Firm-size diversity and active learning through R&D investment:

An ecosystem approach to firm survival in UK data

Eshref Trushin[a], Mehmet Ugur[b,*], and Edna Solomon[b]

Abstract

In the management and industrial organisation literature, firm survival features prominently but with little attention to its relationship with intra-industry diversity. To bridge this gap, we propose an ecosystem approach to industry evolution in which diversity and firm learning play important roles. Using discrete-time hazard estimators and an unbalanced panel dataset of 35,136 R&D-active UK firms, we demonstrate that within industry firm size diversity is conducive to higher exit hazard rates. We also demonstrate that R&D expenditures constitute investment in active learning that moderates the adverse effects of diversity on survival. Our findings are robust to step-wise estimations, different specifications and firm cohorts, control for frailty and for a number of firm-, industry-, and macroeconomic factors widely used for in the survival literature.

Key words: Diversity, firm survival, R&D, active learning, niche discovery

[a]*Durham University Business School;* [b]*University of Greenwich Business School*
** Corresponding author:* Professor Mehmet Ugur, M.Ugur@gre.ac.uk

Research for this paper has been funded by the ESRC under grant ES/K004824/1. We thank our funders for their generous support. The views expressed here are those of the authors, who are responsible for any errors or omissions.

Firm-size diversity and active learning through R&D investment: An ecosystem approach to firm survival in UK data

Introduction

Significant and persistent diversity of firm characteristics within an industry is well documented and puzzling (Dosi *et al.*, 1997). Indeed, some contributors to the field argue that intra-industry firm diversity has been a driver for the development of strategic management (Noda and Collis, 2001:922). Yet there has been little attention paid to how intra-industry size diversity affects firm survival. The relationship between intra-industry diversity and firm mortality has remained below the radars of both management and industrial organisation literature.

In bioecological literature, variety is necessary for successful adaptation, but this literature tended to focus on the relationship between the diversity and stability of the ecosystem as a whole, while the diversity's effects on the fortunes of the system's constituents is largely overlooked (Elton, 1958; McCann, 2000; McNaughton, 1993; May, 1974; Yozdis, 1980). The management literature (Canella *et al.*, 2008; Erhrdt *et al.*, 2003; Jiang *et al.*, 2010; Kilduff *et al.*, 2000) focus largely on the relationship between *firm-level* diversity and performance, again often overlooking the likely effects of industry-level diversity on firm survival. In the industrial organisation, the role of intra-industry diversity *does not* feature prominently (see, for example, Audretsch, 1991; Doms *et al.*, 1995; Geroski, 1995; Fritsch *et al.*, 2006).

Intra-industry size diversity may be an important exit hazard. First, firm-size diversity within an industry may be the outcome of low entry costs, which encourage heterogeneous firm entry, but also lead to higher exit rates as relatively inefficient firms are selected out (Hopenhayn, 1992). Secondly, intra-industry size diversity reflects variation in the choice of the generic specialization and differentiation strategies (Porter, 1980), some of which are less likely to stand competitive pressures or adverse shocks. Thirdly, greater firm-size diversity within the industry may be the outcome of a dispersed firm fitness/efficiency distribution that induces high levels of firm turnover (Jovanovic, 1982). Lastly, the positive association between intra-industry size diversity and firm mortality may be due to industry shakeouts and selection that unfolds as heterogeneous firms discover their fitness status and efficient technology in the industry (Utterback and Suarez, 1993).

This paper addresses the following research questions: if intra-industry size diversity constitutes a hazard factor for firms beyond the direct effect of size itself? Given the level of intra-industry diversity, is investment in research and development (R&D) conducive to active learning that enables the firm to survive longer? With respect to interaction between R&D investment and diversity, we draw on Cohen and Levinthal (1990) and Ericson and Pakes (1989, 1995) among others to argue that they may be substitutes in that higher levels of R&D effort may enable firms to engage in active learning and discover the product and technology niches that will enable them to survive. This line of reasoning connects diversity with the organizational learning literature that point out the role of ecosystem in firm learning, competence formation and acquisition of dynamic capabilities (see, Anderson and Tushman, 2001; Levitt and March, 1993; Teece, 2007; Zahra and George, 2002).

We test the relevance of these theoretical insights by applying panel discrete-time hazard models, which are suitable for interval-censored annual observations in firm-level data. In addition, we control for unobserved firm heterogeneity and use lagged covariates to eliminate competing causal attributions. We check the robustness of our findings to different estimators and step-wise estimations, left truncation, different firm cohorts, and a wide range of firm-, industry- and macro-level factors usually controlled for in the firm survival literature. We also test the robustness of the results to such measures of diversity as the coefficient of variation and the Theil entropy index of firm sizes in an industry.

We find two novel results: (i) intra-industry firm size diversity is associated with higher rates of exit hazard; (ii) after controlling for the direct effect of R&D intensity, intra-industry firm size diversity and firm-level R&D intensity emerge as substitutes in that firms with higher levels of R&D intensity are less likely to exit at each level of intra-industry diversity. The paper also confirms the findings in Ugur, Trushin, and Solomon (2016a): the survival-enhancing effect of R&D investment decreases as the level of R&D intensity increases, with the implication that R&D investment beyond the optimal level may eventually become a hazard factor. Combining the findings above, we suggest that firm survival is a function of two types of learning: (i) *passive learning* that involves the discovery of the firm's position in the fitness distribution within the industry (see Jovanovic, 1982); and (ii) *active learning* that involves the discovery of product/technology niches that would reduce firm mortality at each level of intra-industry diversity, which is in line with the dynamic capabilities literature.

The remainder of the paper is organised as follows. In section 2, we draw on several strands of literature and formulate our testable hypotheses. Section 3 presents the data and estimation methodology. Section 4 reports the results from several discrete-time hazard models, with the preferred model chosen on the basis of likelihood ratio (LR) tests and Akaike and Bayesian information criteria (AIC, BIC). In the conclusion section, we summarise the main findings and discuss their relevance for business practice and future research.

Related literature and research hypotheses

Given the diversity of the intellectual traditions in the field, we draw on several research agendas to derive testable hypotheses. We start with the *bioecological* literature, which examines the effects of ecosystem diversity on system stability. In early work of Elton (1958), diverse ecosystems can maintain stable populations and prevent invasion of other species. This insight has been supported by field experiments on plants and insects, which established that more diverse ecosystems were able to maintain composition and population densities (McNaughton, 1993). Nevertheless, these findings have been contested. For example, May (1974) utilizes Liapunov stability parameters and demonstrates that a growing number of differing species leads to larger fluctuations in ecosystem populations and may result in extinctions. This is because diversity is associated with higher competition in the face of adverse shocks, leading to higher selection.

Gardner and Ashby (1970) adopts a *cybernetic* approach with Liapunov stability measures to demonstrate that a self-generating catastrophe is more likely to happen in more diverse (complex) systems. Similarly, an interpretation of Ashby's law of requisite variety (Ashby, 2011) could be that variety in an industry's shocks must be reflected in adequate variety of firms to respond to shocks, i.e., diversity is a result of heterogeneous shocks in the system. Yozdis (1980) draws on evidence from a sample of food webs in a biological ecosystem and demonstrates that complex ecosystems might be more fragile due to high specialization and smaller scope for interaction between different species. In between, Lehman and Tilman (2000) develop a nuanced view in which biodiversity may stabilize aggregate ecosystem properties such as biomass, but may also destabilize the dynamics of the system's population. Empirical tests of the association between biosystems' diversity and stability deliver mixed results. According to Ives and

Carpenter (2007), only 23 percent of findings from 13 studies indicate a positive and significant relationship, whereas 46 percent of the findings indicate a negative relationship, and the remaining 31 percent indicate a relationship close to zero.

More relevant concepts and findings for business perspectives are provided in the *organisational ecology* literature (Hannan and Caroll, 1992; Hannan and Freeman, 1977, 1989; and Hannan, 2005) to uncover the relationship between diversity, industry evolution and constituent (firm) mortality. In this literature, industry evolution is related to a wide range of system- and organisation-level factors, including firm density in the industry, founding conditions of firms ("birthmarks"), ageing, size and the legitimation of an organizational form. Here, an important dynamics of firm niche width from ecological literature have been found as a significant factor of industry evolution (Dobrev, Kim, and Hannan, 2001). Interesting results on cohort diversity with piecewise exponential models found by Barnett and Hansen (1996): diversity of bank cohorts increased failure rate of Illinois banks. Majority of these explanatory factors are hardly unique to the organisational ecology approach: age, size, number of plants and founding conditions are also core issues in the industrial organisation (see Geroski, 2001 for a review). However, the concepts of density (the number of firms within the industry), cohort diversity, and legitimation are novel – and they establish a positive association between organisational density and mortality. In Carroll and Hannan (1998, 2000), density tends to increase exit hazard because of the selection processes that unfold within diverse industries. Yet the evidence they provide on brewers, publishers, trades unions, automobile producers and other industries remain somewhat descriptive with limited hazard modelling (Miner, 2004).

For modelling, we draw on the industrial organisation literature, e.g., a model of selection with imperfect information Jovanovic (1982) demonstrates how firms discover their efficiencies in year *t*, conditional on information they receive until *t-1*. Once this information is discovered, firms with higher levels of efficiency will grow faster between year *t* and *t+1* compared to those who discover a lower levels of efficiency. Stated differently, the dispersion of firm-size distribution *observed* in any given is a reflection of the dispersion of the efficiency distribution *discovered* in the previous year. From this perspective, intra-industry size diversity in a given period may lead to higher exit rates in the subsequent period because what we observe as size diversity at any time is underpinned by unobserved fitness dispersion in the previous period(s). Exit in the model is endogenous due to the less efficient firms discover that their market values fall below the expected value of their fixed factor of production. Stated differently, size diversity is a significant predictor of exit rates because, in any year, it reflects the cumulative dispersion of the fitness distribution in prior period(s). In Jovanovic (1982), the mortality effect of the size diversity or the mortality effects of the fitness dispersion in prior periods, is independent of any other firm or industry characteristic, including the contemporaneous fitness dispersion in year *t*.

The positive association between size diversity and mortality outlined above can be due to low entry costs too (Hopenhayn, 1992), low entry costs are conducive to higher size diversity as potential entrants are encouraged to enter, and then they exit due to lack of efficiency, e.g. in scale of production. Dunne *et al*. (1988) had already reported that high overall entry rates in US industries were associated with high exit rates in the same and subsequent periods. Disney *et al*. (2003) also report similar contemporaneous associations between entry and exit rates in the UK industry.

Given these findings, we state our first hypothesis **(H1)** as follows: *Higher levels of firm-size diversity within an industry are the outcomes of dispersion in the firms'* *unobserved efficiency/fitness distribution, which leads to higher turnover rates as less fit firms are selected out.*

H1 is compatible with findings in the product life cycle and technological regime studies. As demonstrated in Porter (1980) and Nelson (1995), the life cycle of an industry/technological regime goes through emergence, growth, maturity and decline. In the early phases, business opportunities and models are abundant and an entrepreneurial innovation regime tends to dominate the routinized regime. However, a dominant business model with a routinized regime eventually emerges, where larger firm size and economies of scale play an important role (McGahan *et al*., 2004; Utterback and Suarez, 1993). There is also evidence that the initial number of firms in diverse industries in the early stages of the product/technology cycle is often halved as unsuccessful innovators exit after a period of experimentation (Klepper and Simons, 1996). Similarly, Utterback and Suarez (1993) document the occurrence of industry shakeouts that lead to a drop in the number of mostly small firms. Finally, in a review of 216 industry/product life-cycle studies Peltoniemi (2011) reports that product variety, firm numbers, and R&D activities decline as the industry matures.

However, **H1** depicts only part of the selection process as it assumes passive firm learning about own fitness. Firms can also actively discover opportunities through R&D, which may be a subject to increasing or diminishing scale effects (Ugur *et al*., 2016a). The stochastic model of entry and exit of Ericson and Pakes (1995) illustrates the point. Here, heterogeneous firms faced with idiosyncratic productivity shocks invest in R&D as the firm's success depends on: (i) the stochastic outcome of its investment; (ii) the success of other firms in the industry; and (iii) the competitive pressure from outside the industry. In their model, the high mortality rates in the initial learning period, however, will be followed by longer survival for firms investing in R&D.

Given this analysis, we state our second hypothesis **(H2)** as follows: *Higher levels of firm-size diversity and the firm's R&D intensity are substitutes in that firms that invest more in R&D are likely to survive longer as they discover their product/technology niches and increase the probability of their profitability.*

H2 resonates with the arguments in Audretsch, Howeling, Thurik (2004), who points out firm survival is a result of market firm learning processes and selection driven by experimentation with various business ideas. It also shares common grounds with the strategic management views on firm capacities, e.g. Anderson and Tushman (2001) report that firm failures to cope with demand and technological discontinuities were the key factor of organizational mortality in the US cement and minicomputer industries. Also, Levitt and March (1993) points out to competence traps and myopia that result from experience-based narrow learning and the higher mortality rates due to overreliance on specific niches instead of exploring new processes and products.

A similar thread is evident in Teece (2007), Probst and Raisch (2005) and Cohen and Levinthal (1990) who drove attention to uncertainty in diverse industries, where the likelihood of "competence traps" or "constrained dynamic capabilities" can be reduced by R&D investment. Such investment increases the scope for discovering, reconfiguring, and developing opportunities, integrating assets and "sensing, seizing, and managing" threats. Furthermore, the risk of exit can be ameliorated by the firm's innovation effort that facilitates adaptation to shifts in the technology regime and to the achievement of optimal industry equilibria (Safarzynska and van der Bergh, 2010).

Finally, **H2** also shares common ground with the Schumpeterian models of innovation and selection (Aghion, Akcigit, Howitt 2013; Ugur *et al*., 2016a), where firm's R&D investment advances it towards the technology frontier. The Schumpeterian perspective implies that firms in diverse industries would undertake relatively higher levels of R&D investment to discover their optimal technology niche and to catch up with the movements of the technology frontier. This was already indicated in Klepper (1996), who reports that companies tend to exit primarily because of failure to develop successful routines that would enable them to survive in an environment of volatile technology shifts underpinned by diversity in the industry.

By testing for **H1** and **H2**, we aim to contribute to existing knowledge on industry evolution and firm exits by verifying if: (i) intra-industry size diversity is a significant hazard factor; (ii) R&D investment constitutes active learning and enables firms to survive longer in diverse industries; (iii) if the adverse effects of diversity on survival is neutralised when R&D expenditures are above 9 percent of turnover.

Empirical strategy

First, we define our diversity measures – the Theil entropy index and the coefficient of variation of intra-industry employment sizes within 3-digit SIC industry codes. Here we discuss the strengths and weaknesses of these diversity measures and the ways in which we control for factors that would minimise the risk of bias associated with their potential weaknesses. In the next section we explain our dataset followed by specification of the hazard models, controls, and choice between estimators.

Measuring firm-size diversity

As indicated in Stirling (2010) and Solanas *et al*. (2012), there is no perfect measure of diversity. We measure intra-industry firm-size diversity with two scale-invariant indicators: the Theil's entropy index (TI) and the coefficient of variation (CV) of the firm-specific employment size within 265 industries at the 3-digit SIC level. The Theil's entropy index is often used to measure income inequality, segregation, trade and industry diversity (Cadot *et al.*, 2013). We calculate a *TI* for each industry and every year as follows:

$$TI_{jt} = \frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \frac{L_{ijt}}{L_{jt}} * ln \frac{L_{ijt}}{L_{jt}} \quad (1a)$$

Here L_{ijt} is employment size of the i^{th} firm in industry j and year t; $\overline{L_{jt}}$ is average employment in industry j in year t; and N_{jt} is the number of firms in industry j and year t. Higher *TI* values indicate higher levels of firm-size diversity in the industry in a year. Our choice of Theil index is informed by its excellent property of being invariant not only to unit of measurement, but also to any scale factor. Other advantageous properties are that it is comparable over time and between industries; and it is also additive, symmetric, statistically testable, and decomposable (Theil, 1972; Haughton and Khandker, 2009). Nevertheless, the Theil index is sensitive to the left end of the size distribution, i.e. it may better reflect the diversity among smaller firms compared to larger firms (Haughton and Khandker, 2009).

Our second diversity measure is the coefficient of variation (*CV*) for firm employment within the industry, which is calculated as:

$$CV_{jt} = \sqrt{\frac{\sum_{i=1}^{N_{jt}} (L_{ijt} - L_{jt})^2}{N_{ijt} - 1}} \frac{1}{L_{jt}} = S_{Ljt} \frac{1}{L_{jt}} \quad (1b)$$

Here S_{Ljt} is the standard deviation of employment and $\frac{1}{L_{jt}}$ is the inverse of the mean employment per firm in the industry. All other variables are as defined above. Like *TI*, the *CV* is also invariant to multiplicative scale factors and units of measurement. The drawback here is that it is an interaction term between two variables: the standard deviation of employment and the inverse of the mean employment in the industry, hence, we must additionally control for mean employment in the industry (Sørensen, 2002; Bedeian and Mossholder, 2000; Solanas *et al*., 2012). Both TI and CV are monotonically increasing with employment size diversity.

Data

We merged two firm-level datasets compiled by Office of National Statistics (ONS) in the UK: the Business Structure Database (BSD) and the Business Expenditure on Research and Development (BERD) database. The BSD provides annual demographic data on births, deaths, employment, turnover, number of live local units, foreign/domestic ownership, etc. We excluded companies that exited due to mergers and acquisitions. Hence exit in our dataset refers to liquidations or bankruptcies. This was possible because both firm and local unit (plant) references disappear from the register, if exit is due to liquidation/bankruptcy, but only firm identifiers disappear, if the exit was due to mergers or acquisitions. We have constructed the exit year as the earliest of the death year recorded by the ONS, or the first year when the firm employment and turnover are zero for three consecutive years[1]. We also excluded firms with birth date before 1974 as firms were given the same birth year when the registrar was first introduced in 1973.

BERD consists of repeated annual surveys with stratified sampling of firms known to be R&D-active. We merged BERD with BSD, using the unique firm (enterprise) identifier[2]. In our dataset, the ratio of R&D to turnover is greater than 1 from the 95th percentile onwards. We have considered firms in the top 5 percent of the R&D intensity distribution as outliers and set our baseline estimation sample for firms with R&D intensity less than 1. Our estimation sample consists of 35,136 firms and 158,316 observations from 1998 to 2012. Summary statistics presented in Table A1 in the

---

[1] The second criterion is used because we have identified delays in the ONS assignment of a death code in some cases even though the firm's return for employment and turnover is zero for several years.

[2] Further information on BSD and BERD is provided in Ugur, Trushin, and Solomon (2016b).

Appendix relate to the estimation sample with R&D intensity restricted to less than 1. However, we find strong stability of our results to different cut-off points of R&D intensity, step-wise robust estimations, for different firm birth cohorts, and to exclusion of finance and defence industries with SIC codes 64-66 and 84 (these results are available from the authors).

Scatter plots (Figure 1) are based on data within 3-digit SIC industry codes and allow for visual inspection of the relationship between: (i) firm-size diversity and firm survival time (top panel), and (ii) firm-size diversity and R&D intensity (bottom panel). Plots in the top panel indicate a negative relationship between firm-size diversity and survival time. This is equivalent to a positive association between firm-size diversity and exit hazard rate – as proposed in **H1** above. Those in the bottom panel provide information pertinent to **H2**: as firm-size diversity within an industry increases, the mean R&D intensity in the industry tends to increase. This relationship suggests that firms in more diverse industries tend to invest more in R&D, i.e., their mean active-learning effort is increasing as industry firm size diversity increases. These relationships hold with both measures of diversity (CV in the left and TI in the right panel).

'Insert Figure 1 here'

Modeling exit hazard: Main variables of interest and controls

We follow the general specification for the hazard rate function (Jenkins, 1995), but we use lagged, hence, predetermined or weakly exogenous covariates x_{it} to deal with simultaneity bias. The probability (Pr) of exit in year *t+1* conditional on observable covariates can be stated as follows:

$$Pr(y_{it+1}|x_{it,}v_i) = Pr(x_{it}\beta + M_{it}\alpha + v_i + \gamma_{t+1} + \varepsilon_{it+1}) \quad (2)$$

Here, i and t are firm and year indices; x_{it} is a vector of observable firm- and industry-level covariates that affect firm exit with an estimated vector of β parameters; M_{it} is a vector of macroeconomic and technology (Pavitt) class variables that affect firm exit with an estimated vector of α parameters; γ_{t+1} are year dummies; and ε_{it+1} is the disturbance term. Unobserved heterogeneity between firms is captured by the independently and identically distributed (i.i.d.) normal random variable $v_i|x_{it}, M_{it+1} \sim N(0,\sigma_v^2)$. The strong and very common assumption in estimation of such models is that unobserved heterogeneity (v_i) and the disturbance term ε_{it+1} are independent of the firm, industry, and macroeconomic covariates.

How important is the assumption that unobserved heterogeneity v_i is i.i.d. Normal? Nicoletti and Rondinelli (2010) have evaluated biases in estimated parameters of the discrete time hazard models caused by omitting or misspecifying the unobserved heterogeneity distribution using Monte-Carlo simulations. Their results demonstrate that neglect or misspecification of the unobserved heterogeneity are unlikely to lead to a significant bias in the estimated parameters.

The variables of main interest in x_{it} include the *TI* or *CV* measures of intra-industry diversity and the interactions of the latter with firm-level R&D intensity. The remaining firm/industry covariates in x_{it} and the macroeconomic and technology class covariates in M_{it} are specified in accordance with the best practice in survival analysis. Definitions of all covariates and the literature that justifies their inclusion in the model are presented in Table 1.

The correlation between *TI* and *CV* is 0.44 and significant, but their correlation with the Herfindahl index is low and insignificant. The latter property reduces the risk of collinearity and indicates that the informational contents of intra-industry employment size diversity and sales concentration do reflect different ecosystem properties. We also control for interaction between diversity and R&D intensity to verify whether investment in R&D reduces exit hazard at each level of intra-industry size diversity.

'Insert Table 1 here'

The survival effect of input and output measures of innovation such as R&D expenditures and patents have been studied widely. Existing studies tend to adopt a linear specification, but report conflicting findings. Although survival-enhancing effects of innovation are reported more frequently, Ugur et al. (2016a) demonstrate that a quadratic specification could be more plausible both theoretically and empirically, and this relationship was empirically confirmed by Sharapov *et al*. (2011) and by Zhang and Mohnen (2013).

Another set of firm characteristics that have been investigated in the survival literature consists of firm age and size. Both theoretical and empirical work indicates that: (i) age and size are correlated positively with survival; and (ii) small firms that survive tend to grow faster than larger firms, but have a higher probability of exiting (Geroski, 1995; Klette and Kortum, 2004; Aghion *et al*., 2013). Therefore we control for age and size to also disentangle the effect of diversity from the effect of size itself.

Labour productivity and firm growth rate relative to industry growth are reported as significant determinants of firm survival (Audretsch, 1991, 1995; Hopenhayn, 1992; Ericson and Pakes, 1995; Mata, Portugal, Guimaraes, 1995; Cefis and Marsili, 2005; and Ugur *et al*., 2016a). The fourth set of firm-level characteristics includes number of plants, whether the firm is engaged in civil R&D only, and domestic versus foreign ownership. Inclusion of these observable characteristics in a hazard model is consistent with Audretsch (1991), Audretsch and Mahmood (1995), Mata *et al*. (1995), Fernandes and Paunov (2015), and Sharapov, Kattuman, and Sena (2011).

Of the industry-level covariates, the effect of entry rate has been studied by Baldwin and Rafiquzzaman (1995) who reported that higher entry rates tend to reduce firm survival. Positive association between entry and exit rates at the industry level has been reported by Dunne et al. (1988) and Sidney *et al*. (2003). The effect of market concentration has diverge empirical findings or are reported to be insignificant (see Baldwin and Rafiquzzaman, 1995; McCloughan and Stone, 1998; Ugur *et al*., 2016a).

We control for average number of firm employees at 3-digit SIC industry level in order to address the risk of omitted variable bias, particularly when coefficient of variation is used as diversity measure. We also control for median R&D intensity in the industry to verify whether higher level of creative destruction affects firm mortality (Aghion *et al*., 2013; Fernandes and Paunov (2015); Ugur *et al*., 2016a). Finally, we check, if technology classes matter using the Pavitt (1984) taxonomy.

The final set of covariates relates to macro-level indicators such as onsets of financial crises, real effective exchange rate of the British pound, and GDP growth. Whilst currency appreciation may affect mortality because of decline in international cost competitiveness, the crisis dummy accounts for changes in the business and credit environment. Finally, GDP growth captures the effect of business cycle on firm survival (Goudie and Meeks, 1991; Bhattacharjee *et al*., 2009; Ugur *et al*., 2016).

Estimation methodology

Our estimation methodology follows Wooldridge (2010) on grouped duration data, where firm exit time is known within one year. The discrete-time hazard rate h_{it} that firm *i* exits in *Te* years conditional on survival for *Te-1* years can be expressed as conditional probability (Pr) of firm survival for T_i years as follows:

$$h_{it} = \frac{\Pr(T_{e-1} < T_i \leq T_e)}{\Pr(T_i > T_{e-1})} \tag{3}$$

Bauer and Agarwal (2014: 432) provide evidence that hazard models are "superior to the alternatives" in the context of estimating bankruptcy hazards. The parameters are estimated by maximizing the logarithm of the likelihood function. Whereas the *logit* specification assumes a logistic distribution for the hazard, the *probit* assumes a standard Normal distribution. Given the panel structure of the data we choose random effect estimations as it helps to correct for omitted variable bias (Fernandes and Paunov, 2015), whereas fixed-effect estimations often lead to large biases in all estimated parameters with relatively small number of periods in the dataset due to incidental parameters problem (Lancaster, 2000; Wooldridge, 2010, Bester and Hansen, 2009). The dependent variable is an indicator taking the value of 1 if the firm exits in year *Te,* and zero otherwise. To partially eliminate competing causal attributions, we use one-year forwarded firm exit as our dependent variable (see model 2 above).

Our panel random-effect model estimator controls for unobserved firm heterogeneity. Geroski, Mata, Portugal (2010) emphasize the importance of such control. Wooldridge (2010) demonstrates that a $\sqrt{N}$ consistent estimator in this case, the population-averaged parameters, can be obtained by maximization of the log-likelihood function $\log L$:

$$log\, L = \sum_{i=1}^{N} \sum_{t=1}^{Te} \{y_{it} log\, h_{it} + (1 - y_{it}) log(1 - h_{it})\} \tag{4}$$

For the random effects model the maximum log-likelihood estimations are based on Gauss-Hermite quadrature approximation (see Naylor and Smith, 1982) with a corresponding probability distribution hazard function *Pr(z)*. To check for robustness, we use both non-proportional hazard functions (Logit and Probit) and the proportional specification through Complementary log-log (Clog-log). Although Jenkins (1995) notes that both estimators tend to converge when hazard rates are small, it is appropriate to use both types of estimators as the hazard functions are not known *ex ante*.

The typical distribution specifications are discussed by Wooldridge (2010). For the random-effects Probit estimation the conditional probability distribution is given by the standard Normal Φ cumulative density function in 5a-5c:

$$Pr(z)=\Phi(z) = \Phi(x_{it}\beta + M_{it+1}\alpha + v_i + \gamma_{t+1} + \varepsilon_{it+1}) \quad (5a)$$

For the Logit, the probability distribution function is:

$$Pr(z) = 1/(1 + exp(-z)) \quad (5b)$$

Finally, for the Complementary log-log (Clog-log) random-effects estimator, the conditional probability function is given by:

$$Pr(z) = 1 - exp(-exp(z)) \quad (5c)$$

We use likelihood ratio test to check if the panel random-effects estimators deliver similar results with pooled estimators, i.e. if panel level variance is insignificant and the ratio $\rho = \frac{\sigma_v{}^2}{\sigma_\varepsilon{}^2 + \sigma_v{}^2}$ is different from zero by sampling chance[3]. We also use robust standard errors of the estimated parameters, which provide consistency when the disturbances are not correlated across firms.

Results

We estimate the hazard model with six different estimators: pooled Probit, pooled Logit, pooled Complementary log-log (Clog-log) and their panel random-effects versions. AIC/BIC values suggests that pooled Probit is preferable compared to pooled Logit or pooled Clog-log; and the random-effects Probit estimator is preferable against random-

[3] Stata reports panel level variance $\ln(\sigma_v{}^2)$ in form of lnsig2u_const.

effects Logit or Clog-log. Hence, we report pooled and panel random-effects Probit results in the main text in Table 2. The results are for both measures of diversity: the Theil entropy index (columns 1 and 2) and the coefficient of variation (columns 3 and 4). Results from other pooled and panel estimators are reported in Table A3a (with the Theil entropy index) and A3b (with the coefficient of variation) in the *Appendix*. We have also conducted a likelihood ratio (LR) test to check whether the panel random-effects estimators are preferable to pooled estimators as they allow for taking account of unobserved firm heterogeneity. Because the test favours the random-effects estimators, our preferred results in Table 2 are in columns 2 and 4.[4]

'**Insert Table 2 here**'

Post-estimation results for pooled Probit (bottom three rows in Table 2) indicate that: (i) the model fits the data well as the Pearson χ^2 does not reject the null hypothesis of good fit; (ii) the overall rate of correct classification is high: 95.66 percent in the estimation based on the Theil index and 95.67 percent in the estimation with the coefficient of variation respectively; and (iii) the model has good power to discriminate between exiting and surviving firm as the area under the ROC curve is 0.69 and 0.68, respectively. In addition, there is a high degree of sign and significance consistency across six estimators and two diversity measures (compare Tables 2, A3a, and A3b). The consistency is evident with respect to covariates of main interest (diversity measures, their interactions with R&D intensity, and the latter's linear and quadratic terms) and the wide range of controls discussed in Table 1.

We find strong evidential support for **H1**, which postulates that higher levels of

[4] LR test results are not reported here, but can be provided on request for Table 2 and Tables A3a and A3b.

intra-industry size diversity are associated with higher levels of exit hazard. The coefficient estimates are significant at 1 percent in preferred estimations and in others reported as robustness checks. This finding is observed after controlling for firm size and its square, and for the mean employment in the industry. Hence, the intra-industry size diversity is an exit hazard in its own right. Notice that parameter estimates for the Theil entropy index are larger than those of the coefficient of variation as the mean of the coefficient of variation in the sample is 5.0 compared to a mean Theil index of 1.9. Controlling for firm size and for mean firm size in the industry implies that the diversity-mortality relationship we report is not subject to the risk of omitted variable bias, which is highlighted by Stirling, 2010 and Solanas *et al*., 2012.

Results in Table 2 also lend support to **H2** in that an increase in R&D intensity reduces exit hazard as R&D investment enables firms to engage in active learning (Ericson and Pakes, 1995) and discover its optimal market/technology niche (Anderson and Tushman, 2001; Audretsch *et al*., 2004; Cohen and Levinthal, 1990; Probst and Raisch, 2005; and Teece, 2007). R&D investment moderates the exit hazard associated with diversity by increasing the scope for discovering, reconfiguring, and developing opportunities, and by facilitating adaptation to shifts in the technology regime (Nelson, 1995) or shifts in the technology frontier (Aghion *et al*., 2013; Ugur *et al*., 2016). Beyond these novel findings, we are also able to confirm that the relationship between R&D intensity and firm survival is subject to decreasing scale effects. As R&D intensity increases exit hazard falls at increasingly slower rates; with a turning point beyond which increased R&D intensity increases exit hazard. This finding confirms earlier results from continuous-time hazard models reported in Ugur *et al*. (2016a).

Based on estimates from the preferred model – Probit random effects models (columns 2 and 4 in Table 2) - we have estimated the conditional effects of the diversity measures conditional on different levels of R&D intensity (Table 3). Fixing the covariates at their sample means, we varied the level of R&D intensity from the bottom 5th to the top 95th percentile values. The results indicate that the hazard-increasing effect of diversity declines as R&D intensity increases. The diversity's hazard effect becomes insignificant between the 70th and 75th percentiles (at R&D intensity of approximately 0.09 – 0.11) and is eventually reversed at the top R&D intensity.

'**Insert Table 3 here**'

The results from the preferred estimators and samples above are strongly consistent with the wide range of results reported as sensitivity checks in the Appendix. Specifically, they are consistent with results obtained from: (i) alternative estimators (Table A3a and A3b); (ii) based on top R&D intensity cut-off points; (iii) stepwise regressions; (iv) accounting for left truncation by investigating firms born in 2000 or 2003 and after; (v) a sample that exclude firms in the financial and defence industries (Tables A3-A6).

Our findings for the wide range of control variables are in line with the literature. Larger size and age are associated with lower exit hazard, albeit the effect of size (employment) is subject to decreasing returns. These findings are consistent with the theoretical and empirical literature, which indicates that new entrants have shorter survival time, but those that survive grow faster and survive longer (Klette and Kortum, 2004; Aghion *et al*., 2013; Cefis and Marsili, 2005; and Evans, 1987 among others). The U-shaped relationship between size and exit hazard we report is in line with evidence on size distribution and survival among Portuguese firms (Cabral and Mata, 2003), which suggests that a large size beyond an efficient scale may be a hazard factor in firm dynamics. More importantly, the relationship between size diversity and survival is not mimicking the well-known relationship between small size and exit hazard.

We report that multi-plant firms are less likely to exit as they are better able to diversify risk and restructure similarly to Audretsch and Mahmood (1995) and Fernandes and Paunov (2015). We also report that productivity (real turnover per employee) higher firm growth rates relative to median growth in the industry are associated with lower exit hazard. These findings are in line with Doms *et al*. (1995), Mata *et al*. (1995), and Griliches and Regev (1995) among others.

Of the industry-level covariates, we report that the relationship between exit hazard and market concentration is insignificant. This is in line with prior studies, which offer two explanations: (i) market concentration may be less important than market niches in determining monopoly rents (Gerroski, 1995; Wagner, 1994); (ii) industries with similar concentration ratios often show significant variation in the overall firm-size distribution (Carroll, 1985: 1264). Four Pavitt classes are associated with lower exist hazard relative to the excluded class (Pavitt 4), which consists of firms that depend on import of technology from other industries. This finding is in line with Agarwal and Audretsch (2001) and Cefis and Marsili (2005), who indicate that the nature of the technology in the industry matters. Our finding concerning the effect of the average R&D intensity in the industry (as a proxy for intra-industry creative destruction) suggests that exit hazard increases as creative destruction increases confirming the Schumpeterian innovation models (Aghion *et al*., 2013; Ugur *et al*., 2016).

With respect to macroeconomic variables, we report that real currency appreciation (reduced competitiveness) and the onset of a financial crisis tend to increase exit hazard; whereas annual GDP growth rates have a negative relationship with exit hazard. These findings are in line with those reported in prior survival studies that control for macroeconomic variables (Bhattacharjee *et al*., 2009; Goudie and Meeks, 1991; Ugur *et al*., 2016).

Conclusion

In this paper, we have addressed two novel research questions: (i) does intra-industry size diversity constitute a hazard factor for firms beyond the direct effect of size itself? (ii) does firm investment in R&D moderate the effect of diversity on exit hazard by enabling the firm to engage in active learning about its optimal market/technology niche? We have empirically demonstrated affirmative answers to both questions. The patterns between diversity, R&D and survival time have been confirmed by results from six discrete-time hazard estimators applied to four different samples and two different diversity measures.

The main theoretical foundations may be associated with three lines in the literature. First, the diversity can be caused by low entry costs, which encourage heterogeneous firm entry and lead to higher exit rates as inefficient firms are selected out. Secondly, size diversity reflects the generic specialization and differentiation strategies (Porter, 1980), which may hinder firm dynamic capabilities by locking them into more narrow niches (see discussion in Anderson and Tushman, 2001; Levitt and March, 1993; Teece, 2007; Zahra and George, 2002). Third, greater firm-size diversity within the industry may be the outcome of a highly-dispersed firm fitness/efficiency distribution (Jovanovic, 1982) that induces high levels of firm turnover stochastically.

We also find a strong support that intra-industry firm-size diversity and R&D intensity at the firm level are substitutes in that higher levels of R&D intensity are associated with lower levels of exit hazard at each level of diversity, which could be explained by firm active learning. This suggests that firms in more diverse industries invest more in R&D to discover their optimal market/technology niches. Building on these theoretical linkages between diversity, R&D investment and firm survival, we have found strongly consistent and stable empirical support for our research hypotheses.

Our findings suggest that firms engage in two types of learning in diverse industries: (i) *passive learning* that involves the discovery of the firm's position in the efficiency/fitness distribution; and (ii) *active learning* with R&D that allows firms to identify their optimal market/technology niches and survive longer in the face of adverse shocks. Therefore we argue that intra-industry diversity is an important ecosystem (industry) property that should be taken into account in firm survival models. Overlooking diversity may lead to omitted variable bias and involve inefficient use of information that is readily observable. The practical implication is that firms in diverse industries can reduce exit hazard by investing in process and product innovation.

References

Agarwal R, Audretsch DB. 2001. Does entry size matter? The impact of the life cycle and technology on firm survival. *The Journal of Industrial Economics* **49**: 21-43.

Agarwal R, Gort M. 1996. The Evolution of Markets and Entry, Exit and Survival of Firms. *Review of Economics and Statistics* **78**(3): 489-498.

Agarwal R, Gort M. 2002. Firm and product life cycles and firm survival. *The American Economic Review* **92**(2): 184-190.

Aghion P, Harris C, Howitt P, Vickers J. 2001. Competition, imitation and growth with step-by-step innovation. *Review of Economic Studies* **68**(3): 467-492.

Aghion P, Akcigit U, Howitt P. 2013. What do we learn from Schumpeterian growth theory? Penn Institute of Economic Research Working Papers, no. 13-026. http://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/13-026.pdf

Ashby RW. 2011. Variety, constraints, and the Law of Requisite Variety, *Emergence: Complexity and Organization* **13**(1-2): 190-207.

Agarwal R, Audtretsch D. 1999. The two views of small firms in industry dynamics: A reconciliation. *Economics Letters* **62**: 245-251.

Anderson P, Tushman M. 2001. Organizational Environments and Industry Exit: the Effects of Uncertainty, Munificence and Complexity. *Industrial and Corporate Change* **10**(3): 675-711.

Audretsch D. 1991. New-Firm Survival and the Technological Regime. *Review of Economics and Statistics* **60**(3): 441-450.

Audretsch DB. 1995. Innovation, growth and survival. *International Journal of Industrial Organization* **13**(4): 441-457.

Audretsch DB, Houweling P, Thurik AR. 2004. Industry evolution: Diversity, selection and the role of learning. *International Small Business Journal* **22**(4): 331-348.

Audretsch D, Mahmood T. 1995. New Firm Survival: New Results Using a Hazard Function. *The Review of Economics and Statistics* **77**(1): 97-103.

Audretsch DB, Houweling P, Thurik A. 2004. Industry evolution: Diversity, selection, and the role of learning. *International Small Business Journal* **22**(4): 331-348.

Baldwin J, Rafiquzzaman M. 1995. Selection versus evolutionary adaptation: learning and post-entry performance. *International Journal of Industrial Organization* **13**: 501-522.

Barnett W, Hansen M. 1996. The Red Queen in organizational evolution. *Strategic Management Journal* **17**: 137-157.

Bauer J, Agarwal V. 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance* **40**: 432-442.

Bester A, Hansen C. 2009. A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects. *Journal of Business & Economic Statistics* **27**(2): 131-48.

Bhattacharjee A, Higson C, Holly S, Kattuman P. 2009. Macroeconomic instability and business exit: Determinants of failures and acquisitions of UK firms. *Economica* **76**(301): 108-131.

Cadot O, Carrere C, Strauss-Kahn V. 2013. Trade diversification, income, and growth: What do we know? *Journal of Economic Surveys* **27**(4): 790-812.

Carroll GR, Hannan MT 2000. *The demography of corporations and industries*. Chapter 12. Princeton University Press: Princeton.

Carroll GR. 1985. Concentration and specialization: Dynamics of niche width in populations of Organizations. *American Journal of Sociology* **90:** 1262-283.

Cefis E, Marsili O. 2005. A matter of life and death: innovation and firm survival. *Industrial and Corporate Change* **14**(6): 1167-1192.
Cohen W, Klepper S. 1992. The trade-off between firm size and diversity in the pursuit of technological progress. *Small Business Economics* **4**(1): 1-14.
Cohen W, Levinthal D. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly* **35**(1): 128-152.
Disney R, Haskel J, Heden Y. 2003. Entry, exit and establishment survival in UK manufacturing. *Journal of Industrial Economics* **51**(1): 91-112
Dobrev SD, Kim TY, Hannan MT. 2001. Dynamics of niche width and resource partitioning. *American Journal of Sociology* **106**(5): 1299-337.
Doms M, Dunne T, Roberts MJ. 1995. The role of technology use in the survival and growth of manufacturing plants. *International Journal of Industrial Organization* **13**: 523-542.
Dosi G, Malerba F, Marsili O, Orsenigo L. 1997. Industrial Structures and Dynamics: Evidence, Interpretations and Puzzles. *Industrial and Corporate Change* **6**(1): 3-25.
Dunne T, Roberts MJ, and Samuelson L. 1988. Patterns of firm entry and exit in U.S. manufacturing industries. *RAND Journal of Economics* **19**(4): 495-515.
Elton CS. 1958. *The Ecology of Invasions by Plants and Animals*. Methuen & Co. Ltd.: London.
Ericson R, Pakes A. 1995. Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies* **62**(1): pp. 53-82.
Evans DS. 1987. The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The Journal of Industrial Economics* **35**: 567-581.
Fernandes A, Paunov C. 2015. The risks of innovation: Are innovating firms less likely to die? *The Review of Economics and Statistics* **97**(3): 638-53.
Gardner MR, Ashby WR. 1970. Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature* **228**: 784.
Garmestani A, Allen C, Mittelstaedt J, Stow C, Ward W. 2006. Firm size diversity, functional richness, and resilience. *Environment and Development Economics* **11**: 533-551.
Geroski PA. 1995. What do we know about entry? *International Journal of Industrial Organization* **13**(4): 421-440.
Geroski PA. 2001. Exploring the niche overlaps between organizational ecology and industrial economics. *Industrial and Corporate Change* **10**(2): 507-540.
Geroski P, Mata J, Portugal P. 2010. Founding conditions and the survival of new firms. *Strategic Management Journal* **31**: 510-529.
Goudie AW, Meeks G. 1991. The exchange rate and company failure in a macro-micro model of the UK company sector. *Economic Journal* **101**: 444-457.
Hannan MT, Freeman J. 1977. The population ecology of organizations. *American Journal of Sociology* **82**(5): 929-964.
Hannan MT, Freeman J. 1989. *Organizational Ecology*. Harvard University Press: Cambridge, MA.
Hannan M. 2005. Ecologies of organizations: Diversity and identity. *Journal of Economic Perspectives* **19**(10): 51-70.
Haughton J, Khandker S. 2009. *Handbook on Poverty and Inequality*. The World Bank: Washington, D.C.
Hopenhayn H. 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* **60** (5): 1127–1150.
Ives AR, Carpenter SR. 2007. *Science* **317** (6 July): 58-62.

Jenkins S. 1995. Easy ways to estimate discrete time duration models. *Oxford Bulletin of Economics and Statistics* **57**: 129-138.
Jovanovic B. 1982. Selection and the Evolution of Industry. *Econometrica* **50**(3): 649-670.
Klepper S. 1996. Entry, exit, growth, and innovation over the product life cycle. *The American Economic Review* **86**(3): 562-584.
Klepper S, Simon K. 1996. Innovation and industry shakeouts. *Business and Economic History* **25**(1): 81-89.
Klette TJ, Kortum S. 2004. Innovating firms and aggregate innovation. *Journal of Political Economy* **112**(5): 986–1018.
Lancaster T. 1990. *The Econometric Analysis of Transition Data*. Cambridge University Press: Cambridge.
Lancaster T. 2000. The incidental parameter problem since 1948. *Journal of Econometrics* **95**(2): 391-413.
Lehman CL, Tilman D. 2000. Biodiversity, stability, and productivity in competitive communities. *The American Naturalist* **156**(5): 534-552.
Levitt B, March JG. 1993. Myopia of learning. *Strategic Management Journal* **14**: 95-112.
Mata J, Portugal P. 1994. Life duration of new firms. *The Journal of Industrial Economics* **42**(3): 227-245.
Mata J, Portugal P, Guimaraes P. 1995. The survival of new plants: Start-up conditions and post-entry evolution. *International Journal of Industrial Organization* **13**: 459-481.
Mata J, Portugal P. 2002. The survival of new domestic and foreign owned firms. *Strategic Management Journal* **23**: 323-43.
May RM. 1974. Stability and Complexity in Model Ecosystems. 2nd ed. Princeton University Press: NJ.
McCann KS. 2000. The diversity-stability debate. *Nature* **405**(11 May): 228-233.
McCloughan P, Stone I. 1998. Life Duration of Foreign Multinational Subsidiaries: Evidence from UK Northern Manufacturing Industry 1970-93. *International Journal of Industrial Organization* **16**(6): 719-747.
McGahan A, Argyres N, Baum J. 2004. Context, technology and strategy: Forging new perspectives on the industry life cycle. In: *Advances in strategic management: Business strategy over the industry life cycle*, Baum J, McGahan A (eds.). Vol.21. Elsevier: Amsterdam; 1-21.
McNaughton SJ. 1993. Biodeiversity and functioning of grazing ecosystems. In *Biodiversity and Ecosystem Function*, Schulze E, Mooney H (eds). Springer-Verlag: Berlin; 361-383.
Naylor J, Smith AF. 1982. Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society, Series C* **31**: 214-225.
Nelson R, 1995. Co-evolution of industry structure, technology and supporting institutions, and the making of comparative advantage. *International Journal of the Economics of Business* **2**(2): 171-184.
Nelson R, Winter S. 1982. *An Evolutionary Theory of Economic Change*. Harvard University Press: Cambridge, MA.
Nicoletti C, Rondinelli C. 2010. The (mis)specification of discrete duration models with unobserved heterogeneity: A Monte Carlo study. *Journal of Econometrics* **159**(1): 1-13.
Noda T, Collis DJ. 2001. The evolution of intra-industry firm heterogeneity: Insights from a process study. *Academy of Management Journal* **44**(4): 897-925.
Pavitt K. 1984. Sectoral patterns of technical change: towards a taxonomy and a theory. *Research Policy* **13**(6): 343-373.

Peltoniemi M. 2011. Reviewing industry life-cycle theory: Avenues for future research. *International Journal of Management Reviews* **13**: 349-375.
Porter M. 1980. *Competitive Strategy*. Free Press: NY.
Probst G, Raisch S. 2005. Organizational crisis: The logic of failure. *Academy of Management Perspectives* **19**(1): 90-105.
Safarzynska K, van der Bergh J. 2010. Evolutionary models in economics: A survey of methods and building blocks. *Journal of Evolutionary Economics* **20**: 329-373.
Schumpeter JA. 1942. *Capitalism, Socialism and Democracy*. Harper & Row: New York.
Sharapov D, Kattuman P, Sena V. 2011. Technological environments, R&D investment, and firm survival. MICRO-DYN Working Paper no. 34/10.
Stirling A. 2010. Multicriteria diversity analysis: A novel heuristic framework for apprising energy portfolios. *Energy Policy* **38**: 1622-1634.
Teece D. 2007. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal* **28**: 1319-1350.
Theil H. 1972. *Statistical Decomposition Analysis*. North-Holland Publishing: Amsterdam.
Thomspon P. 2005. Selection and firm survival: evidence from shipbuilding industry, 1825-1914. *Review of Economics and Statistics* **87**: 26-36.
Utterback J, Suarez F. 1993. Innovation, competition and industry structure. *Research Policy* **22**: 1-21.
Ugur M, Trushin E, Solomon E. 2016a. Inverted-U relationship between R&D intensity and survival: Evidence on scale and complementarity effects in UK data. *Research Policy* **45**(7): 1474–1492.
Ugur M, Trushin E, Solomon E. 2016b. A firm-level dataset for analyzing entry, exit, employment and R&D expenditures in the UK: 1997-2012. *Data in Brief* **8**: 153–157.
Yodzis P. 1980. The connectance of real ecosystems. *Nature* **284**: 544 – 545.
Williamson OE. 2005. Hierarchies, markets and power in the economy: An economic perspective. In *Understanding Industrial and Corporate Change*. Dosi G, Teese D, Chytry J. (eds). Oxford University Press: Oxford; 109-136.
Wooldridge J. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. The MIT Press: Cambridge, MA.
Zahra S, George G. 2002. Absorptive capacity: a review, reconceptualization, and extension. *Academy of Management Review* **27**: 185-203.
Zhang M, Mohnen P. 2013. Innovation and survival of new firms in Chinese manufacturing, 2000-2006. UNU-MERIT Working Paper 2013-057. www.merit.unu.edu/publications/wppdf/2013/wp2013-057.pdf.

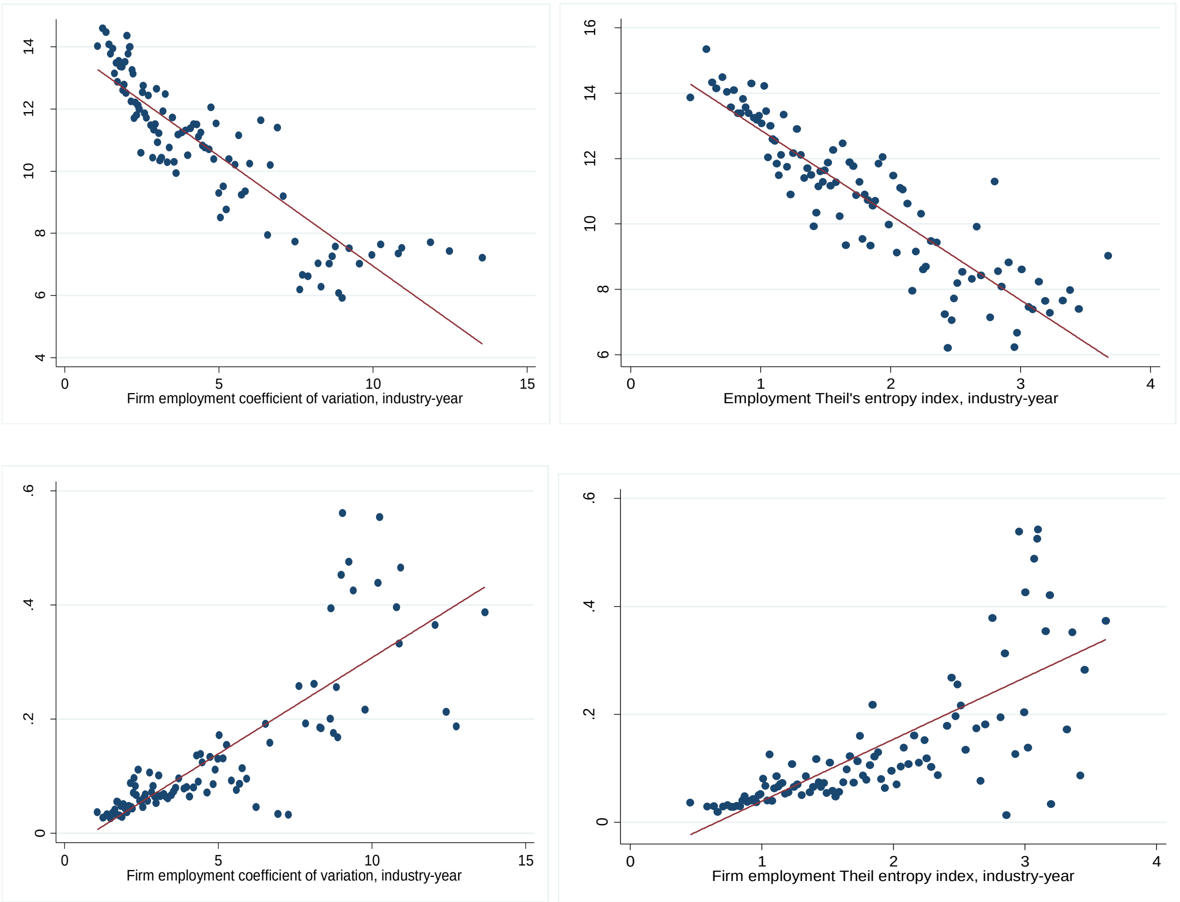

Figure 1: Within-industry size diversity, survival times and R&D intensity.

Note: Coefficient of variation, the Theil entropy index, survival times and R&D intensity are all averages at the 3-digit industry level.

Table 1: Covariates and expected effects on exit hazard

Covariate	Description (expected sign of effect on exit hazard)	Related literature
	Covariates of main interest	
Firm-size diversity (*TI. or C.V.*)	Theil's entropy (*TI*) and coefficient of variation (*CV*) of firm employment in 3-digit industries **(+)**	Not tested before for firm survival
Interactions *TI*log(R&Dint.+1), CV*log(R&Dint.+1)*	Interaction of firm size diversity measures with firm R&D intensity **(-)**	Not tested before for firm survival
	Other firm-level covariates	
R&D intensity		
Ln(R&Dint.+1)	Logarithm of firm R&D intensity **(-)**	Aghion *et al.* (2001; 2013); Ericson and Pakes (1995)
Ln(R&Dint.+1) sq.	Squared logarithm of R&D intensity (+)	Aghion *et al.* (2013); Ericson and Pakes (1995); Sharapov et al. (2011); Zhang and Mohnen (2013), Ugur *et al.* (2016a).
Age *log(firm Age)*	Logarithm of firm age in years **(-)**	Hopenhayn (1992); Ericson and Pakes (1995); Geroski, 1995; Cefis and Marsili (2005); Doms *et al.* (1995); Disney *et al.* (2003)
Age squared *log(firm Age) sq.*	Squared logarithm of firm age **(+)**	Agarwal and Gort (2002); Ericson and Pakes (1995); Cefis and Marsili (2005); Evans (1987)
Size		
Ln(Employment)	Logarithm of firm employees **(-)**	Hopenhayn (1992); Ericson and Pakes (1995); Geroski, 1995; Cefis and Marsili (2005); Doms *et al.* (1995); Disney *et al.* (2003)
Size squared *Ln(Empl.) squared*	Squared log. of firm employees **(+)**	Bhattacharjee *et al.* (2009); Cefis and Marsili (2005)
Local units *Ln(Plants)*	Logarithm of firm's local units (plants) **(+)**	Audretsch and Mahmood (1995); Fernandes and Paunov (2015); Audretsch (1991); Griliches and Regev (1995); Mata *et al.* (1995)
Productivity *Ln(Rturn./empl.)*	Logarithm of deflated turnover per employee **(-)**	Audretsch, 1991; Hopenhayn (1992); Ericson and Pakes (1995)
Growth differential *(Growth_dmed)*	Growth rate of firms' deflated turnover minus median industry growth rate **(-)**	Audretsch, 1991; Hopenhayn (1992); Ericson and Pakes (1995); Cefis and Marsili (2005); Mata *et al.* (1995), Audretsch (1995), Ugur *et al.* (2016a)
Civil R&D *(Civilian R&D only)*	Dummy variable indicating that firm is engaged in civilian R&D only **(+/ -)**	Ugur *et al.* (2016a), Sharapov *et al.* (2011)
UK-owned *(UK owned)*	Dummy variable indicating that firm is UK-owned **(+ /-)**	Ugur *et al.* (2016a), Sharapov *et al.* (2011)
Industry covariates		
Concentration *Herfhindahl index*	Herfindahl-Hirschman index of firm shares in industry turnover at 3-digit industry level **(+/-)**	McCloughan and Stone (1998); Baldwin and Rafiquzzaman (1995); Wagner (1994); Geroski (1995)
Pavitt technology class* *(Pavitt #)*	Dummy variables for Pavitt classes 1 to 5, excluded category is Pavitt 4 **(+/-)**	Pavitt (1984); Agarwal and Audretsch (2001); Cefis and Marsili (2005), Ugur *et al.* (2016a)
Entry rate *log(% entry rate)*	Logarithm of firm entry rate (in %) at 3-digit SIC industry level **(+)**	Hannan and Freeman (1989); Fernandes and Paunov (2015)
Median industry R&D intensity *Ln(Med. R&D int.)*	Logarithm of industry median ratio of total R&D to turnover at 3-digit SIC level **(-)**	Audretsch and Mahmood (1995); Fernandes and Paunov (2015), Ugur *et al.* (2016a)
Average firm size in the industry *Ln(Mean empl.)*	Logarithm of average employees per firm in 3-digit SIC industry level **(+/-)**	Fernandes and Paunov (2015); Mata and Portugal (2002); Audretsch *et al.* (2004)
Macroeconomic indicators		

Crisis year *(Arisis)*	A dummy variable equal 1 for the Asian crisis year of 1998; *dot.com* bubble crisis of 2001; and the recent financial crisis in 2008 **(+)**	Ugur et al. (2016a); Bhattacharjee *et al*. (2009) report higher hazard rates in periods of crises
Average real effective exchange rate *(Areer)*	Average effective exchange rate against a basket of currencies - an increases in *Areer* indicates appreciation **(+)**	Bhattacharjee *et al*. (2009); Goudie and Meeks (1991)
GDP growth rate	Growth rate of GDP, annual % **(-)**	Business cycle literature; Thompson (2005) for industry output, Mata and Portugal (2002) for employment growth

Note: * Pavitt technology classes are from Pavitt (1984), as revised slightly by Bogliacino and Pianta (2010). Pavitt1 consists of firms in science-based industries such as chemicals, office machinery, precision, medical and optical instruments industries, ICT. Pavitt2 includes specialized suppliers of technology - mechanical engineering industries, manufacturers of electrical machinery, equipment, etc. Pavitt3 includes scale-intensive industries such as pulp and paper, transport vehicles, mineral oil refining industries. Pavitt4 includes industries dominated by technology suppliers, e.g., textiles & clothing, food & drink, fabricated metals. Finally, Pavitt5 consists of unclassified industries.

Table 2: Intra-industry firm size diversity and exit hazard

	Theil entropy index		Coefficient of variation	
Dependent variable: exit in year t+1	(1)	(2)#	(3)	(4)#
Diversity	0.0377***	0.0398***	0.0103***	0.0112***
	(0.0109)	(0.0116)	(0.0030)	(0.0032)
Diversity*log(R&Dint.+1)	-0.244***	-0.247***	-0.0581***	-0.0588***
	(0.0588)	(0.0617)	(0.0143)	(0.0149)
Log(R&Dint.+1)	-0.578***	-0.599***	-0.755***	-0.779***
	(0.200)	(0.205)	(0.181)	(0.185)
Log(R&Dint.+1) sq.	1.250***	1.266***	1.274***	1.290***
	(0.271)	(0.282)	(0.272)	(0.283)
Log(firm Age)	-0.222***	-0.214***	-0.222***	-0.215***
	(0.0107)	(0.0114)	(0.0107)	(0.0114)
Log(Employment)	-0.197***	-0.217***	-0.197***	-0.218***
	(0.0121)	(0.0137)	(0.0121)	(0.0137)
Log(Employm.) sq.	0.0191***	0.0209***	0.0191***	0.0209***
	(0.0017)	(0.0019)	(0.0017)	(0.0019)
Log(Real turnover /	-0.0992***	-0.104***	-0.0993***	-0.104***
employees)	(0.0079)	(0.0079)	(0.0079)	(0.0079)
Firm growth relative	-0.0674***	-0.0681***	-0.0674***	-0.0681***
to industry median growth	(0.0102)	(0.0096)	(0.0102)	(0.0096)
Log(Plants)	-0.0226	-0.0202	-0.0211	-0.0186
	(0.0157)	(0.0162)	(0.0157)	(0.0162)
Civil R&D only	-0.0920***	-0.0951***	-0.0925***	-0.0957***
	(0.0130)	(0.0138)	(0.0130)	(0.0138)
UK-owned	-0.0672***	-0.0723***	-0.0665***	-0.0717***
	(0.0211)	(0.0222)	(0.0211)	(0.0222)
Log(% entry rate)	-0.0673	-0.0743	-0.0595	-0.0643
	(0.0595)	(0.0629)	(0.0599)	(0.0635)
Log(Mean industry	0.0160**	0.0152*	0.0193**	0.0187**
employment)	(0.0079)	(0.0085)	(0.0078)	(0.0083)
Log(Median R&D int.	0.633***	0.647***	0.630***	0.642***
in industry)	(0.114)	(0.120)	(0.112)	(0.119)
Herfindahl index	-0.0434	-0.0540	-0.0387	-0.0503
	(0.0614)	(0.0651)	(0.0602)	(0.0642)
Pavitt 1	-0.0637***	-0.0578**	-0.0656***	-0.0608***
	(0.0216)	(0.0232)	(0.0220)	(0.0235)
Pavitt 2	-0.101***	-0.103***	-0.101***	-0.104***
	(0.0181)	(0.0195)	(0.0182)	(0.0195)
Pavitt 3	-0.0253	-0.0218	-0.0264	-0.0228
	(0.0234)	(0.0252)	(0.0234)	(0.0251)
Pavitt 5	-0.0494*	-0.0539**	-0.0463*	-0.0505*
	(0.0257)	(0.0274)	(0.0257)	(0.0274)
Average effective real	0.0120***	0.0119***	0.0120***	0.0119***
exchange rate	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Crisis dummy	0.0660***	0.0620***	0.0661***	0.0622***
	(0.0154)	(0.0159)	(0.0154)	(0.0160)
GDP growth (%)	-0.0250***	-0.0266***	-0.0251***	-0.0266***
	(0.0038)	(0.0036)	(0.0038)	(0.0036)
Constant	-1.389***	-1.379***	-1.382***	-1.374***
	(0.116)	(0.120)	(0.116)	(0.120)
Log($\sigma_v{}^2$)		-2.351***		-2.351***
		(0.0460)		(0.0462)
N	158,316	158,316	158,313	158,313
AIC	53,695.9	53,704.9	53,692.6	53,701.1
BIC	53,935.2	53,954.2	53,931.9	53950.4
Log-likelihood	-26,824.0	-26,827.4	-26,822.3	-26,825.6
chi2	2,712.0	2,466.3	2,710.4	2,464.3
3-digit industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Correctly classified	95.66%	N/A	95.67%	N/A
p > Pearson χ^2	0.98	N/A	0.81	N/A
Area under ROC cirve	0.69	N/A	0.68	N/A

Notes: #: Preferred model. Top R&D intensity is less than 1. Estimators: (1) and (3) – pooled Probit; (2) and (4) – panel Probit with random effects. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 3. Conditional effects of diversity on exit hazard depending on R&D intensity

Percentile of R&D intensity	R&D intensity	Effects at mean Theil entropy of firm employment	Effects at mean coefficient of variation of firm employment
5th	0.0009	0.0395***	0.0111 ***
		(0.0116)	(0.0032)
15th	0.0049	0.0385***	0.0108***
		(0.0116)	(0.0032)
25th	0.0101	0.0373***	0.0105***
		(0.0116)	(0.0032)
35th	0.0169	0.0356***	0.0101***
		(0.0116)	(0.0032)
45th	0.0261	0.0334***	0.0096***
		(0.0116)	(0.0032)
55th	0.0404	0.0298 ***	0.0087***
		(0.0116)	(0.0032)
65th	0.0656	0.0236**	0.0073**
		(0.0116)	(0.0032)
70th	0.0859	0.0186	0.0061*
		(0.0116)	(0.0032)
75th	0.1122	0.0121	0.0045
		(0.0116)	(0.0032)
80th	0.1467	0.0036	0.0026
		(0.0116)	(0.0032)
85th	0.1927	-0.0077	-0.0001
		(0.0116)	(0.0032)
95th	0.3812	-0.0542***	-0.0112***
		(0.0116)	(0.0032)

Note: Other covariates are taken at their mean. Standard errors are in parentheses. Top R&D intensity is less than 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1. Summary statistics

	Survival firms		Exiter firms		Full sample	
	mean	st.d.	mean	st.d.	mean	st.d.
Theil entropy	1.8428	0.8095	2.0504	0.8489	1.8632	0.8155
lrdint_theil	0.1917	0.3379	0.2707	0.3932	0.1992	0.3442
empl_cv	4.8792	3.0757	5.6944	3.4126	4.9610	3.1215
empl_cv_lrnd	0.5729	1.1047	0.8235	1.2988	0.5970	1.1262
lenterate	0.6409	0.0967	0.6327	0.0999	0.6399	0.0980
lrdint	0.0830	0.1247	0.1123	0.1440	0.0858	0.1270
lrdint2	0.0224	0.0623	0.0333	0.0751	0.0235	0.0637
Log (age + 1)	2.6522	0.6810	2.1294	0.7415	2.6077	0.6244
Log (employment + 1)	3.0155	1.6735	2.3002	1.6126	2.9451	1.6887
Log employment squared	11.8941	12.0226	7.8913	10.2346	11.5252	11.9818
Log (live local units + 1)	0.6941	0.5670	0.5182	0.5503	0.6719	0.5706
Productivity: (LogRturn_empl)	4.3776	.9846	4.0777	1.0315	4.3492	0.9949
growth_dmed	0.04889	0.5867	-0.0297	0.7227	0.0430	0.6063
mean_empl	145.889	505.678	133.277	491.224	145.548	518.905
herfindahl	0.09821	0.1082	0.0967	0.1051	0.0982	0.1080
civil_dummy	0.4304	0.4951	0.3501	0.4770	0.4248	0.4943
uk_owner	0.8747	0.3310	0.9272	0.2596	0.8787	0.3264
RnDint_med	0.0836	0.1118	0.1201	0.1339	0.0973	0.1061
lmean_empl	4.3838	0.9524	4.0748	1.0791	4.3635	0.9655
lRnDint_med	0.0679	0.0882	0.0907	0.1045	0.0701	0.0902
Pavitt1	0.3149	0.4644	0.3974	0.4894	0.3233	0.4677
Pavitt2	0.2244	0.4171	0.2065	0.4049	0.2233	0.4164
Pavitt3	0.1005	0.3007	0.0753	0.2639	0.0981	0.2975
Pavitt5	0.0654	0.2473	0.0694	0.2543	0.0657	0.2478
Average effective exchange rate index	92.0826	9.4775	93.9319	9.1565	92.2903	9.4681
dummy for crisis years: 1998, 2001, and 2008	0.1539	0.3608	0.1936	0.3952	0.1556	0.3625
GDP growth rate, %	1.5457	2.1824	1.6766	2.3733	1.5571	2.1834
Number of firms	28,287		6,849		35,136	
Observations/counts	151,467				158,316	

* Note: minimum and maximum values are suppressed to comply with non-disclosure requirements of the data hosts, UK Data Service. Pavitt technology classes: 1 - science-based industries; 2 - specialised suppliers of technology; 3 - scale-intensive industries; 4 - industries dominated by suppliers of technology; 5 – unclassified. Onset of crisis dummy takes value of 1 if year is either 1998, 2001 or 2008. Turnover is deflated by 2-digt output deflator with base year at 2010. The Herfindahl index is based on firm turnover at 3-digit industry level. The GDP growth rate and the average effective real exchange rate are from the World Bank Development Indicators 2016 (www.data.worldbank.org/data-catalog/world-development-indicators).

Table A2a. Intra-industry firm size diversity (Theil index) and exit hazard

	(1)	(2)	(3)	(4)
Theil index (TI)	0.0820***	0.0793***	0.0827***	0.0795***
	(0.0240)	(0.0234)	(0.0243)	(0.0235)
TI*log(R&Dint.+1)	-0.522***	-0.500***	-0.522***	-0.499***
	(0.122)	(0.116)	(0.122)	(0.115)
Log(R&Dint.+1)	-1.253***	-1.192***	-1.256***	-1.192***
	(0.428)	(0.412)	(0.430)	(0.411)
Log(R&Dint.+1) sq.	2.623***	2.483***	2.623***	2.481***
	(0.568)	(0.543)	(0.570)	(0.542)
Log(firm Age)	-0.463***	-0.444***	-0.459***	-0.442***
	(0.0231)	(0.0224)	(0.0237)	(0.0230)
Log(Employment)	-0.426***	-0.412***	-0.432***	-0.415***
	(0.0264)	(0.0256)	(0.0282)	(0.0270)
Log(Employm.) sq.	0.0411***	0.0398***	0.0417***	0.0400***
	(0.0037)	(0.0036)	(0.0039)	(0.0038)
Log(Real turnover /	-0.220***	-0.212***	-0.221***	-0.213***
employees)	(0.0173)	(0.0166)	(0.0173)	(0.0166)
Firm growth relative	-0.136***	-0.127***	-0.136***	-0.126***
to industry median growth	(0.0213)	(0.0199)	(0.0214)	(0.0200)
Log(Plants)	-0.0656*	-0.0662*	-0.0637*	-0.0654*
	(0.0353)	(0.0345)	(0.0358)	(0.0347)
Civil R&D only	-0.200***	-0.194***	-0.201***	-0.194***
	(0.0285)	(0.0277)	(0.0286)	(0.0277)
UK-owned	-0.158***	-0.155***	-0.159***	-0.155***
	(0.0484)	(0.0474)	(0.0489)	(0.0477)
Log(% entry rate)	-0.150	-0.146	-0.151	-0.146
	(0.129)	(0.125)	(0.130)	(0.126)
Log(Mean industry	0.0364**	0.0352**	0.0358**	0.0349**
employment)	(0.0173)	(0.0167)	(0.0172)	(0.0165)
Log(Median R&D int.	1.413***	1.364***	1.414***	1.364***
in industry)	(0.236)	(0.225)	(0.239)	(0.226)
Herfindahl index	-0.0779	-0.0718	-0.0824	-0.0736
	(0.137)	(0.133)	(0.138)	(0.134)
Pavitt 1	-0.142***	-0.139***	-0.140***	-0.138***
	(0.0478)	(0.0464)	(0.0486)	(0.0469)
Pavitt 2	-0.213***	-0.205***	-0.214***	-0.205***
	(0.0403)	(0.0392)	(0.0408)	(0.0395)
Pavitt 3	-0.0565	-0.0557	-0.0549	-0.0551
	(0.0531)	(0.0518)	(0.0539)	(0.0523)
Pavitt 5	-0.116**	-0.114**	-0.118**	-0.114**
	(0.0558)	(0.0540)	(0.0561)	(0.0539)
Average effective real	0.0266***	0.0259***	0.0265***	0.0258***
exchange rate	(0.0020)	(0.0020)	(0.0021)	(0.0020)
Crisis dummy	0.140***	0.134***	0.138***	0.133***
	(0.0330)	(0.0318)	(0.0333)	(0.0321)
GDP growth (%)	-0.0589***	-0.0580***	-0.059***	-0.058***
	(0.0087)	(0.0085)	(0.0088)	(0.0086)
Constant	-2.477***	-2.541***	-2.480***	-2.543***
	(0.260)	(0.253)	(0.262)	(0.254)
lnsig2u			-2.648**	-3.569
_cons			(1.184)	(2.595)
N	158,316	158,316	158,316	158,316
AIC	53,709.5	53,715.4	53,710.9	53,717.2
BIC	53,948.8	53,954.7	53,960.2	53,966.5
3-digit industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Notes: Notes: Top R&D intensity is less than 1. Estimators: (1) – pooled Logit; (2) - pooled Clog-log; (3) - Logit random effects; (4) – Clog-log random effects. See Table 2 for other notes.

Table A2b. Intra-industry firm size diversity (coefficient of variation) and exit hazard

	(1)	(2)	(3)	(4)
Employment CV	0.0224***	0.0217***	0.0227***	0.0218***
	(0.0066)	(0.0064)	(0.0067)	(0.0064)
CV*Log(R&Dint.+1)	-0.124***	-0.119***	-0.124***	-0.119***
	(0.0299)	(0.0286)	(0.0299)	(0.0284)
Log(R&Dint.+1)	-1.633***	-1.556***	-1.637***	-1.556***
	(0.389)	(0.374)	(0.391)	(0.374)
Log(R&Dint.+1) sq.	2.674***	2.534***	2.674***	2.532***
	(0.569)	(0.544)	(0.572)	(0.544)
Log(firm Age)	-0.464***	-0.445***	-0.459***	-0.443***
	(0.0231)	(0.0223)	(0.0238)	(0.0230)
Log(Employment)	-0.427***	-0.413***	-0.434***	-0.416***
	(0.0263)	(0.0256)	(0.0282)	(0.0269)
Log(Employm.) sq.	0.0410***	0.0397***	0.0416***	0.0399***
	(0.0037)	(0.0036)	(0.0039)	(0.0038)
Log(Real turnover /	-0.220***	-0.212***	-0.221***	-0.213***
employees)	(0.0173)	(0.0166)	(0.0173)	(0.0166)
Firm growth relative	-0.136***	-0.126***	-0.136***	-0.126***
to industry median growth	(0.0213)	(0.0199)	(0.0214)	(0.0200)
Log(Plants)	-0.0625*	-0.0632*	-0.0603*	-0.0623*
	(0.0354)	(0.0345)	(0.0360)	(0.0348)
Civil R&D only	-0.202***	-0.195***	-0.202***	-0.195***
	(0.0285)	(0.0277)	(0.0287)	(0.0277)
UK-owned	-0.157***	-0.154***	-0.158***	-0.154***
	(0.0484)	(0.0474)	(0.0489)	(0.0477)
Log(% entry rate)	-0.131	-0.127	-0.131	-0.127
	(0.130)	(0.126)	(0.131)	(0.126)
Log(Mean industry	0.0435**	0.0421**	0.0430**	0.0419***
employment)	(0.0170)	(0.0164)	(0.0169)	(0.0162)
Log(Median R&D int.	1.405***	1.357***	1.406***	1.357***
in industry)	(0.232)	(0.221)	(0.236)	(0.223)
Herfindahl index	-0.0715	-0.0664	-0.0770	-0.0685
	(0.134)	(0.130)	(0.135)	(0.131)
Pavitt 1	-0.148***	-0.145***	-0.145***	-0.144***
	(0.0487)	(0.0474)	(0.0496)	(0.0478)
Pavitt 2	-0.216***	-0.208***	-0.217***	-0.208***
	(0.0404)	(0.0393)	(0.0410)	(0.0396)
Pavitt 3	-0.0591	-0.0583	-0.0574	-0.0577
	(0.0530)	(0.0517)	(0.0538)	(0.0522)
Pavitt 5	-0.110**	-0.107**	-0.111**	-0.108**
	(0.0559)	(0.0541)	(0.0562)	(0.0539)
Average effective real	0.0265***	0.0258***	0.0264***	0.0258***
exchange rate	(0.0020)	(0.0020)	(0.0021)	(0.0020)
Crisis dummy	0.140***	0.134***	0.138***	0.133***
	(0.0330)	(0.0318)	(0.0333)	(0.0321)
GDP growth (%)	-0.0592***	-0.0583***	-0.0595***	-0.0584***
	(0.0088)	(0.0086)	(0.0088)	(0.0086)
Constant	-2.463***	-2.527***	-2.466***	-2.529***
	(0.259)	(0.252)	(0.262)	(0.254)
lnsig2u			-2.511**	-3.540
_cons			(1.043)	(2.527)
N	158,313	158,313	158,313	158,313
AIC	53,706.2	53,712.0	53,707.6	53,713.8
BIC	53,945.5	53,951.4	53,956.9	53,963.1
3-digit industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Notes: Notes: Top R&D intensity is less than 1. Estimators: (1) – pooled Logit; (2) - pooled Clog-log; (3) - Logit random effects; (4) – Clog-log random effects. See Table 2 for other notes.

Table A3a. Intra-industry firm size diversity and exit hazard: Result consistency by estimator (*Theil index*)

	(1)	(2)	(3)	(4)
Theil index (TI)	0.0820***	0.0793***	0.0827***	0.0795***
	(0.0240)	(0.0234)	(0.0243)	(0.0235)
TI*log(R&Dint.+1)	-0.522***	-0.500***	-0.522***	-0.499***
	(0.122)	(0.116)	(0.122)	(0.115)
Log(R&Dint.+1)	-1.253***	-1.192***	-1.256***	-1.192***
	(0.428)	(0.412)	(0.430)	(0.411)
Log(R&Dint.+1) sq.	2.623***	2.483***	2.623***	2.481***
	(0.568)	(0.543)	(0.570)	(0.542)
Log(firm Age)	-0.463***	-0.444***	-0.459***	-0.442***
	(0.0231)	(0.0224)	(0.0237)	(0.0230)
Log(Employment)	-0.426***	-0.412***	-0.432***	-0.415***
	(0.0264)	(0.0256)	(0.0282)	(0.0270)
Log(Employm.) sq.	0.0411***	0.0398***	0.0417***	0.0400***
	(0.0037)	(0.0036)	(0.0039)	(0.0038)
Log(Real turnover /	-0.220***	-0.212***	-0.221***	-0.213***
employees)	(0.0173)	(0.0166)	(0.0173)	(0.0166)
Firm growth relative	-0.136***	-0.127***	-0.136***	-0.126***
to industry median growth	(0.0213)	(0.0199)	(0.0214)	(0.0200)
Log(Plants)	-0.0656*	-0.0662*	-0.0637*	-0.0654*
	(0.0353)	(0.0345)	(0.0358)	(0.0347)
Civil R&D only	-0.200***	-0.194***	-0.201***	-0.194***
	(0.0285)	(0.0277)	(0.0286)	(0.0277)
UK-owned	-0.158***	-0.155***	-0.159***	-0.155***
	(0.0484)	(0.0474)	(0.0489)	(0.0477)
Log(% entry rate)	-0.150	-0.146	-0.151	-0.146
	(0.129)	(0.125)	(0.130)	(0.126)
Log(Mean industry	0.0364**	0.0352**	0.0358**	0.0349**
employment)	(0.0173)	(0.0167)	(0.0172)	(0.0165)
Log(Median R&D int.	1.413***	1.364***	1.414***	1.364***
in industry)	(0.236)	(0.225)	(0.239)	(0.226)
Herfindahl index	-0.0779	-0.0718	-0.0824	-0.0736
	(0.137)	(0.133)	(0.138)	(0.134)
Pavitt 1	-0.142***	-0.139***	-0.140***	-0.138***
	(0.0478)	(0.0464)	(0.0486)	(0.0469)
Pavitt 2	-0.213***	-0.205***	-0.214***	-0.205***
	(0.0403)	(0.0392)	(0.0408)	(0.0395)
Pavitt 3	-0.0565	-0.0557	-0.0549	-0.0551
	(0.0531)	(0.0518)	(0.0539)	(0.0523)
Pavitt 5	-0.116**	-0.114**	-0.118**	-0.114**
	(0.0558)	(0.0540)	(0.0561)	(0.0539)
Average effective real	0.0266***	0.0259***	0.0265***	0.0258***
exchange rate	(0.0020)	(0.0020)	(0.0021)	(0.0020)
Crisis dummy	0.140***	0.134***	0.138***	0.133***
	(0.0330)	(0.0318)	(0.0333)	(0.0321)
GDP growth (%)	-0.0589***	-0.0580***	-0.059***	-0.058***
	(0.0087)	(0.0085)	(0.0088)	(0.0086)
Constant	-2.477***	-2.541***	-2.480***	-2.543***
	(0.260)	(0.253)	(0.262)	(0.254)
lnsig2u			-2.648**	-3.569
_cons			(1.184)	(2.595)
N	158316	158316	158316	158316
AIC	53709.5	53715.4	53710.9	53717.2
BIC	53948.8	53954.7	53960.2	53966.5
Log-likelihood	-26830.7	-26833.7	-26830.4	-26833.6
chi2	2883.2	3001.7	2298.6	2384.7
3-digit industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Notes: Notes: Top R&D intensity is less than 1. Estimators: (1) – pooled Logit; (2) - pooled Clog-log; (3) - Logit random effects; (4) – Clog-log random effects. For other notes, see Table 2 in the main text.

Table A3b. Intra-industry firm size diversity and exit hazard: consistency by estimator (*Coefficient of variation*)

	(1)	(2)	(3)	(4)
Employment CV	0.0224***	0.0217***	0.0227***	0.0218***
	(0.0066)	(0.0064)	(0.0067)	(0.0064)
CV*Log(R&Dint.+1)	-0.124***	-0.119***	-0.124***	-0.119***
	(0.0299)	(0.0286)	(0.0299)	(0.0284)
Log(R&Dint.+1)	-1.633***	-1.556***	-1.637***	-1.556***
	(0.389)	(0.374)	(0.391)	(0.374)
Log(R&Dint.+1) sq.	2.674***	2.534***	2.674***	2.532***
	(0.569)	(0.544)	(0.572)	(0.544)
Log(firm Age)	-0.464***	-0.445***	-0.459***	-0.443***
	(0.0231)	(0.0223)	(0.0238)	(0.0230)
Log(Employment)	-0.427***	-0.413***	-0.434***	-0.416***
	(0.0263)	(0.0256)	(0.0282)	(0.0269)
Log(Employm.) sq.	0.0410***	0.0397***	0.0416***	0.0399***
	(0.0037)	(0.0036)	(0.0039)	(0.0038)
Log(Real turnover /	-0.220***	-0.212***	-0.221***	-0.213***
employees)	(0.0173)	(0.0166)	(0.0173)	(0.0166)
Firm growth relative	-0.136***	-0.126***	-0.136***	-0.126***
to industry median growth	(0.0213)	(0.0199)	(0.0214)	(0.0200)
Log(Plants)	-0.0625*	-0.0632*	-0.0603*	-0.0623*
	(0.0354)	(0.0345)	(0.0360)	(0.0348)
Civil R&D only	-0.202***	-0.195***	-0.202***	-0.195***
	(0.0285)	(0.0277)	(0.0287)	(0.0277)
UK-owned	-0.157***	-0.154***	-0.158***	-0.154***
	(0.0484)	(0.0474)	(0.0489)	(0.0477)
Log(% entry rate)	-0.131	-0.127	-0.131	-0.127
	(0.130)	(0.126)	(0.131)	(0.126)
Log(Mean industry	0.0435**	0.0421**	0.0430**	0.0419***
employment)	(0.0170)	(0.0164)	(0.0169)	(0.0162)
Log(Median R&D int.	1.405***	1.357***	1.406***	1.357***
in industry)	(0.232)	(0.221)	(0.236)	(0.223)
Herfindahl index	-0.0715	-0.0664	-0.0770	-0.0685
	(0.134)	(0.130)	(0.135)	(0.131)
Pavitt 1	-0.148***	-0.145***	-0.145***	-0.144***
	(0.0487)	(0.0474)	(0.0496)	(0.0478)
Pavitt 2	-0.216***	-0.208***	-0.217***	-0.208***
	(0.0404)	(0.0393)	(0.0410)	(0.0396)
Pavitt 3	-0.0591	-0.0583	-0.0574	-0.0577
	(0.0530)	(0.0517)	(0.0538)	(0.0522)
Pavitt 5	-0.110**	-0.107**	-0.111**	-0.108**
	(0.0559)	(0.0541)	(0.0562)	(0.0539)
Average effective real	0.0265***	0.0258***	0.0264***	0.0258***
exchange rate	(0.0020)	(0.0020)	(0.0021)	(0.0020)
Crisis dummy	0.140***	0.134***	0.138***	0.133***
	(0.0330)	(0.0318)	(0.0333)	(0.0321)
GDP growth (%)	-0.0592***	-0.0583***	-0.0595***	-0.0584***
	(0.0088)	(0.0086)	(0.0088)	(0.0086)
Constant	-2.463***	-2.527***	-2.466***	-2.529***
	(0.259)	(0.252)	(0.262)	(0.254)
lnsig2u			-2.511**	-3.540
_cons			(1.043)	(2.527)
N	158313	158313	158313	158313
AIC	53706.2	53712.0	53707.6	53713.8
BIC	53945.5	53951.4	53956.9	53963.1
Log-likelihood	-26829.1	-26832.0	-26828.8	-26831.9
chi2	2881.9	3000.1	2279.4	2369.9
3-digit industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Notes: Notes: Top R&D intensity is less than 1. Estimators: (1) – pooled Logit; (2) - pooled Clog-log; (3) - Logit random effects; (4) – Clog-log random effects. For other notes, see Table 2 in the main text.

Table A4: Intra-industry firm size diversity and exit hazard: consistency with top R&D intensity at the 98th percentile

	(1)	(2)
Diversity	0.0316***	0.0091***
	(0.0110)	(0.0031)
Diversity*Log(R&Dint.+1)	-0.0843**	-0.0279**
	(0.0378)	(0.0114)
Log(R&Dint.+1)	-0.538***	-0.617***
	(0.127)	(0.143)
Log(R&Dint.+1) sq.	0.516***	0.616***
	(0.0870)	(0.155)
Log(firm Age)	-0.198***	-0.206***
	(0.0111)	(0.0112)
Log(Employment)	-0.226***	-0.221***
	(0.0134)	(0.0135)
Log(Employm.) sq.	0.0220***	0.0213***
	(0.0019)	(0.0019)
Log(Real turnover /	-0.104***	-0.104***
employees)	(0.0076)	(0.0078)
Firm growth relative	-0.0538***	-0.0588***
to industry median growth	(0.0089)	(0.00922)
Log(Plants)	-0.0185	-0.0184
	(0.0160)	(0.0161)
Civil R&D only	-0.0920***	-0.0937***
	(0.0135)	(0.0137)
UK-owned	-0.0721***	-0.0723***
	(0.0219)	(0.0220)
Log(% entry rate)	-0.0877	-0.0727
	(0.0620)	(0.0629)
Log(Mean industry	0.0120	0.0162*
employment)	(0.0084)	(0.00827)
Log(Median R&D int.	0.425***	0.478***
in industry)	(0.112)	(0.114)
Herfindahl index	-0.0446	-0.0365
	(0.0640)	(0.0635)
Pavitt 1	-0.0544**	-0.0596***
	(0.0224)	(0.0231)
Pavitt 2	-0.102***	-0.103***
	(0.0192)	(0.0193)
Pavitt 3	-0.0183	-0.0242
	(0.0249)	(0.0250)
Pavitt 5	-0.0507*	-0.0503*
	(0.0270)	(0.0272)
Average effective real	0.0116***	0.0117***
exchange rate	(0.0009)	(0.0009)
Crisis dummy	0.0583***	0.0590***
	(0.0156)	(0.0158)
GDP growth (%)	-0.0263***	-0.0261***
	(0.0036)	(0.0036)
Constant	-1.351***	-1.359***
	(0.117)	(0.118)
lnsig2u	-2.355***	-2.352***
_cons	(0.0447)	(0.0459)
N	162,869	162,869
AIC	56148.1	55114.9
BIC	56398.2	55364.6
Log-likelihood	-28049.1	-27532.4
chi2	2538.8	2482.3
3-dist industry dummies	Yes	Yes
Year dummies	Yes	Yes

Notes: Estimations from panel Probit with random effects. (1): Estimation with Theil entropy index; (2) Estimation with coefficient of variation. Fully consistent with results from other hazard estimators, which are not reported here but can be provided on request. Also fully consistent with estimations based on 75th and 90th percentile values for R&D intensity, which are not reported but available on request. For other notes, see Table A2 in the main text. The number of firms is 35,569.

Table A5a: Intra-industry firm size diversity and exit hazard: Step-wise result consistency (*Theil entropy index*)

	(1)	(2)	(3)	(4)	(5)
Theil	0.0415***	0.0418***	0.0312***	0.0421***	0.0398***
	(0.0101)	(0.0104)	(0.0111)	(0.0116)	(0.0116)
Theil*log(R&Dint.+1)	-0.177***	-0.115*	-0.196***	-0.224***	-0.247***
	(0.0569)	(0.0599)	(0.0615)	(0.0618)	(0.0617)
Log(R&Dint.+1)	0.553***	-0.421**	-0.715***	-0.632***	-0.599***
	(0.173)	(0.191)	(0.199)	(0.205)	(0.205)
Log(R&Dint.+1) sq.	-0.175	0.696***	1.232***	1.205***	1.266***
	(0.240)	(0.257)	(0.276)	(0.281)	(0.282)
Log(Firm age)	-0.226***	-0.237***	-0.235***	-0.235***	-0.214***
	(0.0095)	(0.0112)	(0.0113)	(0.0113)	(0.0114)
Log(Employment)	-0.211***	-0.208***	-0.212***	-0.216***	-0.217***
	(0.0131)	(0.0135)	(0.0137)	(0.0137)	(0.0137)
Log(Empl.) sq.	0.0203***	0.0210***	0.0202***	0.0209***	0.0209***
	(0.0018)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Log(% entry rate)	-0.137**	-0.116*	-0.131**	-0.145**	-0.0743
	(0.0607)	(0.0628)	(0.0634)	(0.0638)	(0.0629)
Log(Real turnover / employees)		-0.0956***	-0.110***	-0.112***	-0.104***
		(0.0076)	(0.0079)	(0.0079)	(0.0079)
Firm growth relative to industry median		-0.0656***	-0.0661***	-0.0652***	-0.0681***
		(0.0095)	(0.0096)	(0.0096)	(0.0096)
Log(Plants)		-0.0054	-0.0024	-0.0023	-0.0202
		(0.0160)	(0.0161)	(0.0161)	(0.0162)
Civil R&D only		-0.0687***	-0.0587***	-0.0730***	-0.0951***
		(0.0131)	(0.0133)	(0.0136)	(0.0138)
Herfindahl			-0.0990	-0.125*	-0.0540
			(0.0631)	(0.0650)	(0.0651)
UK owned			-0.0784***	-0.0768***	-0.0723***
			(0.0221)	(0.0221)	(0.0222)
Log(Mean industry employment)			0.0177**	0.0116	0.0152*
			(0.0083)	(0.0085)	(0.0085)
Log(Median R&D int. in industry)			0.586***	0.455***	0.647***
			(0.108)	(0.119)	(0.120)
Pavitt 1				-0.0209	-0.0578**
				(0.0230)	(0.0232)
Pavitt 2				-0.102***	-0.103***
				(0.0195)	(0.0195)
Pavitt 3				-0.0059	-0.0218
				(0.0251)	(0.0252)
Pavitt 5				-0.0551**	-0.0539**
				(0.0274)	(0.0274)
Average effective real exchange rate					0.0119***
					(0.0009)
Crisis dummy					0.0620***
					(0.0159)
GDP growth (%)					-0.0266***
					(0.0036)
Constant	-0.833***	-0.340***	-0.255***	-0.180**	-1.379***
	(0.0541)	(0.0669)	(0.0777)	(0.0792)	(0.120)
lnsig2u	-2.291***	-2.302***	-2.303***	-2.305***	-2.351***
_cons	(0.0782)	(0.0709)	(0.0698)	(0.0684)	(0.0460)
N	158316	158316	158316	158316	158316
AIC	53021.1	53987.7	53951.9	53925.8	53704.9
BIC	54121.2	54127.3	54131.4	54145.2	53954.2
Log-likelihood	-29000.6	-26979.9	-26957.9	-26940.9	-26827.4
chi2	2310.9	2202.0	2240.6	2273.7	2466.3
3-dist industry dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes

Notes: panel Probit with random effects. Fully consistent with results from other estimators, which are not reported here but available on request. Number of firms is 35,136. See also notes for Table 2 in the main text.

Table A5b: Intra-industry firm size diversity and exit hazard: Step-wise result consistency (*Coefficient of variation*)

	(1)	(2)	(3)	(4)	(5)
Employment CV	0.0119***	0.0132***	0.0099***	0.0124***	0.0112***
	(0.0028)	(0.0029)	(0.0030)	(0.0032)	(0.0032)
CV*log(R&Dint.+1)	-0.0434***	-0.0308**	-0.0446***	-0.0537***	-0.0588***
	(0.0138)	(0.0145)	(0.0148)	(0.0149)	(0.0149)
Log(R&Dint.+1)	0.388***	-0.559***	-0.877***	-0.796***	-0.779***
	(0.149)	(0.166)	(0.180)	(0.185)	(0.185)
Log(R&Dint.+1) sq.	-0.101	0.802***	1.256***	1.228***	1.290***
	(0.245)	(0.263)	(0.279)	(0.284)	(0.283)
Log(Firm age)	-0.225***	-0.237***	-0.234***	-0.235***	-0.215***
	(0.0095)	(0.0112)	(0.0112)	(0.0113)	(0.0114)
Log(Employment)	-0.211***	-0.208***	-0.212***	-0.216***	-0.218***
	(0.0131)	(0.0135)	(0.0136)	(0.0137)	(0.0137)
Log(Empl.) sq.	0.0204***	0.0209***	0.0202***	0.0208***	0.0209***
	(0.0018)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Log(% entry rate)	-0.117*	-0.0872	-0.111*	-0.127**	-0.0643
	(0.0615)	(0.0637)	(0.0641)	(0.0644)	(0.0635)
Log(Real turnover / employees)		-0.0969***	-0.110***	-0.112***	-0.104***
		(0.0076)	(0.0079)	(0.0079)	(0.0079)
Firm growth relative to industry median		-0.0663***	-0.0665***	-0.0654***	-0.0681***
		(0.0095)	(0.0096)	(0.0096)	(0.0096)
Log(Plants)		-0.0030	-0.0012	-0.0009	-0.0186
		(0.0160)	(0.0161)	(0.0161)	(0.0162)
Civil R&D only		-0.0691***	-0.0594***	-0.0742***	-0.0957***
		(0.0131)	(0.0133)	(0.0136)	(0.0138)
Herfindahl			-0.0992	-0.120*	-0.0503
			(0.0624)	(0.0641)	(0.0642)
UK owned			-0.0777***	-0.0763***	-0.0717***
			(0.0221)	(0.0222)	(0.0222)
Log(Mean industry employment)			0.0203**	0.0155*	0.0187**
			(0.0082)	(0.0084)	(0.0084)
Log(Median R&D int. in industry)			0.554***	0.449***	0.642***
			(0.108)	(0.117)	(0.119)
Pavitt 1				-0.0262	-0.0608***
				(0.0233)	(0.0235)
Pavitt 2				-0.104***	-0.104***
				(0.0195)	(0.0195)
Pavitt 3				-0.0069	-0.0228
				(0.0251)	(0.0251)
Pavitt 5				-0.0517*	-0.0505*
				(0.0274)	(0.0274)
Average effective real exchange rate					0.0119***
					(0.0009)
Crisis dummy					0.0622***
					(0.0160)
GDP growth (%)					-0.0266***
					(0.0037)
Constant	-0.827***	-0.340***	-0.268***	-0.188**	-1.374***
	(0.0525)	(0.0650)	(0.0777)	(0.0794)	(0.120)
lnsig2u	-2.286***	-2.298***	-2.301***	-2.304***	-2.351***
_cons	(0.0825)	(0.0734)	(0.0709)	(0.0691)	(0.0462)
N	158,313	158,313	158,313	158,313	158,313
AIC	53999.4	53967.0	53935.0	53909.1	53701.1
BIC	54199.4	54106.6	54114.5	54128.5	53950.4
Log-likelihood	-28989.7	-26969.5	-26949.5	-26932.6	-26825.6
chi2	2303.2	2200.1	2236.4	2269.9	2464.3
3-dist industry dummies	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes

Notes: panel Probit with random effects. Fully consistent with results from other estimators, which are not reported here but available on request. Number of firms is 35,136. See also notes for Table 2 in the main text.

Table A6: Intra-industry firm size diversity and exit hazard: Firms born in 2000 or 2003 and thereafter

	Firms born in 2000+		Firms born in 2003+	
Diversity measure	Theil	CV	Theil	CV
Diversity	0.0599***	0.0185***	0.0721**	0.0218***
	(0.0222)	(0.0058)	(0.0304)	(0.0081)
Diversity*log(R&Dint.+1)	-0.232**	-0.0665***	-0.323**	-0.0816**
	(0.108)	(0.0256)	(0.143)	(0.0349)
Log(R&Dint.+1)	-1.003***	-1.126***	-0.635	-0.858*
	(0.377)	(0.336)	(0.506)	(0.451)
Log(R&Dint.+1) sq.	1.573***	1.659***	1.192*	1.250**
	(0.484)	(0.485)	(0.627)	(0.628)
Log(firm Age)	-0.205***	-0.205***	-0.240***	-0.239***
	(0.0273)	(0.0273)	(0.0376)	(0.0376)
Log(Employment)	-0.242***	-0.242***	-0.228***	-0.229***
	(0.0284)	(0.0283)	(0.0415)	(0.0415)
Log(Employm.) sq.	0.0263***	0.0262***	0.0247***	0.0246***
	(0.0047)	(0.0047)	(0.0074)	(0.0074)
Log(Real turnover /	-0.119***	-0.120***	-0.103***	-0.103***
employees)	(0.0153)	(0.0152)	(0.0207)	(0.0207)
Firm growth relative	-0.0524***	-0.0521***	-0.0638***	-0.0633***
to industry median	(0.0153)	(0.0152)	(0.0201)	(0.0201)
Log(Plants)	-0.0370	-0.0355	-0.0021	-0.00007
	(0.0342)	(0.0342)	(0.0488)	(0.0489)
Civil R&D only	-0.110***	-0.112***	-0.122***	-0.125***
	(0.0278)	(0.0278)	(0.0398)	(0.0398)
UK-owned	-0.0244	-0.0226	0.0244	0.0269
	(0.0566)	(0.0566)	(0.0813)	(0.0813)
Log(% entry rate)	-0.118	-0.0757	-0.320*	-0.251
	(0.120)	(0.123)	(0.175)	(0.183)
Log(Mean industry	0.0179	0.0239	0.0324	0.0387**
employment)	(0.0153)	(0.0146)	(0.0204)	(0.0195)
Log(Median R&D int.	0.865***	0.872***	1.116***	1.103***
in industry)	(0.209)	(0.205)	(0.270)	(0.265)
Herfindahl index	-0.0834	-0.0713	-0.294*	-0.283*
	(0.120)	(0.118)	(0.168)	(0.167)
Pavitt 1	-0.0571	-0.0708	-0.114*	-0.130*
	(0.0469)	(0.0480)	(0.0649)	(0.0666)
Pavitt 2	-0.0816**	-0.0871**	-0.0671	-0.0736
	(0.0383)	(0.0384)	(0.0513)	(0.0514)
Pavitt 3	-0.0157	-0.0212	-0.127	-0.131
	(0.0602)	(0.0599)	(0.0896)	(0.0893)
Pavitt 5	-0.0551	-0.0529	0.0020	0.0055
	(0.0498)	(0.0498)	(0.0676)	(0.0676)
Average effective real	0.0102***	0.0101***	0.0069***	0.0067***
exchange rate	(0.0017)	(0.0017)	(0.0025)	(0.0025)
Crisis dummy	0.115***	0.115***	0.146***	0.146***
	(0.0368)	(0.0369)	(0.0503)	(0.0503)
GDP growth (%)	-0.0373***	-0.0375***	-0.0403***	-0.0403***
	(0.0060)	(0.0060)	(0.0075)	(0.0075)
Constant	-1.232***	-1.249***	-0.999***	-1.017***
	(0.237)	(0.237)	(0.349)	(0.349)
lnsig2u	-2.337***	-2.337***	-2.330***	-2.330***
_cons	(0.0869)	(0.0868)	(0.112)	(0.111)
N	35,226	35,226	19,193	19,193
Number of firms	11,622	11,622	7,612	7,612
AIC	15553.8	15550.8	8921.5	8920.2
BIC	15765.5	15762.6	9118.1	9116.8
Log-likelihood	-7751.9	-7750.4	-4435.7	-4435.1
chi2	597.1	600.6	334.5	336.5
3-digit industry dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes

Notes: panel Probit with random effects. Fully consistent with results from other hazard estimators, which are not reported here but can be provided on request. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.