em determinadas áreas geográficas (e.g. espaço urbano vs espaço rural; divisão do espaço em regiões) é também considerada em alguns estudos de sobrevivência (e.g. Strotmann, 2007; Huiban, 2011; Fotopoulos e Louri, 2000; Tödtling e Wanzenböck, 2003). Alguns destes autores concluem que a probabilidade de morte é maior em áreas urbanas do que em áreas rurais (Tödtling e Wanzenböck, 2003; Strotmann, 2007; Huiban, 2011); contudo, outros encontram uma conclusão contrária (Fotopoulos e Louri, 2000).

Neste contexto, este estudo pretende responder às seguintes questões: Terá o espaço influência na sobrevivência dos estabelecimentos em Portugal? Em particular, qual a influência das economias de aglomeração (localização e urbanização) sobre o

risco de morte? Será que a localização em determinadas regiões influencia a sobrevivência dos estabelecimentos?

Assim, os principais objetivos deste trabalho são: (1) recolher e analisar a sobrevivência dos estabelecimentos em Portugal Continental (coorte de estabelecimentos de 1995 a operar na Indústria Transformadora) e detetar eventuais diferenças por região; (2) recolher e analisar informação estatística sobre variáveis explicativas da sobrevivência das empresas (ou estabelecimentos), em particular, de natureza espacial; (3) recorrendo ao modelo de Cox que permite a estimação semi-paramétrica da função de risco, concluir sobre a influência das economias de aglomeração e da localização nas regiões NUTS II (Norte, Centro, Lisboa, Alentejo e Algarve) na sobrevivência dos estabelecimentos em Portugal.

Desta forma, a estrutura desta dissertação é a seguinte. No Capítulo 1 é apresentada uma breve contextualização da análise de sobrevivência na ciência económica, no âmbito da Teoria da Empresa, Ecologia das Organizações e Teoria Evolucionista. No Capítulo 2 procede-se à revisão de literatura sobre a modelização da sobrevivência de empresas, apresentando as metodologias utilizadas nos estudos da sobrevivência bem como os determinantes de sobrevivência, internos e externos à empresa. No Capítulo 3 são analisados estudos que particularizam a relação entre variáveis espaciais (economias de aglomeração e divisão do espaço) e análise de sobrevivência. No Capítulo 4 dá-se resposta às questões de investigação anteriormente referidas, através da estimação semi-paramétrica do modelo de Cox para o caso Português. Por fim, são retiradas as principais conclusões sobre influência do espaço na sobrevivência dos estabelecimentos em Portugal e apontadas pistas para investigação futura.

Capítulo 1: A sobrevivência de empresas e a ciência económica: breve contextualização

A Teoria Neoclássica define a empresa como uma entidade que transforma recursos (*inputs*) em bens ou serviços (*output*), tendo como objetivo a maximização do lucro. De acordo com esta visão, a dimensão e o número de empresas que existem no mercado são explicadas pela relação entre a procura e a estrutura de custos (Church e Ware, 2000). Contudo, de acordo com Penrose (1959), a Teoria Neoclássica não explica os determinantes do crescimento das empresas ou os seus limites. Ainda, os autores neoclássicos não explicam por que razão determinadas transações ocorrem no mercado, enquanto outras são realizadas no interior das empresas. É neste contexto que surge a Teoria da Empresa, procurando ultrapassar algumas das limitações da Teoria Neoclássica (e.g. Coase, 1937; Williamson, 1985).

Williamson (1985) (cfr Silva *et al.*, 2004) procura explicar a diferença entre o mercado e a empresa sustentando que o objetivo da empresa é a poupança de custos de transação. Coase (1937) explica que o papel da empresa emerge quando as operações dentro da empresa possuem um custo inferior ao custo de serem efetuadas através do sistema de mercado. De acordo com Coase (1937), as empresas são entidades em que as atividades estão organizadas segundo princípios distintos da lógica do mercado: a organização no interior da empresa obedece ao poder hierárquico do gestor, abdicando do sistema de preços do mercado. Logo, a organização no interior da empresa é realizada pela mão visível do gestor, que decide o que produzir, quanto e como. Em alternativa, no mercado as transações são organizadas segundo o sistema de preços (mão invisível do mercado). A decisão de organizar a atividade económica na empresa ou de recorrer ao mercado resulta da comparação dos benefícios e custos.¹ Mata (2007) explica que a atividade económica é concentrada no interior da organização quando os custos de recorrer ao mercado excedem os benefícios. Contudo, há limites à dimensão

¹ Podemos identificar como benefícios de usar o mercado o aproveitamento de economias de escala, a adaptação eficiente e a minimização de custos (Church e Ware, 2000; Mata, 2007). Para além do preço pago pelo bem ou serviço, existem outros custos de recorrer ao mercado. Estes custos, denominados de custos de transação, estão associados à pesquisa do produto (*search costs*), negociação do preço e da transação (*costs of bargaining*) e custos de estabelecimento de contratos de compra e venda e de acordos em situação de incumprimento de contratos (*costs of monitoring and enforcing contractual agreements*) (Mata, 2007).

de uma empresa que decorrem de, frequentemente, a propriedade e a gestão estarem descentralizadas e, enquanto o objetivo do proprietário é a maximização de lucros, o do gestor poderá ser a maximização da sua utilidade.²

Vários estudos sobre a sobrevivência das empresas incorporam princípios enunciados pela Teoria da Empresa. Esta fornece conceitos que estão subjacentes em estudos como Mata e Portugal (1994) (que identifica a aplicação de técnicas menos intensivas em capital, elevadas taxas de entrada e escala mínima eficiente elevada como associados a uma elevada taxa de mortalidade das empresas) ou Persson (2004) (que identifica a escala sub-ótima como determinante da morte de empresas), entre outros.

Para além da Teoria da Empresa, o estudo da empresa e da sua sobrevivência é também tratado no âmbito da Ecologia das Organizações. De acordo com Singh e Lumsden (1990), esta área procura estudar a diversidade das organizações, isto é, os processos de nascimento, morte e mudança das empresas, dando ênfase à questão de como as condições sociais influenciam estes processos. Para tal, recorre a conhecimentos de outras ciências, como é o caso da biologia e da sociologia, e investiga o processo de seleção natural das organizações.

A análise da ecologia das organizações pode ser feita a três níveis: nível organizacional (desenvolvimento), nível populacional (seleção) e nível da comunidade (aspectos macroeconómicos) (Carroll, 1984). Tal como Singh e Lumsden (1990) referem, a maior parte da investigação debruça-se sobre a análise ao nível da organização e população, em particular, dos processos de formação, mortalidade e mudança das organizações.

Relativamente à formação de organizações, e em concordância com Singh e Lumsden (1990), tem sido dada uma grande relevância ao papel da densidade e da dinâmica populacional. Assim, o número de organizações no mercado (densidade) influencia a taxa de formação das organizações via processos de legitimação e concorrência (Hannan, 1986 cfr Singh e Lumsden, 1990). A dinâmica da densidade e a alteração dos seus níveis, por sua vez, provoca efeitos na dinâmica da população (Tucker *et al.*, 1988 cfr Singh e Lumsden, 1990).

² Segundo Church e Ware (2000) a discrepância de objetivos entre o gestor e o proprietário é limitada por fatores internos e externos à empresa. Os gestores, internamente, estão limitados por contratos de incentivos (em que por exemplo o seu rendimento depende dos lucros). Externamente, as limitações advêm dos mercados de trabalho de gestores, dos mercados de capital, do perigo de falência e da concorrência no mercado do produto.

Ainda de acordo com os mesmos autores, as teorias que sustentam a mortalidade das organizações são: *fitness set theory*, *liability of newness* (responsabilidade da novidade), *resource partitioning theory* (teoria da partilha dos recursos), *liability of smallness* (responsabilidade da pequenez), a influência das condições de formação das organizações e a dependência da densidade e dinâmica da população. A *fitness set theory* relaciona-se com a teoria da concorrência e com a teoria da largura do nicho que distingue entre organizações especializadas e generalistas. A teoria da partilha de recursos parte desta distinção e analisa como os especialistas e generalistas operam em diferentes áreas de recursos. A responsabilidade da novidade e da pequenez prende-se com as desvantagens das organizações novas e pequenas. A influência das condições iniciais analisa o ambiente social do momento da formação das organizações. Por fim, a dependência da densidade e a dinâmica da população, uma vez mais, relaciona a taxa de mortalidade com questões de legitimação e concorrência das organizações.

Quanto à mudança das organizações, pouca atenção tem sido dada a este tópico. Neste âmbito, Hannan e Freeman (1977) sustentam que as empresas estão sujeitas a inércia no seu processo de mudança e, como tal, as mudanças vêm sobretudo de nascimentos e mortes.

Alguns conceitos apresentados nesta abordagem são utilizados no âmbito da análise de sobrevivência das empresas. Os mais frequentemente referidos são a “responsabilidade da novidade” (Persson, 2004; Jovanovic, 1982, Ericson e Pakes, 1995 cfr Manjón-Antolín e Arauzo-Carod, 2008; Geroski, 1995 cfr Esteve-Pérez e Mañez-Castillejo, 2008) e a “responsabilidade da pequenez” (Mata e Portugal, 1994; Jovanovic, 1982 e Ericson e Pakes, 1995 cfr Manjón-Antolín e Arauzo-Carod, 2008; Box, 2008; Geroski, 1995 cfr Esteve-Pérez e Mañez-Castillejo, 2008).

Uma terceira perspetiva de análise é a Teoria Evolucionista. De acordo com Silva *et al.* (2004), esta é vista como um subconjunto do grupo das teorias baseadas em competência e centra-se nos conceitos de rotina, racionalidade limitada e seleção natural.

De acordo com Nelson e Winter (1982) (cfr Silva *et al.*, 2004), a rotina é definida como o conjunto de comportamentos previsíveis e regulares que engloba uma série de especificidades da empresa possuindo um papel análogo ao que os genes têm na biologia. Esta é adquirida através da acumulação de processos de aprendizagem que

fomentam o conhecimento. Contudo, os próprios autores consideram que existe uma parte do comportamento que não é atribuído à rotina. Estas questões não rotineiras são, assim, atribuídas a elementos aleatórios.

Nelson e Winter (1982) (cfr Silva *et al.*, 2004) defendem que as empresas não são maximizadoras do lucro, mas apresentam um comportamento motivado pelo lucro. Esta visão implica uma abordagem de racionalidade limitada impulsionada por fenómenos económicos reais complexos. A presença de incerteza e a própria dificuldade em compreender os fenómenos complexos leva a múltiplas interpretações pelas empresas o que implica que as suas decisões sejam tomadas com base num conjunto limitado de informação.

Neste seguimento, surge o conceito de seleção natural. Os mesmos autores mostram que as empresas possuindo como objetivo a motivação pelo lucro (ao contrário da sua maximização), aplicam certas regras de decisão que, não sendo ideais, satisfazem o objetivo. Ao longo do tempo a seleção natural atua como o mecanismo de mercado distinguindo empresas rentáveis de empresas não rentáveis. As primeiras vão empurrando as segundas para fora do mercado num processo de seleção natural.

Também os conceitos aqui apresentados são utilizados na análise de sobrevivência das empresas. É o caso de Nelson e Winter (1982) (cfr Esteve-Pérez e Mañez-Castillejo, 2008), que identificam a idade como determinante da sobrevivência das empresas, na medida em que estas necessitam de um período de estabelecimento que permita o desenvolvimento de certas competências organizacionais.

De acordo com Silva *et al.* (2004), existem alguns pontos de convergência entre a Ecologia das Organizações e a Teoria Evolucionista, em particular, a ênfase nos processos de seleção e a adoção de conceções cognitivas semelhantes. Constata-se ainda que a própria Teoria Neoclássica tem, recentemente, vindo a reconhecer que o pressuposto da informação perfeita é insatisfatório, tendo vindo a incorporar o risco na tomada de decisões pelas empresas através da atribuição de probabilidades a diferentes resultados possíveis (ainda que nem sempre seja possível definir essas probabilidades).

Após este breve enquadramento teórico, e com o objetivo de avaliar a sobrevivência das empresas em Portugal, o capítulo seguinte pretende descrever as metodologias utilizadas no estudo da sobrevivência de empresas, identificando, em particular, os seus determinantes e modelos de sobrevivência.

Capítulo 2: Modelização da sobrevivência de empresas: revisão de literatura

A análise da sobrevivência das empresas tem sido alvo de vários estudos que têm como finalidade compreender o processo de entrada, permanência e saída das empresas do mercado, bem como identificar os fatores que explicam que algumas empresas tenham, *ceteris paribus*, mais hipóteses de sobreviver do que outras (Manjón-Antolín e Arauzo-Carod, 2008).

No seguimento da literatura, o conceito de sobrevivência é apresentado como o tempo decorrente entre a entrada e a saída de uma determinada empresa (ou estabelecimento) da indústria, ou seja, entre o início e o termo da sua atividade (Mata e Portugal, 1994; Acs *et al.*, 2007; Manjón-Antolín e Arauzo-Carod, 2008).

Neste trabalho, será dado um enfoque particular à modelização da sobrevivência realizada através de modelos econométricos de duração. Significa assim que se excluem outras abordagens metodológicas como análises descritivas ou estimações lineares uni ou multivariadas. Ainda, com o intuito de aferir sobre a sobrevivência das empresas, são analisados os principais determinantes ou variáveis que influenciam a duração da vida das empresas (Mata e Portugal, 1994).

Neste sentido, são apresentados de seguida os principais determinantes (que podem ser internos e externos à empresa) e metodologias utilizadas na análise de sobrevivência.

2.1. Determinantes da sobrevivência

Mata e Portugal (1994) enumeram quatro factos estilizados sobre os movimentos de entrada e saída de empresas do mercado: (1) as taxas de entrada são elevadas na maioria dos países; (2) os movimentos de entrada são acompanhados de posteriores movimentos de saída; (3) quer as empresas que entram, quer as que saem, têm normalmente pequena dimensão; (4) das entrantes poucas têm capacidade para sobreviver e adquirir quota de mercado. Assim, torna-se crucial analisar não apenas as características relevantes das empresas no momento de entrada no mercado, mas

também o que se passa após essa entrada ocorrer e que dita a sua permanência ou abandono da indústria.

Utilizando a terminologia adotada por Manjón-Antolín e Arauzo-Carod (2008), os determinantes da sobrevivência das empresas podem classificar-se como internos ou externos à empresa. No primeiro grupo incluem-se todos aqueles que são específicos à empresa como a dimensão, a idade, as atividades de investigação e desenvolvimento (I&D) levadas a cabo e a propriedade da empresa. No segundo grupo inserem-se fatores relacionados com a envolvente da empresa em três vertentes: específicos da indústria; relacionados com o ciclo económico; e espaciais (a desenvolver no capítulo seguinte).

2.1.1. Determinantes internos à empresa

Em primeira instância, analisam-se os determinantes que à empresa dizem respeito: dimensão, idade, investigação e desenvolvimento e propriedade.

Relativamente à dimensão, a literatura defende, em geral, que as empresas mais pequenas estão associadas a menores probabilidades de sobrevivência (Mata e Portugal, 1994; Box 2008; Geroski, 1995 cfr Esteve-Pérez e Mañez-Castillejo, 2008; Jovanovic, 1982 e Ericson e Pakes, 1995 cfr Manjón-Antolín e Arauzo-Carod, 2008), o que alguns denominaram de “responsabilidade da pequenez” (Strotmann, 2007; Box, 2008). Esta regularidade justifica-se porque empresas mais pequenas atuam, muitas vezes, numa escala sub-ótima possuindo uma desvantagem de custo (Persson, 2004), têm uma gestão menos preparada (Lucas, 1978 cfr Mata e Portugal, 1994), prenunciam menores expectativas de sucesso (Frank, 1988 cfr Mata e Portugal, 1994) revelando a sua fraca capacidade empreendedora à entrada (Strotmann, 2007), têm mais dificuldades no acesso ao mercado de capitais (Mata e Portugal, 1994; Strotmann, 2007), aplicam técnicas menos intensivas em capital (Mata e Portugal, 1994) e possuem maiores dificuldades em conseguir trabalho qualificado (Strotmann, 2007). Uma outra perspetiva defende que a “responsabilidade da pequenez” está mais ligada a saídas voluntárias, e apresenta o conceito de “responsabilidade da média” sugerindo que o risco de saída por motivo de insolvência é superior em empresas de média dimensão (Woywode, 1998 cfr Strotman, 2007).

Quanto à idade, a maioria dos autores defende que quanto mais antiga é uma empresa, menor é o seu risco de morte, ou seja, maior é a probabilidade de sobreviver (Persson, 2004; Geroski, 1995 cfr Esteve-Pérez e Mañez-Castillejo, 2008; Jovanovic, 1982 e Ericson e Pakes, 1995 cfr Manjón-Antolín e Arauzo-Carod, 2008). À hipótese que o risco de insucesso de uma empresa diminui com o passar do tempo chama-se “responsabilidade da novidade” (Strotmann, 2007). Esta pressupõe que uma empresa necessita de um período de estabelecimento que permita o desenvolvimento de certas competências organizacionais (Carroll e Hannan, 2000, Nelson e Winter, 1982 e Stinchcombe, 1965 cfr Esteve-Pérez e Mañez-Castillejo, 2008). Uma visão alternativa, “responsabilidade da adolescência”, defende que o risco de morte aumenta após a entrada, atinge um máximo e depois diminui, ou seja, tem a configuração de U invertido (Strotmann, 2007; Bruderl e Schussler, 1990 e Fichman e Levinthal, 1991 cfr Esteve-Pérez e Mañez-Castillejo, 2008). Isto indica que há um período inicial de utilização de recursos e tentativa de sucesso (Strotmann, 2007) em que as taxas de risco de morte aumentam até um máximo, atingido no momento em que os recursos se esgotam, e que depois diminuem com a saída das empresas menos eficientes do mercado (Esteve-Pérez e Mañez-Castillejo, 2008). Por fim, surge o conceito de “responsabilidade da senescência” em que a partir de uma determinada idade a probabilidade de saída aumenta devido à desadequação das empresas ao mercado (Santarelli e Lotti, 2005; Baum, 1989 e Hannan, 1998 cfr Esteve-Pérez e Mañez-Castillejo, 2008).

O desenvolvimento de atividades de investigação e desenvolvimento (I&D) apresenta, também, um importante papel na taxa de sobrevivência das novas empresas. É de esperar que as empresas que mais investem em I&D aumentem a sua probabilidade de sobrevivência (Esteve-Pérez e Mañez-Castillejo, 2008). Segundo Esteve-Pérez e Mañez-Castillejo (2008), e de acordo com a teoria baseada nos recursos da empresa, as empresas que apostam em I&D e publicidade adquirem uma vantagem competitiva, ao desenvolverem inovações que não podem ser imitadas pelos concorrentes, o que melhora a sua eficiência e aumenta a probabilidade de sobrevivência. Contudo, ainda de acordo com os mesmos autores, estas inovações podem ser produzidas pela empresa, usufruindo da propriedade de não poderem ser imitadas, ou adquiridas através de compra de tecnologia, o que é acessível a qualquer concorrente. Logo, espera-se que empresas que produzam o seu próprio I&D possuam maior probabilidade de sobreviver

(Almus e Nerlinger, 1999, Calvo, 2006 e Cefis e Marsili, 2006 cfr Esteve-Pérez e Mañez-Castillejo, 2008). Por fim, o “prêmio de inovação” é superior quando a inovação é feita ao nível do processo face à inovação do produto (Cefis e Marsili, 2005 cfr Manjón-Antolín e Arauzo-Carod, 2008).

Analisando o efeito da propriedade na probabilidade de sobrevivência das novas empresas, Manjón-Antolín e Arauzo-Carod (2008) fazem distinção entre: estabelecimentos criados de novo e estabelecimentos criados a partir de empresas já existentes; multi-estabelecimentos e estabelecimentos únicos; e empresas domésticas e empresas estrangeiras. Em relação às primeiras, é de esperar que estabelecimentos criados a partir de empresas já existentes detenham uma maior probabilidade de sobrevivência do que os estabelecimentos criados de novo (Audretsch e Mahmood, 1995 cfr Manjón-Antolín e Arauzo-Carod, 2008). Mata e Portugal (2004) concluem que, por um lado, novas empresas têm desvantagens em termos de aprendizagem e podem despoletar ações agressivas por parte de empresas já estabelecidas no mercado; mas, por outro lado, quando se tratam de empresas associadas a estrangeiros cuja empresa mãe possui superioridade tecnológica e/ou organizacional é preferível iniciar uma empresa de raiz do que adquirir uma empresa já existente porque, assim, é possível incorporar essas vantagens. Quanto à segunda dicotomia, é mais provável a sobrevivência de multi-estabelecimentos do que de um estabelecimento único, uma vez que os primeiros estão, geralmente, associados a empresas de maior dimensão, sinalizam-se como mais otimistas e com menos restrições financeiras (Mata e Portugal, 1994). Para comparação do terceiro conjunto surge a necessidade de análise das diferentes nacionalidades das empresas (Manjón-Antolín e Arauzo-Carod, 2008). Quanto a este assunto, Esteve-Pérez e Mañez-Castillejo (2008) consideram que a literatura é inconclusiva e, portanto, não é possível concluir sobre a sobrevivência com base neste argumento. Por um lado, o capital estrangeiro numa empresa pode ser percebido como qualidade não observada. Por outro lado, empresas estrangeiras podem ter problemas com a coordenação e adaptação a outros mercados (Hymer, 1976 cfr Esteve-Pérez e Mañez-Castillejo, 2008). Mata e Portugal (2004) analisam no seu estudo as diferenças de sobrevivência entre empresas domésticas e estrangeiras aferindo sobre as suas características no momento da entrada e o seu desempenho/crescimento pós-entrada. A sua conclusão aponta para que as empresas domésticas têm menos

probabilidade de sobreviver, ou seja, têm mais propensão em sair do mercado, que as empresas estrangeiras. Isto acontece tanto em empresas criadas de novo, como em empresas criadas a partir de existentes.

2.1.2. Determinantes externos à empresa

a) Específicos da indústria

São várias as características da indústria que podem condicionar o tempo decorrido entre a entrada e a saída de uma empresa do mercado. Questões como tecnologia, taxas de entrada, escala mínima eficiente e dinâmica do sector de atividade podem ter um papel significativo na sobrevivência das empresas.

Quanto à tecnologia, é de esperar que empresas *high-tech* possuam menor probabilidade de sobrevivência do que as *low-tech* (Audretsch, 1995 e Agarwal e Audretsch, 2001 cfr Manjón-Antolín e Arauzo-Carod, 2008) devido à rápida obsolescência verificada nas primeiras (Agarwal e Gort, 2002 cfr Manjón-Antolín e Arauzo-Carod, 2008). Na sua investigação, Agarwal e Audretsch (2001) relacionam a dimensão de uma empresa à entrada, a sua tecnologia e a sua probabilidade de sobrevivência. As conclusões retiradas são: as taxas de sobrevivência de produtos *low-tech* são superiores em empresas de grande dimensão à entrada do que em empresas de pequena dimensão, não sendo essa vantagem da dimensão tão evidente quanto a produtos *high-tech*; e as taxas de sobrevivência de empresas de pequena dimensão são superiores em produtos *high-tech* face aos *low-tech*, diferença que não é tão significativa no caso das empresas de grande dimensão.

Em relação às taxas de entrada, quanto maiores estas forem, menor a probabilidade de sobrevivência (Mata e Portugal, 1994; Manjón-Antolín e Arauzo-Carod, 2008). Mata e Portugal (1994) justificam esta regularidade com o acréscimo de concorrência e desafios que uma empresa está sujeita em consequência do maior fluxo de entradas.

A evidência aponta ainda para uma menor probabilidade de insucesso de empresas que operem em indústrias com uma escala mínima eficiente menor (Audretsch e Mahmood, 1991; Mata e Portugal, 1994; Agarwal e Audretsch, 1995, Tveteras e Eide,

2000, Strotmann, 2007 cfr Manjón-Antolín e Arauzo-Carod, 2008). Mata e Portugal (1994) explicam que isto acontece porque, geralmente, a entrada ocorre em pequena dimensão sendo necessário a empresa crescer para atingir a escala mínima eficiente. Assim, quanto maior esta for, mais árdua é a tarefa de a alcançar. Mata e Portugal (1994) referem, ainda, que o caminho até à escala mínima eficiente é mais importante quando as desvantagens em operar numa subescala são maiores e que, as novas empresas têm maior probabilidade de sobrevivência quando a maioria das empresas estabelecidas operam numa escala inferior à ótima. No mesmo sentido, Agarwal e Audretsch (2001) mencionam no seu estudo uma implicação da teoria da seleção da empresa, isto é, empresas que entram numa escala sub-ótima, se possuírem um bom desempenho crescem, o que é favorável à sua sobrevivência. Assim, as que são bem-sucedidas expandem-se, as que não são mantêm-se pequenas e podem ser forçadas a sair por operarem numa escala inferior à ótima.

Por fim, relativamente à dinâmica do sector de atividade, verifica-se que o risco de morte é maior em fases mais avançadas do ciclo de vida da indústria (Agarwal e Audretsch, 2001 e Agarwal *et al.*, 2002 cfr Manjón-Antolín e Arauzo-Carod, 2008), porque em fases iniciais há maior margem para inovação (Winter, 1984 cfr Manjón-Antolín e Arauzo-Carod, 2008). Agarwal e Audretsch (2001) mostram que as empresas em fases de formação concorrem com o intuito de produzirem um melhor produto que lhes permita crescerem e, desta forma, aumentarem as suas hipóteses de sobrevivência, o que não acontece em fases mais avançadas do ciclo. Mata e Portugal (1994) explicam que é mais fácil sobreviver em indústrias de crescimento rápido, crescimento este que se relaciona intimamente com as fases iniciais do ciclo de vida da indústria onde são introduzidas inovações, desde que as empresas cresçam sem despoletarem respostas agressivas das concorrentes, nomeadamente devido a perdas de quota de mercado.

b) Relacionados com o ciclo económico

A probabilidade de sobrevivência está, ainda, relacionada com o ciclo económico, ou seja, é superior em períodos de expansão económica e menor em períodos de abrandamento económico (Box, 2008; Geroski, 1995 e Caves, 1998 cfr Manjón-Antolín e Arauzo-Carod, 2008). Box (2008) procura no seu estudo analisar não

apenas a sobrevivência sob a perspectiva microeconómica como, também, envolver a vertente macroeconómica. Assim, conclui que em períodos de expansão macroeconómica ou quando as taxas de juro são baixas o risco de morte diminui; e, quando se está perante períodos de crise, a taxa de sobrevivência diminui. Verifica-se, ainda, que em períodos de baixo crescimento económico, baixas taxas de juro e altas taxas de desemprego, as taxas de entrada na indústria aumentam (Highfield e Smiley, 1987 cfr Box, 2008). Por fim, é em períodos em que as taxas de desemprego são menores que as empresas criadas têm mais hipóteses de sobreviverem (Audretsch e Mahmood, 1995 cfr Manjón-Antolín e Arauzo-Carod, 2008).

2.2. Metodologias de análise de sobrevivência

a) Análise de sobrevivência: generalidades

Quando se fala em análise de sobrevivência de uma empresa, fala-se no estudo do tempo decorrido entre a entrada e saída de uma empresa do mercado, ou seja, entre a entrada e a ocorrência do acontecimento de interesse. Apesar de o enfoque neste trabalho ser na área de Economia, a análise de sobrevivência pode ser aplicada a outras áreas como é o caso da medicina, ciências biomédicas, engenharia, sociologia, psicologia, etc.

Nesta secção, e com base em Cox e Oakes (1984), Manjón-Antolín e Arauzo-Carod (2008), Rocha e Papoila (2009) e Cleves *et al.* (2010), será apresentada a metodologia de análise de sobrevivência sob o ponto de vista estatístico, isto é, o estudo da análise de sobrevivência através da aplicação de modelos e métodos estatísticos a uma dada amostra de empresas.

Um aspeto importante prende-se com a determinação dos indivíduos que integram a amostra. De acordo com Rocha e Papoila (2009), estes podem ser censurados, sempre que, para uma dada empresa, não for observada a ocorrência do acontecimento de interesse, neste caso, a morte. A censura pode ser à direita (o mais comum), à esquerda, intervalar, dupla ou independente e não informativa. A censura à direita acontece quando a ocorrência do acontecimento de interesse ainda não se verificou no termo da observação; já a censura à esquerda ocorre quando o

acontecimento de interesse tem lugar antes do início da observação. A censura intervalar, por sua vez, acontece quando apenas se sabe que o acontecimento de interesse ocorreu num certo intervalo de tempo, mas não o momento exato em que ocorreu. A censura dupla pode-se referir a dois casos: um primeiro em que há censura à esquerda e à direita e, um outro, em que é censurada a origem e o tempo até a ocorrência do acontecimento de interesse. Por fim, a censura independente e não informativa diz respeito a amostras em que o tipo de censura não está especificado e, como tal, recorre-se a ferramentas estatísticas que recorrendo à utilização de probabilidades de risco, garantem a independência condicional dos mecanismos de morte e censura. Pode, ainda, estar presente o fenómeno de truncatura, ou seja, quando se excluem empresas da amostra. Em casos de censura e truncatura é importante adotar os métodos de estimação mais adequados.

O objetivo final da aplicação das técnicas estatísticas de análise de sobrevivência é aferir quanto à sobrevivência das empresas tendo em conta os seus determinantes representados por variáveis explicativas.

Em primeiro lugar, é necessário definir sob a forma de uma função a sobrevivência e o risco no contexto de análise de sobrevivência.

Segundo Rocha e Papoila (2009), pode definir-se a função sobrevivência de uma empresa, $S(t)$,³ como a probabilidade de uma empresa com o tempo de vida T sobreviver para além do instante t , isto é:

$$(2.1) \quad S(t) = P(T > t), t \geq 0$$

A função risco/função *hazard*, $h(t)$,⁴ por sua vez, descreve a evolução temporal da variação infinitesimal da probabilidade de morte de uma empresa, isto é, a taxa instantânea de falência de uma empresa no momento t condicionada pela sobrevivência até esse momento. Representa-se da seguinte forma:

$$(2.2) \quad h(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t+dt | T \geq t)}{dt}$$

³ $S(t)$ é monótona decrescente e contínua com $S(t) = 0$ e $S(+\infty) = \lim_{t \rightarrow \infty} S(t) = 0$. A função sobrevivência é decrescente. T é uma variável aleatória não negativa, absolutamente contínua.

⁴ A função risco representa como o risco de morte se comporta ao longo do tempo e tem como características: $h(t) \geq 0$ e $\int_0^{+\infty} h(t)dt = +\infty$; sendo $f(t)$ a função densidade de probabilidade representada por $f(t) = -S'(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t+dt)}{dt}$, a função risco apresenta a seguinte relação: $h(t) = f(t)/S(t)$.

As funções risco podem ter diversas configurações. De acordo com Rocha e Papoila (2009), as mais comuns são: monótona crescente, monótona decrescente, constante, *bathtub-shaped* e *hump-shaped*.

Para proceder à análise de sobrevivência de uma empresa, é necessário relacionar os determinantes da sobrevivência com o tempo de sobrevivência. Uma forma de proceder a esta relação é estabelecida através de um modelo de regressão em que se relacionam variáveis explicativas (variáveis independentes ou covariáveis) (determinantes da sobrevivência) com uma variável dependente (tempo de sobrevivência). Estes modelos dividem-se em categorias, cujas mais frequentemente utilizadas são os modelos com funções de risco proporcionais (PH) (do inglês “*Proportional Hazard*”) e modelos de tempo de vida acelerado (AFT) (do inglês “*Accelerated Failure Time*”).

Nos modelos PH acredita-se que há proporcionalidade entre as funções de risco de empresas com diferentes valores para as variáveis independentes, não dependendo esta relação do tempo t . A função de risco condicional nestes modelos é, assim, dada por:

$$(2.3) \quad h(t; \mathbf{z}) = h_0(t)\varphi(\mathbf{z})$$

em que \mathbf{z} representa o vector de variáveis independentes associadas a uma dada empresa; $h_0(t)$ é um patamar que corresponde à função de risco de uma empresa cujas variáveis explicativas se associam ao vector $\mathbf{z} = \mathbf{0}$; e $\varphi(\mathbf{z})$ é uma função que relaciona o vector de variáveis independentes com a sobrevivência de uma empresa, isto é, representa o “risco relativo” de uma empresa com um vector de covariáveis \mathbf{z} face ao risco de uma empresa com o vector $\mathbf{z} = \mathbf{0}$.

Nos modelos AFT existe um fator multiplicativo t associado às variáveis explicativas que provoca um efeito de aceleração (ou travão) do tempo decorrido até à morte de uma empresa. Esta característica é visível na função risco condicional associada que se define da seguinte forma:

$$(2.4) \quad h(t; \mathbf{z}) = h_0(t\psi(\mathbf{z}))\psi(\mathbf{z})$$

em que $\psi(\mathbf{z})$ é uma função que relaciona o vector de covariáveis \mathbf{z} com a sobrevivência de uma empresa.

Em Economia, é frequente encontrarem-se situações que se adequam mais a um tipo de modelos do que outros. Na análise de sobrevivência este aspeto também ocorre,

sendo necessário verificar qual o tipo de modelos, PH ou AFT, é o mais adequado ao problema em causa.

b) Estimação dos modelos

De acordo com Manjón-Antolín e Arauzo-Carod (2008), uma vez escolhido o modelo base mais apropriado, PH ou AFT, procede-se à respetiva estimação. Subentendendo-se a existência de um modelo de regressão procede-se a uma estimação paramétrica ou semi-paramétrica através do modelo de máxima verosimilhança. Existe, ainda, uma terceira vertente de estimação que é não paramétrica (Rocha e Papoila, 2009).

Começando pela estimação não paramétrica da função de sobrevivência, esta é feita através de: tabelas de mortalidade, que representam dados relativos à sobrevivência de um coorte (do inglês *cohort*)⁵, e através das quais se pode estimar a “probabilidade condicional de morte num intervalo de tempo, dada a sobrevivência no início desse intervalo e, a probabilidade de sobrevivência para além de um dado intervalo” (Rocha e Papoila, 2009, p. 31); ou pelo estimador de Kaplan-Meier que estima a função de sobrevivência em t através da proporção de empresas que sobreviveram para além de t .⁶

A estimação paramétrica permite a aplicação direta de métodos de inferência estatística na estimação dos parâmetros do vetor β , vetor constituído por coeficientes de regressão que exprimem a influência das covariáveis na sobrevivência das empresas, assumindo uma distribuição estatística particular para o tempo de sobrevivência. Para tal, recorre-se à estimação por máxima verosimilhança.

Na estimação semi-paramétrica, o modelo de Cox (1972) (cfr Rocha e Papoila, 2009) tornou-se amplamente reconhecido e utilizado no âmbito da análise de sobrevivência devido à sua flexibilidade. Este modelo define-se como semi-paramétrico uma vez que tem a sua função risco constituída por duas partes: a primeira parte que corresponde à função de risco subjacente, $h_0(t)$, que não é especificada; e a segunda

⁵ *Cohort* representa no caso dos estudos de sobrevivência de empresas um grupo de empresas criadas no mesmo instante temporal, seguidas durante o estudo e para as quais é construída a tabela de mortalidade.

⁶ Kaplan-Meier (1958) sugeriram um estimador não paramétrico da função de sobrevivência quando existe censura.

que parametriza o efeito das covariáveis na sobrevivência das empresas. Assim, define-se a função de risco da seguinte forma:

$$(2.5) \quad h(t; \mathbf{z}) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}) = h_0(t) \exp(\beta_1 z_1 + \dots + \beta_p z_p)$$

em que β_1, \dots, β_p representam os coeficientes de regressão a estimar que exprimem a influência das variáveis independentes na sobrevivência das empresas; e $h_0(t)$ corresponde à função risco da empresa com o vector $\mathbf{z} = \mathbf{0}$ ou a função associada a certas condições padrão.

Como as funções de risco de morte de empresas com valores diferentes nas variáveis explicativas não depende de t , o modelo de Cox corresponde a um modelo de riscos proporcionais (PH). Caso se especificasse $h_0(t)$ obter-se-ia um modelo PH paramétrico.

Com algumas operações matemáticas verifica-se que o modelo de Cox corresponde a um modelo linear para o logaritmo do risco relativo, ou seja:

$$(2.6) \quad \log\left(\frac{h(t; \mathbf{z})}{h_0(t)}\right) = \beta_1 z_1 + \dots + \beta_p z_p$$

Assim, $\exp(\beta_j)$ é interpretado como o risco relativo de ocorrência do acontecimento de interesse, morte de uma empresa, que difere em uma unidade na variável z_j sendo tudo o resto igual à empresa com o vector $\mathbf{z} = \mathbf{0}$ /condições padrão.

Cox formulou uma função de verosimilhança que, uma vez não dependendo de $h_0(t)$, tornou possível a inferência estatística sobre $\boldsymbol{\beta}$. Assim, tem-se:

$$(2.7) \quad L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_{(i)})}{\sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{z}_l)}$$

em que, $\mathbf{z}_{(i)}$ é o vector de covariáveis associado à empresa que “morre” em $t_{(i)}$; e R_i o conjunto de risco em $t_{(i)}$.

Representando esta função de forma mais usual, isto é, espelhando a probabilidade de ocorrência de um acontecimento observável, tem-se:

$$(2.8) \quad L[\boldsymbol{\beta}, h_0(t)] = \prod_{i=1}^n [h_0(t_i) \exp(\boldsymbol{\beta}' \mathbf{z}_i) S_0(t_i) \exp(\boldsymbol{\beta}' \mathbf{z}_i)]^{\delta_i} [S_0(t_i) \exp(\boldsymbol{\beta}' \mathbf{z}_i)]^{1-\delta_i} = \prod_{i \in D} \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_i)}{\sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{z}_l)} \prod_{i \in D} (h_0(t_i) \sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{z}_l)) \prod_{i=1}^n S_0(t_i) \exp(\boldsymbol{\beta}' \mathbf{z}_i)$$

em que D é o conjunto de empresas para as quais se observa a ocorrência do acontecimento de interesse.

c) Extensões

No seguimento de Manjón-Antolín e Arauzo-Carod (2008) são apresentadas algumas extensões que se relacionam com a consideração de tempo discreto, incorporação da heterogeneidade não observada e existência de riscos competitivos.

Manjón-Antolín e Arauzo-Carod (2008) expõem que a consideração de tempo discreto relaciona-se com a natureza dos dados económicos que, normalmente, são agrupados em períodos discretos mesmo que a natureza subjacente se encontre em tempo contínuo. Este é o caso dos dados correspondentes à sobrevivência das empresas. Contudo, apesar desta evidência, muitas vezes assume-se tempo contínuo ou, então, recorre-se à opção de não assumir explicitamente qual a natureza considerada.

Manjón-Antolín e Arauzo-Carod (2008) explicam, também, o conceito de heterogeneidade não observada. Este prende-se com a dificuldade de incluir nas variáveis independentes toda a variação verificada na sobrevivência das empresas. Há fatores que não são observáveis e, portanto, não são captados pelas variáveis que se pretende que expliquem a sobrevivência das empresas. Isto provoca problemas na qualidade dos resultados estimados. Contudo, os autores que incorporam esta questão na sua estimação não encontram, geralmente, diferenças substanciais face à omissão da mesma.

O conceito de riscos competitivos prende-se com a existência de mais do que uma forma de ocorrência do acontecimento de interesse, isto é, a morte (Manjón-Antolín e Arauzo-Carod, 2008; Rocha e Papoila, 2009). No âmbito da análise de sobrevivência associada às empresas esta questão está presente através da existência de diversas formas de sair do mercado, sendo pertinente substituir o modelo “*single-spell*” utilizado por uma especificação “*multiple-spell*” (Manjón-Antolín e Arauzo-Carod, 2008). Contudo, a literatura tem-se concentrado mais no fenómeno de saída do mercado do que na consideração de caminhos alternativos para essa saída (Manjón-Antolín e Arauzo-Carod, 2008).

Nesta dissertação, e no âmbito do modelo de Cox, são apresentadas outras extensões no Capítulo 4 devidamente explicitadas e justificadas.

Capítulo 3: Espaço e sobrevivência de empresas

Como foi anteriormente referido, as variáveis espaciais têm influência na sobrevivência das empresas apresentando-se, a par de fatores específicos da indústria e de fatores relacionados com o ciclo económico, como determinantes externos à sobrevivência das empresas.

Verifica-se contudo que a influência dos determinantes espaciais na sobrevivência não é clara, o que pode ser explicado, em parte, pelo facto dos determinantes espaciais possuírem, por parte da investigação científica, um tratamento menos exaustivo que outras áreas da análise de sobrevivência. Manjón-Antolín e Arauzo-Carod (2008) corroboram esta afirmação, referindo que os determinantes espaciais afetam o desempenho das empresas (Fujita *et al.*, 1999 cfr Manjón-Antolín e Arauzo-Carod, 2008) ressaltando, porém, que a evidência empírica apresenta resultados inconclusivos nesta matéria.

Analisando alguns artigos que a literatura apresenta, é possível identificar diversas variáveis que se acomodam na categoria de determinantes espaciais, sendo o seu impacto na sobrevivência das empresas diverso. Na Tabela 1, é apresentada uma síntese de vários estudos que introduzem a dimensão espacial no estudo da sobrevivência das empresas, enfatizando os principais objetivos, a amostra utilizada, a metodologia aplicada, as variáveis (dependente e independentes) e o seu efeito estimado, bem como as principais conclusões.

Como é possível observar, de entre as metodologias utilizadas (análise não paramétrica, modelo de Cox, regressão logística, modelo AFT paramétrico, modelo de duração de Prentice-Gloeckler, modelo de risco exponencial e modelo de variação geográfica), a mais frequente é o modelo de Cox, descrita com detalhe na secção 2.2. A maioria dos estudos mede a sobrevivência através do período de sobrevivência/idade das empresas, seja em meses ou em anos (Fotopoulos e Louri, 2000; Tödtling e Wanzenböck, 2003; Boschma e Wenting, 2007; Wennberg e Lindqvist, 2010), existindo ainda alguns que recorrem a variáveis binárias (Shaver e Flyer, 2000; Staber, 2001), taxas de sobrevivência/morte (Acs *et al.*, 2007; Strotmann, 2007) e probabilidade de sobrevivência (Huiban, 2011). Quanto às variáveis explicativas, os estudos incorporam várias opções, sendo dado destaque na análise que se segue às variáveis espaciais.

Tabela 1 – Espaço e sobrevivência das empresas: estudos empíricos

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Fotopoulos e Louri (2000)	Analisar a influência da localização na sobrevivência das empresas, comparando a área urbana de Atenas com o resto da Grécia	209 empresas do sector da indústria transformadora; Grécia; coorte de 1982, 1983 e 1984 até 1992	Análise não paramétrica; Modelo de Cox com <i>time-varying covariates</i>	Tempo de vida de uma empresa /sobrevivência em anos (os efeitos estimados analisam a influência na taxa de risco/morte das empresas)	- <i>SIZE</i> : dimensão em termos de emprego (a variar com o tempo)	(-)	As empresas localizadas na área urbana de Atenas têm vantagens de sobrevivência face às empresas localizadas no resto do país. Isto é particularmente relevante em empresas de pequena dimensão.	
					Variáveis específicas da empresa	- <i>GROWTH</i> : crescimento do emprego		(-)
						- <i>PROFIT</i> : rácio do lucro líquido sobre o total de ativos (a variar com o tempo)		(-)
						- <i>LEVERAGE</i> : rácio entre passivos de médio e longo prazo sobre o ativo total (a variar com o tempo)		(+)
						- <i>CAPITAL</i> : rácio dos ativos fixos sobre o total de ativos (a variar com o tempo)		(-)
					Variável específica da indústria	- <i>SUNK</i> : proxy de custos afundados, definida como 1 menos o rácio de máquinas e equipamento comprados em segunda mão sobre o total de investimento em máquinas e equipamento (a variar com o tempo)		(-)
					Variável espacial	- <i>ATHENS</i> : 1 se a empresa é localizada em Atenas, 0 caso contrário		(-)
Interação dimensão e localização	- <i>INTERACTION</i> : produto entre a variável espacial <i>ATHENS</i> e uma <i>dummy</i> que tem valor 1 se a empresa tem menos de 30 trabalhadores	(-)						
Momento de entrada – coorte <i>dummies</i> (categoria de referência: 1984)	- <i>COHORT_82</i> : 1 se entrou em 1982, 0 caso contrário	(-)						
	- <i>COHORT_83</i> : 1 se entrou em 1983, 0 caso contrário	(-) ⁷						

⁷ Os resultados apresentados são estatisticamente significativos a um nível de significância de pelo menos 10%.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Shaver e Flyer (2000)	Analisar de que forma a aglomeração influencia a decisão de localização e o posterior desempenho (medido através da sobrevivência) dos estabelecimentos	101 estabelecimentos estrangeiros criados de raiz (com origem no Canadá, França, Alemanha, Japão ou Reino Unido) que entraram no sector da indústria transformadora nos EUA em 1987	Modelo Logit condicional para a decisão de localização; Regressão logística binária para a análise de sobrevivência ⁸	<i>SURVIVAL</i> : variável <i>dummy</i> com 1 se o estabelecimento continua em funções/sobrevive até ao final de 1995 e 0 caso contrário	Medida de aglomeração	- <i>PROPORTION</i> : proporção de estabelecimentos industriais no estado em que o estabelecimento se localiza	(-) (-)	A heterogeneidade entre as empresas conduz a diferentes contributos para as economias de aglomeração existindo seleção adversa na escolha de localização. A aglomeração diminui a probabilidade de sobrevivência e este efeito é mais significativo para indústrias geograficamente mais concentradas.
					Outras variáveis: fatores específicos da empresa e da indústria	- <i>INDUSTRY GROWTH</i> : crescimento da indústria (1987-1993)	n.s. n.s.	
						- <i>US- FOREIGN</i> : proporção de empresas dos EUA, por indústria, com operações no estrangeiro em 1987	(+) (+)	
						- <i>FINANCIAL DATA</i> : 1 se a empresa mãe estrangeira tem dados financeiros disponíveis, 0 caso contrário	n.s. n.s.	
						- <i>SUBSIDIARIES</i> : nº de filiais que a empresa estrangeira controla nos EUA em 1986	n.s. n.s.	
					Outras variáveis espaciais	- <i>GEOGRAPHIC CONCENTRATION</i> : medida de concentração geográfica da indústria de Ellison e Glaeser's (1997) em 1987	n.s. (+) ⁹	
						- <i>COUNTRY DUMMY</i> : identifica as entrantes de acordo com as 5 nacionalidades consideradas		

⁸ Nas colunas seguintes são analisados os resultados da segunda metodologia, isto é, aqueles que se referem à sobrevivência dos estabelecimentos.

⁹ A primeira coluna refere-se à estimação utilizando a amostra completa (101 estabelecimentos) enquanto a segunda coluna refere-se à estimação incluindo apenas estabelecimentos em indústrias geograficamente concentradas (56 estabelecimentos). A abreviatura n.s, significa que o resultado não é estatisticamente significativo a um nível de significância de pelo menos 10%.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões
Staber (2001)	Determinar se a proximidade geográfica das empresas, do mesmo sector ou sectores têxteis complementares, tem influência na sua sobrevivência	1213 empresas de roupa de malha; Baden-Württemberg, Alemanha; período 1960-1998	Modelo PH de Cox	<i>Termination</i> : 1 se a empresa acaba antes do fim do período de observação, 0 caso contrário	Proximidade geográfica	- <i>Same-industry cluster</i> : nº de empresas de roupa de malha localizadas a menos de 10 km no ano em que a empresa encerrou ou em 1998 (+)	A localização em <i>clusters</i> especializados diminui a sobrevivência das empresas. Por outro lado, a localização em <i>clusters</i> diversificados com empresas que operam em sectores complementares aumenta a sobrevivência das empresas.
						- <i>Cluster diversity</i> : índice de Herfindahl para empresas em indústrias têxteis relacionadas, a menos de 10 km, no ano em que a empresa encerrou ou em 1998. Este tem o valor 1 se as empresas operam em todos os sectores do ramo têxtil (diversificação), 0 se as empresas operam num só sector (concentração) (-)	
					Especialização dos negócios	- <i>Horizontal specialization</i> : 1 se a empresa está ativa no máximo em dois grupos de produtos, 0 caso contrário n.s. - <i>Vertical specialization</i> : 1 se a empresa participa em apenas uma etapa de valor acrescentado, 0 caso contrário (+)	
					Dimensão do emprego	- <i>Organization size</i> : nº de trabalhadores (em logaritmo) (-)	
					Orientação do mercado	- <i>Exports</i> : variável codificada com 0 se a empresa vende apenas para a Alemanha, com 1 se vende para países adjacentes, com 2 se vende para toda a Europa e com 3 se vende para todo o mundo (-)	
					Influência da localização	- <i>Albstadt location</i> : 1 se a empresa se localiza em Albstadt, 0 caso contrário (-)	
					Influência do período de formação	- <i>Founded before 1946</i> : 1 se a empresa foi criada antes do fim 2ª Guerra Mundial (antes de 1946), 0 caso contrário n.s. ¹⁰	

¹⁰ A abreviatura n.s. significa que o resultado não é estatisticamente significativo a um nível de significância de pelo menos 5%.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Tödtling e Wanzenböck (2003)	Investigar o papel do grau de aglomeração e da estrutura sectorial das regiões nas características, atividade e sobrevivência das empresas	1360 <i>start-ups</i> ; Áustria; período 1990-1997	Regressão de Cox	Período de sobrevivência (em meses)	Nº licenças	- <i>Number of licences</i> : identifica entre 2 categorias: uma licença ou mais do que uma licença	(+)	A mudança no tipo de área não provoca mudanças ao nível da sobrevivência. O impacto do tipo de área é indireto, isto é, através de variáveis como o nº de licenças e o nível de experiência na indústria.
					Género do fundador	- <i>Gender</i> : identifica entre 2 categorias: masculino e feminino	(+)	
					Experiência na indústria	- <i>Industry experience</i> : identifica entre 2 categorias: se tem experiência no sector ou se não tem	(-)	
					Tempo dedicado	- <i>Proportion of time devoted by founder to start-up</i> : identifica entre 2 categorias: <i>part-time start-up</i> e <i>full-time start-up</i>	(-)	
Acs <i>et al.</i> (2007)	Determinar a influência do capital humano regional e das externalidades do conhecimento na sobrevivência das empresas	Sector dos serviços; EUA - 394 <i>Labour Market Areas</i> ; 1990-1992 (período de recessão) e 1993-1995 (período de crescimento)	Modelo de variação geográfica com estimação OLS	Taxa de sobrevivência de empresas no sector dos serviços	Capital Humano	- Diploma universitário (% de adultos com habilitação universitária em 1990) - Desistências no ensino secundário (% de adultos sem habilitações de ensino secundário em 1990)	(-) (+) (-) (-)	O capital humano regional influencia positivamente a taxa de sobrevivência das empresas no período 1993-1995. A especialização da região no sector dos serviços influencia negativamente a taxa de sobrevivência.
					Externalidades do conhecimento	- Nº de estabelecimentos do sector dos serviços na região por cada 1000 habitantes na região - Nº de estabelecimentos do sector privado na região por cada 1000 habitantes na região	(-) (-) (+) (+)	
					Características Regionais	- Crescimento populacional - Crescimento do rendimento <i>per capita</i>	(-) (-) (+) (+)	
						- Logaritmo da população - Taxa de desemprego - Dimensão do estabelecimento	(+) (+) (+) (+) ¹¹	

¹¹ A primeira coluna corresponde ao período 1990-1992 e a segunda coluna ao período 1993-1995. Os resultados apresentados são estatisticamente significativos a 5%.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Boschma e Wenting (2007)	Investigar se dinâmicas de <i>spinoff</i> , as economias de aglomeração e o momento de entrada na indústria têm impacto na sobrevivência das empresas	308 empresas do sector automóvel; Grã-Bretanha; período 1895-1968	Modelo PH de Cox	Idade da empresa como <i>proxy</i> da sua <i>performance</i> (os efeitos estimados analisam a influência na taxa de risco/morte das empresas)	Economias de localização em indústrias relacionadas	- RREIN: mede para cada empresa, no seu ano de entrada, o nº de pessoas empregadas em indústrias relacionadas na região, em escala logarítmica	(-)	A proximidade com indústrias relacionadas afeta positivamente a sobrevivência, ao contrário da concorrência local. Empresas que entraram entre 1895-1906 (primeiras entrantes) possuem vantagens de sobrevivência. A existência de experiência passada, nomeadamente de sucesso, tem impacto positivo na sobrevivência
					Economias de urbanização	- RPOPU: mede para cada empresa, no seu ano de entrada, a população da região, em escala logarítmica	n.s.	
					Economias de localização no mesmo sector	- RCOMP: mede para cada empresa, no seu ano de entrada, o nº empresas do sector automóvel na região	(+)	
					Identificação do momento de entrada	- ENTR1: 1 se entrou entre 1895-1906	(-)	
						- ENTR2: 1 se entrou entre 1907-1919	n.s.	
					Cenário pré-entrada das empresas	- EXPEF: 1 se empresa é experiente, 0 caso contrário	(-)	
						- SPINOF: 1 se a empresa é “ <i>spinoff</i> ”, 0 caso contrário	n.s.	
	- YRPAR: nº anos de produção da empresa-mãe	(-) ¹²						

¹² A abreviatura n.s. significa que o resultado não é estatisticamente significativo a um nível de significância de pelo menos 5%.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Strotmann (2007)	Determinar o impacto de determinantes específicos da empresa, da indústria, da aglomeração e do ciclo de negócios na sobrevivência dos estabelecimentos	2605 estabelecimentos com menos de 50 trabalhadores a operar na indústria transformadora; Baden-Württemberg (Alemanha); período de 1981-1994	Modelo PH de Cox; Modelo AFT paramétrico; Modelo de duração de Prentice-Gloeckler	Taxa de risco/morte das empresas	Dimensão	- Dimensão da <i>start-up</i> : logaritmo do emprego nas <i>start-up</i> - (Dimensão da <i>start-up</i>) ² : utilizada para aferir não linearidades	(-) (+)	Através da variável grau de aglomeração (em que 1 corresponde a uma área rural e 0 a uma área urbana), verifica-se que há maior probabilidade de morte das <i>start-ups</i> localizadas em áreas urbanas do que em áreas rurais.
					Tipo de estabelecimento	- 1 se estabelecimento único, 0 caso contrário (multi-estabelecimento)	(-)	
					Características da indústria	- Crescimento da indústria: taxa média de crescimento anual do volume de negócios do sector - Concentração da indústria: logaritmo do índice de Herfindahl	(-)	
						- Dimensão da indústria: logaritmo do nº inicial de estabelecimentos no mercado - Taxa de entrada na indústria: nº de <i>start-ups</i> ou a proporção destas no nº total de empresas - Economias de escala da indústria: <i>proxy</i> de Comanor-Wilson (1967)	(-) n.s. (+)	
						- Heterogeneidade da indústria: logaritmo da taxa de excesso de trabalho	n.s.	
					Condições tecnológicas	- Regime tecnológico: 1 se <i>high/medium tech</i> , 0 caso contrário	n.s.	
					Impacto da aglomeração	- Grau de aglomeração: 1 se a empresa se localiza numa área rural, 0 se a empresa se localiza numa área urbana	(-)	
					Influência do ciclo de negócios	- Coorte 1982: 1 se coorte de 1982, 0 caso contrário - Coorte 1983: 1 se coorte de 1983, 0 caso contrário - Coorte 1984: 1 se coorte de 1984, 0 caso contrário	n.s. (-) n.s. ¹³	

¹³ A abreviatura n.s. significa que o resultado não é estatisticamente significativo a um nível de significância de pelo menos 5%. Apenas os resultados da estimação do modelo PH de Cox são apresentados.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Wennberg e Lindqvist (2010)	Analisar o efeito dos <i>clusters</i> na sobrevivência das novas empresas	3799 empresas na área das telecomunicações, equipamento eletrónico, serviços financeiros, tecnologia de informação, equipamento médico e indústrias farmacêuticas; Suécia; período entre 1993-2002	Modelo de risco exponencial	<i>Survival</i> : o tempo decorrido entre o início e a morte da empresa (os efeitos estimados analisam a influência na taxa de risco/morte das empresas)	Forma legal	- Forma legal = incorporação: <i>dummy</i> que identifica se a empresa começou como incorporação	(-)	A localização num <i>cluster</i> tem efeitos positivos na sobrevivência das empresas.
					Economias de urbanização	- Densidade populacional: nº de habitantes por km ² na região	(-)	
						- Emprego noutros sectores: emprego total da região exceto o emprego no <i>cluster</i> (logaritmo)	n.s.	
					Diferenças regionais no custo de vida	- Universidades locais: nº instituições de pesquisa médica, universidades, escolas técnicas e escolas de negócios na região em cada ano	n.s.	
						- Índice de preços da habitação (logaritmo)	n.s.	
					Capital humano	- Capital humano das empresas: proporção de trabalhadores com diploma universitário por empresa	(-)	
						- Capital humano especial/específico: proporção de trabalhadores com diploma de engenharia ou ciências por empresa	(-)	
Emprego	- Trabalhadores (logaritmo)	(-)						
	- Emprego no <i>cluster</i> (logaritmo)	(-) ¹⁴						

¹⁴ A abreviatura n.s. significa que o resultado não é estatisticamente significativo a um nível de significância de pelo menos 5%.

Autor(es)	Principal objetivo	Amostra	Metodologia	Variável Dependente	Variáveis Explicativas	Efeito Estimado	Conclusões	
Huiban (2011)	Investigar as diferenças de sobrevivência e criação de estabelecimentos entre áreas rurais, periurbanas, urbanas e a região de Paris	1 025 252 estabelecimentos; França; período 1993-2002	Regressão logística	Probabilidade de sobrevivência do estabelecimento i	- IDF: 1 se o estabelecimento se localiza em Paris, 0 caso contrário	(-)	Nas áreas rurais e periurbanas, a probabilidade de sobrevivência é superior. Nas áreas urbanas, e especialmente em Paris a probabilidade de sobrevivência é menor. A criação de estabelecimentos, porém, é mais fácil em regiões urbanas sobretudo Paris.	
					Localização do estabelecimento	- PERI: 1 se o estabelecimento se localiza numa área periurbana, 0 caso contrário		(+)
						- RURAL: 1 se o estabelecimento se localiza numa área rural, 0 caso contrário		(+)
					Dimensão	- SIZE: nº de trabalhadores iniciais		(+)
					Intensidade da concorrência local	- LCOMP: logaritmo do inverso do índice de Herfindahl regional		(-) ¹⁵
	Influência por indústria	- Indústria: <i>dummies</i> associadas aos vários sectores da indústria						

¹⁵ Os resultados apresentados são estatisticamente significativos a um nível de significância de pelo menos 5%.

Da tabela anterior pode identificar-se um conjunto de estudos que têm por base o conceito de economias de aglomeração bem como as externalidades subjacentes a estas e a sua influência na sobrevivência das empresas. Staber (2001) e Acs *et al.* (2007) introduzem a dicotomia especialização/diversificação de uma região; Shaver e Flyer (2000) relacionam economias de aglomeração e escolha de localização com a heterogeneidade das empresas; Boschma e Wenting (2007) adicionam os conceitos de *spinoff*, economias de localização e economias de urbanização; e, por fim, Wennberg e Lindqvist (2010) clarificam a noção de *cluster*.

Acs *et al.* (2007) procuram no seu estudo compreender o papel do capital humano regional e das externalidades do conhecimento na sobrevivência das empresas recorrendo a um modelo de variação geográfica com estimação OLS. Desta forma, analisam se uma região beneficia do facto de possuir melhor capital humano e se a partilha do conhecimento, e consequentemente a especialização ou diversificação da região, possui vantagens para as empresas em termos de sobrevivência. Os autores encontram como principais resultados que o capital humano regional influencia positivamente a taxa de sobrevivência das empresas americanas no período de crescimento de 1993-1995. Contudo, a especialização da região no sector dos serviços influencia negativamente a taxa de sobrevivência sugerindo, ainda, que a diversificação detém um papel importante na sobrevivência das empresas.

Staber (2001) analisa a proximidade geográfica das empresas e a sua sobrevivência dando, também, relevância à especialização/diversificação das empresas aglomeradas. A diversificação, aqui, ocorre quando as empresas operam em mais do que uma atividade do sector têxtil (atividades complementares). Para tal, o autor investiga a sobrevivência das empresas de roupa de malha no distrito industrial de Baden-Württemberg.¹⁶ Os principais resultados alcançados indicam que a localização em *clusters* especializados diminui a sobrevivência das empresas, enquanto a localização em *clusters* diversificados com empresas que operam em sectores complementares aumenta a sobrevivência das empresas.

Porter (1998) (cfr Shaver e Flyer, 2000) clarifica o conceito de economias de aglomeração, isto é, externalidades positivas que resultam da concentração geográfica de empresas e que incentivam que estas, com o objetivo de incrementar e melhorar o

¹⁶ Sinteticamente, o distrito industrial é definido como um local onde se concentram empresas, normalmente de pequena dimensão e especializadas num sector e sectores relacionados (Staber, 2001).

seu desempenho, preferiram localizar-se em proximidade com outras empresas. De acordo com Marshall (1920) (cfr Shaver e Flyer, 2000), estas externalidades podem advir da partilha de conhecimento entre empresas, da criação de uma área de trabalho ou fornecedores especializados. Shaver e Flyer (2000), contudo, defendem que esta relação recíproca de externalidades entre as empresas nem sempre é proporcional e relacionam a diferença dos benefícios conseguidos na aglomeração com a heterogeneidade das empresas. Assim, as “melhores” empresas, as mais eficientes, ganham menos com as externalidades da aglomeração do que as “piores”, as menos eficientes. Isto acontece porque as empresas que possuem, por exemplo, melhores tecnologias, fornecedores ou capital humano contribuem muito mais para o efeito externalidade que os seus concorrentes com condições mais fracas. Ao verem que o seu contributo é muito mais significativo do que os benefícios que podem retirar, as melhores empresas optam, desta forma, por não se aglomerarem. Assim, os autores concluem que a heterogeneidade entre as empresas conduz à existência de seleção adversa na escolha de localização. Para além da escolha de localização, Shaver e Flyer (2000) estudam a sobrevivência das empresas concluindo que a aglomeração diminui a probabilidade de sobrevivência sendo este efeito mais significativo em indústrias geograficamente mais concentradas.

Boschma e Wenting (2007), paralelamente ao trabalho de Klepper (2007), analisam os efeitos da aprendizagem entre as empresas, nomeadamente os processos de *spinoff* e as economias de aglomeração, sobre a sobrevivência das empresas da indústria automóvel na Grã-Bretanha, recorrendo ao modelo de Cox. Os autores definem *spinoff* como uma empresa cujo fundador possui experiência anterior no sector. De acordo com Arthur (1994) e Klepper (2002) (cfr Boschma e Wenting, 2007), as dinâmicas subjacentes a esta categoria de empresas pressupõem que há transferência de experiência e rotinas entre a empresa-mãe e o *spinoff*. Por outro lado, o mecanismo de economias de aglomeração permite a transferência de conhecimento entre empresas numa determinada região. As economias de aglomeração podem ser divididas em duas categorias (Hoover, 1948 cfr Boschma e Wenting, 2007). A primeira, economias de localização, refere-se a externalidades conseguidas devido à concentração geográfica de empresas do mesmo sector ou sectores relacionados. A segunda categoria, economias de urbanização, corresponde a externalidades alcançadas com a concentração geográfica

de empresas, independentemente do sector em que operam. Os autores verificam que, ao contrário da concorrência local, a proximidade com indústrias relacionadas e a existência de experiência passada, nomeadamente de sucesso, possui influência positiva na sobrevivência.

Wennberg e Lindqvist (2010) investigam o papel dos *clusters* na sobrevivência das empresas suecas recorrendo a um modelo de risco exponencial. Um *cluster* é definido como concentração de empresas relacionadas no mesmo espaço geográfico (Porter, 1998 cfr Wennberg e Lindqvist, 2010). Os autores defendem que os *clusters* estão associados a taxas de sobrevivência mais elevadas e melhores desempenhos atraindo, desta forma, a atividade empreendedora.

Da análise dos estudos apresentados, é possível concluir que as economias de localização têm, segundo alguns autores, um impacto negativo na sobrevivência das empresas (Staber, 2001; Acs *et al.* 2007; Boschma e Wenting, 2007), enquanto as economias de localização que resultam de sectores relacionados possuem influência positiva na sobrevivência (Staber, 2001; Boschma e Wenting, 2007, Wennberg e Lindqvist, 2010). Por seu lado, os estudos apresentados evidenciam uma influência positiva das economias de urbanização na sobrevivência (Acs *et al.* 2007, Wennberg e Lindqvist, 2010).

A Tabela 1 inclui, também, um outro conjunto de estudos em que se considera a influência da localização em regiões com diferentes características na sobrevivência das empresas. Começando pela dicotomia entre espaço rural e espaço urbano sugerida por Strotmann (2007), estende-se a análise a outros artigos que sugerem diferentes divisões do espaço. Huiban (2011) distingue entre áreas urbanas, áreas periurbanas, áreas rurais e a região de Paris; Fotopoulos e Louri (2000) distinguem entre a área urbana de Atenas e as restantes regiões da Grécia; e, por fim, Tödtling e Wanzenböck (2003) consideram sete áreas distintas da Áustria: região de Viena, outros centros terciários, áreas industriais, áreas industriais antigas, áreas rurais industrializadas, áreas turísticas e áreas rurais.

Strotmann (2007) estuda o efeito da aglomeração regional na sobrevivência das empresas através dos modelos PH de Cox, AFT paramétrico e duração de Prentice-Gloeckler. Assim, através de uma variável *dummy* que mede o grau de aglomeração, distingue entre áreas rurais e urbanas na região de Baden-Württemberg (Alemanha),

concluindo que empresas que se localizam em áreas rurais possuem maiores taxas de sobrevivência. De facto, áreas densamente povoadas e com salários mais elevadas poderão dificultar a sobrevivência das novas empresas. Porém, a proximidade a clientes e fornecedores e a disponibilidade de trabalhadores mais qualificados podem conduzir a um efeito oposto, melhorando a sobrevivência nas áreas urbanas.

Huiban (2011), por sua vez, prolonga esta classificação e identifica quatro categorias de regiões francesas onde os estabelecimentos se podem localizar: áreas urbanas (aquelas que possuem pelo menos 5000 postos de trabalho), áreas periurbanas (aquelas em que pelo menos 40% dos moradores trabalham em áreas urbanas), áreas rurais (que não entram nas duas categorias anteriores) e a região de Paris. O autor mostra que a criação de estabelecimentos é mais propícia em regiões concentradas, nomeadamente as áreas urbanas e, em especial, a região de Paris, uma vez que estas beneficiam de externalidades associadas à aglomeração como é o caso do trabalho mais qualificado. Todavia, apesar de as taxas de entrada serem mais elevadas nestas áreas, a sobrevivência é mais difícil devido, sobretudo, à maior intensidade concorrencial que caracteriza estas regiões.

Fotopoulos e Louri (2000) testam, também, a influência da capital na sobrevivência comparando as empresas localizadas na área urbana de Atenas com as empresas situadas nas restantes regiões da Grécia. Ao contrário de Huiban (2011), estes autores encontram uma influência positiva da capital na sobrevivência das empresas.

Tödting e Wanzenböck (2003) analisam a influência das características sectoriais das regiões nas características, atividade e sobrevivência das empresas. Para tal, distinguem sete tipos de regiões austríacas: região de Viena, outros centros terciários, áreas industriais, áreas industriais antigas, áreas rurais industrializadas, áreas turísticas e áreas rurais. Apesar do modelo de Cox apontar para a não influência do tipo de área geográfica, será de esperar taxas de sobrevivência das empresas mais baixas na região de Viena, nas antigas áreas industriais e nas áreas rurais industrializadas.

Conclui-se assim que os estudos apresentados mostram a divisão do espaço por áreas com diferentes características e a sua influência na sobrevivência das empresas. Alguns autores concluem que a probabilidade de morte é maior em áreas urbanas do que em áreas rurais (Tödting e Wanzenböck, 2003; Strotmann, 2007; Huiban, 2011); contudo, outros autores encontram uma conclusão contrária (Fotopoulos e Louri, 2000).

Após estes três capítulos de revisão de literatura passa-se, no capítulo seguinte, à análise estatística e econométrica da sobrevivência que visa dar resposta às questões de investigação formuladas.

Desta forma, são utilizados os conceitos fornecidos nesta primeira fase de trabalho, nomeadamente os determinantes e metodologias de sobrevivência, com vista a analisar a influência do espaço na sobrevivência dos estabelecimentos em Portugal. Uma vez mais, é de sublinhar a importância das variáveis espaciais tratadas no decorrer deste capítulo, visto que é na análise destas que reside o objetivo central desta dissertação.

Capítulo 4: A influência do espaço na sobrevivência dos estabelecimentos em Portugal

Pretende-se, neste capítulo, responder à seguinte questão: em que medida o espaço tem influência na sobrevivência dos estabelecimentos em Portugal?

A revisão de literatura efetuada nos capítulos anteriores permitiu identificar os principais determinantes que influenciam a sobrevivência das empresas bem como as metodologias utilizadas no estudo da sobrevivência. As variáveis identificadas agrupam-se em dois grupos: determinantes internos à empresa (ou estabelecimentos) e determinantes externos à empresa (ou estabelecimentos), onde se incluem, entre outros, as variáveis de natureza espacial. Dos determinantes apresentados, são selecionados os mais relevantes e mais comumente utilizados de forma a serem integrados como variáveis explicativas de um modelo que pretende explicar a sobrevivência dos estabelecimentos em Portugal.

De seguida, são apresentadas e sustentadas as principais escolhas quanto ao modelo e variáveis. Por fim, efetua-se a estimação e apresentam-se os principais resultados.

4.1. Variável dependente

A sobrevivência dos estabelecimentos no período 1995-2006 é determinada com base nos *Quadros de Pessoal* (Gabinete de Estratégia e Planeamento do Ministério do Trabalho e da Solidariedade Social (GEP-MTSS, 1995-2006), Portugal). Esta base de dados resulta de um inquérito obrigatório realizado anualmente pelo Ministério da Solidariedade e Segurança Social a todas as empresas a operar em Portugal (exceto administração pública, trabalho doméstico ou empresas sem trabalhadores remunerados). O inquérito recolhe informação ao nível da empresa, estabelecimento e trabalhador, de 1985 até 2009, incluindo informação sobre a localização das empresas, sector de atividade, estrutura do capital, número de estabelecimentos, qualificação e remuneração dos trabalhadores, entre outras. De acordo com Mata e Portugal (1994), esta base de dados apresenta a vantagem de permitir o acompanhamento das empresas

durante a sua vida, bem como de constituir o único recurso em Portugal que interliga as empresas com os seus estabelecimentos e trabalhadores.

Para este estudo, a população-alvo corresponde aos estabelecimentos da indústria transformadora (CAE – Rev. 2.1: 15 a 37 (Anexo 1))¹⁷ em Portugal Continental que iniciaram a sua atividade em 1995, considerando-se o seu percurso até 2006 (ou seja, correspondente ao coorte de 1995). Utilizam-se dados anuais, ao nível do estabelecimento, provenientes dos *Quadros de Pessoal* (GEP - MTSS, 1995-2006), estando a escolha do período temporal de análise (1995-2006) relacionada com mudanças registadas ao nível da CAE.¹⁸ Para a adequação da base de dados ao objetivo pretendido, foram excluídos estabelecimentos que tendo iniciado as suas operações em 1995 tiveram uma interrupção na sua atividade (saíram e regressaram à base de dados no período considerado).

De acordo com Mata e Portugal (1994), a sobrevivência de uma empresa (ou estabelecimento) corresponde à sua duração de vida, isto é, ao tempo decorrido entre o seu nascimento e a sua morte. Neste estudo, os nascimentos dizem respeito aos estabelecimentos criados em 1995¹⁹ e as mortes correspondem à saída permanente dos estabelecimentos da base de dados no período considerado (1995-2006). Há, ainda, uma outra opção que corresponde a estabelecimentos que, tendo iniciado atividade em 1995, se encontravam ainda a operar em 2006. Neste caso, não há ocorrência do acontecimento de interesse, morte, existindo portanto censura à direita.

Por questões de conveniência, considera-se como referência para cada ano da base de dados o primeiro dia desse ano. Ou seja, indicar que um estabelecimento nasceu e morreu em 1995 significa que existia no dia 1 de Janeiro de 1995 e morreu algures entre o dia 1 de Janeiro de 1995 e o dia 1 de Janeiro de 1996.

Feitas estas considerações iniciais, o objetivo é analisar o percurso de 4917 estabelecimentos da indústria transformadora nascidos em 1995 em Portugal Continental até 2006. Sob o ponto de vista de sobrevivência, um estabelecimento pode

¹⁷ Considerou-se a CAE – Rev. 2.1 (Decreto-lei n.º 197/2003, de 27 de Agosto), que corresponde a uma atualização da CAE – Rev. 2 (decreto-lei n.º 112/1993, de 14 de Maio). De forma a controlar possíveis reclassificações entre a CAE – Rev. 2 e a CAE – Rev. 2.1, considerou-se para cada estabelecimento a CAE a dois dígitos correspondente ao último ano em que esse estabelecimento consta na base de dados.

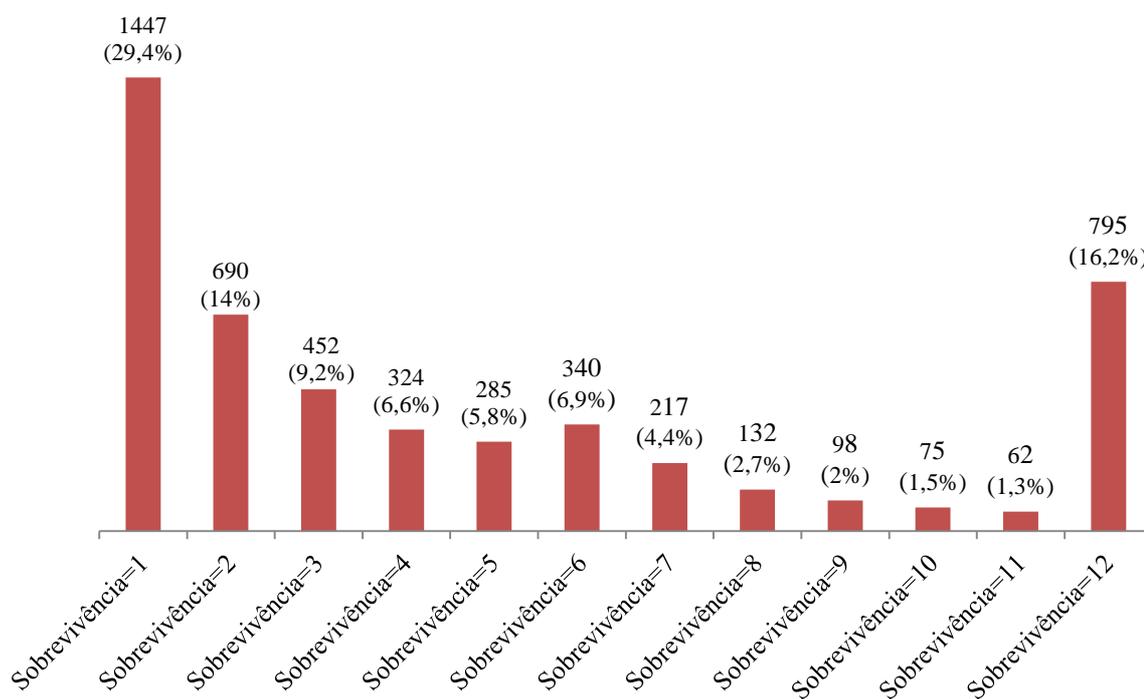
¹⁸ Em Dezembro de 1991 entrou em vigor a CAE – Rev. 2 e em Março de 2007 a CAE – Rev. 3 dificultando, desta forma, a análise para um período de tempo mais alargado.

¹⁹ Considera-se que um estabelecimento nasceu em 1995 caso se tenha encontrado ausente da base de dados desde 1991.

sobreviver entre um mínimo de um ano (nasceu e morreu em 1995) e um máximo de doze anos ou mais (nasceu em 1995 e ainda operava em 2006).

A Figura 1 a seguir representada pretende mostrar a evolução da sobrevivência dos estabelecimentos (com base nos dados constantes do Anexo 2).

Figura 1 – Sobrevivência dos estabelecimentos (em número e em percentagem dos nascimentos totais) (1995-2006)



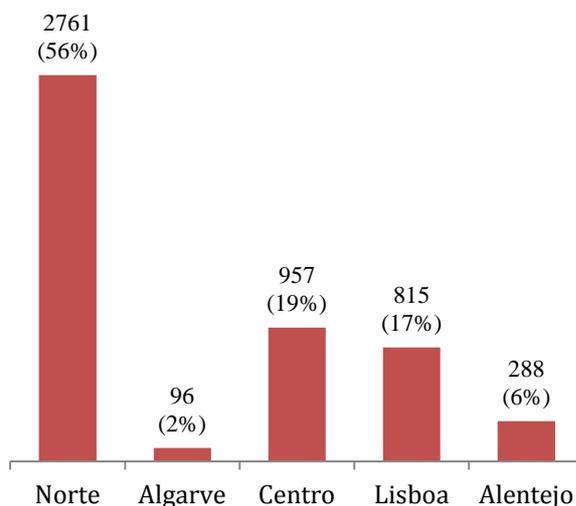
Dos 4917 estabelecimentos nascidos em 1995, 29% (1447 estabelecimentos) sobreviveram apenas um ano, isto é, morreram em 1995; 14% (690 estabelecimentos) sobreviveram dois anos, ou seja, morreram em 1996, e assim sucessivamente. Da análise direta da figura, é ainda possível retirar que a mortalidade diminui com o tempo, isto é, à medida que o período de análise aumenta o número de estabelecimentos que morrem é cada vez menor estando a maior vaga de mortes concentrada nos primeiros anos de vida dos estabelecimentos. Finalmente é de notar que “Sobrevivência=12” corresponde a estabelecimentos para os quais ainda não ocorreu o acontecimento de interesse, a morte, o que significa que ainda operavam em 2006. Logo, podemos

verificar que 16% dos estabelecimentos (795 estabelecimentos) registaram uma sobrevivência de pelo menos 12 anos.

De seguida é efetuada uma análise os nascimentos (Figuras 2 e 3) e da sobrevivência (Figuras 4 e 5) a nível regional (NUTS II) e por sector de atividade (CAE Rev. 2.1 a dois dígitos) (Anexo 1).

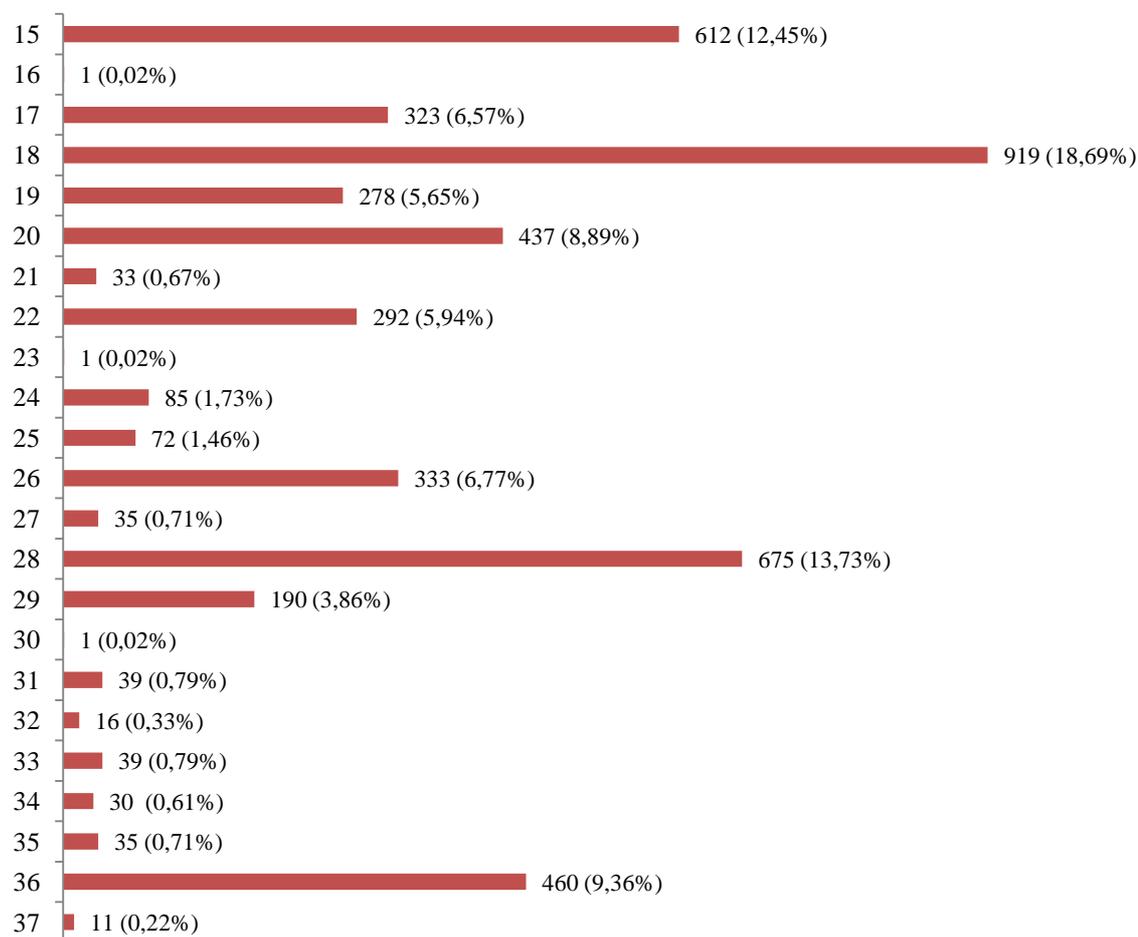
Da Figura 2 é possível retirar que o Norte registou a maior percentagem de nascimentos (56%, correspondente a 2761 estabelecimentos), e o Algarve registou o valor mais baixo (2%, correspondente a 96 estabelecimentos).

Figura 2 – Estrutura dos nascimentos por NUTS II (em número e em percentagem dos nascimentos totais) (1995)



Por sua vez, a Figura 3 ilustra que é a *Indústria do vestuário: preparação, tingimento e fabricação de artigo e peles com pelo* (CAE 18) que regista o maior número de nascimentos de estabelecimentos no ano de 1995 (18,69% dos nascimentos, correspondente a 919 estabelecimentos). No outro extremo encontram-se os sectores da *Indústria do tabaco* (CAE 16), *Fabricação do coque, produtos petrolíferos refinados e tratamento de combustível nuclear* (CAE 23) e *Fabricação de máquinas de escritório e de equipamento para o tratamento automático de informação* (CAE 30) que registaram apenas um nascimento (0,02% dos nascimentos totais).

Figura 3 – Estrutura dos nascimentos por CAE a dois dígitos (em número e em percentagem dos nascimentos totais) (1995)

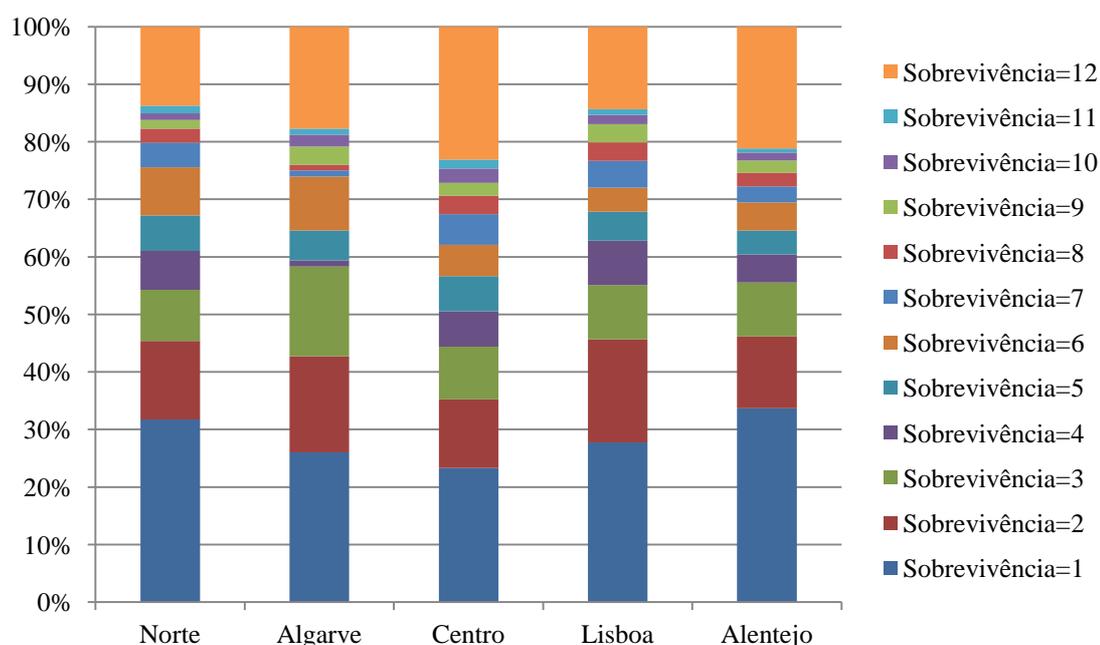


As Figuras 4 e 5 analisam a sobrevivência dos estabelecimentos por NUTS II e CAE a dois dígitos.

Da Figura 4 extrai-se que também a nível regional o número de mortes diminui com o tempo. Note-se que o Centro é, das cinco regiões apresentadas, aquela cuja sobrevivência parece ser mais prolongada. No final de três anos, esta região apresentava, ainda, uma percentagem acumulada de falências de aproximadamente 44%, quando todas as outras regiões já tinham ultrapassado 50%. No final do período de análise, é esta também a região que possui uma maior percentagem de estabelecimentos ainda a operar (23,1%). Por outro lado, o Algarve foi a região com maior percentagem acumulada de falências nos três primeiros anos (cerca de 58%) e o

Norte a região com menor percentagem de estabelecimentos ainda em atividade em 2006 (13,7%).

Figura 4 – Sobrevivência de estabelecimentos por NUTS II (em percentagem dos nascimentos por NUTS II) (1995-2006)



Através do teste generalizado de Mann-Whitney-Wilcoxon e do teste log-rank (Rocha e Papoila, 2009; Cleves *et al.*, 2010) pretende-se aferir quanto à evidência de diferenças regionais na sobrevivência. Desta forma, foi testada a hipótese de igualdade das funções sobrevivência por regiões, ou seja, NUTS II:

$$H_0: S_{Norte}(t) = S_{Centro}(t) = S_{Lisboa}(t) = S_{Alentejo}(t) = S_{Algarve}(t)$$

Os resultados obtidos para o teste de Wilcoxon generalizado (corroborado pelo teste log-rank), apresentado na Tabela 2, indicam que a evidência estatística é favorável à existência de diferenças significativas de sobrevivência por região.²⁰

²⁰ Em alternativa, foram realizados testes de Wilcoxon generalizados e testes log-rank com o objetivo de comparar cada região com o restante território de Portugal Continental. A evidência estatística mostrou que as diferenças regionais que são significativas residem sobretudo na região Norte e na região Centro.

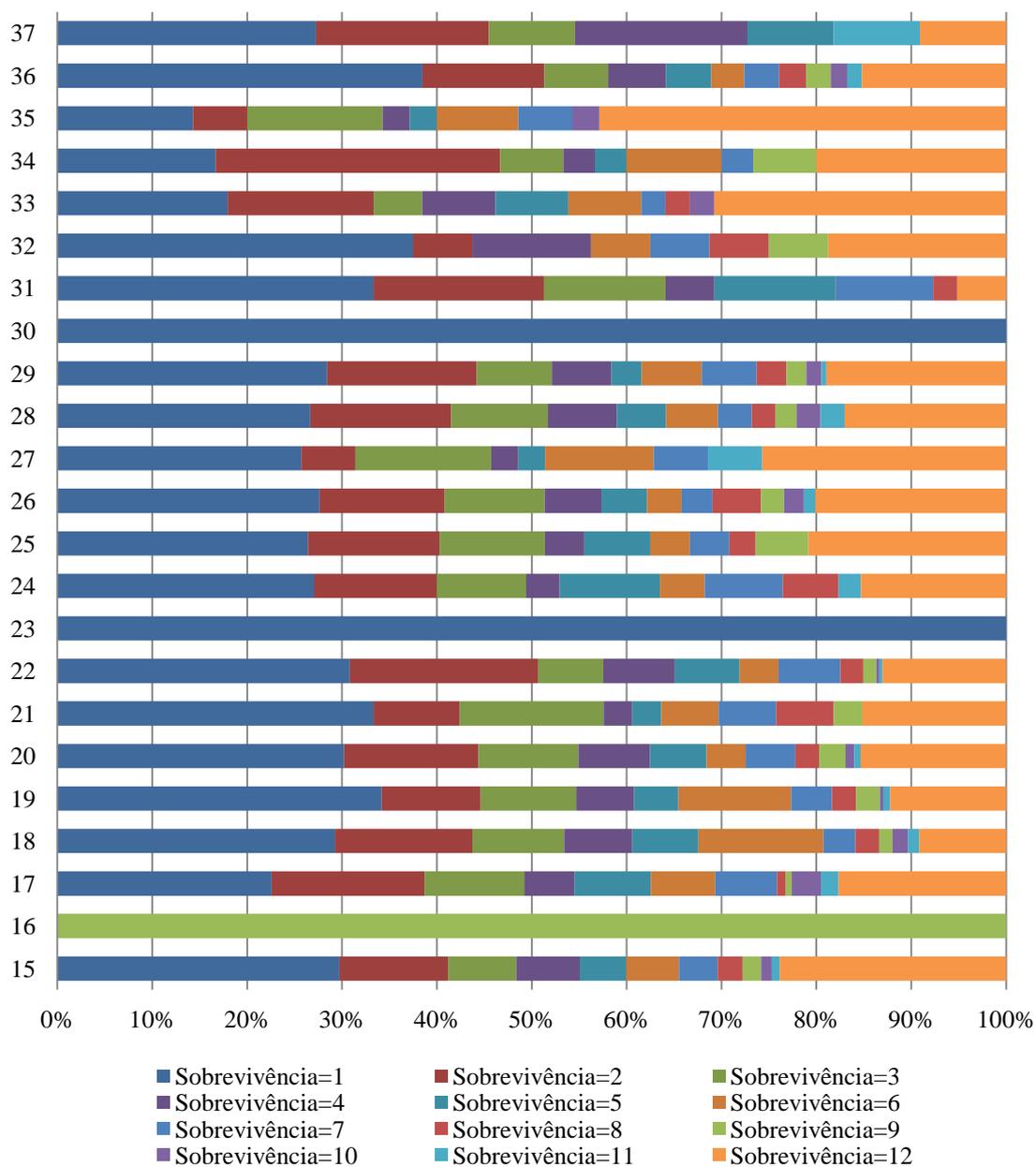
Tabela 2 – Teste de Wilcoxon generalizado para a igualdade das funções sobrevivência por regiões

	Eventos observados	Eventos esperados	<i>Sum of ranks</i>	
1 – Norte	2382	2221,99	489716	Chi ² (4) = 51,95 (Pr>chi2 = 0,0000)
2 – Centro	736	909,08	-558490	
3 – Lisboa	698	668,05	58864	
4 – Alentejo	227	241,97	15454	
5 – Algarve	79	80,91	-5544	
	4122	4122	0	

A Figura 5, por sua vez, evidencia a relação entre o sector de atividade (CAE Rev. 2.1, Anexo 1) e a sobrevivência dos estabelecimentos.

Excluindo da análise os estabelecimentos que apenas registaram um nascimento em 1995, verifica-se que foi no sector identificado com a CAE a dois dígitos 31 (*Fabricação de máquinas e aparelhos elétricos, n. e.*) que se registou uma menor sobrevivência ao fim de três anos (cerca de 58% dos estabelecimentos já tinham falecido nesta altura). De notar, ainda, que apenas os sectores das CAE 15 (*Indústrias alimentares e das bebidas*), 17 (*Fabricação de têxteis*), 24 (*Fabricação de produtos químicos*), 27 (*Indústrias metalúrgicas de base*), 32 (*Fabricação de equipamento e de aparelhos de rádio, televisão e comunicação*), 33 (*Fabricação de aparelhos e instrumentos médico-cirúrgicos, ortopédicos, de precisão, de ótica e de relojoaria*) e 35 (*Fabricação de outro material de transporte*) registaram uma percentagem acumulada de falências ao fim de três anos inferior a 50%, com um mínimo para o sector de *Fabricação de outro material de transporte* com a percentagem acumulada de cerca de 34%. Analisando os estabelecimentos que ainda operavam em 2006, regista-se um máximo no sector com a CAE 35 (*Fabricação de outro material de transporte*), em que cerca de 43% dos estabelecimentos ainda se encontravam em funções em 2006, e um mínimo para o sector com CAE 31 (*Fabricação de máquinas e aparelhos elétricos, n. e.*) com uma percentagem de apenas cerca de 5%.

Figura 5 – Sobrevivência de estabelecimentos por CAE a dois dígitos (em percentagem dos nascimentos por CAE) (1995-2006)



4.2. Descrição do modelo

No capítulo 2 foram apresentadas as principais metodologias utilizadas na análise de sobrevivência de empresas.

Dos métodos expostos na secção 2.2 e confrontando estes com as escolhas dos autores (capítulos 2 e 3), a opção para este estudo recai sobre o modelo de riscos proporcionais (PH) de Cox. De acordo com Rocha e Papoila (2009), a abordagem de Cox é das mais utilizadas em análise de sobrevivência devido à sua especificação inovadora que lhe permite adequar-se a diversas situações. A preponderância do modelo de Cox justifica-se, assim, pela sua flexibilidade. Este modelo de estimação semi-paramétrica define a função risco em duas partes: uma que parametriza o efeito das variáveis explicativas na sobrevivência das empresas e outra, função risco subjacente, que não é especificada. A inferência estatística relativamente aos coeficientes de regressão torna-se possível, uma vez que a estimação se faz através da função de verosimilhança que é independente da função risco subjacente.

Ainda, e considerando o estudo de Manjón-Antolín e Arauzo-Carod (2008) que reúne os principais contributos literários sobre a análise de sobrevivência das empresas, verifica-se uma predominância do modelo de Cox como principal metodologia utilizada: dos vinte e seis estudos sintetizados pelos autores, apenas dois não utilizam a especificação de Cox.

Por fim, tendo em conta os estudos que recorrem a variáveis espaciais para explicar a sobrevivência das empresas (Capítulo 3), conclui-se através da análise direta da Tabela 1 que o modelo de Cox é também a escolha mais recorrente. Esta é a opção de Fotopoulos e Louri (2000), Staber (2001), Tödting e Wanzenböck (2003), Boschma e Wenting (2007) e Strotmann (2007).

Relembrando, o modelo Cox pretende estimar a seguinte função de verosimilhança:

$$(2.9) \quad L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\beta}' \mathbf{z}_{(i)})}{\sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{z}_l)}$$

em que $\boldsymbol{\beta}$ corresponde aos coeficientes de regressão a estimar; $\mathbf{z}_{(i)}$ é o vetor de variáveis explicativas associado ao estabelecimento que morre em t_i , podendo estas estar incluídas na categoria de variáveis espaciais ou na categoria de variáveis de controlo; e R_i o conjunto de risco em $t_{(i)}$.

Para efeitos de estimação é utilizada a aproximação feita por Breslow (1974) (cfr. Rocha e Papoila, 2009):

$$(4.1) \quad L(\beta) = \prod_{i=1}^k \frac{\exp(\beta' s_i)}{[\sum_{l \in R_i} \exp(\beta' z_l)]^{d_i}}$$

em que $s_i = \sum_{j=1}^{d_i} z_{ij}$.

É, ainda, de salientar que esta análise comporta censura à direita uma vez que para alguns estabelecimentos ainda não se verificou o acontecimento de interesse, isto é, a morte, aquando do termo da observação.

Para além do modelo de riscos proporcionais (PH) de Cox são também consideradas algumas extensões ao modelo de Cox.

Em primeiro lugar, introduzem-se algumas *time-varying variables*, isto é, variáveis explicativas cujos valores variam ao longo do tempo. São também introduzidas interações de certas variáveis explicativas como uma função tempo.

Por fim, considera-se a incorporação de efeitos de grupo através da análise do modelo estratificado e de modelos com fragilidade que visam corrigir a correlação entre elementos de um grupo e a heterogeneidade não observada.

4.3. Variáveis explicativas

A revisão de literatura efetuada nos capítulos 2 e 3 permitiu identificar os principais determinantes de sobrevivência das empresas que são, aqui, integrados como variáveis explicativas. A Tabela 4 reúne, de entre os determinantes de sobrevivência analisados, as variáveis que se aplicam a este estudo bem como o seu sinal esperado e estatísticas descritivas. Note-se ainda que as variáveis explicativas apresentadas são definidas ao nível do estabelecimento.

Tabela 3 – Variáveis explicativas

	Variável	Descrição	Sinal esperado (risco de morte)	Estatísticas descritivas			
				Média	Desvio-padrão	Mín.	Máx.
(1)	Dimensão (<i>size</i>)	Número de trabalhadores do estabelecimento em 1995 (em logaritmo)	(-)	1,423	1,046	0	5,986
	Número de estabelecimentos (<i>nrest</i>)	Número de estabelecimentos da empresa a que pertence o estabelecimento em 1995	(-)	1,447	2,895	1	69
(2)	Regime tecnológico (<i>hightech</i>)	<i>Dummy</i> : 1 se <i>high tech</i> , 0 caso contrário, 1995	(+)	0,095	0,293	0	1
	Entrada (<i>entradas</i>)	Número de novos estabelecimentos por ano e por CAE a dois dígitos (em logaritmo), 1995-2006	(+)	6,049	0,863	0	7,536
	Escala mínima eficiente (<i>eme</i>)	Mediana do número de trabalhadores por estabelecimento por CAE a dois dígitos em 1995	(+)	3,659	0,771	3	20
	Dinâmica do sector de atividade (<i>txcres</i>)	Taxa de crescimento média anual por CAE a dois dígitos, 1995-2006	(-)	0,047	0,062	-0,25	0,333
(3)	Economias de localização (<i>ecloc</i>)	Número total de estabelecimentos por CAE a dois dígitos e NUTS III em 1995 (em logaritmo)	(+)	5,470	1,384	0	7,429
	Economias de urbanização (<i>ecurb</i>)	Número total de estabelecimentos da indústria e serviços por NUTS III em 1995 (em logaritmo)	(-)	9,192	0,936	6,561	10,87
(4)	Norte (<i>norte</i>)	<i>Dummy</i> : 1 se estabelecimento se localiza no Norte, 0 caso contrário	?	0,529	0,499	0	1
	Lisboa (<i>lisboa</i>)	<i>Dummy</i> : 1 se estabelecimento se localiza no Lisboa, 0 caso contrário	?	0,160	0,367	0	1
	Centro (<i>centro</i>)	<i>Dummy</i> : 1 se estabelecimento se localiza no Centro, 0 caso contrário	?	0,231	0,422	0	1
	Alentejo (<i>alentejo</i>)	<i>Dummy</i> : 1 se estabelecimento se localiza na Alentejo, 0 caso contrário	?	0,060	0,238	0	1
	Algarve (<i>algarve</i>)	<i>Dummy</i> : 1 se estabelecimento se localiza nos Algarve, 0 caso contrário	?	0,020	0,139	0	1

Da análise da tabela 2, é possível extrair que as variáveis explicativas possuem quatro propósitos distintos. Esses propósitos são os seguintes: (1) estudar a influência das características específicas do estabelecimento (dimensão e estrutura da propriedade) na sobrevivência; (2) analisar o impacto das especificidades da indústria na sobrevivência (regime tecnológico, entrada, escala mínima eficiente e dinâmica do sector de atividade); (3) analisar a influência das economias de aglomeração na sobrevivência dos estabelecimentos (economias de localização e economias de urbanização); (4) avaliar se a sobrevivência dos estabelecimentos está relacionada com a região onde estas se localizam (Norte, Lisboa, Centro, Alentejo e Algarve).

Note-se que as variáveis englobadas no ponto (1) dizem respeito aos determinantes internos à empresa, enquanto as restantes estão incluídas na categoria de determinantes externos à empresa, correspondendo as variáveis englobadas nos pontos (3) e (4) a variáveis espaciais. Repare-se ainda que algumas das variáveis consideradas são inspiradas na Teoria da Empresa (regime tecnológico, as taxas de entrada e a escala mínima eficiente), existindo outras que são abordadas na Ecologia das Organizações (“responsabilidade da pequenez” associada à dimensão dos estabelecimentos). Por fim, é importante referir que apesar de termos incorporado as variáveis referidas nos pontos (1) e (2), muito referenciadas na literatura, é nos pontos (3) e (4) que reside a resposta à pergunta de investigação formulada.

(1) Influência das características específicas do estabelecimento

No ponto (1) são incorporados os determinantes de sobrevivência internos ao estabelecimento. No seguimento de Mata e Portugal (1994), Fotopoulos e Louri (2000), Staber (2001), Strotmann (2007) e Huiban (2011), a dimensão de uma empresa (ou estabelecimento) é medida em termos de emprego. Assim, a medida adotada é o número de trabalhadores iniciais do estabelecimento (em logaritmo), sendo de esperar um efeito positivo da dimensão na probabilidade de sobrevivência das empresas.

Hipótese 1: Quanto maior a dimensão do estabelecimento, maior a sua probabilidade de sobrevivência (e menor o risco de morte).

A propriedade pode ser analisada de diferentes maneiras, tendo-se optado por considerar a distinção entre multi-estabelecimentos e estabelecimentos únicos.

De acordo com Strotmann (2007), a distinção entre multi-estabelecimentos e estabelecimentos únicos pode ser feita através de uma *dummy* que assume o valor 1 para o caso de estabelecimentos únicos e 0 para multi-estabelecimentos, sendo de esperar uma maior probabilidade de sobrevivência em multi-estabelecimentos. Já Mata e Portugal (1994) utilizam como variável explicativa o número de estabelecimentos por empresa, sendo também de esperar que a probabilidade de sobrevivência aumente com o número de estabelecimentos.

Hipótese 2: Quanto maior o número de estabelecimentos de uma empresa, maior a probabilidade de sobrevivência do estabelecimento associado a essa empresa (e menor o risco de morte).

Como acima referido, os pontos (2), (3) e (4) correspondem aos determinantes externos ao estabelecimento, isto é, que se relacionam com a sua envolvente.

(2) Impacto das especificidades da indústria

De acordo com Agarwal e Audretsch (2001) e Strotmann (2007), o regime tecnológico pode ser controlado utilizando uma variável *dummy* codificada com o valor 1 para o caso de estabelecimentos *high and medium-high tech* e 0 caso contrário.

A construção desta variável é feita de acordo com a OCDE (2011).²¹ O sinal esperado aponta para que os estabelecimentos *high and medium-high tech* possuam menor probabilidade de sobrevivência que os estabelecimentos *low and medium-low tech*.

Hipótese 3: Um estabelecimento que opere num sector de intensidade tecnológica elevada, *high and medium-high tech*, tem menor probabilidade de

²¹ De acordo com a OCDE (2011), *ISIC Rev. 3 Technology Intensity Definition – Classification of manufacturing industries into categories on R&D intensities*, são classificados como *high and medium high tech* estabelecimentos a operar nas seguintes categorias da CAE a dois dígitos: 24, 29, 30, 32, 33, 31, 34, 35. As restantes são incluídas na categoria *low and medium low tech*.

sobrevivência (e maior risco de morte) do que um estabelecimento que opere num sector *low and medium-low tech*.

Mata e Portugal (1994) e Strotmann (2007) sugerem a inclusão das taxas de entrada no sector em análise de sobrevivência. Como foi anteriormente referido, Mata e Portugal (1994) justificam esta escolha devido ao aumento de concorrência que advém do acréscimo do número de entradas. A medida aqui escolhida é o número de novos estabelecimentos na indústria (em logaritmo). A construção desta variável faz-se através da contabilização dos nascimentos totais por CAE a dois dígitos para cada ano no período considerado (1995-2006). É de esperar que quanto maior o número de entradas, menor a probabilidade de sobrevivência.

Hipótese 4: Quanto mais estabelecimentos entrarem no mercado, menor a probabilidade de sobrevivência de um estabelecimento (e maior o risco de morte).

Mata e Portugal (1994) controlam a escala mínima eficiente em duas partes. Em primeiro lugar, calculam a escala mínima eficiente da indústria e, de seguida, verificam se as empresas operam numa escala sub-ótima utilizando a medida proposta por Lyons (1980). Neste trabalho, uma vez que se trata de um estudo ao nível do estabelecimento, a medida proposta por Mata e Portugal (1994) não é aplicável.²² Por outro lado, de acordo com (Lyons, 1980) existem outras *proxies* utilizadas no cálculo da escala mínima eficiente como é o caso da técnica do “*mid-point*” (proposta por Weiss, 1963), a média e a mediana. Deste modo, e tendo consciência das suas limitações²³, e a exemplo de Görg e Strobl (2003), a medida utilizada é a mediana da dimensão (em número de trabalhadores) dos estabelecimentos em 1995 por CAE a dois dígitos.²⁴

²² A medida proposta pelos autores calcula a escala mínima eficiente através do logaritmo de metade da dimensão média das empresas que, em média, operam com 1,5 estabelecimentos. No nosso caso, como se está a calcular a escala mínima eficiente ao nível do estabelecimento, esta relação não é passível de ser adotada.

²³ De acordo com Weiss (1963) esta medida é enviesada pela existência de numerosos estabelecimentos de pequena dimensão.

²⁴ Também foi experimentada a média, medida utilizada por exemplo no artigo de Görg *et al.* (2000). No entanto, as diferenças encontradas não foram muito significativas.

Hipótese 5: Quanto maior a escala mínima eficiente de um sector de atividade, menor será a probabilidade de sobrevivência de um estabelecimento (e maior o risco de morte).

Quanto à dinâmica do sector de atividade, é utilizada como medida a taxa de crescimento da indústria, variável considerada em Mata e Portugal (1994), Shaver e Flyer (2000) e Strotmann (2007). Neste estudo, calcula-se esta variável através da taxa de crescimento média anual relativa à variação do número de estabelecimentos que operam em cada ano por CAE a dois dígitos. É esperado um efeito positivo desta variável na sobrevivência dos estabelecimentos.²⁵

Hipótese 6: Quanto maior a taxa de crescimento média anual de um sector de atividade, maior a probabilidade de sobrevivência do estabelecimento (e menor o risco de morte).

(3) Influência das economias de aglomeração

Como anteriormente referido, as economias de aglomeração, isto é, as externalidades que resultam da concentração geográfica de empresas (ou estabelecimentos), podem ser divididas em duas categorias: economias de localização e economias de urbanização.

De acordo com Boschma e Wenting (2007), as economias de localização medem-se através do número de trabalhadores no mesmo sector (ou relacionados) na região considerada. Contudo, considera-se que a medida mais adequada deverá ser em função do número de estabelecimentos e não trabalhadores, uma vez que a segunda não reflete a concentração do emprego por estabelecimentos (Figueiredo *et al.*, 2009). Assim, a medida escolhida será o número de estabelecimentos no ano de 1995 na mesma CAE a dois dígitos por NUTS III²⁶ (em logaritmo) e espera-se uma influência

²⁵ Neste estudo considera-se que a variável taxa de crescimento da indústria já capta, em certa parte, o efeito do ciclo económico. Como tal, não foi incluída nenhuma variável que identificasse períodos de expansão e de recessão no período de análise considerado.

²⁶ Segundo INE (2005), *Nomenclatura das unidades territoriais para fins estatísticos*, as NUTS III dividem Portugal Continental em vinte e oito unidades territoriais estatísticas. São elas: Alto Trás-os-Montes, Ave, Cávado, Douro, Entre Douro e Vouga, Grande Porto, Minho-Lima, Tâmega, Baixo

negativa na sobrevivência dos estabelecimentos, o que poderá ser explicado pela concorrência entre estabelecimentos do mesmo sector e é confirmado pela literatura.

Hipótese 7: As economias de localização têm uma influência negativa na probabilidade de sobrevivência do estabelecimento (aumentando, assim, o risco de morte).

As economias de urbanização, por sua vez, podem ser medidas através do número de habitantes da região (em logaritmo) (Boschma e Wenting, 2007; Wennberg e Lindqvist, 2010) ou o nível de atividade económica da região (e.g. Guimarães *et al.*, 2004) Neste estudo, seguimos a última aproximação, considerando como medida de economias de urbanização o número de estabelecimentos da indústria e serviços na região no ano de 1995 (em logaritmo), sendo a unidade territorial utilizada as NUTS III. Espera-se um sinal positivo destas na sobrevivência dos estabelecimentos, explicada pelo acesso a um mercado diversificado e confirmado pelos estudos empíricos.

Hipótese 8: As economias de urbanização têm uma influência positiva na probabilidade de sobrevivência do estabelecimento (diminuindo, assim, o risco de morte).

(4) Influência da região onde as empresas se localizam

No seguimento de estudos que dividem o espaço em áreas/regiões com diferentes características (Fotopoulos e Louri, 2000; Tödting e Wanzenböck, 2003; Strotmann, 2007; Huiban, 2011) são incorporadas variáveis *dummy* que decompõem o território nacional nas cinco regiões NUTS II: Norte, Centro, Lisboa, Alentejo e Algarve. O objetivo é tentar captar especificidades das regiões que possam influenciar a sobrevivência dos estabelecimentos, não captadas por outras variáveis.

Mondego, Baixo Vouga, Beira Interior Norte, Beira Interior Sul, Cova da Beira, Dão-Lafões, Médio Tejo, Oeste, Pinhal Interior Norte, Pinhal Interior Sul, Pinhal Litoral, Serra da Estrela, Grande Lisboa, Península de Setúbal, Alentejo Central, Alentejo Litoral, Alto Alentejo, Baixo Alentejo, Lezíria do Tejo, Algarve.

Do teste de Wilcoxon e log-rank (Tabela 2) verificou-se que há evidência estatística de existência de diferenças de sobrevivência regionais sendo possível formular a seguinte hipótese:

Hipótese 9: A localização de um estabelecimento numa região influencia a sua probabilidade de sobrevivência e, logo, o risco de morte.

Antes mesmo de partir para a estimação, estudou-se a correlação entre as variáveis explicativas escolhidas para o modelo com recurso ao *STATA/MP 12.0*. Na Tabela 4 é possível verificar que não existem valores que impossibilitem a sua incorporação no modelo.

Tabela 4 – Matriz de correlação entre as variáveis explicativas

	<i>size</i>	<i>nrest</i>	<i>txcres</i>	<i>entradas</i>	<i>eme</i>	<i>ecloc</i>	<i>ecurb</i>
<i>size</i>	1,000						
<i>nrest</i>	0,046	1,000					
<i>txcres</i>	-0,033	-0,017	1,000				
<i>entradas</i>	-0,106	-0,125	0,125	1,000			
<i>eme</i>	0,185	0,108	0,012	-0,548	1,000		
<i>ecloc</i>	-0,004	-0,100	0,066	0,518	-0,172	1,000	
<i>ecurb</i>	0,049	0,044	0,034	-0,071	0,045	0,506	1,000

4.4. Resultados

Formuladas as hipóteses, recorrendo ao *software STATA/MP 12.0* e a várias referências de análise de sobrevivência e econométricas (Baltagi, 2005; Cameron e Trivedi, 2005; Cameron e Trivedi, 2009; Cleves *et al.*, 2010; Greene, 2012), foram ensaiadas três especificações do modelo de Cox (Tabela 5):

Especificação I: Utilizam-se todas as variáveis definidas – (1), (2), (3) e (4);

Especificação II: São incluídas as variáveis tipo (1) e (2) e as economias de aglomeração – (3);

Especificação III: São incluídas as variáveis (1) e (2) e as *dummies* por NUTS II²⁷ – (4).

Tabela 5 – Determinantes do risco de morte dos estabelecimentos em Portugal (1995-2006): Modelo de Cox

Variáveis	Especificação		
	I	II	III
<i>size</i>	-0,0915*** (0,000)	-0,0927*** (0,000)	-0,0888*** (0,000)
<i>nrest</i>	-0,0254*** (0,003)	-0,0257*** (0,003)	-0,0252*** (0,003)
<i>txcres</i>	-1,2805*** (0,002)	-1,2552*** (0,002)	-1,2746*** (0,002)
<i>entradas</i>	0,0779** (0,025)	0,0727** (0,036)	0,0755** (0,011)
<i>hightech</i>	0,0036 (0,960)	0,0027 (0,969)	0,0099 (0,890)
<i>eme</i>	0,0850*** (0,001)	0,0974*** (0,000)	0,0822*** (0,001)
<i>ecloc</i>	0,0013 (0,947)	0,0263 (0,141)	
<i>ecurb</i>	0,0734** (0,015)	0,0729*** (0,001)	
<i>norte</i>	0,1914*** (0,000)		0,2576*** (0,000)
<i>lisboa</i>	0,1162 (0,133)		0,2743*** (0,000)
<i>alentejo</i>	0,1789** (0,020)		0,1477* (0,052)
<i>algarve</i>	0,1186 (0,329)		0,1805 (0,128)
<i>Log-likelihood</i>	-32773,478	-32782,415	-32777,433
<i>Chi square</i>	118,46	100,59	110,55
Nº observações	23373	23373	23373

Notas: *P-values* dentro de parênteses; * *p-value* < 0,1; ** *p-value* < 0,05; *** *p-value* < 0,01

Da Tabela 5, é possível retirar as primeiras conclusões sobre a influência dos determinantes no risco de morte dos estabelecimentos portugueses.

Da especificação I verifica-se que as variáveis não espaciais são significativas e apresentam o sinal esperado, com exceção da variável *hightech* que é não significativa. Das variáveis com influência negativa tem-se, por exemplo, que a variação de 1% no número de trabalhadores do estabelecimento provoca uma redução no risco de morte de 0,09%, enquanto o aumento em uma unidade no número de estabelecimentos associado à empresa à qual o estabelecimento faz parte (*nrest*) provoca uma diminuição do risco

²⁷ A região *centro* corresponde à base de comparação por regiões.

de morte em cerca de 3% ($\exp(-0,0254) = 0,97$). Das variáveis com influência positiva, destaca-se a escala mínima eficiente: um aumento em uma unidade na mediana do número de trabalhadores por estabelecimento por sector de atividade (*eme*) provoca um acréscimo no risco de morte de cerca de 9% ($\exp(0,0850) = 1,09$).

Quanto às variáveis espaciais, verifica-se que as economias de localização (*ecloc*) não são estatisticamente significativas, enquanto as economias de urbanização (*ecurb*) apresentam um sinal contrário ao esperado, ou seja, o aumento em 1% no número de estabelecimentos da indústria e serviços por NUTS III provoca um acréscimo no risco de morte de 0,07%.

Considerando a influência da localização numa região na sobrevivência, verifica-se que o *norte* e o *alentejo* revelaram possuir um risco de morte superior ao *centro*, ou seja, um estabelecimento localizado na região *norte* enfrenta um risco de morte superior em 21% face a um estabelecimento situado no *centro* ($\exp(0,1914) = 1,21$), e um estabelecimento localizado no *alentejo* apresenta um acréscimo de risco de cerca de 20% face ao *centro* ($\exp(0,1789) = 1,20$). Por outro lado, a localização em *lisboa* e *algarve* não tem um efeito estatisticamente distinto da localização no *centro*.

Na especificação II excluem-se as variáveis *dummy* por região, verificando-se que os sinais obtidos para as variáveis consideradas mantêm-se iguais aos obtidos na especificação I.

A especificação III exclui as economias de localização e de urbanização, alcançando, também, resultados idênticos à especificação I e II para as variáveis não espaciais. Contudo, nesta especificação a localização de um estabelecimento em *lisboa* aumenta o risco de morte face a um estabelecimento localizado na região *centro* em cerca de 32% ($\exp(0,2743) = 1,32$).

Nas secções seguintes, propõe-se (i) proceder ao teste do pressuposto do *Proportional Hazard* e, em caso de necessidade, à sua correção (4.4.1); (ii) apresentar extensões ao modelo de Cox considerando efeitos de grupo que têm como objetivo estudar a influência na sobrevivência da localização dos estabelecimentos em regiões de diferentes perspectivas (4.4.2); (iii) apresentar testes diagnósticos de forma a aferir a adequabilidade dos modelos apresentados (4.4.3); (iv) finalmente, proceder uma breve comparação dos resultados (4.4.4).

4.4.1. Pressuposto da *Proportional Hazard* e Modelo de Cox com variáveis explicativas em interação com o tempo

Para testar o pressuposto do *Proportional Hazard* (PH) recorreu-se a dois métodos a seguir descritos.

Em primeiro lugar, através do teste baseado na reestimação que afere quanto à correta especificação do modelo (Cleves *et al.*, 2010), concluiu-se que nenhuma das três especificações apresentadas cumpria o pressuposto do PH. Ainda através deste método, incluindo separadamente as variáveis em interação com uma função tempo (para o caso $g(t) = t$), verificou-se que as variáveis que não cumpriam o pressuposto do PH a um nível de significância de 5% eram: *size*, *eme* e *alentejo*.

De seguida, confrontou-se o resultado obtido com o teste que utiliza os resíduos de Schoenfeld. Esta medida que permite avaliar a correlação entre os resíduos das variáveis explicativas e uma função tempo (Rocha e Papoila, 2009; Cleves *et al.*, 2010) confirmou que o modelo na sua globalidade não cumpre o pressuposto do PH e, desta feita, as variáveis onde se detetam os principais problemas são: *size*, *entradas*, *eme*, *txcres* e *alentejo*.

Conclui-se então que o modelo não cumpre o pressuposto do *Proportional Hazard*. Rocha e Papoila (2009) sugerem diversas estratégias para resolver o problema da não verificação dos riscos proporcionais (pressuposto PH). Uma das sugestões avançadas por estas autoras corresponde à introdução de variáveis explicativas dependentes do tempo e é de seguida apresentada.

Tendo como objetivo resolver o problema da não verificação dos riscos proporcionais, são estimadas as três especificações do modelo incluindo variáveis transformadas, isto é, multiplicadas por uma função tempo. Dos dois testes levados a cabo e acima resumidos, foi possível determinar quais as variáveis que não cumprem o pressuposto do PH. Visto que os dois testes indicam diferentes conjuntos de variáveis, foram testadas várias tentativas²⁸ com vista a encontrar as melhores especificações do modelo, isto é, que maximizam o seu *log-likelihood*. Desta forma, optou-se pela

²⁸ Foram levadas a cabo várias tentativas onde se experimentaram: diferentes grupos de variáveis em interação com o tempo, dentro das cinco identificadas como problemáticas; diferentes formas funcionais para a função tempo (nomeadamente $g(t) = \ln(t)$); especificações alternativas para as variáveis referentes às economias de aglomeração (consideração do número de trabalhadores em detrimento do número de estabelecimentos e consideração de municípios no lugar de NUTS III).

inclusão das variáveis *size*, *txcres*, *entradas*, *eme* e *alentejo* multiplicadas por uma função tempo, $g(t) = t$.

Os resultados adquiridos estão expostos na Tabela 6 onde são apresentadas três especificações: a especificação IV, V e VI por comparação com as especificações I, II e III, respetivamente.

Tabela 6 – Determinantes do risco de morte dos estabelecimentos em Portugal (1995-2006): Modelo de Cox com variáveis em interação com o tempo

Variáveis	Especificação		
	IV	V	VI
<i>size_t</i>	-0,0230*** (0,000)	-0,0231*** (0,000)	-0,0225*** (0,000)
<i>nrest</i>	-0,0249*** (0,004)	-0,0257*** (0,003)	-0,0247*** (0,004)
<i>txcres_t</i>	-0,3363*** (0,007)	-0,3402*** (0,007)	-0,3331*** (0,008)
<i>entradas_t</i>	0,0209*** (0,002)	0,0207** (0,002)	0,0203*** (0,001)
<i>hightech</i>	-0,0051 (0,938)	0,0029 (0,964)	0,0040 (0,950)
<i>eme_t</i>	0,0243*** (0,000)	0,0270*** (0,000)	0,0236*** (0,000)
<i>ecloc</i>	-0,0009 (0,960)	0,0227 (0,160)	
<i>ecurb</i>	0,0669** (0,022)	0,0745*** (0,000)	
<i>norte</i>	0,1509*** (0,002)		0,2108*** (0,000)
<i>lisboa</i>	0,0882 (0,251)		0,2318*** (0,000)
<i>alentejo_t</i>	-0,0018 (0,918)		-0,0060 (0,733)
<i>algarve</i>	0,0865 (0,473)		0,1458 (0,216)
<i>Log-likelihood</i>	-32769,314	-32774,929	-32772,494
<i>Chi square</i>	126,79	115,56	120,43
Nº observações	23373	23373	23373

Notas: *P-values* dentro de parênteses; * *p-value* < 0,1; ** *p-value* < 0,05; *** *p-value* < 0,01

Da especificação IV conclui-se que, uma vez mais, as variáveis não espaciais, com exceção da variável *hightech*, cumprem o sinal previsto na literatura. Desta forma, as variáveis *size*, *nrest* e *txcres* influenciam positivamente a sobrevivência, enquanto as variáveis *entradas* e *eme* têm uma influência negativa. Apesar desta interpretação direta, é necessário ter um especial cuidado com a interpretação das variáveis em interação com o tempo. Assim, considerando o caso da variável *size_t* e *txcres_t*, verifica-se que a

influência negativa no risco de morte se torna cada vez mais negativa com o tempo. Ou seja, a razão das funções de risco vai diminuir com o tempo de forma linear (Rocha e Papoila, 2009). Por outro lado, no caso da variável *entradas_t* a razão das funções de risco irá aumentar de forma linear sendo a influência positiva no risco de morte aumentada com o tempo. Quanto às variáveis espaciais, verifica-se que *ecloc* continua a não ser significativa, enquanto *ecurb* continua com um sinal contrário ao esperado. Em termos de *dummies* regionais, apenas a variável *norte* é significativa agravando o risco de morte para estabelecimentos localizados nesta região em 16% face a estabelecimentos do *centro* do país ($\exp(0,1509) = 1,16$).

A especificação V, por sua vez, repete os sinais obtidos na especificação II, desta vez, contudo, considerando esta pequena modificação na influência das variáveis *size_t*, *txcres_t* e *entradas_t* referida na especificação IV.

Por fim, a especificação VI vai de encontro aos resultados da especificação III, exceto na variável correspondente ao *alentejo*. Note-se que quer na especificação IV quer na especificação VI, a única diferença em termos de sinal com a especificação I e III, respetivamente, é o facto da variável *alentejo* passar a não significativa devido à introdução da multiplicação pela função tempo.

Em conclusão, verifica-se que os resultados obtidos com as especificações I, II, III e IV, V, VI, respetivamente, não são muito distintos (pelo menos em sinal), exceto para o caso do *alentejo*. De facto, as principais diferenças residem na influência marginal dos coeficientes para os quais foi registada a alteração via interação com a função tempo. Esta conclusão é reforçada pela sugestão apresentada por Rocha e Papoila (2009) que explica que o não cumprimento do pressuposto do *Proportional Hazard*, por vezes, não tem influência na interpretação em amostras de dimensão alargada (como é o caso em estudo).

4.4.2. Resultados do Modelo de Cox com efeitos de grupo

Até ao momento, de acordo com Cleves *et al.* (2010) apenas foram considerados modelos de efeitos fixos em que todos os estabelecimentos usufruem da mesma função risco subjacente. Segundo Baltagi (2005) os modelos de efeitos fixos são adequados

quando o foco é no conjunto de estabelecimentos considerados e não na inferência que é feita a partir desta amostra para o total de estabelecimentos existentes.

Neste estudo, pretende-se ir mais longe e compreender se o facto de os estabelecimentos serem provenientes de diferentes regiões (Norte, Centro, Lisboa, Alentejo ou Algarve) e a forma como isto é incorporado no modelo tem algum efeito nos resultados. Desta forma, foi incorporado o conceito de efeitos de grupo (Cleves *et al.*, 2009) nas três especificações, de seguida, apresentadas.

a) Modelo estratificado

Nesta especificação considera-se que os estabelecimentos não sustentam todos a mesma *baseline* (função risco subjacente), existindo diferentes *baselines* de acordo com os grupos definidos (Rocha e Papoila, 2009; Cleves *et al.*, 2010). Uma vez que o objetivo é inspecionar os efeitos de grupo por região, os cinco estratos/grupos considerados referem-se às cinco regiões definidas. Assim, a função de risco a estimar passa a ser:

$$(4.2) \quad h(t; \mathbf{z}) = h_{0j}(t) \exp(\boldsymbol{\beta}'\mathbf{z})$$

com $j = \text{norte, centro, lisboa, alentejo, algarve}$.

b) Modelos com fragilidade

De acordo com Rocha e Papoila (2009) e Cleves *et al.* (2010), os modelos com fragilidade correspondem a modelos com efeitos aleatórios. Baltagi (2005) explica que modelos com efeitos aleatórios são adequados quando é escolhida uma amostra aleatória de um conjunto total de estabelecimentos. Estes modelos são induzidos pela presença de correlação entre os estabelecimentos, que pode relacionar-se com o tempo de vida ou com a heterogeneidade não observada. Considerando a hipótese de possível correlação entre os estabelecimentos de uma dada região, a função risco que se estima é:

$$(4.3) \quad h(t; \mathbf{z}, v) = h_0(t) \exp(\boldsymbol{\beta}'\mathbf{z} + v_i)$$

em que $i = \text{norte, centro, lisboa, alentejo, algarve}$; $v_i = \log(\alpha_i)$; α_i representa a fragilidade do grupo i .²⁹

²⁹ Assume-se que a variável que exprime a fragilidade de cada grupo é não negativa, segue a distribuição gama com média 1 e variância θ (Cleves *et al.*, 2010).

b.1) Modelo com correção de correlação entre elementos de um grupo: Segundo Rocha e Papoila (2009), considera-se que os métodos utilizados supõem tempos de vida independentes. Contudo, pode existir correlação entre os elementos de um dado grupo deixando cair, desta forma, a hipótese de independência entre os tempos de vida dos estabelecimentos desse grupo. Neste caso, é feita uma correção dos efeitos de *cluster* (para o caso, as regiões) de forma a obter uma estimação robusta da variância (Cleves *et al.*, 2010).

b.2) Modelo com correção da heterogeneidade não observada: De acordo com Rocha e Papoila (2009), muitas vezes os estabelecimentos com os mesmos valores para as variáveis explicativas diferem entre si existindo, portanto, uma heterogeneidade entre eles. Nestes casos, ao modelar a heterogeneidade não observada ao nível da região controla-se para fatores de risco não explicados pelo modelo. Ao contrário de b.1), o efeito de pertencer a uma região é logo incorporado nos resultados (Cleves *et al.*, 2010).

Tabela 7 – Determinantes do risco de morte dos estabelecimentos em Portugal (1995-2006): Modelo de Cox com efeitos de grupos

Variáveis	Especificação		
	Modelo estratificado	Modelos com fragilidade	
		Com correção de correlação entre elementos de um grupo	Com correção da heterogeneidade não observada
<i>size</i>	-0,0911*** (0,000)	-0,0927*** (0,000)	-0,0918*** (0,000)
<i>nrest</i>	-0,0252*** (0,003)	-0,0257*** (0,000)	-0,0253*** (0,003)
<i>txcres</i>	-1,2164*** (0,003)	-1,2552** (0,030)	-1,2767*** (0,002)
<i>entradas</i>	0,0735** (0,035)	0,0727** (0,040)	0,0773** (0,026)
<i>hightech</i>	-0,0021 (0,976)	0,0027 (0,960)	0,0033 (0,963)
<i>eme</i>	0,0824*** (0,001)	0,0974*** (0,002)	0,0870*** (0,001)
<i>ecloc</i>	0,0018 (0,928)	0,0263* (0,057)	0,0052 (0,787)
<i>ecurb</i>	0,0732** (0,015)	0,0729** (0,026)	0,0751*** (0,005)
<i>theta</i>			0,0039*** (0,003)
<i>Log-likelihood</i>	-28002,85	-32782,415	-32778,53
<i>Chi square</i>	67,63	150,58	71,25
Nº observações	23373	23373	23373
Nº grupos	5	5	5

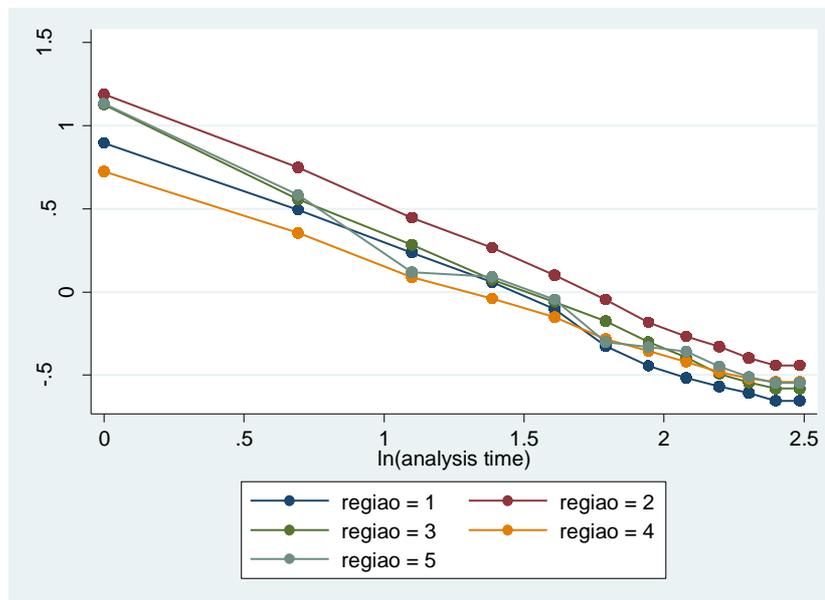
Notas: *P-values* dentro de parênteses; * *p-value* < 0,1; ** *p-value* < 0,05; *** *p-value* < 0,01

Na Tabela 7 é possível comparar os resultados das três aproximações acima indicadas.

Quanto ao modelo estratificado, os resultados obtidos vão de encontro com os resultados alcançados na especificação I. Ou seja, exceto *hightech* e *ecloc* que não são significativas e *ecurb* que tem um sinal contrário ao esperado, todas os outros coeficientes estimados apresentam um sinal de acordo com o esperado. Quando se compara com a especificação II que não considera a divisão do espaço, também não há grandes diferenças nos coeficientes das restantes variáveis. Assim, não há grandes motivos para considerar que as *baselines* devem variar livremente entre regiões. Para uma conclusão mais fidedigna, recorre-se à seguinte representação gráfica que indica que se as curvas forem relativamente paralelas, pode-se concluir que os riscos são proporcionais e que basta adicionar as variáveis *dummy* ao modelo de Cox (Rocha e Papoila, 2009).

Da Figura 6, parece que as diferentes regiões têm *baselines* não muito distintas, à exceção do Alentejo e do Algarve.

Figura 6 – Teste da proporcionalidade dos riscos por regiões



Notas: região=1 (Norte), região=2 (Centro), região=3 (Lisboa), região=4 (Alentejo) e região=5 (Algarve).

Analisando agora os modelos com fragilidade, verifica-se que, segundo Cleves *et al.* (2010), se efetivamente existir correlação entre os estabelecimentos, a especificação b.1) não está correta nem sequer permite uma interpretação dos coeficientes. Porém, a estimação robusta da variância permite recolher informações interessantes em termos de inferência estatística. Desta feita, é possível verificar que a principal diferença advém da significância da variável *ecloc*. Isto é, relativamente aos restantes modelos, caso exista correlação entre os estabelecimentos, a estimação robusta da variância produz que a variável *ecloc* passe a ser significativa.

Centrando as atenções em b.2), dado o valor obtido para o *p-value* de *theta*, verifica-se que há evidência estatística significativa de que há correlação dentro dos grupos definidos. A interpretação dos coeficientes deve ser feita condicional à fragilidade (Cleves *et al.*, 2010), isto é, para um dado nível de fragilidade, um estabelecimento adicional numa empresa provoca que um dado estabelecimento daquela empresa possua um risco inferior em cerca de 2% ($\exp(-0,0253) = 0,9750$). Tendo em conta esta ressalva, verifica-se que também estes resultados não estão muito longe das outras especificações ensaiadas.

4.4.3. Testes diagnósticos de qualidade do ajustamento

Para aferir quanto à qualidade do ajustamento, recorre-se à inspeção dos resíduos de Cox-Snell. Os modelos estarão bem definidos se a estimativa de Nelson-Aalen da função de risco cumulativa dos resíduos se aproximar de uma reta de declive um e ordenada na origem igual a zero (Rocha e Papoila, 2009; Cleves *et al.*, 2010).

Os seis gráficos seguintes permitem tirar conclusões quanto a adequabilidade do modelo da especificação I, II, III, IV, V e VI.

Dos seis gráficos a seguir apresentados (Figuras 7 a 12), verifica-se que quando se incorporam todas as variáveis explicativas (especificações I e IV), a especificação I, isto é, aquela em que não há interação de nenhuma variável com uma função tempo, é a que apresenta uma melhor qualidade do ajustamento. Nos restantes modelos, que incorporam apenas uma categoria de variáveis espaciais, comparando diretamente II com V e III com VI, verifica-se que os modelos em que há interação de algumas variáveis com funções tempo são substancialmente melhores (especificações V e VI).

Figura 7 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (I)

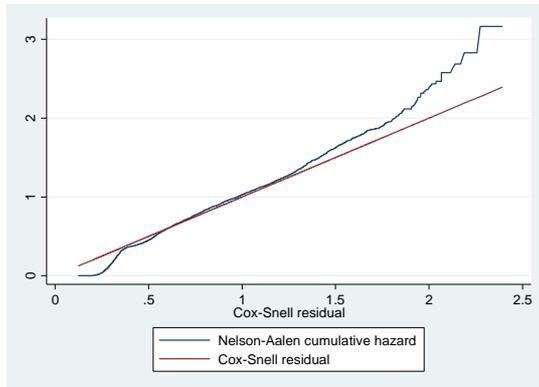


Figura 8 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (II)

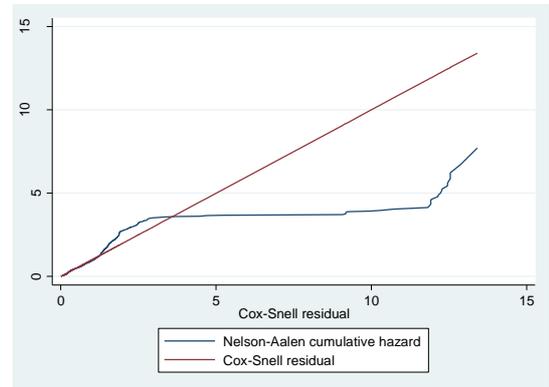


Figura 9 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (III)

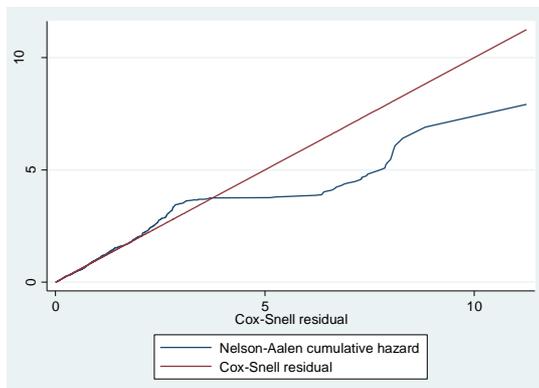


Figura 10 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (IV)

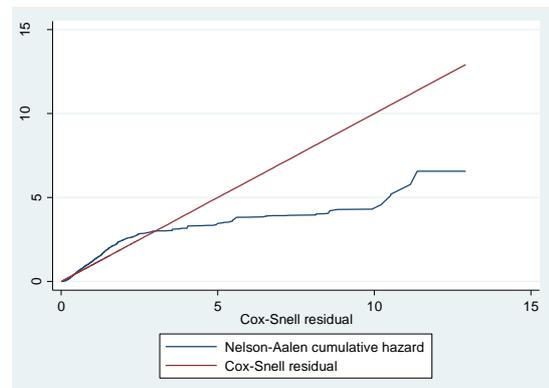


Figura 11 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (V)

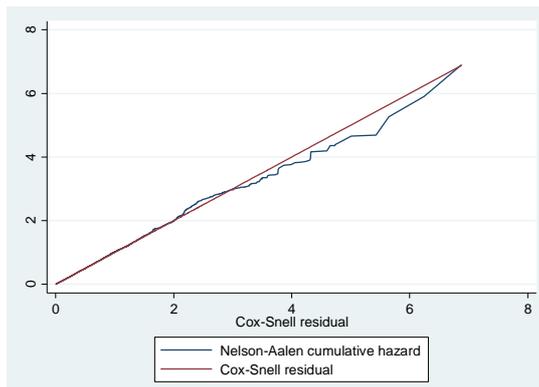
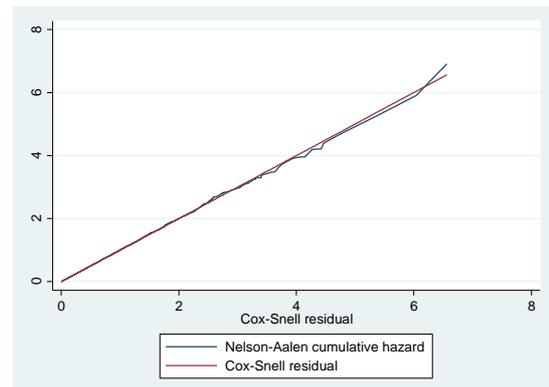


Figura 12 – Resíduos de Cox-Snell e função risco cumulativa de Nelson-Aalen (VI)



4.4.4. Comparação de resultados

Por fim, comparam-se os resultados obtidos das diferentes especificações com o sinal esperado da literatura. Isto permite verificar a volatilidade dos resultados face às

diversas especificações, bem como retirar as conclusões finais quanto ao cumprimento, ou não, das hipóteses formuladas.

Tabela 8 – Verificação das hipóteses: resultados obtidos *versus* sinal esperado

Hipóteses	Sinal esperado	Especificação								
		I	II	III	IV	V	VI	a)	b.1)	b.2)
1 – <i>size</i>	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
2 – <i>nrest</i>	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
3 – <i>hightech</i>	(+)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
4 – <i>entradas</i>	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
5 – <i>eme</i>	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
6 – <i>txcres</i>	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
7 – <i>ecloc</i>	(+)	n.s.	n.s.		n.s.	n.s.		n.s.	(+)	n.s.
8 – <i>ecurb</i>	(-)	(+)	(+)		(+)	(+)		(+)	(+)	(+)
9 – <i>regiões</i>	*	*		*	*		*			

Notas: A abreviatura n.s. significa que o resultado é não estatisticamente significativo a um nível de significância de pelo menos 10%. * A localização de um estabelecimento numa região influencia a sua probabilidade de sobrevivência e, logo, o risco de morte.

Da Tabela 8, verifica-se que as hipóteses 1, 2, 4, 5 e 6 vão, de forma unanime, de encontro ao sinal esperado na literatura. Isto é, quanto maior a dimensão, o número de estabelecimentos e a taxa de crescimento do sector, menor o risco de morte de um estabelecimento. Por outro lado, quanto maior o número de entradas e a escala mínima eficiente, maior o risco de morte. Note-se que estas são variáveis não espaciais, para as quais os resultados obtidos são confirmados em vários estudos. Verifica-se ainda que das variáveis não espaciais, apenas a hipótese 3 não vai de encontro com a literatura, tendo-se revelado não significativa nas sucessivas especificações.

Das variáveis espaciais, a hipótese 7 em quase todas as especificações parece não ser significativamente diferente de zero, exceto na especificação a) onde vai de encontro com o sinal esperado. A hipótese 8, por seu lado, tem o sinal sistematicamente contrário ao esperado, ilustrando que as economias de urbanização têm um impacto negativo na sobrevivência de um estabelecimento.

Quanto à hipótese 9, pode concluir-se pela influência da localização em regiões na sobrevivência dos estabelecimentos. Na especificação I, a evidência aponta para um maior risco no Norte e Alentejo face à região Centro. Na especificação III, este

acréscimo de risco face ao Centro estende-se, também, a Lisboa. Nas especificações IV e VI, devido à alteração da variável correspondente ao Alentejo, esta deixa de apresentar um agravamento do risco de morte face ao Centro em comparação com as especificações I e III, respetivamente. Quanto aos efeitos de grupos, de forma mais ou menos clara, parece também existir alguma evidência que os estabelecimentos de cada região têm características distintas. Isto advém, sobretudo, da evidência de existência de heterogeneidade não observada entre as regiões – especificação b.2).

Por fim, através das diferentes especificações apresentadas, pode concluir-se sobre a robustez do modelo. Das variáveis não espaciais, os resultados obtidos para os coeficientes de estimação são muito similares para todas as especificações, exceto quando há uma alteração da própria variável explicativa como é o caso das variáveis em interação com uma função tempo. Das variáveis espaciais que são estatisticamente significativas, o coeficiente associado às economias de urbanização também não sofre uma alteração relevante nas várias especificações, residindo a maior instabilidade nas *dummies* regionais. Isto permite concluir que, com as ressalvas indicadas, os resultados obtidos não são, de forma geral, muito sensíveis à alteração das especificações.

Conclusões e desenvolvimentos futuros

Nesta dissertação pretendeu-se analisar a sobrevivência dos estabelecimentos em Portugal dando especial ênfase ao papel do espaço. Em particular, pretendeu-se responder às seguintes questões: Qual a influência das economias de aglomeração (localização e urbanização) sobre o risco de morte? Será que a localização em determinadas regiões influencia a sobrevivência dos estabelecimentos?

Em primeiro lugar, foram revisitadas três teorias que fornecem conceitos utilizados em análise de sobrevivência: Teoria da Empresa, Ecologia das Organizações e Teoria Evolucionista. Após esta breve contextualização desenvolveu-se uma revisão de literatura em que se identificaram os principais determinantes da sobrevivência das empresas (ou estabelecimentos) bem como as metodologias utilizadas.

Tendo como objetivo responder às questões de investigação formuladas, aplicou-se o modelo de Cox no estudo da sobrevivência de 4917 estabelecimentos do coorte de 1995 a operar na indústria transformadora em Portugal Continental. De forma a aprofundar esta investigação, foram ensaiadas nove especificações: três com diferentes combinações de variáveis (I, II e III); três que corrigem o pressuposto do *Proportional Hazard* através da introdução de variáveis em interação com uma função tempo (IV, V e VI); e as três últimas que introduzem extensões ao modelo incorporando efeitos de grupo. A opção de proceder a nove estimações, mais que concluir sobre qual a mais correta, permitiu concluir sobre a robustez dos resultados alcançados.

Os resultados obtidos permitem concluir que as variáveis não espaciais são significativas e apresentam o sinal de acordo com o esperado, à exceção do regime tecnológico. Assim, verifica-se que quanto maior a dimensão do estabelecimento, o número de estabelecimentos ou a taxa de crescimento do sector, menor o risco de morte de um estabelecimento. Estes resultados corroboram, por isso, a hipótese da “responsabilidade da pequenez”, referida pela Ecologia das Organizações, que identifica as desvantagens de custo, a gestão menos preparada, a maior dificuldade de acesso ao mercado de capitais, entre outras, como causa da desvantagem de sobrevivência associada a pequenas empresas/estabelecimentos. Por sua vez, a influência positiva da taxa de crescimento na sobrevivência pode ser explicada pela maior margem de inovação associada a fases de crescimento mais acelerado. Por outro lado, quanto maior

o número de entradas e a escala mínima eficiente, maior o risco de morte, o que pode ser justificado, no primeiro caso, com o acréscimo de concorrência, enquanto o segundo prende-se com o facto da entrada normalmente ocorrer em pequena dimensão e quanto maior a escala mínima eficiente, mais árdua a tarefa de a alcançar.

Quanto à influência das variáveis espaciais, objetivo central desta dissertação, os resultados não são tão conclusivos.

Centrando a atenção na primeira questão formulada, as economias de localização revelaram-se não significativas enquanto as economias de urbanização apresentaram um sinal contrário ao esperado, isto é, um impacto negativo na sobrevivência dos estabelecimentos.

Verificou-se ainda que, em relação à segunda questão, a localização do estabelecimento numa região influencia a sua sobrevivência. A aplicação do teste de Wilcoxon e log-rank revelou a existência de diferenças regionais de sobrevivência dos estabelecimentos. Com a estimação do modelo de Cox, a especificação I evidenciou um maior risco no Norte e Alentejo face ao Centro, sendo esse acréscimo de risco face ao Centro estendido a Lisboa na especificação III. Isto repete-se nas especificações IV e VI, exceto para a variável Alentejo que passa a não significativa. Quanto aos efeitos de grupos, de forma mais ou menos clara, parece também existir alguma evidência de que os estabelecimentos de cada região têm características distintas. Mais do que a forma como é incorporada, verifica-se que há, de facto, uma influência da localização que, no entanto, não é muito estável nas várias especificações. Note-se, contudo, que face às primeiras especificações (I, III, IV e VI), o Centro possui sempre melhores níveis de sobrevivência do que pelo menos uma das regiões, com exceção do Algarve.

Como é possível observar, existem várias possibilidades de investigação futura: analisar a sobrevivência dos estabelecimentos incorporando diferentes medidas de economias de localização e de economias de urbanização; incluir vários momentos de entrada na base de dados e não apenas uma coorte; confrontar os resultados obtidos com estimações paramétricas e não paramétricas.

Referências bibliográficas

Acs, Z. J., C. Armington e T. Zhang (2007), “The Determinants of New-Firm Survival across Regional Economies: The Role of Human Capital Stock and Knowledge Spillover”, *Papers in Regional Science*, Vol. 86, Nº 3, pp. 367-391

Agarwal, R. e D. B. Audretsch (2001), “Does entry size matter? The impact of the life cycle and technology on firm survival”, *The Journal of Industrial Economics*, Vol. 49, Nº 1, pp. 21-43

Audretsch, D. B. e T. Mahmood (1991), “The hazard rate of new establishments: a first report”, *Economics Letters*, Vol. 36, pp. 409-412

Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, 3ª edição, Chichester, John Wiley & Sons, Ltd

Boschma, R. A. e R. Wenting (2007), “The spatial evolution of the British automobile industry: Does location matter?”, *Industrial and Corporate Change*, Vol. 16, Nº 2, pp. 213-238

Box, M. (2008), “The death of firms: exploring the effects of environment and birth cohort on firm survival in Sweden”, *Small Business Economics*, Vol. 31, pp. 379-393

Classificação Portuguesa das Atividades Económicas (CAE) – Rev. 2: Decreto-Lei nº 182/93 de 14 de Maio, *Diário da República nº 112 – I Série – A*, Ministério do Planeamento e da Administração do Território, Lisboa

Classificação Portuguesa das Atividades Económicas (CAE) – Rev. 2.1: Decreto-Lei nº 197/2003 de 27 de Agosto, *Diário da República nº 197 – I Série – A*, Presidência do Conselho de Ministros, Lisboa

Cameron, A. C. e P. K. Trivedi (2005), *Microeconometrics – Methods and applications*, New York, Cambridge University Press

Cameron, A. C. e P. K. Trivedi (2009), *Microeconometrics Using Stata*, Lakeway Drive, Stata Press

Carroll, G. R. (1984), “Organizational Ecology”, *Annual Review of Sociology*, Vol. 10, pp. 71-93

Church, J. e R. Ware (2000), *Industrial Organization*, United States of America, McGraw-Hill

- Cleves, M., R. G. Gutierrez, W. Gould e Y. V. Marchenko (2010 [2002]), *An Introduction to Survival Analysis Using Stata*, 3ª edição, Lakeway Drive, Stata Press
- Coase, R. H. (1937), “The Nature of the Firm”, *Economica*, Vol. 4, Nº 16, pp. 386-405
- Cox, D. R. e D. Oakes (1984), *Analysis of Survival Data*, London, Chapman and Hall
- Esteve-Pérez, S. e J. A. Mañez-Castillejo (2008), “The Resource-Based Theory of the Firm and Firm Survival”, *Small Business Economics*, Vol. 30, pp. 231-249
- Figueiredo, O., P. Guimarães e D. Woodward (2009), "Localization economies and establishment size: was Marshall right after all?", *Journal of Economic Geography*, Vol. 9, Nº 6, pp. 853-868
- Fotopoulos, G. e H. Louri (2000), “Location and Survival of New Entry”, *Small Business Economics*, Vol. 14, pp. 311-321
- Gabinete de Estratégia e Planeamento do Ministério do Trabalho e da Solidariedade Social (GEP-MTSS) (1995-2006), *Quadros de Pessoal*, Lisboa
- Görg, H., E. Strobl e F. Ruane (2000), “Determinants of Firm Start-Up Size: An Application of Quantile Regression for Ireland”, *Small Business Economics*, Vol. 14, pp. 211-222
- Görg, H. e E. Strobl (2003), “Footloose multinationals?”, *The Manchester School*, Vol. 71, Nº 1, pp. 1-19
- Greene, W. H. (2012), *Econometrics Analysis*, 7ª edição, New Jersey, Prentice Hall
- Guimarães, P., O. Figueiredo e D. Woodward (2004), “Industrial location modeling: Extending the random utility framework”, *Journal of Regional Science*, Vol. 44, Nº 1, pp. 1 -20.
- Hannan, M. T. e J. Freeman (1977), “The Population Ecology of Organizations”, *American Journal of Sociology*, Vol. 82, Nº5, pp. 929-964
- Huiban, J-P. (2011), “The spatial demography of new plants: urban creation and rural survival”, *Small Business Economics*, Vol. 37, pp. 73-86
- Instituto Nacional de Estatística (INE) (2005), *Nomenclatura das unidades territoriais para fins estatísticos*, Lisboa

Kaplan, E. L. e P. Meier (1958), “Nonparametric estimation from incomplete observations”, *Journal of the American Statistics Association*, Vol. 53, Nº 282, pp. 457–481

Klepper, S. (2007), “Disagreements, Spinoffs, and the evolution of Detroit as the Capital of the U.S. Automobile Industry”, *Management Science*, Vol. 53, Nº 4, pp. 616-631

Lyons, B. (1980), “A New Measure of Minimum Efficient Plant Size in UK Manufacturing Industry”, *Economica*, Vol. 47, pp. 19-34

Manjón-Antolín, M. C. e J.-M Arauzo-Carod (2008), “Firm survival: methods and evidence”, *Empirica*, Vol. 35, pp. 1-24

Mata, J. e P. Portugal (1994), “Life Duration of New Firms”, *The Journal of Industrial Economics*, Vol. 42, Nº 3, pp. 227-245

Mata, J. e P. Portugal (2004), “Patterns of Entry, Post-Entry Growth and Survival”, *Small Business Economics*, Vol. 22, pp. 283-298

Mata, J. (2007), *Economia da Empresa*, 4ª edição, Lisboa, Fundação Calouste Gulbenkian

Organização para a Cooperação e Desenvolvimento Económico (OCDE) (2011), *ISIC Rev. 3 Technology Intensity Definition – Classification of manufacturing industries into categories on R&D intensities*

Penrose, E. (1959), *The Theory of the Growth of the Firm*, Oxford, Basil Blackwell

Persson, H. (2004), “The Survival and Growth of New Establishments in Sweden, 1987–1995”, *Small Business Economics*, Vol. 23, pp. 423-440

Rocha C. e A. L. Papoila (2009), “Análise de Sobrevivência”, *Atas do XVII Congresso da Sociedade Portuguesa de Estatística*, 30 Setembro – 3 Outubro, Sesimbra

Santarelli, E. e F. Lotti (2005), “The Survival of Family Firms: The Importance of Control and Family Ties”, *International Journal of the Economics of Business*, Vol. 12, Nº 2, pp. 183-192

Shaver, J. M. e F. Flyer (2000), “Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States”, *Strategic Management Journal*, Vol. 21, pp. 1175-1193

Singh J. V. e C. J. Lumsden (1990), “Theory and Research in Organizational Ecology”, *Annual Review of Sociology*, Vol. 16, pp.161-195

Silva, S. T., A. A. C. Teixeira, M. R. Silva (2004), “Economics of the Firm and Economic Growth. An hybrid theoretical framework of analysis”, *Working Paper*, Nº 158, Faculdade de Economia do Porto

Staber, U. (2001), “Spatial Proximity and Firm Survival in a Declining Industrial District: The Case of Knitwear Firms in Baden-Württemberg”, *Regional Studies*, Vol. 35, Nº 4, pp. 329-341

Strotmann, H. (2007), “Entrepreneurial Survival”, *Small Business Economics*, Vol. 28, pp. 87-104

Tödttling, F. e H. Wanzenböck (2003), “Regional differences in structural characteristics of start-ups”, *Entrepreneurship & Regional Development*, Vol. 15, pp. 351-370

Weiss, L. W. (1963), “Factors in changing concentration”, *The Review of Economics and Statistics*, Vol. 45, Nº1, pp. 70-77

Wennberg, K. e G. Lindqvist (2010), “The effect of clusters on the survival and performance of new firms”, *Small Business Economics*, Vol. 34, pp. 221-241

Williamson, O. (1985), *The Economic Institutions of Capitalism*, New York, The FreePress

Anexo 1 – CAE Rev. 2.1 a dois dígitos

- 15 – Indústrias alimentares e das bebidas;
- 16 – Indústria do tabaco;
- 17 – Fabricação de têxteis;
- 18 – Indústria do vestuário: preparação, tingimento e fabricação de artigo e peles com pelo;
- 19 – Curtimenta e acabamento de peles sem pelo: fabricação de artigos de viagem, marroquinaria, artigos de correeiro, seleiro e calçado;
- 20 – Indústria da madeira e da cortiça e suas obras, exceto mobiliário: fabricação de obras de cestaria e de espartaria;
- 21 – Fabricação de pasta, de papel e cartão e seus artigos;
- 22 – Edição, impressão e reprodução de suportes de informação gravados;
- 23 – Fabricação do coque, produtos petrolíferos refinados e tratamento de combustível nuclear;
- 24 – Fabricação de produtos químicos;
- 25 – Fabricação de artigos de borracha e de matérias plásticas;
- 26 – Fabricação de outros produtos minerais não metálicos;
- 27 – Indústrias metalúrgicas de base;
- 28 – Fabricação de produtos metálicos, exceto máquinas e equipamento;
- 29 – Fabricação de máquinas e de equipamentos, n. e.;
- 30 – Fabricação de máquinas de escritório e de equipamento para o tratamento automático de informação;
- 31 – Fabricação de máquinas e aparelhos elétricos, n. e.;
- 32 – Fabricação de equipamento e de aparelhos de rádio, televisão e comunicação;
- 33 – Fabricação de aparelhos e instrumentos médico-cirúrgicos, ortopédicos, de precisão, de óptica e de relojoaria;
- 34 – Fabricação de veículos automóveis, reboques e semi-reboques;
- 35 – Fabricação de outro material de transporte;
- 36 – Fabricação de mobiliário: outras indústrias transformadoras, n. e.;
- 37 – Reciclagem.

Anexo 2 – Nascimentos e sobrevivência (NUTS II e CAE a dois dígitos)

Relação entre nascimentos e sobrevivência (1995-2006)

	Nascimentos	
	Nº	%
Sobrevivência=1	1447	29,43%
Sobrevivência=2	690	14,03%
Sobrevivência=3	452	9,19%
Sobrevivência=4	324	6,59%
Sobrevivência=5	285	5,80%
Sobrevivência=6	340	6,91%
Sobrevivência=7	217	4,41%
Sobrevivência=8	132	2,68%
Sobrevivência=9	98	1,99%
Sobrevivência=10	75	1,53%
Sobrevivência=11	62	1,26%
Sobrevivência=12	795	16,17%
	4917	100%

Nascimentos e sobrevivência por NUTS II (1995-2006)

	Norte		Algarve		Centro		Lisboa		Alentejo	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
Nascimentos	2761	56,15%	96	1,95%	957	19,46%	815	16,57%	288	5,86%
Sobrevivência=1	876	31,73%	25	26,04%	223	23,30%	226	27,73%	97	33,68%
Sobrevivência=2	378	13,69%	16	16,67%	114	11,91%	146	17,91%	36	12,50%
Sobrevivência=3	245	8,87%	15	15,63%	88	9,20%	77	9,45%	27	9,38%
Sobrevivência=4	187	6,77%	1	1,04%	59	6,17%	63	7,73%	14	4,86%
Sobrevivência=5	169	6,12%	5	5,21%	58	6,06%	41	5,03%	12	4,17%
Sobrevivência=6	231	8,37%	9	9,38%	52	5,43%	34	4,17%	14	4,86%
Sobrevivência=7	119	4,31%	1	1,04%	51	5,33%	38	4,66%	8	2,78%
Sobrevivência=8	67	2,43%	1	1,04%	31	3,24%	26	3,19%	7	2,43%
Sobrevivência=9	42	1,52%	3	3,13%	21	2,19%	26	3,19%	6	2,08%
Sobrevivência=10	32	1,16%	2	2,08%	24	2,51%	13	1,60%	4	1,39%
Sobrevivência=11	36	1,30%	1	1,04%	15	1,57%	8	0,98%	2	0,69%
Sobrevivência=12	379	13,73%	17	17,71%	221	23,09%	117	14,36%	61	21,18%
	100%		100%		100%		100%		100%	

Nascimentos e sobrevivência por CAE a dois dígitos (1995-2006)

	15		16		17		18		19	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
Nascimentos	612	12,45%	1	0,02%	323	6,57%	919	18,69%	278	5,65%
Sobrevivência=1	182	29,74%	0	0,00%	73	22,60%	269	29,27%	95	34,17%
Sobrevivência=2	70	11,44%	0	0,00%	52	16,10%	133	14,47%	29	10,43%
Sobrevivência=3	44	7,19%	0	0,00%	34	10,53%	89	9,68%	28	10,07%
Sobrevivência=4	41	6,70%	0	0,00%	17	5,26%	66	7,18%	17	6,12%
Sobrevivência=5	30	4,90%	0	0,00%	26	8,05%	64	6,96%	13	4,68%
Sobrevivência=6	34	5,56%	0	0,00%	22	6,81%	121	13,17%	33	11,87%
Sobrevivência=7	25	4,08%	0	0,00%	21	6,50%	31	3,37%	12	4,32%
Sobrevivência=8	16	2,61%	0	0,00%	3	0,93%	23	2,50%	7	2,52%
Sobrevivência=9	12	1,96%	1	100,00%	2	0,62%	13	1,41%	7	2,52%
Sobrevivência=10	7	1,14%	0	0,00%	10	3,10%	15	1,63%	1	0,36%
Sobrevivência=11	5	0,82%	0	0,00%	6	1,86%	11	1,20%	2	0,72%
Sobrevivência=12	146	23,86%	0	0,00%	57	17,65%	84	9,14%	34	12,23%
	100%		100%		100%		100%		100%	

	20		21		22		23		24	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
Nascimentos	437	8,89%	33	0,67%	292	5,94%	1	0,02%	85	1,73%
Sobrevivência=1	132	30,21%	11	33,33%	90	30,82%	1	100,00%	23	27,06%
Sobrevivência=2	62	14,19%	3	9,09%	58	19,86%	0	0,00%	11	12,94%
Sobrevivência=3	46	10,53%	5	15,15%	20	6,85%	0	0,00%	8	9,41%
Sobrevivência=4	33	7,55%	1	3,03%	22	7,53%	0	0,00%	3	3,53%
Sobrevivência=5	26	5,95%	1	3,03%	20	6,85%	0	0,00%	9	10,59%
Sobrevivência=6	18	4,12%	2	6,06%	12	4,11%	0	0,00%	4	4,71%
Sobrevivência=7	23	5,26%	2	6,06%	19	6,51%	0	0,00%	7	8,24%
Sobrevivência=8	11	2,52%	2	6,06%	7	2,40%	0	0,00%	5	5,88%
Sobrevivência=9	12	2,75%	1	3,03%	4	1,37%	0	0,00%	0	0,00%
Sobrevivência=10	4	0,92%	0	0,00%	1	0,34%	0	0,00%	0	0,00%
Sobrevivência=11	3	0,69%	0	0,00%	1	0,34%	0	0,00%	2	2,35%
Sobrevivência=12	67	15,33%	5	15,15%	38	13,01%	0	0,00%	13	15,29%
	100%		100%		100%		100%		100%	

	25		26		27		28		29	
	Nº	%								
Nascimentos	72	1,46%	333	6,77%	35	0,71%	675	13,73%	190	3,86%
Sobrevivência=1	19	26,39%	92	27,63%	9	25,71%	180	26,67%	54	28,42%
Sobrevivência=2	10	13,89%	44	13,21%	2	5,71%	100	14,81%	30	15,79%
Sobrevivência=3	8	11,11%	35	10,51%	5	14,29%	69	10,22%	15	7,89%
Sobrevivência=4	3	4,17%	20	6,01%	1	2,86%	49	7,26%	12	6,32%
Sobrevivência=5	5	6,94%	16	4,80%	1	2,86%	35	5,19%	6	3,16%
Sobrevivência=6	3	4,17%	12	3,60%	4	11,43%	37	5,48%	12	6,32%
Sobrevivência=7	3	4,17%	11	3,30%	2	5,71%	24	3,56%	11	5,79%
Sobrevivência=8	2	2,78%	17	5,11%	0	0,00%	17	2,52%	6	3,16%
Sobrevivência=9	4	5,56%	8	2,40%	0	0,00%	15	2,22%	4	2,11%
Sobrevivência=10	0	0,00%	7	2,10%	0	0,00%	17	2,52%	3	1,58%
Sobrevivência=11	0	0,00%	4	1,20%	2	5,71%	17	2,52%	1	0,53%
Sobrevivência=12	15	20,83%	67	20,12%	9	25,71%	115	17,04%	36	18,95%
	100%		100%		100%		100%		100%	

	30		31		32		33		34	
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
Nascimentos	1	0,02%	39	0,79%	16	0,33%	39	0,79%	30	0,61%
Sobrevivência=1	1	100,00%	13	33,33%	6	37,50%	7	17,95%	5	16,67%
Sobrevivência=2	0	0,00%	7	17,95%	1	6,25%	6	15,38%	9	30,00%
Sobrevivência=3	0	0,00%	5	12,82%	0	0,00%	2	5,13%	2	6,67%
Sobrevivência=4	0	0,00%	2	5,13%	2	12,50%	3	7,69%	1	3,33%
Sobrevivência=5	0	0,00%	5	12,82%	0	0,00%	3	7,69%	1	3,33%
Sobrevivência=6	0	0,00%	0	0,00%	1	6,25%	3	7,69%	3	10,00%
Sobrevivência=7	0	0,00%	4	10,26%	1	6,25%	1	2,56%	1	3,33%
Sobrevivência=8	0	0,00%	1	2,56%	1	6,25%	1	2,56%	0	0,00%
Sobrevivência=9	0	0,00%	0	0,00%	1	6,25%	0	0,00%	2	6,67%
Sobrevivência=10	0	0,00%	0	0,00%	0	0,00%	1	2,56%	0	0,00%
Sobrevivência=11	0	0,00%	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Sobrevivência=12	0	0,00%	2	5,13%	3	18,75%	12	30,77%	6	20,00%
	100%		100%		100%		100%		100%	

	35		36		37	
	Nº	%	Nº	%	Nº	%
Nascimentos	35	0,71%	460	9,36%	11	0,22%
Sobrevivência=1	5	14,29%	177	38,48%	3	27,27%
Sobrevivência=2	2	5,71%	59	12,83%	2	18,18%
Sobrevivência=3	5	14,29%	31	6,74%	1	9,09%
Sobrevivência=4	1	2,86%	28	6,09%	2	18,18%
Sobrevivência=5	1	2,86%	22	4,78%	1	9,09%
Sobrevivência=6	3	8,57%	16	3,48%	0	0,00%
Sobrevivência=7	2	5,71%	17	3,70%	0	0,00%
Sobrevivência=8	0	0,00%	13	2,83%	0	0,00%
Sobrevivência=9	0	0,00%	12	2,61%	0	0,00%
Sobrevivência=10	1	2,86%	8	1,74%	0	0,00%
Sobrevivência=11	0	0,00%	7	1,52%	1	9,09%
Sobrevivência=12	15	42,86%	70	15,22%	1	9,09%
	100%		100%		100%	