ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

INNOVATION AND THE SURVIVAL OF NEW FIRMS ACROSS BRITISH REGIONS

Christian Helmers and Mark Rogers

Number 416
December 2008

Manor Road Building, Oxford OX1 3UQ

Innovation and the Survival of New Firms Across British Regions*

Christian Helmers
Wolfson College
University of Oxford

Mark Rogers
Harris Manchester College
University of Oxford

December 2008

ABSTRACT

This paper analyses the survival of the complete cohort of more than 162,000 limited companies incorporated in Britain in 2001 over the subsequent five-year period. For this purpose, we estimate firms' hazards of failure and survival functions using nonparametric and semi-parametric techniques. The paper focuses on two important policy-related issues. The first is to what extent survival rates vary across regions in Britain. A second, and related, policy issue concerns innovation. The data available allows us to look at the intellectual property (IP) activity of all British firms, including that of the 162,000 new firms in 2001. The results indicate substantial differences in survival rates across regions, and also that IP activity is associated with a higher probability of survival. These differences across regions, and the importance of IP activity, remain when we condition on a large range of regional, industry and firm-level characteristics shifting firms' hazards of failure.

KEYWORDS: Start-ups, firm survival, region, IP.

JEL Classification: D21, L25, L26, M13

*Authors' email addresses: christian.helmers@wolfson.ox.ac.uk and mark.rogers@hmc.ox.ac.uk

1 Introduction

The objective of this paper is to analyze the survival of the complete cohort of more than 162,000 limited companies incorporated in the United Kingdom in 2001 over the subsequent five-year period. For this purpose, we estimate firms' hazards of failure and survival functions using non-parametric and semi- parametric techniques. Estimates of survival rates are interesting in their own right but we provide new results relating to i) regional factors and ii) intellectual property. The two main variables used to assess regional factors are unemployment rates and house prices. For intellectual property, the data available allows us to look at the patenting and trade marking activity of all UK firms, including the 162,000 new firms in 2001. While empirical studies of patenting and performance are common for larger firms, there are relatively few for small firms and virtually none for start-ups. The use of trade mark data is also novel, although there is recent evidence that trademarks proxy innovative activity (Greenhalgh and Rogers, 2006, 2007). While both the regional and IP aspect of firm survival are of interest to economists and policy makers alike, we are also interested in combining these two aspects. For example, does a firm that is IP active in North East England have the same chance of survival as an IP active firm in South East England?

The paper is structured as follows. The next section provides a short summary of the relevant literature on firm survival, as well as studies that use patent and trade mark data. The third section discusses the process of new firm creation, survival and how this relates to the empirical estimation. The fourth section discusses the Oxford Firm Level IP database, which makes the analysis possible. We then provide an overview of survival rates across the different British regions, followed by the estimates of the Cox proportional hazard model.

2 Related literature

2.1 Survival analysis

Survival analysis is commonly used in the economics literature to analyze the determinants of firm failure. These papers are strikingly similar in their estimation approaches, but differ substantially in terms of the type and depth of the data studied. The main variables that have been found to play a role are (start-up) firm-size, ownership, industry growth, age and the number of firms entering and leaving the industry.

Disney, Haskel and Heden (2003) provide an analysis of around 140,000 manufacturing establishments in the UK from 1986 to 1991. They find that only around 35 percent of new entrants survive after five years. Those that do survive are around four times larger than new entrants (in terms of employment). They use non-parametric estimates and a Cox proportional hazards model and, although they are limited in terms of possible explanators, find that establishments that are part of groups have lower exit rates. This represents, to our knowledge, the only recently published evidence on UK firms. Mata and Portugal (1994) track a cohort of Portuguese manufacturing firms born in 1983 to analyze determinants of their eventual failure. Mata and Portugal's data includes all manufacturing firms with at least five employees. They use the Cox proportional hazards model to find that start-up size of the new-born firms, industry growth and the number of plants operated by the new-born firms reduce the likelihood of failure, while entry into the industry increases the likelihood. These two studies are typical of large sample empirical work; in particular, there are often very few firm level variables available and none that proxy innovation.

Audretsch (1991), who analyzed firm survival at the industry level, did include a variable on innovation. The industry level innovation variables were derived from identifying new products and processes in over 100 industry journals. Audretsch finds that survival rates can be higher in more innovative industries. Audretsch and Mahmood (1995) use this industry level variable in a Cox proportional hazards model of survival for around 12,000 US manufacturing firms founded in 1976. They identify a number of factors associated with firm survival. The larger the new firm's employment, relative to industry's minimum efficiency scale (MES), the higher survival. New firm survival is lower when industry innovation rates are high (and there is no counteracting effect from small firm innovation rates as in Audretsch, 1991). They also find higher unemployment rates to be associated with lower survival rates. They also find that stand-alone companies have a lower likelihood to fail compared to branches or subsidiaries. This finding may appear counterintuitive, as one might expect subsidiaries to receive support from parents. However, subsidiaries may be under pressure to perform and their parents may be quick to close them down if they do not.

Cefis and Marsili (2005) use firm-level dummies to distinguish between innovating and non-innovating firms in a sample of 3,275 new-born Dutch firms over the period 1996-2003. They find that innovators benefit from an innovation premium giving them higher life expectancy (11 percent higher survival time for innovating firms). They also distinguish between product and process innovations,[1] finding that firms introducing process innovations experience a 25 percent increase in survival time, while product innovations do not have any statistically significant effect. Using Pavitt's (1984) sector

[1]The information on innovation comes from the Community Innovation Survey (CIS-2).

classification according to their technological intensity, they also find that firms in technologically more intensive sectors have higher chances of survival, which stands in direct contrast to the findings by Audretsch and Mahmood (2005). According to their results, firm survival is also influenced by age and size. However, size has decreasing effects on firm survival.

Cockburn and Wagner (2007) use a small sample of 356 internet-related firms that made an IPO on the NASDAQ during the dot-com boom between 1998 and 2001. All firms are in Internet or software industries. During this boom period, the US Patent and Trademark Office made it possible to apply for patents on software and notably on business methods, and the authors match patent data to their sample. Hence, they can test whether patents of that category had any different effect on survival compared to patents in 'traditional' categories. They find that patenting is positively related with firm survival. Firms that applied for more patents were less likely to exit the market. However, due to the specific sample used by Cockburn and Wagner, the authors are unable to control for the environment in which the firms operate by using industry-level variables. Finally, Jensen et al. (2006) use a sample of 260,000 Australian firms that were alive at some stage over 1997-2003. They also have patents and trademark data at the firm level. Using a piece-wise exponential hazard function they find that trademarking is associated with greater survival for 'new' firms (post 1997 entry), but that patenting has no significant association.

2.2 Patents and trade marks

Schumpeter (1934) suggested a distinction between inventions, describing new discoveries, and innovation, describing the successful implementation of an invention into a commercial product. Intellectual property in form of patents and trade marks can capture both aspects of Schumpeter's typology.

Since Schmookler (1966), many researchers have used patents as an indicator of innovation, even though it really only indicates invention. As such, patents have been found to be positively correlated with firms' productivity and market value (Hall, 2000, Griliches, 1990, Bloom and Van Reenen, 2002, Klette and Kortum, 2004). Thus, firms with a larger number of patents should be expected to be more innovative and, therefore, have a competitive advantage. Patents may also serve strategic purposes, such as deterring and blocking competitors from entering a certain market (Hall, 2007).

Trade marks have not been extensively used in previous studies, but there is some evidence that they proxy some aspects of the end of the innovation process (just as patents sometimes proxy the start). Fundamentally, trade marks provide a signalling function to indicate a certain level of quality or other characteristics that consumers can expect from a product. As such, trade marks help consumers reduce search costs and hence producers are able to sell larger quantities or charge a higher price (Landes and Posner, 1987). This availability of such a signalling function can be integral to the innovation process, since trade marks generate some protection against imitation. Hence, Greenhalgh and Rogers (2007) argue that trade marks proxy product innovation by firms (see also Mendoca et al., 2004). Overall, trade mark data is likely to proxy some new product innovation and also a range of activities that are associated with product innovation, such as marketing, advertising and design. Perhaps more importantly, trade mark data may capture which firms are better at this bundle of activities.

3 Idea generation, start-ups and modeling survival

As indicated above, empirical studies on firm survival are generally not based on structural models. However, it is valuable to outline the basic processes at work. The process is driven by entrepreneurs who generate a constant stream of new business ideas and then, if the idea is considered above a certain value, set up firms to capitalize on these. These ideas are then tested in the market place and, as is well known, there is a large failure rate. Firms will continuously evaluate their prospects and, when they see no likelihood of success, they will exit.[2] We can think of failure stemming from one, or both, of two principal aspects:

i) The underlying quality of the firm's idea relative to others in the market place

ii) The resources available to the entrepreneur to capitalize on the idea.

Resources could include finance, and the related capital, labour and materials; but they could also include knowledge about production methods or markets. There is also an important role for competitive pressure, by either incumbents or other new entrants, in affecting a firm's survival chances. This competitive pressure can be in various forms, such as substitute goods reducing demand, or competitor firms using up, or raising the costs of, resources (e.g. skilled labor). In a simple situation one can think of competition acting to select the best ideas but, in reality, there is an interplay between ideas and resources. A firm with an idea whose quality is below average may still be able to survive if it has more resources.[3]

These observations mean that an empirical model should allow exit to occur at any time and that the probability of exit should depend on a range of variables proxying

[2]This is, in fact, a legal requirement since it is an offence for directors to allow a firm to continue trading if they adjudge the firm to not be a 'going concern'.

[3]As marketing experts will quickly tell you, the best product idea in the world will get no where if no consumer finds out about it.

the quality of idea, resources available and intensity of competition. Previous studies have focused on competitive conditions, and have also used firm age and size as proxies for a firm's resources. Our data is rich enough to only look at firms of the same age (the 2001 cohort) removing the need to rely on age as a proxy for resources. However, we also include firm size as a further conditioning variable. In addition, we introduce three new variables that capture the resources available to entrepreneurs: whether the firm is located near a university, the number of directors of the firm and the average house price. House prices may be critical since entrepreneurs often use their house as collateral for loans (or remortgage to gain access to funds). Regional differences may be important as availability of resources and market conditions may vary across regions. To investigate these issues we include unemployment rates by county and unitary authority. House prices may also play a role here as they are a proxy for rentals and wage costs. As in previous studies, a range of industry-level competition variables is also included.

Most existing studies, however, do not condition on the quality of the underlying idea. In our analysis IP can be considered as a proxy for a better quality idea hence, ceteris paribus, IP should increase the probability of survival. The IP variables could also proxy better 'resources', such as management quality or human capital at the firm. A further possibility is that the IP variables capture the reduction in competition (which is of course the textbook, legal role of IP). One reason why we are reluctant to stress this last interpretation is that we are using publications of IP, and not grants. Legally, it is only after the grant of a patent (or registration of a trade mark) that the IP right has its full effect. Another reason is that start-up firms may have difficulty in enforcing IP rights, due to the cost of enforcement in the courts. These issues mean that, in this paper at least, the IP variables are best interpreted as proxies for the quality of the idea, as well as management and human capital.

4 The OFLIP database

4.1 General

The data used for the analysis comes from the Oxford Firm Level IP (OFLIP) database. The database draws on the Financial Analysis Made Easy (FAME) data that covers the entire population of registered UK firms (FAME downloads data from Companies House records).[4] OFLIP contains additional information on the IP activity of firms in

[4] In this paper we use firms to mean registered firms. Hence firm refers to the legal entity that organizes production, in contrast to census-type data that uses the plant or production unit.

the form of patents and trade marks. OFLIP has been constructed by matching the FAME database and a number of firm-level IP datasets.[5]

The FAME database is a commercial database provided by Bureau van Dijk.[6] To construct the data set, the December 2006 edition of FAME has been used. It covers around 2.04 million active firms. For all of these firms, basic information, such as name, registered address, firm type, and industry code are available. Availability of financial information varies substantially across firms. The smallest firms are legally required to submit only very basic balance sheet information such as shareholders' funds and total assets. The largest firms provide a large range of profit and loss information as well as detailed balance sheet data. Importantly, the FAME database also lists around 0.9 million so called 'inactive' firms. These inactive firms are those that have exited the market and belong to one of the following categories: dissolved, liquidated, entered receivership or declared non-trading. The fact that FAME tracks inactive firms allows us to identify all firms entering and exiting the market throughout the five-year period observed.[7] FAME gives exact dates for market entry in the form of a firm's incorporation date. To determine date of exit we use the date that the last set of accounts were filed.[8] The firm-level data is augmented by regional data on house prices and unemployment (at the county and unitary authority level). Further details are in Appendix A.

4.2 IP data

The IP data used for the construction of the OFLIP database comes from three different sources: the UK IP Office, Marquesa Ltd. and the European Patent Office (EPO) ESPACE Bulletin. Data on UK patent publications were supplied by the UK IP Office. Marquesa Ltd supplied data on UK trade mark publications and Community (OHIM) marks registered. Data on EPO publications by British entities was downloaded from ESPACE Bulletin DVD 2006/001. For our analysis, we use publications of UK patents, trade marks and EPO patents, as well as registrations for Community trade marks (further details and a discussion of the IP data is in Appendix A).

[5]For details on the matching process and further details on the database see Rogers, Helmers, and Greenhalgh (2007).

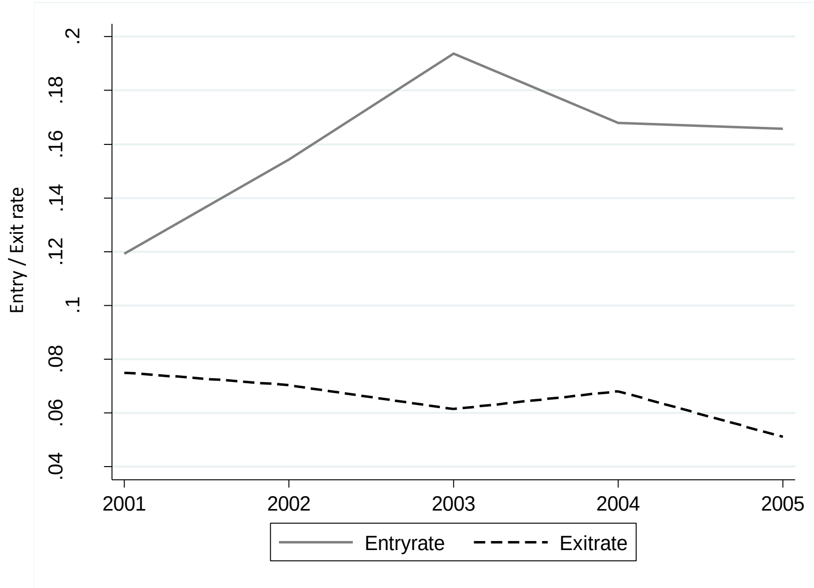
[6]http://www.bvdep.com/en/FAME.html

[7]As such, the data set is much wider in coverage than for example the one used by Bridges and Guariglia (2006) who also use the FAME database for their firm survival analysis.

[8]This is an important advantage of OFLIP over census data sets used in previous work, such as Dunne et al. (1987). In cases where the exit date was missing, the date of last annual return or the date of last transaction at companies house is used instead.

5 Overview of survival

Figure 1: Entry / Exit rates for population of UK firms 2001-2005



Entry and exit rates for the population of registered firms in the UK between 2001 and 2005 are plotted in Figure 1. The exit rate slightly decreased over the period observed, while the entry rate substantially increased between 2001 and 2003. The reason for this was a change in the tax law in 2002. The UK Government introduced a zero percent rate of corporation tax for registered companies with profits up to £ 10,000. The result of this was to rapidly increase the number of sole traders that formed companies (such as tradespeople and consultants). Over the 5-year period shown, entry rates exceed exit rates on average by 9.5 percentage points. This implies an average annual net increase of around 170,000 firms in the UK. The influence of the tax change is therefore substantial and means that entry rates should be treated with caution.[9] In particular, it is thought that many IT consultants and business service sole traders converted into registered companies in 2002 and 2003. To counter this somewhat unintended effect, the government introduced new legislation taxing all the company's profit at 19 percent if the profits earned were distributed to shareholders. Hence, the

[9] It is not clear whether exit rates would also be biased downwards. In general, we might expect some exit because a firm wants to convert back to a sole trader (avoiding the slightly higher administrative costs of a registered company). The tax change after 2002 would appear to outweigh these administrative costs, hence we might expect exit to be slightly lower.

zero percent tax rate applied only up to £ 10,000, if the profits were retained.[10]

Geroski (1991) found there was a positive correlation between entry rates and exit rates across industries in the UK. Using data from 1987 he reported a correlation coefficient of 0.79 for a sample of 95 industries. Here, we find a correlation coefficient of only 0.20 for the five year period for a sample of 252 industries. Looking at individual years, the correlation coefficients vary between 0.004 in 2003 and 0.28 in 2001. Given the tax driven increase in entry rates in 2002 and 2003, the low correlation might be expected. The possibility that the tax changes had a non-uniform effect on entry across industries, also means that it is not possible to compare the correlation coefficients obtained here with Geroski's.

Our analysis focuses on the cohort of firms incorporated in 2001, this avoids the problem arising from increased entry rates due to the change in tax regime in 2002. Using only firms incorporated in 2001 also avoids problems of left truncation (i.e., all firms are observed from the onset of failure risk). For our data, there were a total of 162,469 new firms registered in 2001. The survival rates for this cohort are as follows. In 2002, 161,493 or 99.4 percent were still in business. The high figure simply reflects that a registered company almost always survives to file its first set of accounts. By 2003, 140,215 or 86.3 percent were still in business, with 76.4 percent left by 2004, and 70.24 percent by 2005.[11]

5.1 Regional differences

Table 1 shows new firm formation and failure rates by region, where regions are defined according to the definition of Regional Development Agencies (RDAs) (it also shows high-tech and IP-active firms and failure rates - see below for explanations).[12] The failure ratio is computed as one minus the fraction of newly incorporated firms still alive in 2005 i.e., $1 - \frac{Firms_{2005,rda}}{Firms_{2001,rda}}$. Specifically, the table compares the failure rate for IP-active and non-IP active firms (159,743 in 2001) by RDA.[13] The London and

[10]These somewhat confusing regulations are now to be abolished and only a small company tax rate of 19 percent will apply (see HM Revenue and Customs website http://www.hmrc.gov.uk/).

[11]Note that when the data is used for formal survival analysis, we encounter the problem of right-censoring. This means that for all firms of the 2001 cohort that have not exited by the end of 2005, failure remains unobserved. The only thing that is known about these firms is that failure occurs some time between $[t, \infty)$.

[12]The Regional Development Agencies Act 1998 has led to the establishment of 12 RDA's in Britain. We also disaggregate RDAs into county and unitary authority level data to provide further insight into spatial patterns of survival. In order to allocate firms to counties and unitary authorities within RDAs, we matched firms' post codes with counties' and unitary authorities' post code areas. We are aware of the fact that this may cause problems for multi-location firms. However, since we are analyzing start-up companies, we are confident that postcodes reported in FAME indeed correspond to the actual physical locations of firms. Note that due to data restrictions, we exclude Northern Ireland from our analysis.

[13]A firm is counted as IP active if it has had any form of IP within the period 2001-2005.

North West RDAs have the highest failure rate: in London and in the North West, 35.7 percent and 30 percent respectively of 2001 firms fail by 2005. In contrast, the lowest failure rates are in the Scottish Highlands and the South West regions.

Table 1: Number of new firms and failure rates by RDA

RDA	No. of			Failure rate			
	new firms	IP-active	High-tech	IP-inactive		IP-active	
				all firms	high-tech	all firms	high-tech
South West	11,687	183	46	0.254	0.143	0.082	0
South East	26,273	442	88	0.284	0.218	0.104	0.1
London	46,255	808	66	0.357	0.339	0.113	0
East of England	11,763	198	66	0.288	0.236	0.091	0
East Midlands	7,873	147	25	0.261	0.208	0.122	0
Yorkshire	9,018	169	23	0.273	0.286	0.101	0
North West	14,957	223	50	0.301	0.298	0.148	0
West Midlands	17,482	265	61	0.289	0.241	0.094	0.286
North East	3,112	47	13	0.257	0.385	0.043	n.a.
Wales	4,387	76	25	0.268	0.261	0.079	0
Scotland	9,218	164	21	0.261	0.263	0.128	0.5
Highland	444	4	0	0.221	n.a.	0.5	n.a.

Notes:
1. A firm is counted as IP active if it has had any form of IP within the period 2001-2005.
2. High-tech firms belong to HS categories 244, 353,33,35,30.

Figure 4 also shows failure rates but this time according to the average for counties and unitary authorities (Great Britain has 142 of these covering the country). The figure shows five different 'shades' of failure rates, based on the quintiles of the distribution. For example, the darkest shaded counties have a failure rate of between 31 percent and 47 percent; the white regions have the lowest failure rate (between 5 percent and 23 percent). Clearly, these are major differences in failure rates. Note also that there are differences across counties within RDAs. This indicates that either the support given to new firms varies within RDAs or, more likely, that there are a great many other mechanisms at work in driving failure rates. Figure 4 indicates that the high failure rates occur in London, Kent, Buckingshire and the M4 corridor. One might suggest that the high rates of new firm formation in these regions make this an expected result, but clearly these regions also have greater demand and other characteristics that drive the high entry rates. An alternative explanation is to say that competition is more intense in these regions but this is, in many ways, a tautology. Explaining why competitive intensity varies across regions is a more difficult task. However, it is not only regions with high rates of firm formation that have high failure rates, the figure also shows some other high failure rate areas: Weston super Mare (directly west of London and M4 corridor); some authorities in the South Wales valleys, and also Pembrokeshire; Herefordshire; Liverpool to Manchester corridor; and South Ayrshire in Scotland.

Table 1 shows that there were 2,726 IP-active firms in the 2001 cohort across the

2001-2005 period. The unique aspect of the database used here is that it contains full (population) data of IP active firms. Since financial data on R&D are almost never available for start-up firms, using IP data is one of the few proxies for innovation with population coverage. An alternative possible proxy is the sub-set of firms located in high-tech industries (there are 484 such firms).[14] The columns in the right hand side of the table show the failure rates for IP-inactive and IP-active firms, with each of these split into high-tech and non-high-tech. The failure rate for IP-active firms is lower than for IP-inactive firms for all regions (except the Highlands, where there were only 4 IP-active firms). Within the group of IP-inactive firms, high-tech firms tend to have lower failure rates, although this is not always the case, especially for Yorkshire and the North East RDAs.

The differences highlighted here are interesting as background to innovation policy but we are, ultimately, interested in why differences occur. One major aspect may be that failure rates differ across industries and that regions differ according to their mix of industries. This is something we analyse in more detail in the remainder of the paper.

5.2 Firm-level and industry-level summary statistics

Table 2 summarizes the basic characteristics of all firms at the beginning of the period and of those that survived by 2005. Due to the enormous variance in the data with regard to financial information on firms, we focus on median values. The first column summarizes the data for the full 2001 cohort, while the second column displays the 2001 values for those firms that survived by 2005. Column three presents the values for the survivors in 2005. Comparing the values for firms' financial variables between columns 1 and 3 shows a fourfold increase in median assets and a twofold increase in median turnover. Comparing surviving firms and the complete cohort in 2001 shows that survivors had already in 2001 higher median assets and turnover. In summary, the statistics suggest both strong growth of surviving firms and the exit of smaller firms. The data also include the number of directors of a firm, which is used as a proxy for the managerial skill pool available to a firm. Overall, we observe that the median number of directors falls from 4 to 3 between 2001 and 2005 for surviving firms and the complete cohort which suggests that firms start off with more directors than needed to run their business efficiently.

For the binary variables, we look at mean values to obtain the percentage share within the sample. As such, the number of subsidiaries increased by 2.42 percentage

[14]We follow the standard OECD definition: high tech firms as those in UK SIC 244, 353, 33, 32, and 30.

points between 2001 and 2005 comparing the whole sample in 2001 with the firms still existing in 2005, which indicates a positive relationship between survival and being part of a holding company. This is corroborated by looking at the 2001 data for surviving firms which shows that the share of subsidiaries is significantly higher (8.5 percent) than for the entire sample (6.6 percent). Foreign ownership increased only slightly, which has to be interpreted as weak evidence for drop out of domestically owned firms. The university variable indicates whether a firm is located at British university science park.[15] This variable captures possible effects arising from cooperation agreements with research institutes, as for example knowledge spillovers, but also the availability of skilled human resources. Overall, the number of firms linked to universities is very small for newly incorporated firms. The subsidiary, foreign ownership and university data are derived from the last set of accounts (e.g. 2005 for survivors), hence these figures do not capture any changes in a firm's status through time. Finally, the percentage share of firms that patent or take out a trade mark is tiny for the entire population of firms, although there are substantial differences across sectors. It is evident from Table 2, that IP activity increases substantially over time. This holds true both for the entire sample and the 2001 data for the surviving firms.

Table 2: Firm characteristics of the 2001 cohort (2001 vs. 2005)

	Year 2001 All	Year 2001 Survivors	Year 2005
Variable		Median	
Total assets (1,000 £)	7	13	28
Turnover (1,000 £)	55	65	111
No. of directors	4	4	3
Variable		Mean	
Patent	0.05%	0.06%	0.17%
Trade Mark	0.19%	0.23%	0.52%
Patent / Trade Mark	0.23%	0.28%	0.67%
Subsidiary	6.63%	8.49%	9.05%
Foreign owned	2.39%	2.73%	2.80%
University	0.08%	0.09%	0.10%

To gain further insight about the IP active firms, Table 3 presents a summary of their characteristics and IP activity. The number of IP active firms more than doubles within the five-year period. Yet, the share of IP active firms within the 2001 cohort remains tiny at around 0.7 percent in 2005. UK trade marks are the most common

[15] This is derived from searching all of the firm's current address for the word university, hence any university business park is likely to be picked up by this method.

form of IP used by the firms. The relatively high number of firms with EPO patents is surprising.[16] The increase in the number of firms obtaining Community trade marks outpaced the increase in firms taking out UK trade marks. This might point towards stronger international business orientation of surviving firms.

Table 3: Number of IP active firms / Number of patents and trade marks 2001-2005

Year	No. of firms				
	IP active	UK TM	Community TM	UK patent	EPO patent
2001	372	262	57	50	36
2002	940	744	112	77	58
2003	918	611	186	128	81
2004	759	451	187	108	103
2005	761	436	216	115	114
Total	3,750	2,504	758	478	286
Year		Average no. of TM		Average no. of Patents	
		UK TM	Community TM	UK patent	EPO patent
2001		1.47	1.69	1.62	1.78
2002		1.43	1.50	1.75	1.47
2003		1.47	1.35	1.37	1.49
2004		1.54	1.35	1.65	1.38
2005		1.53	1.52	1.74	1.58
Total		1.48	1.44	1.61	1.50

6 Survival Analysis

The previous sections have shown that about 30 percent of all newly incorporated firms failed over the four year period 2001-2005. In this section, we analyze firm failure rates using survival analysis.

Survival analysis describes the time that elapses from a certain starting point, for example birth, until a specific event occurs, for example death. Hence, the dependent variable in survival regression analysis is time. Since time is obviously positive or zero, the data is usually not normally distributed. This could easily be dealt with using standard econometric techniques. However, besides the conceptual problem of assuming that time is normally distributed, in survival analysis the data is also right-censored. This means that individuals are observed only during a certain interval and the event of interest does not occur for all individuals in the sample within that pe-

[16] Obtaining an EPO is much more expensive than a UK patent and possibly offers protection throughout the EU, hence smaller firms might be thought to use EPO patents infrequently.

riod. Standard statistical techniques are unable to deal with this kind of censoring. In contrast, survival techniques deal well with this problem and address also the problem of non-negativity.

To see more clearly the link between firm behavior and our empirical approach, assume that the number of firms n in the economy evolves over time according to a $M = (M_m)_{m \geq 0}$ discrete-time Markov chain with $M_0 = 0$ and transition matrix P. Firms can influence a vector of firm characteristics denoted as $x_t \in X$. Hence, the maximization problem is to find the optimal policy for $\{x_{t+1}\}_{t=0}^{\infty}$ given initial values of (n_0, x_0) maximizing

$$V(n,x) = E\{\sum_{t=0}^{\infty} \delta^t \pi(n_t, x_t)\} \tag{1}$$

where $0 < \delta < 1$ is a discount factor, subject to

$$n_{t+1} = g(n_t, x_t, \epsilon_{t+1}) \tag{2}$$

With ϵ_t being a sequence of i.i.d. variables. The Bellman equation is therefore for each $i \in [1, ...M]$

$$v(n_i, x) = \max_{x \in X}\{\pi(n,x) + \delta \sum_{j=1}^{m} P_{ij} v(n_j, x')\} \tag{3}$$

Where $P_{ij} = Pr[n_{t+1} = n_j \mid n_t = n_i]$ for $i = 1, ..., M$ and $j = 1, ..., M$, which indicates the probability that transition from state i to j occurs and $P_{ij} \geq 0$ and $\sum_{j=1}^{m} P_{ij} = 1$.

Now, given the transition matrix P, firms spend a length of time T before transition from i to j occurs. In our case, we assume that $M = 2$ such that a state may be either $= 0$ if the firm is alive or $= 1$ if the firm is dead, where death is an absorbing state. This implies that

$$M = \begin{cases} 0 & \text{if } 0 \leq t < T \\ 1 & \text{if } t \geq T \end{cases}$$

In survival analysis, we estimate the probability of a firm making a transition from state i to state j. Alternatively, we can specify the hazard rate of a firm to exit the market, denoted by $h(t)$, as

$$h(t) = Pr(T = t|T \geq t) = \frac{f(t)}{S(t)} \tag{4}$$

where $S(t)$ is the survival function with $S(t) = 1 - F(t) = \Pr(T > t)$. The hazard rate is interpreted as the conditional probability density that failure occurs at time t conditional on the firm having survived up to that point in time. Precisely these hazard rates constitute the non-zero elements of the transition matrix P_{ij} from the Bellman equation (3). Hence, in estimating the hazard rate of failure, we estimate firms' probabilities of transition from state i to j as specified in the optimal control problem in (3).

We first estimate simple hazard and survival functions using nonparametric techniques, only distinguishing between firms that are IP active and those that are IP inactive. In a second step, we move on to include covariates, using a semi-parametric model, namely the Cox proportional hazard model and use more refined IP indicators. It appears to be more appropriate to estimate a proportional hazard model than a accelerated lifetime model (AFT). In a proportional hazard model covariates shift the hazard function, while in the AFT model, the hazard function is identical across firms, but firms move faster along it according to their covariates. For our purpose, shifts in the hazard function lend themselves more easily for interpretation.[17]

6.1 Non-parametric Approach

We start by estimating the survival function $S(t)$ using the Kaplan-Meier (1958) estimator. The Kaplan-Meier estimator is a simple frequency non-parametric estimator, i.e., it does not make any assumptions about the distribution of failure times or how covariates shift the hazard function. We will use it to estimate separate survival functions for IP active and IP inactive firms and for all British regions and test whether they are statistically different.

The Kaplan-Meier estimator is given by

$$\hat{S}(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{5}$$

where n_i denotes the number of firms in the risk set at t_i and d_i the number of failures at t_i. The product is over all observed failure times less than or equal to t.

[17] There are very few examples in the literature employing AFT models. For example Cefis and Marsili (2005) use the AFT model but only because their main variable of interest, the innovation variable, violates the proportional hazards assumption.

Since the Kaplan-Meier estimator estimates the hazard or survival function for each period of risk, we first group firms into year intervals for ease of interpretation. However, since we are grouping continuous data into discrete intervals, we use the so called Lifetable estimator to adjust for grouping. The Lifetable estimator produces an estimate centered on the midpoint of the interval in order to account of firms leaving at different times within the year-interval (Jenkins, 2005). As one major objective of this paper is to estimate the effect of firm's innovative activity on its survival probability, we group the data set into IP active and IP inactive firms and estimate the corresponding survival functions for each group. Since IP activity is a proxy for innovation this can be viewed as a test of survival of innovative firms versus non-innovative firms. We summarize firms' overall IP activity with a single dummy variable taking the value of one if a firm has obtained any form of IP over the period 2001 to 2005.

Figure 2: Survival rates for IP-active and IP-inactive firms

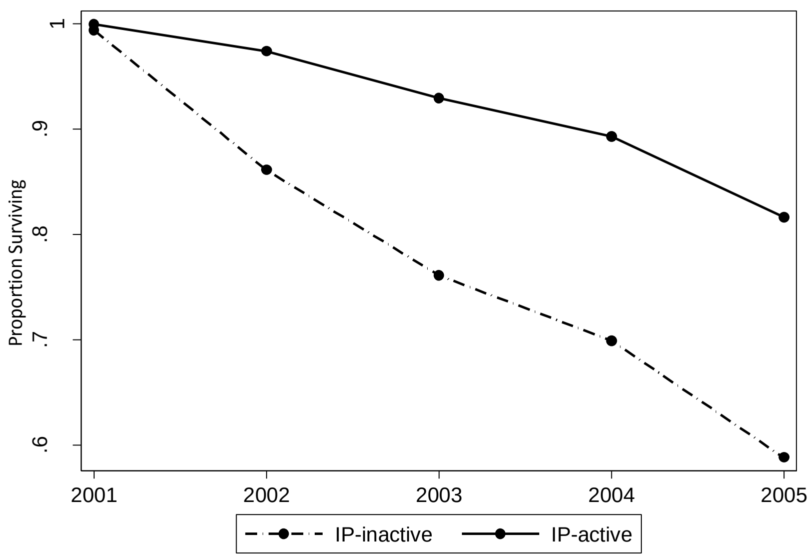


Table 4 shows the results for the Lifetable estimates for both groups. To test more formally, whether there is a difference across the two groups, we use the log rank test. The null hypothesis of the log rank test is that there is no difference between groups.[18]

[18]One common criticism of the use of the log rank test is that it gives too much weight to later event times as the number of observations in the risk sets become relatively small. This is not the case in our sample as 70 percent of the population survive the five-year interval studied which gives more than 114,000 observations in 2005. We also expect the test to be appropriate as it is best suited as a test for differences between groups when the hazards of the groups are proportional to each other, which can be seen in Figure 2 to approximately hold.

It clearly rejects the null hypothesis at the 1 percent level for the survival function of IP active and IP inactive firms to be equal.[19] In addition, Figure 2 plots the survival functions for IP-active and IP-inactive groups. The estimated survival function for IP-active firms lies above the survival function for IP-inactive firms throughout the entire time analyzed. For both groups of firms, however, the figure shows that risk of failure increases rapidly during the first two years of existence, which matches our observation from Section 5 that most firms drop out between their second and third year of life. Subsequently, the survival rate risk slightly flattens to become steeper again during the fourth year after incorporation. Table 4 and Figure 2 provide strong indicative evidence that IP may have an important impact on firms' chances to survive.

Table 4: Lifetable estimates for IP active and IP inactive firms

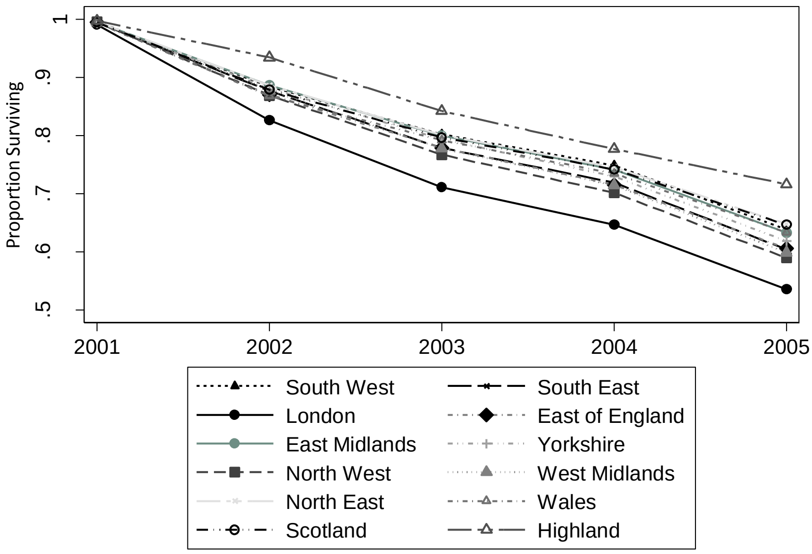
Year	Total no. of firms	No. of failures	S(t)	Std. error
		IP inactive		
2001	159,743	975	0.9939	0.0002
2002	158,768	21,208	0.8611	0.0009
2003	137,560	15,972	0.7611	0.0011
2004	121,587	9,895	0.6992	0.0011
2005	111,688	9,644	0.5881	0.0014
		IP active		
2001	2,726	1	0.996	0.0004
2002	2,725	70	0.9740	0.0031
2003	2,655	121	0.9296	0.0049
2004	2,534	100	0.8929	0.0059
2005	2,432	109	0.8163	0.0089
Log-rank test for equality of survivor functions				
$\chi^2(1) = 499.02$		$\Pr > \chi^2 = 0.0000$		

Figure 3 plots the survival rates by region. It shows that there exist differences across regions. Most noticeably, London and the Highlands stand out. London exhibits are markedly lower survival rate than any other RDA region. Its survival rate for 2005 is 0.5355, while the one for the Highlands is 0.7164. The survival rates for the other regions vary between 0.5895 for the North West and 0.6479 for the North East. We also perform a log rank test for differences in survival rates across RDAs. Also in this case, the rejects the null hypothesis at the 1 percent level for the survival function across regions to be equal.

Using year intervals implies that the only thing that is known about entry and exit

[19]Note that the log rank test yields the same result using continuous rather than grouped data.

Figure 3: Survival rates across RDA regions



is that it occurred at some point within the year interval. This problem is known as interval censoring. To avoid this problem in the subsequent analysis, we construct a continuous time measure by using complete entry and exit dates for all the analysis that follows.

6.2 Proportional hazard Cox model

6.2.1 The model

The descriptive evidence of Section 5, and the non-parametric estimates of Section 6.1, point to substantial differences in firm survival across regions and between IP-active and inactive firms. An important question is what drives these differences?

In order to take account of such covariates influencing a firm's hazard function, we write the hazard function for firm i in a general way as

$$h_i(t) = g(t, \beta_0 + \beta_x x_i) \tag{6}$$

where x_i is a vector of covariates and β are the coefficients to be estimated.

The most popular way to estimate the hazard function is to parameterize it as follows

$$h_i(t) = h_0(t)\exp(\beta_0 + \beta_x x_i) \tag{7}$$

which is called the proportional hazard model. It consists of two components, h_0 is the baseline hazard which depends only on t and not on x_i and the covariates $\exp(\beta_0 + \beta_x x_i)$. The hazard of firm i with covariates x_i is, therefore, multiplicatively proportional to the baseline hazard h_0. Hence, the covariates x_i shift the hazard by a constant proportion relative to the baseline hazard. This also implies that the ratio of firm i's hazard is proportional to any firm j's hazard in the sample

$$\frac{h(t|x_i)}{h(t|x_j)} = \frac{\exp(\beta_x x_i)}{\exp(\beta_x x_j)} = \exp[\beta_x(x_i - x_j)] \tag{8}$$

where $\exp[\beta_x(x_i - x_j)]$ is constant if the covariates do not vary over time.

The functional form of the Cox model (Cox, 1972) is written in the proportional hazard formulation shown above. However, in the Cox model, no specific functional form for the baseline hazard h_0 is assumed. More specifically, due to the way the hazard is computed, h_0 drops out of the calculations. Therefore, the intercept is non-identifiable and the relationship between the hazard rate and the covariates is estimated without taking account of h_0. This implies that only the relative hazard between firms, not the absolute hazard rate, can be estimated. The relative hazard at time t between firms i and j is defined by (8). The coefficients β are estimated using the Partial Likelihood (PL) method instead of the Maximum Likelihood (ML) method. The PL considers only the ordered failure times while the ML estimator focuses on firm spells. Hence, the Cox model is notably ignoring all information available at times when no failure occurs. This is done because the Cox model assumes that spells in which no failure occurs contain little information on the incidence of failure. Cox (1972) has shown that ignoring those spells results only in little efficiency loss.

It is important to test for whether the proportional hazard assumption holds for our data. A range of tests were conducted, including graphical tests plotting the log Kaplan-Meier estimator, $\log(-\log \hat{S}(t))$, against $\log t$ and tests based on the residuals. We find the models estimated below to be appropriate. Finally, as with all such analysis, it is important to stress that the coefficients obtained from estimating (7) should be interpreted as correlations, and cannot be given a causal interpretation.

6.2.2 Results for IP variables

The first column of results in Table 5 are from a parsimonious specification, which simply includes separate dummy variables for a firm having a patent - UK and/or EPO - or a trade mark - UK and/or Community. The results are consistent with the non-parametric estimations.[20] Having a patent is associated with a reduction of the hazard rate by 55 percent, and having a trade mark, with a reduction of the hazard by slightly more than 52 percent.[21] These are substantial effects. The second column of Table 5 presents the results for patent and trade mark counts. These results allow us to distinguish between the effect of the different forms of IP on the hazard of failure. UK and EPO patents show very similar associations with survival, lowering the hazard rate by about 40 percent for each patent. However, there is a marked difference in the effect of UK and Community trade marks. While each UK trade mark is associated with a drop in the hazard rate by about 40 percent, Community trade marks have an effect of only little more than 24 percent. Note, however, that most IP-active firms in the sample only have one patent or one trade mark (see Table 3).

6.2.3 Regional effects

The third specification in Table 5 adds the RDA dummy variables. The Highlands is the benchmark category since we know from Section 6.1 that it is the RDA region with the highest survival rate. Hence, all other regions are measured according to how much more likely firms are to fail than in the Highlands. The results show substantial differences among the RDA coefficients. The coefficient for London is the largest in magnitude, pointing to a 87 percent increased failure rate relative to the Highlands. Another region with a substantially higher failure rate is the North West (55 percent vis-à-vis firms in the Highlands). The magnitude and statistical significance of the patent and trade mark coefficients remain almost unaffected from the inclusion of RDA dummies.

In the last three columns of Table 5 a variety of firm and industry level variables are also added to the specification (see Section 6.2.4), but two further regional variables are also added which we discuss here: unemployment and house prices. The statistically significant coefficient on the unemployment rate indicates that higher unemployment shifts the hazard rate upwards. This is consistent with high unemployment being as-

[20] All estimates in Table 5 have adjusted standard errors to account for within group correlation. Note that the number of firms falls by 28,671 once we include sector-level variables as the corresponding information is missing in FAME for these firms.

[21] Coefficients have to be exponentiated to obtain this interpretation of their magnitude, e.g., $1 - \exp(-0.797) = 0.55$.

sociated with low demand or that high unemployment regions have other structural impediments (such as lack of access to finance, poor infrastructure, etc.). The other region variable is house prices. Since some argue that high house prices may allow entrepreneurs to raise more finance (since their house can act as collateral), high house prices could be associated with higher survival (as struggling firms can borrow more). However, high house prices are also associated with high wage and rental costs, which could reduce survival. Moreover, high house prices may generate greater entry of firms, which could increase subsequent exit rates. After some investigation regarding the functional form of our specification, the use of a quadratic term for house prices seems appropriate. Higher house prices initially shift the hazard rate down, but after a threshold value (£170,000) house prices are associated with an increase in hazard. This is consistent with the idea that house prices are valuable as collateral but at some point the high wage and rental costs offset this advantage.

6.2.4 Industry and firm level effects

The remaining three columns of Table 5 also condition on firm-level and industry-level variables. All industry-level variables have been calculated on the SIC 3-digit level using the *entire* OFLIP database (i.e., *all* firms in the economy). Therefore, they reflect the environment in which the new firms operate. The last column of in Table 5 adds a set of 2-digit industry dummies as a further robustness check.

At the industry level, we include a proxy for market entry and exit costs as proposed by Bernard and Jensen (2007).[22] Capital intensity, computed as the ratio of the amount of firms' assets and labor within each industry, is also added. To measure how important firm size is within each industry, we computed the minimum efficiency scale (MES) as the ratio of average first-year firm size to average firm size within the industry (as proposed by Baldwin and Gellatly, 2003). This provides a better measure of firm size in the industry than the traditional MES measure as used by, for example, Fritsch et al. (2005). To proxy for competition within industries the standard four-firm concentration ratio is included. Alternative proxies for competition were explored, such as the Herfindahl index (Mata and Portugal, 1994) or the price-cost-margin (Aghion et al. 2005); all measures yield very similar results. Finally, the industry growth rate measured by growth in industry-level assets, is also added. Asset growth is used since data coverage is substantially higher than for turnover or employment (for growth variables see Audretsch and Mahmood, 1995).

The firm-level variables have been defined already (see Table 2) and include a uni-

[22] The proxy is constructed as follows Entry/Exit costs = $1 - \min[\text{entryrate}; \text{exitrate}]$.

versity, foreign owned and subsidiary dummies, along with (log of) number of directors plus IP variables.

The coefficients and significance for the industry and firm level variables are very similar across all three specifications (an exception is the coefficient on the four firm concentration ratio which loses significance in the last column of Table 5). The coefficient of the proxy for entry and exit costs is negative, statistically highly significant and very large in magnitude in both specifications (using IP dummies or counts). Bernard and Jensen (2007) also found higher entry/exit costs to lower the failure rate of firms. This is consistent with the idea that high entry/exit costs restrict competition. The positive coefficients on capital intensity implies that increasing capital intensity shifts the failure hazard up. This is consistent with new firms being credit-constrained and therefore find it more difficult to survive in more capital intensive industries. The coefficient on the MES variable is never statistically significant. The four-firm concentration ratio provides an interesting result. According to the negative and statistically significant coefficient, industries with higher concentration warrant a lower hazard of failure. We interpret this result as the outcome of less competition and possibly higher prices (through for example collusion) in more concentrated industries. The significance of this coefficient is lost when industry dummies are added. Contrary to Audretsch (1991), industry growth has a statistically significant effect on the hazard rate in all three specifications.

At the firm-level, firms linked to universities have a lower hazard rate than other firms. The effect is sizeable, with a downward shift of the hazard function by around 59 percent. This result is consistent with spillovers from, and collaboration with, universities, raising survival, and also that small, university-based start-ups may receive assistance during their early years of existence (e.g. subsidized rent). Similar to Audretsch and Mahmood (1995) and Bernard and Jensen (2007), we find that subsidiaries are less likely to exit the market (perhaps due to financial constraints being less binding). Similarly, firms owned by foreign firms are less likely to exit. Finally, increasing the log of the number of directors lowers the hazard of failure by about 28 percent. More directors could indicate more skill or management resources for the firm, or possibly that the quality of the business idea is higher.

Despite adding the above variables, the coefficients on the patent and trade mark dummies remain statistically significant, although coefficient magnitude falls. However, including the industry and firm level variables does reduce significance on IP count variables: only the coefficients of UK patents and trade marks remain statistically significant. Both EPO patents and Community trade marks no longer have a statistically significant effect on the hazard of failure (see last two columns of Table

5). Note also that the coefficients on the RDA dummies remain statistically significant (apart from the coefficient for the North East).

6.3 Differences in the survival rates of IP-active firms across regions

The differences in survival rates across regions are almost always statistically significant even when conditioning on industry and firm level variables in Table 5. Ideally, one would like to know why these differences occur. It is likely that our explanatory variables fail to control completely for important aspects of firm characteristics, such as quality of workers or access to finance. There is also the possibility that the various industry variables fail to capture important aspects of the competitive process in different industries. Generating further variables is an objective of further work in this area and, ultimately, it may be possible to remove the significance of the RDA dummies by achieving statistical independence of survival and regional dummies conditional on these additional variables.

Even with the data at hand we can provide some additional insight into the role of RDAs, and regional factors more generally, by looking at whether IP-active firms experience different survival rates according to their location. In order to do this, the set of RDA dummies is interacted with an IP dummy, and these interaction terms are added to the specification in the last regression in Table 5.[23] Table 6 shows the results for the IP dummy, RDA indicators and their interaction terms. The estimated coefficients of all other covariates have been omitted (as their qualitative implications have not changed relative to the results presented in Table 5).

The first and third columns results in Table 6 do not include the new interaction terms and how the coefficient of the IP dummy is negative and statistically significant. This implies a downward shift of the hazard of failure of approximately 64 percent, confirming the basic findings from Table 5.

Including the interaction terms in column 2, the coefficient of the IP dummy variable turns positive, but is not statistically different from zero. Hence, the effect of IP on a firm's hazard of failure is given solely by the coefficient of the interaction term for the respective RDA in which a firm operates. The coefficients show that IP-activity is negatively correlated with firm failure for all regions, although the coefficients are not statistically significant for the East Midlands and Scotland. These results hold when a set of 2-digit industry dummies is added (column 4).

The results show that IP-active firms in London have lower hazards of failure than

[23]The IP dummy assumes the value of 1 if a firm has either a patent or a trade mark or both. We chose to use an IP dummy instead of separate patent and trade mark dummies due to multicollinearity, which occurs when we included interaction terms of both patent and trade mark dummies with regional dummies.

all other regions, except for the North East and Wales. Overall, these results show that IP-activity has differential effects on the survival of firms across regions. This provides an answer to the question posed in the introduction: does a firm that is IP active in North East England have the same chance of survival as an IP active firm in South East England? Table 6 implies that IP-active firms in the North East have lower hazard rates than firms in the South East.

To summarize these results, Table 7 shows a ranking of regions based on their survival rates (relative to Highlands), with the first column based on survival rates of IP-active firms within that region. The North East is ranked number one since it has the lowest coefficient in Table 6, namely -3.419. The second ranked region based on IP-active firms is Wales (with a coefficient of -2.424) and so on. The second column shows ranking based on conditional survival probabilities (from the coefficients on region dummies in column six in Table 5). The third column shows the unconditional survival rates for IP-inactive firms (from Table 1).

For IP-active firms, column one of Table 7 shows that the North East and Wales have the highest ranking, suggesting that conditions in these regions are conducive to new, innovative firms (although it is possible that new, IP active firms in these regions have better ideas, management or other unobserved characteristic that drives the results). The North East and Wales also have relatively good survival rates using conditional or unconditional ranks for non IP-active new firms (second and third columns). It is interesting to consider the London region. IP-active firms fair relatively well in London, whereas the second and third column show that overall survival rates are poor. This result is consistent with IP-active firms based in London benefiting from the presence of knowledge spillovers, as well as close proximity to specialist skills and finance. At the bottom of the table are Scotland and the East Midlands (although we should note that the difference in coefficients in Table 6 for these regions and others with low ranking is not large).

7 Conclusion

This paper uses a new database to analyze the survival of the complete cohort of all British companies registered in 2001. The database is able to track the outcome of these 162,000 firms to 2005 and also, uniquely, their intellectual property (IP) activity during 2001-2005. The paper focuses on two important policy-related issues. The first is the extent to which IP activity alters the survival outcomes of these start-up firms. The second is whether, and to what extent, survival of start-up firms varies across regions. Intellectual property activity is captured by four measures: UK patent publications, EPO patent publications, UK trade mark publications and Community trade mark registrations. Over the five year period, 3,750 (2.3 percent) of the 2001 start-up firms use one or more of these forms of IP. The most common form of IP used is a UK trade mark, followed by Community trade marks and then UK patents. The dominance of trade marks reflects two factors. First, obtaining a trade mark is cheaper and easier than patent protection. Second, trade marks are used by all sectors in the economy, hence they also capture IP activity by service sector firms. The use of trade marks in empirical research is novel, hence we should clarify what trade mark activity indicates. Our expectation is that a trade mark proxies the launch of a new, or upgraded, product; hence, we expect a trade mark to indicate innovation. The use of patent data in empirical research is more common, although there are few studies on small, start-up firms.

RDAs were set up to coordinate substantial assistance to firms with the core aim of encouraging enterprise, employment and competitiveness (in 2005/6 the budget for the English RDA's was £2.2 billion). Given the existence of these RDAs, it is interesting to ask if and how firm survival varies across them. The answer is that there are large differences. The failure rate of the 2001 cohort of firms by 2005 varies between 36 percent in London to 22 percent in the Highlands (Table 1). Looking at firms in high-tech sectors we also find wide variations in failure rates across RDAs. Table 1 also indicates that the failure rates of IP-active firms varies across RDAs, although in this case the highest rate is 15 percent in the North West (London is now 11 percent). This also indicates that the failure rate for IP active firms are substantially lower than for IP-inactive firms. This is a result that runs through all of our analysis. The differential ranking of regions summarized in Table 7 represents important background for policy makers since it implies the high survival rates of regions like the North East and Wales might have lessons for other regions.

Why might failure rates vary so much across regions? A short answer is that competitive conditions vary. This, in turn, is due to differing industrial structures and economic conditions within each region. Section 5 uses a Cox proportional hazards model to investigate these issues and, at the same time, it includes firm-level IP and other characteristics that might influence survival. The results indicate that higher unemployment reduces rates of survival for new firms. For house prices, the analysis suggests a quadratic relationship, with higher house prices initially showing a positive association with survival, but this relationship stops around £170,000. The results indicate, as expected, that competitive conditions are important. Variables for exit/entry costs, capital intensity and concentration (all defined at 3-digit SIC level) are all significant. IP activity is now disaggregated into its four components and we find that only UK patents and trade marks are significantly positively correlated with survival, with the coefficient for UK trade marks being similar in magnitude to UK patents.

The results also show that being located near a university reduces failure rates, as does being foreign owned or part of a larger group of firms. We also find that firms with more directors have higher survival rates.

The Cox model also includes a set of dummies for the RDAs. If the industry-level and firm-level variables, unemployment rates and house prices across counties and unitary authorities had completely explained the regional differences we would have expected the coefficients on these RDA dummies to be insignificant. This is not the case: there are still significant differences in survival rates across RDAs despite controlling for a range of factors. These differences are due to unobserved factors, some of these could relate to very specific industry factors, but others are likely to relate to the availability of resources or support to start-up firms.

Finally, the paper finds that the survival rates of IP-active firms varies across regions. IP-active firms in the North East, Wales and London have the highest survival rates. These results indicate that the link between survival for a region's IP-active, and its non IP-active firms, is not strong. For example, London ranks third best for survival for IP-active firms, but eleventh in terms of unconditional survival of all firms.

References

[1] Aghion Philippe, Bloom Nick, Blundell Richard, Griffith Rachel, and Howitt Peter (2005): *Competition and Innovation: An Inverted-U Relationship*, Quarterly Journal of Economics, Vol. 120, No. 2, pp. 701-728.

[2] Audretsch David (1991): *New-Firm Survival and the Technological Regime*, Review of Economics and Statistics, Vol. 60(3), pp. 441-450.

[3] Audretsch David and Mahmood Talat (1995): *New Firm Survival: New Results Using a Hazard Function*, The Review of Economics and Statistics, Vol. 77, No. 1., pp. 97-103.

[4] Baldwin John and Gellatly Guy (2003): *Innovation Strategies and Performance in Small Firms*, Edward Elgar.

[5] Bernard Andrew and Jensen Bradford (2007): *Firm structure, multinationals, and manufacturing plant deaths*, Review of Economics and Statistics, Vol. 89, No. 2, pp. 193-204.

[6] Bloom, N. and J. Van Reenen (2002). "Patents, Real Options and Firm Performance." *Economic Journal* 112(478): C97-116.

[7] Bridges Sarah and Guariglia Alessandra (2006): *Financial Constraints, Global Engagement, and Firm Survival in the UK: Evidence from Micro Data*, Working Paper Nottingham University.

[8] Cefis Elena and Marsili Orietta (2005): *A matter of life and death: innovation and firm survival, Industrial and Corporate Change*, Vol. 14, No. 6, pp. 1167-1192.

[9] Cockburn Iain and Wagner Stefan (2007): *Patents and the Survival of Internet-related IPOs*, NBER Working Papers 13146.

[10] Cox David (1972): *Regression Models and Lifetables*, Journal of the Royal Statistical Society, Vol. 34, 187220.

[11] Disney, R., Haskel, J. and Heden, Y. (2003). *Entry, exit and establishment survival in UK manufacturing'*, Journal of Industrial Economics, 51(1), pp. 91-112.

[12] Dunne Timothy, Roberts Mark and Samuelson Larry (1989): *The Growth and Failure of U.S. Manufacturing Plants*, The Quarterly Journal of Economics, Vol. 104(4), pp. 671-98.

[13] Fritsch Michael, Brixy Udo and Falck Oliver (2006): *The Effect of Industry, Region, and Time on New Business Survival - A Multi-Dimensional Analysis*, Review of Industrial Organization, Vol. 28, No. 3, pp. 285-306.

[14] Geroski Paul (1991): *Domestic and foreign entry in the United Kingdom: 1983-1984*, in Geroski P. and Schwalbach J. (Eds.) *Entry and Market Contestability*, Basil Blackwell, Oxford.

[15] Greenhalgh Christine and Rogers Mark (2006): *The value of innovation: The interaction of competition, R&D and IP*, Research Policy, Vol. 35(4), pp. 562-580.

[16] Greenhalgh Christine and Rogers Mark (2007): *Trade marks and performance in UK firms: Evidence of Schumpeterian competition through innovation*, Oxford University, Economics Department Working Paper: http://www.economics.ox.ac.uk/Research/wp/pdf/paper300.pdf.

[17] Griliches Zvi (1990): *Patent Statistics as Economic Indicators: A Survey*, Journal of Economic Literature, Vol. 28(4), pp. 1661-1707.

[18] Hall Bronwyn (2000): *Innovation and Market Value* in Productivity, Innovation and Economic Performance (eds.) R. Barrell, G. Mason and M. O'Mahoney. Cambridge UK, Cambridge University Press, pp. 177-198.

[19] Hall Bronwyn (2007): *Patents and Patent Policy*, Oxford Review of Economic Policy, Forthcoming.

[20] Jenkins Stephen (2005): *Survival Analysis*, University of Essex.

[21] Jensen Paul, Webster Elizabeth and Buddelmeyer Hielke (2006): *Innovation, Technological Conditions and New Firm Survival*, Melbourne Institute Working Paper No. 26/06.

[22] Kaplan E. L. and Meier Paul (1958): *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association, Vol. 53, pp. 457-48.

[23] Klette Tor Jakob and Kortum Samuel (2004): *Innovating Firms and Aggregate Innovation*, Journal of Political Economy, Vol. 112, Issue 5, pp. 986-1018.

[24] Landes William and Posner Richard (1987): *Trademark law: An economic perspective*, Journal of Law and Economics, Vol. 30(2), pp. 265-309.

[25] Mata Jose and Portugal Pedro (1994): *Life Duration of New Firms, The Journal of Industrial Economics*, Vol. 42, No. 3., pp. 227-245.

[26] Mendoca, Pereira, and Godinho (2004): *Trademarks as an indicator of innovation and industrial change*, Research Policy, Vol. 33, pp. 1385-1404.

[27] Pavitt, K. (1984). *Sectoral patterns of technical change*, Research Policy 13: 343-373.

[28] Rogers Mark, Helmers Christian and Greenhalgh Christine (2007): *An analysis of the characteristics of small and medium enterprises that use intellectual property*, Oxford Intellectual Property Research Centre.

[29] Schumpeter Joseph (1934): *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*, Harvard University Press.

[30] Schmookler Jacob (1966): *Invention and economic growth*, Cambridge, Mass., Harvard University Press.

[31] van Zeebroeck Nicolas (2007): *The puzzle of patent value indicators*, CEB Working Paper.

A Data Appendix

In this appendix, we outline details regarding the construction of the data set and describes some issues related to the measurement of the IP variables used in our analysis above.

At the county and unitary authority level, we added house prices measured as the average price of all property types within a county or unitary authority. The data for England and Wales comes from the Land Registry's Residential Property Price Reports.[24] The data for Scotland comes from the Halifax House Price Index.[25] The data for Scotland is available only at a slightly more aggregate level applying the definitions of the former local government regions of Scotland. Whenever the data is available, we use prices of the final quarter of the preceding year as our measure of house prices.

Unemployed rates are measured as the ratio of unemployed over all economically active persons, by county and unitary authority. The data comes from the UK Office for National Statistics' Labour Force Survey (LFS) where we calculated annual averages from the quarterly data available.[26]

The IP data also includes application date, although it is only available for IP that has been published. Equally, for those patents which succeeded in being granted also the grant year is available. IP is commonly regarded as an output measure of innovation and we use it as a proxy for the quality of business ideas. Hence, if the objective is to measure whether it conveys a firm any competitive edge over its competitors, it seems appropriate to use publication date (registration date in case of Community trade marks) as the reference point.

Patent data have been widely used in applied work for very diverse purposes. Yet, the use of patent data has many more or less well-documented pitfalls. A problem with using UK and EPO patents arises from the fact that sometimes applicants file the same or very similar applications at both institutions. For example, firms may apply first for a patent at the UK office and then use the obtained priority date to file for an EPO patent making the same, or very similar, claims. Such patents belong to the same patent family and it is unclear whether they should all be counted as single patents.[27] On the one hand, since these patents are based on the same invention, they do not

[24] http://www.landreg.gov.uk/.

[25] http://www.hbosplc.com/economy/housingresearch.asp.

[26] http://www.statistics.gov.uk/STATBASE/Source.asp?vlnk=358

[27] If the exactly same application is made to both institutions, one of the two has to be withdrawn once one of the institutions grants the patent.

each represent a new innovation by the firm. On the other, as we are interested in whether firms gain any competitive advantage from patenting, such strategic patenting may matter. We therefore count patents pertaining to the same patent family as single patents.

Another concern when using patent data is heterogeneity across patents with respect to their actual value. In particular, it is argued that many patents are of little or no value. The patent literature has developed a number of possible ways to discriminate among patents, including the fact whether a patent was actually granted, the number of citations received, the respective patent family size, renewals, opposition and litigation and direct values reported by the firms themselves in surveys (van Zeebroeck, 2007). In our situation there are various problems with such methods. Using patent grants for new firms would mean the firms may already be two to four years old by the time the patent is granted. In a similar way, the use of citations (or renewals) means waiting for several years after the patent is granted, meaning we could study at the earliest startups in the 1990s. Our data has no information on opposition, litigation and patent family size and, in any event, these may be more important for larger firms. Given this, we adopt a rather crude measure by distinguishing between UK and EPO patents, where EPO patents are considered more valuable. The importance is not only reflected in distinctively higher fees for EPO patents, but due to the international scope of EPO patents. An EPO patent application costs Euro 4000, while a UK patent application costs around Euro 300. In fact, applications for EPO are likely to be much higher since it needs to be submitted in two languages and use of a patent attorney is strongly recommended. For both EPO and UK patents, the application is published, if the application passes an initial examination, after 18 months. Hence, our use of publications means that there is an 18 month delay from submitting original invention. It can be expected that only more valuable patents will be patented in several countries other than the country in which the firm is registered.

In the UK, trade marks can be obtained either through an application to the UK Intellectual Property Office for a UK Trade Mark, or through an application to the Office of Harmonization for the Internal Market for a Community Trade Mark. Fees differ substantially, as the UK trade mark costs about 300 Euros while the Community trade mark costs around 2000 Euros. For trade marks, the difference between using publication data and registration date is small (around 90 percent of published trade marks are registered within a few months).

Table 5: Cox Regression

Variables	Coefficient					
Sector level						
Entry/Exit costs	..	..	..	-14.446***	-14.442***	-13.010***
				(0.238)	(0.238)	(0.511)
Capital intensity	..	..	..	0.006*	0.006*	0.042***
				(0.003)	(0.003)	(0.006)
MES	..	..	..	3.005	3.019	-1.048
				(6.061)	(6.061)	(6.509)
4-Firm Concentration Ratio	..	..	..	-0.074**	-0.075**	0.032
				(0.032)	(0.032)	(0.044)
Industry Growth Rate	..	..	..	0.098**	0.098**	0.097*
				(0.043)	(0.043)	(0.057)
Firm level						
Patent Dummy	-0.797***	..	..	-0.438**	..	..
	(0.182)			(0.211)		
UK Patent Count	..	-0.534***	-0.517***	..	-0.507**	-0.539**
		(0.184)	(0.182)		(0.229)	(0.233)
EPO Patent Count	..	-0.546**	-0.538**	..	-0.066	-0.095
		(0.221)	(0.221)		(0.217)	(0.223)
Trade Mark Dummy	-0.742***	..	..	-0.657***	..	..
	(0.083)			(0.101)		
UK TM Count	..	-0.516***	-0.523***	..	-0.525***	-0.538***
		(0.087)	(0.088)		(0.096)	(0.097)
CTM Count	..	-0.275**	-0.301**	..	-0.208	-0.224
		(0.119)	(0.122)		(0.152)	(0.153)
University	..	..	..	-0.891***	-0.896***	-0.894***
				(0.294)	(0.294)	(0.295)
Ln no. of Directors	..	..	..	-0.334***	-0.335***	-0.294***
				(0.018)	(0.018)	(0.018)
Foreign Owned	..	..	..	-0.816***	-0.816***	-0.827***
				(0.048)	(0.048)	(0.049)
Subsidiary	..	..	..	-2.049***	-2.048***	-2.038***
				(0.056)	(0.056)	(0.056)
County / Unitary Authority						
Unemployment rate	..	..	..	4.737***	4.735***	4.829***
				(0.469)	(0.469)	(0..470)
Ln Housing prices	..	..	..	-4.011***	-4.011**	-3.755**
				(1.439)	(1.439)	(1.458)
Ln (Housing prices)^2	..	..	..	0.165***	0.165**	0.156**
				(0.060)	(0.060)	(0.061)
RDAs						
South West	..	..	0.252***	0.282**	0.282**	0.312**
	..	..	(0.095)	(0.123)	(0.123)	(0.123)
South East	..	..	0.374***	0.309**	0.309**	0.339***
	..	..	(0.094)	(0.123)	(0.123)	(0.124)
London	..	..	0.626***	0.532***	0.532***	0.557***
	..	..	(0.094)	(0.128)	(0.128)	(0.129)
East of England	..	..	0.379***	0.409***	0.409***	0.433***
	..	..	(0.095)	(0.122)	(0.122)	(0.123)
East Midlands	..	..	0.274***	0.253**	0.253**	0.276**
	..	..	(0.096)	(0.123)	(0.123)	(0.123)
Yorkshire	..	..	0.326***	0.322***	0.322***	0.347***
	..	..	(0.095)	(0.122)	(0.122)	(0.123)
North West	..	..	0.437***	0.335***	0.335***	0.369***
	..	..	(0.095)	(0.122)	(0.122)	(0.122)
West Midlands	..	..	0.389***	0.424***	0.424***	0.455***
	..	..	(0.095)	(0.121)	(0.121)	(0.122)
North East	..	..	0.240**	0.118	0.118	0.146
	..	..	(0.099)	(0.127)	(0.127)	(0.128)
Wales	..	..	0.295***	0.239*	0.239**	0.255**
	..	..	(0.097)	(0.125)	(0.125)	(0.125)
Scotland	..	..	0.258***	0.272**	0.272**	0.293**
	..	..	(0.096)	(0.122)	(0.122)	(0.123)
Highland	..	..	dropped	dropped	dropped	dropped
Industry dummies	..	..	..	..	..	included
Number of firms:	162,374	162,374	162,374	133,703	133,703	133,703
Number of observations:	701,101	701,101	701,101	591,182	591,182	591,182
Number of failures:	58,043	58,043	58,043	35,756	35,756	35,756

Notes:
1. Adjusted standard errors for within correlation in parentheses.
2. * indicates significance at 10%; * at 5%; ** at 1%.

Table 6: Cox Regression: IP-Region Interactions

Variables	Coefficient			
IP Dummy	-1.013***	0.975	-1.041***	0.889
	(0.062)	(0.966)	(0.062)	(0.960)
South West	0.289**	0.303**	0.319**	0.333***
	(0.123)	(0.123)	(0.123)	(0.124)
South East	0.317**	0.329***	0.347***	0.359***
	(0.123)	(0.124)	(0.124)	(0.125)
London	0.540***	0.555***	0.566***	0.581***
	(0.128)	(0.129)	(0.129)	(0.129)
East of England	0.417***	0.430***	0.439***	0.453***
	(0.122)	(0.123)	(0.123)	(0.124)
East Midlands	0.262**	0.272**	0.285**	0.294**
	(0.122)	(0.123)	(0.123)	(0.124)
Yorkshire	0.331***	0.344***	0.357***	0.369***
	(0.122)	(0.123)	(0.123)	(0.124)
North West	0.342***	0.354***	0.376***	0.388***
	(0.121)	(0.122)	(0.122)	(0.123)
West Midlands	0.431***	0.443***	0.462***	0.475***
	(0.121)	(0.122)	(0.122)	(0.123)
North East	0.126	0.144	0.154	0.173
	(0.127)	(0.128)	(0.128)	(0.129)
Wales	0.248**	0.264**	0.264**	0.279**
	(0.125)	(0.125)	(0.125)	(0.126)
Scotland	0.281**	0.291**	0.303**	0.312**
	(0.122)	(0.123)	(0.123)	(0.124)
Highland	dropped	dropped	dropped	dropped
IP dummy × South West	..	-2.081**	..	-2.029**
	..	(1.001))	..	(0.996)
IP dummy × South East	..	-1.813*	..	-1.757*
	..	(0.978)	..	(0.972)
IP dummy × London	..	-2.234**	..	-2.166**
	..	(0.973)	..	(0.967)
IP dummy × East of England	..	-1.999**	..	-1.944**
	..	(0.995)	..	(0.989)
IP dummy × East Midlands	..	-1.632	..	-1.575
	..	(0.995)	..	(0.989)
IP dummy × Yorkshire	..	-1.919*	..	-1.863*
	..	(0.997)	..	(0.991)
IP dummy × North West	..	-1.817*	..	-1.759*
	..	(0.987)	..	(0.981)
IP dummy × West Midlands	..	-1.925*	..	-1.875*
	..	(0.985)	..	(0.979)
IP dummy × North East	..	-3.419**	..	-3.365**
	..	(1.380)	..	(1.375)
IP dummy × Wales	..	-2.424**	..	-2.386**
	..	(1.090)	..	(1.083)
IP dummy × Scotland	..	-1.603	..	-1.538
	..	(0.988)	..	(0.983)
IP dummy × Highland	dropped	dropped	dropped	dropped
Industry dummies	..	..	included	included
Number of firms:	133,703	133,703	133,703	133,703
Number of observations:	591,182	591,182		591,182591,182
Number of failures:	35,756	35,756	35,756	35,755

Notes:

1. Adjusted standard errors for within correlation in parentheses.
2. * indicates significance at 10%; * at 5%; ** at 1%.
3. Covariates include: Entry/exit costs, capital intensity, MES, 4-firm concentration ratio, industry growth rate, university indicator, ln no. of directors, foreign-owned indicator, subsidiary indicator, unemployment rate, ln housing prices, ln (housing prices)^2

Table 7: Ranking of regions based on survival probabilities of new firms

	IP-active firms	All firms (conditional)	All firms (unconditional)
North East	1	1	2
Wales	2	2	4
London	3	11	10
South West	4	5	1
East of England	5	9	7
West Midlands	6	10	8
Yorkshire	7	7	5
North West	8	8	9
South East	9	6	6
East Midlands	10	3	3
Scotland	11	4	3

Notes:
1. Table shows ranking of regions according to relative firm survival rates with respect to Highlands. First column shows ranking for IP-active firms (based on Table 5). Second column shows ranking based on conditional survival probabilities (from coefficients on region dummies in column six in Table 5). The third column shows unconditional survival probabilities based on region for all IP-inactive firms (see Table 1)

Figure 4: Failure rates of IP-inactive firms by county / unitary authority