Small Business Economics (2006) 26: 103–116
DOI 10.1007/s11187-004-7813-9

Why do Most Firms Die Young?

Robert Cressy

ABSTRACT. A model is developed to explain why most firms die in the first few years of trading. A risk averse entrepreneur with initial capital endowment faces a Brownian motion in net worth over time. To balance return (profits growth) and risk (variance of profits) she adopts a portfolio strategy, choosing market positioning to achieve an optimal combination of risk and return at each instant, given her financial and human capital endowments and attitude towards risk. Failure occurs when the firm's value falls below the opportunity cost of staying in business. The resulting distribution of failure is Inverse Gaussian, implying, for specific parameter values, a positively skewed failure curve of the type observed in practice. In addition the model presents a novel-measure of management human capital (MHC) which implies that high MHC entrepreneurs will have higher absolute and marginal profits growth than low MHC entrepreneurs at given levels of risk.

KEY WORDS: entrepreneur, failure curve, human capital, risk-aversion, inverse Gaussian, port folio

1. Introduction

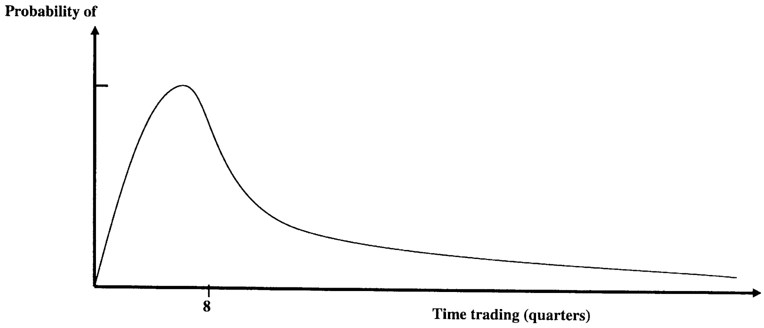
It is a well-established empirical fact that (at least in Manufacturing industry) young firms tend to be more failure prone than older ones (see e.g., Hall, 1987, Evans, 1987a, b, Dunne et al., 1989). Less well-studied is how the failure rate of firms varies with time trading within the first year or so of life. Whilst in the small business economics literature several recent studies suggest that around half of a randomly selected cohort of startups die in the first two and a half years of trading, a less known but equally remarkable fact is that the firm failure distribution over time trading is positively skewed with a mean that appears relatively constant in calendar time (Ganguly, 1985; Bruderl et al, 1992; Cressy, 1996);[1] see Figure 1. To the layperson in this area the explanation for the shape of the curve is not immediately obvious: Why should the failure rate of firms should first *rise steeply* to a peak at 18–24 months of age and then *fall gradually* to converge on a *low long run value*?[2]

Recent studies have examined firm-level determinants of failure. For example, early innovation might enhance a startup's survival chances in an industry dominated by large firms, or differences in human and financial capital at start-up (or prior to closure) and the availability of outside opportunities to entrepreneurship may be instrumental in the survival of small business (Evans and Jovanovic, 1989; Bates, 1989; Holtz-Eakin et al., 1994a, b; Mata, 1992; Cressy, 1996). For example, human capital in the form of team starts run by more experienced owners with vocational qualifications and less outside options seem to survive longer. Finally, theoretical literature suggests that entrepreneurs may be more risk-tolerant than wage workers and so choose the riskier income stream associated with entrepreneurship (Kihlstrom and Laffont, 1979). It is a short argument from there to conclude that small firms' owners may differ in their attitudes to risk, and that this itself may impact their failure rate, particularly at the startup stage.[3] Consistently with the failure curve described above, some studies have found the determinants of survival to vary with the time the firm has been trading for firm- or industry-level reasons (Audretsch, 1991; Bruderl et al., 1992; Argarwal and Audretsch, 2001). However, none of these studies provide a formal optimising model generating the observed failure curve.

By contrast the work of Jovanovic (1982) and Ericson and Pakes (1995) generates dynamic

Final version accepted on December 17, 2004

CASS Business School
London
UK
E-mail: r.c.cressy@city.ac.uk



Source: NatWest Startun Tracking Exercise

Figure 1. Business closure rates by time trading.

optimising models of the firm and in principle failure rates that vary along its lifecyle. Jovanovic assumes that human capital (efficiency) is unknown but can be learnt about by entering business. Ericson and Pakes allow for the role of learning by doing in addition to learning pure and simple. What none of these models do is to take into account the role of human capital evolution in conjunction with risk and risk aversion, letting the entrepreneur to adjust his portfolio over time. Little attention has been paid in particular to *attitudes* to and the *management* of uncertainty.[4] This is contrary to much observed reality, and, whilst empirical work in other disciplines than economics allows for the role of uncertainty, a.perusal of that literature reveals no formal optimising model of the firm as the basis for proper testing (see e.g., Gimeno et al., 1997 and references therein).

The finance literature has however long recognised that by portfolio rebalancing an individual can *optimise* exposure to risk. This fact conforms well with results from recent empirical studies of successful entrepreneurs who typically aim to *reduce* the amount of market uncertainty they face in engaging in business (e.g., Knight, 1965; Prais, 1983; Delmar, 1995). Risk reduction for the firm is achieved by diversification across products and services produced, across clients supplied and across input producers.[5] This practice clearly pays dividends by reducing fluctuation in the firm's revenue so that the chances of it falling below the entrepreneur's exit threshold (minimum equity) are thereby lowered. But 'there is no such thing as a free lunch' and we would expect for given levels of entrepreneurial ability that this lower risk would be purchased at the expense of a lower return. Likewise, higher risk is associated with higher return: for example, if a niche marketing strategy allows the business to grow faster than the market average, we should expect this to be associated with higher variability of profits. A specialist product or bespoke service will more often than not be one for which the demand is more variable as tastes switch capriciously to other product niches or a larger firm uses economies of scale to force down price.

Thus the external environment must be managed effectively for the business to optimise its response to risk. The ability to do this will be determined in part by factors internal to the firm, including the entrepreneur's own managerial ability and her attitudes to risk. In economic terms, entrepreneurial ability can be understood in terms of the effect of inherent traits or attributes on the cost, measured by the extra risk involved, of achieving a given increment to the firm's growth rate. We identify management human capital (MHC) as the ability to achieve growth at lower cost in terms of firm-specific risk.[6] Higher MHC individuals should have lower costs of growth in this sense. Attitudes towards risk can also be expected to play a role in determining the entrepreneur's selection of projects (see Kihlstrom and Laffont, 1979) with less cautious entrepreneurs likely to risk exposure to projects with more variable returns. This choice in turn has implications for failure rates. Firms run by headstrong entrepreneurs whose growth aspirations outstrip their capacity to manage them–are likely to fail early as they 'grow too fast too soon'. In general, risk aversion should be associated with lower failure rates but also with slower growth.

The present paper provides a theoretical model to explain how in an uncertain environment human and financial capital together with entrepreneurial risk aversion and human capital interact with endogenous market positioning to affect the risk-return distribution faced by an entrepreneur. It pulls together three strands of the literature, namely, portfolio theory as a strategy for balancing risk and return; a variant on Gibrat's law as a description of the growth process; and human capital and market positioning as a determinant of entrepreneurial productivity which is allowed to vary (exogenously) with the lifecycle of the entrepreneurial team. The outcome is a model consistent with the observed firm failure distribution, with the failure rate first rising steeply and then tailing off to a low long run value, predicting that indeed 'most firms die young'.

The optimal growth-risk model resulting from this process determines the systematic part of the firm's growth. However, it allows that part of a firm's growth is determined outside the entrepreneur's control by random proportional shocks to the firm's startup equity value. Finally, we let the entrepreneur make an optimal decision as to when, if at all, to quit business. This will involve the comparison of the likely returns to continuing as compared with the value of an outside alternative (Jovanovic, 1982; Evans and Jovanovic, 1989). The entrepreneur's value function will accordingly contain a stopping rule allowing the 'switching' decision to be made optimally. Recent developments in the theory of investment under uncertainty (Dixit and Pindyck, 1994) drawing on the early work of Merton (1969)[7] provide the tools by which this phenomenon can be treated analytically. In this framework market research is a control variable of the entrepreneur.[8]

2. The model

More formally, suppose, then, that an entrepreneur's profits are affected in two ways by a choice (market positioning) parameter θ a higher value of which raises the firm's growth rate and simultaneously its variance. A risk-averse entrepreneur will therefore find that there is a utility cost to higher growth in the form of higher risk and will therefore wish to balance the two. In the following model we allow both effects to operate.

Profits, $x(t)$, in any period, starting from some initial value x(0), are assumed to be subject to random proportional shocks, i.e. we assume Gibrat's law holds for the firm *conditional on entrepreneurial human capital*.[9] In continuous time this implies that the profit function x follows a Geometric Brownian motion (GBM):

$$dx = \mu_x x\,dt + \sigma_x x\,dz \tag{1}$$

where μ_x and σ_x are known time-invariant functions that represent the expected growth of profits and the variance of profits over a small time interval.[10] These functions are assumed to take the form:

$$\mu_x = H\mu\theta^{\beta}, \quad \sigma_x = \sigma\theta^{\alpha}, \quad H, \sigma, \alpha, \beta, \theta > 0 \tag{2}$$

Note that $\theta = 1 = \mathrm{B}$ implies that firm and market parameters are identical so the firm is 'centrally positioned' in the market measured by risk and return i.e. $\mu_x = \mu$ and $\sigma_x = \sigma$.

We shall assume in this model that the *systematic* component of the market growth rate is non-negative, which implies that the *systematic* component of a firm's growth, μ_x, is also non-negative. This does not of course imply that the *realised* growth rate of both market and firm are negative.[11]

From equation (2) a higher θ is associated with a higher expected return *and* a higher variance of return. With an additional assumption we can show that this will generate the familiar positively sloped mean-variance Efficient Frontier of portfolio analysis. Solving (2) for μ_x as a function of σ_x we get:

$$\mu_x = \mu H\left(\frac{\sigma_x}{\sigma}\right)^k, \quad k = \beta/\alpha > 0 \tag{3}$$

If θ is chosen optimally (i.e. to maximise expected utility) we can identify this expression with the Efficient Frontier which shows how firm-specific risk is associated with growth when the entrepreneur chooses different market positioning strategies θ. Before we do this we state

Definition 1: *The Efficient Frontier* (or *EF*) *is the curve of equation* (*3*) *for different* θ *plotted in* (μ_x, σ_x) *space.*

Now, using Assumption 1 above and making

Assumption 2: $\beta < \alpha$ *or equivalently* $k < 1$ we will find that the EF is upward-sloping and concave in positioning space:

$$\frac{d\mu_x}{d\sigma_x} = kK\sigma_x^{k-1} > 0, \quad K = \mu H/\sigma^k \tag{4}$$

and

$$\frac{d^2\mu_x}{d\sigma_x^2} = k(k-1)K\sigma_x^{k-2} < 0 \quad \text{for } k < 1 \tag{5}$$

See Figure 2.

In Figure 2 we have two EFs parameterised by different levels of human capital, H. Each is upward sloping reflecting the essential cost of growth as additional risk. The entrepreneur's expected utility indifference curves are also plotted showing the optimal combination of risk and return given her preferences, human capital endowment etc. Tangency points as usual pick out the optima as H varies. We emphasise here that μ is to be interpreted as *market* return and σ *market* risk. The entrepreneur's choice of theta determines both his firm-specific risk and firm-specific return (μ_x and σ_x)[12] and thus 'positions' her in the market. Whilst μ and σ are parameters from the point of view of the entrepreneur, the actual level of return and risk she faces (μ_x and σ_x) will vary across firms via θ and its determining parameters. The optimal value of θ, θ^*, in Figure 2 maximises the entrepreneur's discounted expected utility *over time,* $V\{x(t) \mid \mu_x(\theta), \sigma_x(\theta)\}$ – see below.

2.1. *Role of management human capital*

Management capital plays a natural role here by raising the entrepreneur's marginal return for *given* level of risk or, equivalently, by reducing the marginal cost of growth. Thus any changes in a parameter that increase the height and slope of the efficient frontier can be identified as economically efficient. Management human capital (i.e. managerial ability, particularly in the management of risk), henceforth MHC, is an obvious candidate for this role.

Definition 2: *Management Human capital is any parameter more of which reduces the marginal cost of growth measured along the EF, i.e. reduces the value of* $\frac{d\sigma_x}{d\mu_x}$

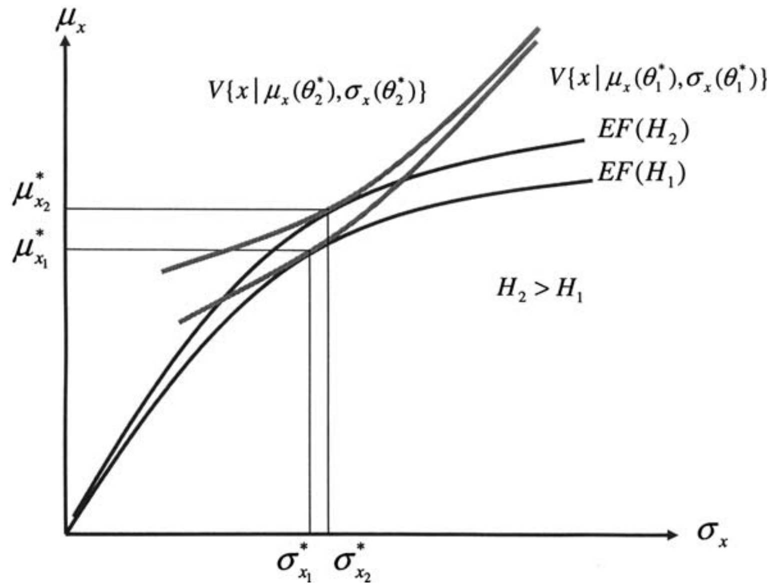


Figure 2. Efficient frontier for firm-specific risk and return and the entrepreneur's optimum market positioning.

Remark: *An example of such a parameter is H in equation* (3) *above.*

Proposition 1: *Management human capital (MHC) measured by the parameter H of equation 3 increases the height and slope of the Efficient Frontier in* (μ_x, σ_x)*-space.*

Proof. Differentiating the level of the EF of equation 3 and its slope of equation 4 respectively w.r.t. H we get

$$\frac{d\mu_x}{dH} = \frac{K}{H}\sigma_x^k = \mu\left(\frac{\sigma_x}{\sigma}\right)^k > 0 \tag{6}$$

$$\frac{\partial^2\mu_x}{\partial H\partial\sigma_x} = k\frac{K}{H}\sigma_x^{k-1} = k\frac{\mu}{\sigma}\left(\frac{\sigma_x}{\sigma}\right)^{k-1} > 0 \tag{7}$$

This proposition is illustrated in Figure 1 above. As H increases EF pivots upward from EF(H_1) to EF(H_2). Thus the slope of EF has increased at σ_{x_1} implying more growth for a given level of risk. The new equilibrium (tangency point) occurs to the left of the old implying a higher equilibrium growth rate and a higher level of equilibrium risk.

2.2. *The entrepreneur's optimisation strategy*

The entrepreneur's optimisation strategy has several components:

1. Choose the market positioning parameter $\theta = \theta * (\alpha, k, H, \mu, \sigma, \gamma)$ to maximise expected utility $V\{x(t) \mid \mu_x(\theta), \sigma_x(\theta)\}$ from following an optimal policy i.e. to optimally continue or close the business at any point in time t.
2. This results in a choice of the expected growth rate and risk of the firm relative to the market $\{\mu_x(\theta), \sigma_x(\theta)\}$ and places the entrepreneur at a position on the EF, given the parameter set $(\alpha, k, H, \mu, \sigma, \gamma)$. This position may change through time as e.g. H evolves.[13]
3. Note that the all-important management capital of the entrepreneur (H) plays a part in determining determine the slope and position of the EF at this initial stage. Attitudes to risk will determine the curvature of the utility function and the tangency optimum.
4. We can think of H (among other things) as also influencing initial resources of the entrepreneur (x_0). Initial resources combined with a decision on market positioning will determine the expected growth rate of the firm through time. Growth is measured alternatively by profits, equity value or utility as we shall see later.
5. The entrepreneur also chooses at startup an *ex post* exit *rule* which allows her to compare at each point in time the value of continuing in business, $V(x(t)|.)$, with the value of switching to an alternative activity worth W. This enables her to decide optimally when to 'fail' i.e. to close the business solvently.[14]
6. After these decisions have been made, chance factors then take over to alter the actual trajectory experienced, and the actual decision, if any, to exit the business at some point in time.t

2.3. *Entrepreneur's optimisation problem*

The entrepreneur is an expected utility maximiser and chooses a function $V(x(t)|\ \theta)$ to solve the following optimisation problem given his MHC:

$$V(x|\theta) = max\left\{u(x) + \frac{1}{1+r}E[V(x+dx|\theta), W\right\} \tag*{(8)[15]}$$

where r is the appropriate risk-adjusted discount rate, $u(x)$ is the utility of current (known) profits x, the expectation is taken over the *random* increment dx given by equation (1), and W is the present discounted value (p.d.v.) of the alternative activity open to the individual. The left hand part of the term in curly brackets in (8) is the value of continuing in business, given that market positioning is optimal; and the right hand part (W) is the value of ceasing to trade. Equation (8) says simply that if, once x is realised in any period, the value of the return to the alternative activity exceeds that to the current business, the entrepreneur closes her business, i.e. the business 'fails'.[16] Otherwise she continues with the current activity.

The entrepreneur's utility function $u(x)$ is assumed to satisfy the conditions for constant relative risk aversion (CRRA):

$$u(x) = x^{1-\gamma}/(1-\gamma) \Rightarrow \frac{-x}{u'}u'' = \gamma \in (0,1) \qquad (9)$$

We thus define failure as a situation where the entrepreneur's equity (or equivalently, discounted entrepreneurial utility) falls to the opportunity cost of staying in business.The relevant form of the discounted utility of an owner of an existing firm *and whose value V(x) exceeds W*, is obtained by a straightforward application of Ito's lemma (se Dixit and Pindyck, 1994) to (8) as:

$$\begin{aligned} V(x) &= \max_\theta\{(1/2)\sigma_x^2 x^2 V''(x) + \mu_x x V'(x) + u(x)\}/r \\ &= \max_\theta\{(1/2)\sigma^2\theta^{2\alpha}x^2 V''(x) + \mu\theta^\beta x V'(x) \\ &\quad + u(x)\}/r \qquad (10) \end{aligned}$$

using (2). This maximisation process yields the first order condition

$$\alpha v\theta^{2\alpha-1}xV'' + \beta H\mu\theta^{\beta-1}V' = 0, v = \sigma^2 \qquad (11)$$

an expression for optimal effort θ^*:

$$\theta^* = \left(\frac{-\beta H\mu V'(x)}{\alpha v V''(x)x}\right)^{\frac{1}{2\alpha-\beta}} \qquad (12)$$

The second order condition for a maximum requires that

$$\alpha(2\alpha-1)\theta^{2\alpha-2}vxV''(x) + H\beta(\beta-1)\theta^{\beta-2}\mu V' < 0, \quad v = \sigma^2 \qquad (13)$$

It is straightforward to show that, under the assumption that $V' > 0$, this reduces to the simpler condition

$$\beta < 2\alpha \quad \text{or} \quad \alpha > \beta/2 \qquad (14)$$

and so Assumptions 1 and 2 guarantee the second order condition (13) is fulfilled.[17] As the basis for our first proposition we now make

Assumption 3:[18] $r - \left(\mu_x^* - \gamma\frac{v_x^*}{2}\right)(1-\gamma) > 0$ Where $\mu_x^* = H\mu\theta^{*\beta}$, $v_x^* = \sigma^2\theta^{*\beta}$. Inserting θ^* back in to the optimal value function (10), we can demonstrate

Proposition 2: *The optimal value function V(x) satisfying the maximisation problem of* (10) *above is of the form* $V(x) = Cx^{1-\gamma}$, $C > 0$ *which implies V is concave increasing in profits. More specifically,* $C = [r - (\mu_x^* - \gamma\frac{v_x^*}{2})(1-\gamma)]^{-1}(1-\gamma)^{-1} > 0$.

Proof. We guess the solution for V to be $V(x) = Cx^{1-\gamma}$ from the general form of the first order condition. Differentiating this expression w.r.t. x we get

$$V'(x) = V(x)(1-\gamma)/x > 0 \qquad (15)$$

$$V''(x) = -V(x)\gamma(1-\gamma)/x^2 < 0 \qquad (16)$$

Substituting these back into equation (10) we get, after some manipulation,

$$V(x) = x^{1-\gamma}/(1-\gamma)\left[r - (1-\gamma)\left(\mu_x^* - \frac{\gamma}{2}v_x^*\right)\right] \qquad (17)$$

This implies that C is given by:

$$C = \left[r - \left(\mu_x^* - \gamma\frac{v_x^*}{2}\right)(1-\gamma)\right]^{-1}(1-\gamma)^{-1} \qquad (18)$$

From Assumption 3 we know that this expression is positive so proving that $V(x) = Cx^{1-\gamma}$ for $C > 0$, with C given by (18); which completes the proof. □

Corollary 2.1: *The reduced form of the optimal market positioning parameter is* $\theta^* = \left(\frac{kH\mu}{v\gamma}\right)^{\frac{1}{(2-k)\alpha}}$ *where H is the measure of management human capital stated in Definition 2 above.*

Proof. Substitute the form for V given in Proposition 2 into equation (12). □

Thus the optimal value function representing the entrepreneur's discounted utility value is increasing concave in current profits x. It will therefore satisfy the second order condition with appropriate assumptions about the parameters, as we shall see below.

The general form for θ^* as a function of model parameters is thus

$$\theta = \theta^*(\alpha, k, H, \mu, \sigma, \gamma)$$

The comparative statics of this critical effort level with respect to profit mean growth rate (μ) and variance (v) and risk aversion (γ) are presented in Table I.[19]

From Table I we note, firstly, that for zero growth $\mu = 0$ we have $\theta^* = 0$ and so effort, and the effect of any parameter on effort, is zero.[20] We observe also that a higher growth of the market increases the amount of effort in market positioning as the entrepreneur, for given market risk, will now get a higher marginal return from effort in creating growth. On the other hand, an increase in market risk (v) will decrease the amount of effort devoted to market positioning and the firm's growth rate, since the marginal cost of growth measured by the higher risk it incurs (σ_x), has now increased. Likewise, if the entrepreneur becomes more risk averse (γ increases) the entrepreneur puts *less* effort into market positioning and will therefore grow more slowly since the associated increase in business risk (σ_x) is now more costly when measured in utility terms. Finally, higher management human capital entrepreneurs will put more effort into market positioning since this increases growth opportunities at lower marginal cost.

2.4. *Firm's equity value*

To characterise the optimal switching rule and the failure rate functions to be generated later we need to know the nature of the process determining the firm's equity value through time. It turns out that under a GBM for profits the firm's equity value will like profits follow a GBM process. This is the content of

TABLE I
Effects of market growth, market risk and entrepreneurial risk aversion on optimal firm growth (μ_x^*) and risk (σ_x^*)

Parameter	μ_x^*	σ_x^*
μ	+	+
v	–	?
γ	–	–
H	+	+

Proposition 3: *The firm's equity value $S(t)$ follows a GBM dS/S with mean $\mu_s = \mu_x$ and variance $\sigma_s^2 = \sigma_x^2$.*

Proof. The proof for dS/S is a straightforward application of the properties of the Lognormal distribution. The expected level of the firm's profits at time t in the future given that profits x_0 are earnt at startup (now) is:

$$Ex(t) = x_0 e^{\mu_x t} = x_0 e^{\mu_x (t-0)} \tag{19}$$

(Aitcheson and Brown, 1957). This immediately generalises to the level of profits at time τ in the future starting from $x(t)$ now:

$$Ex(\tau) = x(t) e^{\mu_x(\tau - t)}, \quad \tau \geq t \tag{20}$$

Now the firm's equity value at any time t, $S(t)$, will be the expected present discounted value of the stream of profits $x(\tau)$ in the future:

$$S(t) = E \int_t^\infty x(\tau) e^{-r(\tau - t)} d\tau \tag{21}$$

where r is the risk-adjusted discount rate.[21] Using (17) and integrating we get

$$S(t) = \frac{x(t)}{r - \mu_x}, \quad r > \mu_x \tag{22}$$

From equation (1), and since S is proportional to x from (22), we must have

$$\frac{dS}{S} = \frac{dx}{x} = \mu_x \, dt + \sigma_x \, dz = \mu_s \, dt + \sigma_s \, dz \tag{23}$$

so that S follows a GBM with

$$\mu_s = \mu_x \quad \text{and} \quad \sigma_s = \sigma_x \tag{24}$$

□

2.5. *Entrepreneur's discounted utility*

We also need to know the nature of the process determining the behaviour of the firm's expected utility from entrepreneurship through time, $V(x)$. This will enable us to derive a very intuitive form for the function $V(x)$ in terms of the expected growth rate of utility through time. Once more, a simple GBM describes the process.

Proposition 4: *Entrepreneur's discounted utility $V(x)$ under CRRA utility follows a GBM dV/v with mean $\mu_v = (1-\gamma)\left(\mu_s - \frac{\gamma}{2}\sigma_s^2\right)$ and variance $\sigma_v^2 = (1-\gamma)^2\sigma_s^2$.*

Proof. See Appendix.[22] □

Remark. *This last result is intuitive: For given risk attitudes, a higher growth rate of utility is possible the higher the firm's equity growth and the lower the risk associated with it. However, concern about risk (γ) will reduce the growth of utility from any given market positioning strategy by giving greater negative weight to growth variability component in that strategy. Caution also acts indirectly on utility growth by preventing the entrepreneur growing her business as fast as her more reckless colleagues with the objective of avoiding extra risk associated with that growth.*

Proposition 4 also enables us to derive a different and intuitively appealing form for the entrepreneur's expected utility function under continuation. We have

Corollary 4.1: *The expected utility function for continuing in business is $V(x) = u(x)/(r-\mu_v)$ where $x(t)$ is the value of x at time t and μ_v is the growth rate of expected utility given in Proposition 4 above.*

Proof. Use Propositions 4 and 2 together with the definition of u(x) and the result follows. □

Remark 1. *It is a well known feature of the Random Walk that the best predictor of tomorrow's value is today's. Thus one might expect that in utility terms the best predictor of tomorrow's satisfaction is today's. This is precisely what Corollary 4.1 says when utility is expected to be constant over time ($\mu_v = 0$). For $u(x)$ is not only current utility at time t, it is also the expected one period utility of the individual, viewed from period t, for any period in the future. Hence if x is constant (or as here, is expected to be) then $u(x)/r$ is the p.d.v. of the stream of utility from x. Allowing for growth in utility through time at an expected proportional rate μ_v we get the adjusted discount rate r-μ_v of the formula.*[23]

2.5. *Stopping sets*

We have assumed that the stopping set is defined by V falling to W or below. It is well known, however, that Brownian motion is continuous. This implies that if utility crosses the line $V(t) = W$ it will do so continuously rather than in a discontinuous jump, and the entrepreneur will switch to the alternative activity the instant that $V(t) = W$. Also, since we have shown that there is a monotonic relationship between V, x and S we can define the following switch point equivalently in terms of x and S.

Let the entrepreneur's reservation value in terms of profits x be defined as x^* (W, γ, r, μ_v) where

$$x^* = [W(1-\gamma)(r-\mu_v)]^{\frac{1}{1-\gamma}} \text{ solves } V(x^*) = Cx^{*1-\gamma} = W \qquad (25)$$

and C is defined by equation (17) above. Also let the corresponding point in terms of equity be S^* where

$$S^* = \frac{x^*}{r-\mu_x} \qquad (26)$$

and x^* is given by equation (27).

2.6. *Stopping rule*

We can state the above stopping rule in words as:

CEASE TRADING (I.E. SWITCH TO THE ALTERNATIVE ACTIVITY) WHEN AND ONLY WHEN $V(x) = W$ (EQUIVALENTLY $x = x^*$ or $S = S^*$); OTHERWISE CONTINUE IN BUSINESS (27)

We can now prove a proposition about the form of the failure time density implied by the present model. It is a special case of a more general result on GBM processes.

2.7. *Distribution of failure times*

We wish to derive the distribution of firm failure times using the equation of evolution of the firm's equity which is GBM as shown in Proposition 3. We begin with a

Definition 3: *A firm is said to fail the first time the entrepreneur's equity value $S(t)$ equals s^* defined by equation (26).*

Remark. *In the language of the random walk literature, the instant T that $S(t)$ crosses the boundary S^* for the first time is known as the first passage time.*

We now present the results on the failure density without proof (available from the Author on request). They are straightforward following Cox and Miller (1965).

The probability that the firm will fail by time t, $G(t|.)$, must now be of the form:

$$G(t|R_0, 0) = \Phi\left(\frac{R_0 - \mu_R t}{\sigma_R \sqrt{t}}\right) - \exp\left\{\frac{-2\mu_R R_0}{\sigma_R^2}\right\} \times \Phi\left(\frac{-R_0 - \mu_R t}{\sigma_R \sqrt{t}}\right) \tag{28}$$

where $\Phi(z)$ is the Standard Normal c.d.f (see Cox and Miller, 1965). Then the density of the time to failure, $g(t|.)$, is obtained by partially differentiating this expression with respect to t. We are now ready to state

Proposition 5: *The firm failure time density under GBM, $g(t\backslash.)$ will be Inverse Gaussian.*

Proof. Note that V follows a GBM so that $R = \log(S)$ will follow an Absolute Brownian Motion (ABM) with parameters $\mu_R = E[\log(S)]$, $\sigma_R^2 = \mathrm{var}[\log S]$. Differentiation of equation (28) then yields:

$$g(t|R_0) = \frac{R_0}{t} \frac{1}{\sqrt{2\pi\sigma_R^2 t}} \exp\left\{-\frac{1}{2}\left(\frac{R_0 + \mu_R t}{\sigma_R \sqrt{t}}\right)^2\right\}, \quad 0 < t < \infty \tag{29}$$

This expression will integrate to less than one if μ is positive since then eventual failure is not certain (see below). The density (29) is known as the Inverse Gaussian density (Folks and Chhikara, 1978) and shows the probability of failure in the interval $(t,\ t + dt)$ given initial equity $R_0 > 0$.[24] □

We can identify the conditions under which this distribution has a maximum and under which it is positively skewed. However, to see that the density of equation (36) is positively skewed under certain parameter values we illustrate with some simulations. (see Figure 3).

2.8. *Long run failure rate*

We define the long run failure rate of the firm as the limiting value of $g(t|.)$ as t tends to infinity.

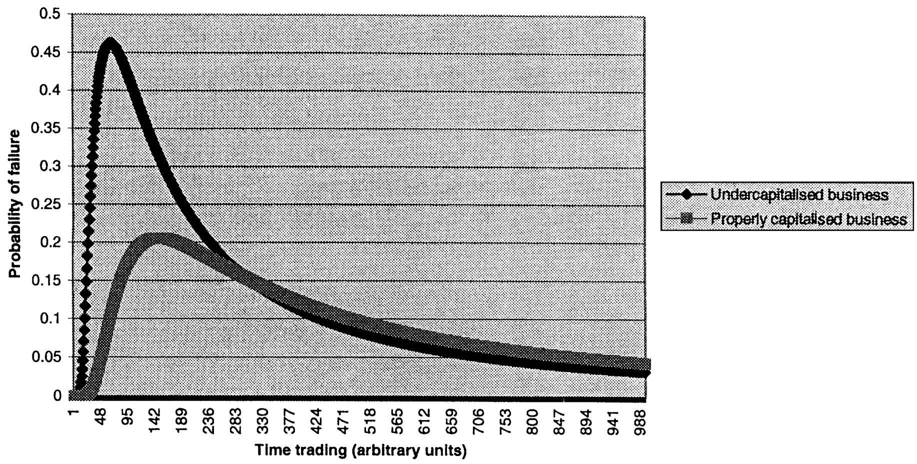


Figure 3a. Capitalisation and failure.

This limit is zero regardless of the value of the parameters. Thus starting with

Definition 5: *The long run failure rate (LFR) of firms is defined as* $g(\infty|R_0)$
we have

Proposition 6: *The long run failure rate (LFR) of all firms is zero.*

Proof. Take limits in equation (29). □

2.9. *Effects of initial capitalisation*

Figure 3 shows the effect of varying the initial equity level of the firm R_0 on the failure curve. Note that higher initial capitalisation is associated with (a) a delay in experiencing any risk at all (the period before the probability of failure becomes positive); (b) lower peak failure rates; (c) no differences in the 'long run' probability of failure. This suggests that only in the 'short term' is startup finance influential, and particularly in the period after the 'honeymoon' (zero risk) period is over. For low capitalisation firms the honeymoon period is very short, and the risk rises to very high levels relative to the long run. We shall prove some exact results to support these claims below.

2.10. *Differences in growth rates*

Figure.3b (not included) shows the variation of the failure curve with growth rates. Note that the higher growth is associated with a higher peak failure rate though the honeymoon period is unaffected. Again, in the long run the failure rates for all firms converge on zero, though this is not obvious from the figure.

2.11. *Effects of risk*

Figure 3c (not included) shows the effect of varying the risk of the firm. Such variation is consistent with human capital differences that *persist* through time and concern the management of the firm's returns. The graphs indicate that (a) peak (short run) failure rates rise monotonically with risk; (b) the peak failure rate for a higher-σ firm occurs *earlier* than that of a low-σ firm; (c) the long run rates of failure converge once more on zero.

2.12. *Long run failure rate (LFR) and initial conditions*

Proposition 5 showed that as t tends to infinity the chances of failure, g(t|.), tend to zero. Thus in the long run the initial conditions of the firm become irrelevant to its success, and, provided it survives a long enough period (particularly the first two years), it can be almost certain to continue further. This will depend critically on its risk-return combination discussed above.

2.13. *Lifetime failure probability (LFP)*

Given the Inverse Gaussian form of the failure density it is important to know the chances of failure of the firm *viewed from the outset* (rather than from a given time t from inception, considered in the previous paragraph). Supposing we can formulate this probability, we want to know how far the firm's chances are influenced by its parameters set at the startup stage.

Definition 6: *The lifetime failure probability (LFP),* Γ, *is the integral of the failure density* $g(t|.)$ *from* $t = 0$ *to* $t = \infty$:

$$\Gamma = \int_0^\infty g(t|.)\,dt \tag{30}$$

It can be shown that the LFP depends on the sign of μ_R. Stationary or growing firms are not certain to fail, i.e., for $\mu_R \geq 0$ we have $\Gamma < 1$ (Proposition 7 below).[25] The chances of failure over a complete lifetime will moreover depend significantly on the firm's initial conditions, growth rate and riskiness, and these in turn as we shall show, are related to measures of human capital. The results are stated formally in the next two propositions.

Proposition 7: *For growing or stationary firms,* $\mu_R \geq 0$, *the lifetime failure probability (LFP) is given by* $\Gamma(R_0, R^*, \mu_R, \sigma_R) = \exp\left\{\frac{-2(R_0 - R^*)\mu_R}{\sigma_R^2}\right\} \leq 1$ *This is increasing in firm specific variance* $({\sigma^2}_R)$

and the opportunity cost of entrepreneurship ($R^* = \log(S^*)$), *and decreasing in the initial capital of the firm* (R_0) *and the firm-specific growth rate* (μ_R). *Finally the LFP is ambiguous in managerial human capital, being positively related to initial capitalisation, risk, return and the opportunity cost of entrepreneurship.*

Proof. For proof of the form of $\Gamma(.)$ see Cox and Miller (1965) or Ingersoll (1987). For the comparative statics of the proposition differentiate Γ with respect to the relevant parameters, bearing in mind the dependence of the firm-specific growth rate and variance on the other parameters via θ and stated in Table II. □

Remark 1: *Table II shows that the LFR is decreasing in the firm's growth rate and increasing in its risk. This reflects that fact that a greater expected growth rate moves the firm away from the failure threshold whereas a greater variance of growth makes it more likely to cross it. The LFR is also decreasing in initial capitalisation, as the greater the initial capital buffer the less the likelihood that the firm will run out of money through chance depletions ('mistakes'). Finally, it is ambiguous in human capital of the entrepreneur since on the one hand the greater initial capitalisation and superior management of risk that this affords will make any given growth rate less likely to result in failure, but on the other, higher MHC will increase the opportunity cost of entrepreneurship, and increase the risk of the business.*

Remark 2: *The fact that start-up capital is positively related to survival does not necessarily imply that it is finance itself that ensures survival. Businesses run by more talented entrepreneurs will on average start with higher financial capital.*[26] *A policy of 'throwing money at the problem' will not then be guaranteed to work, since the randomly selected recipients of such finance may not know how to use it effectively.*

TABLE II
Parameter effect on the LFP (Γ)

Parameter	Effect on lifetime failure probability ($\mu_R > 0$)
R_0	–
R^*	+
μ_R	–
v_R	+
H	?

3. Summary and conclusions

The paper developed a theoretical model of firm growth under uncertainty to explain how managerial and financial capital together with entrepreneurial risk aversion influence the entrepreneur's decision on market positioning and consequent risk- return distribution faced. This in turn was shown to determine the lifetime failure probability of the business. Drawing on three strands of the existing literature, namely, the portfolio approach to balancing risk and return, the Random Walk as a description of growth under uncertainty and human capital as a determinant of entrepreneurial, skills in risk-management, we developed a model consistent with the observed firm failure distribution. This distribution implies that the chances of failure first rise steeply and then tail off gradually to converge on a small long run failure rate. As a result, 'most firms die young'.

One explanation of the initial rise in the failure rate was that initial financial resources were depleted through time as a result of trading losses and just bad luck. Another was the role of managerial human capital which enabled the more talented entrepreneur to grow faster at lower cost measured by the increase in her firm's equity risk. This in turn reduced the chances of equity falling below acceptable levels. However, initial resources in the form of start-up equity and managerial skills in the form of firm-specific risk exposure were themselves fundamentally a function of the *human* capital possessed by the entrepreneur, as in Cressy (1996). As in Jovanovic (1982) entrepreneurs in our model learn over time trading but this occurs in an exogenous and deterministic way rather than by updating of priors. At the forefront of our model is the choice of a risk-return combination for the business at each point on the entrepreneurial leaming curve. Low human capital types tend to start with too little human capital and to decide on the 'wrong' combination (too high risk relative to growth) and fail early. ill firms where the initial financial capital was simply a windfall, the theory predicts that the correlation of survival and finance to be weaker, as the entrepreneur's management capital would be unaltered by such windfall's, and thus the costs of growth, if high, would remain so. Therefore the firm's subsequent fortunes might wane over time.

The theory assumed that firms' equity value was subject to random proportional shocks from some initial level. Were all firms to grow at the same (given) rate, this would mean that the riskiness of the firm's value would' track the market and increase over time. However, niche market opportunities allowed the firm to expend research effort to alter the risk-return relationship experienced by the firm itself. Entrepreneurial risk aversion then interacted with this to produce an optimal tradeoff between the firm's actual growth and risk.

The model emphasised the importance of risk in firm failure. However, closure under risk is modelled as a *rational* decision by the entrepreneur and is not therefore cause for alarm. The closure decision of the entrepreneur was characterised (as in Jovanovic,1982) by a cutoff value that equated the firm's current equity to the entrepreneur's outside opportunity. We showed that the entrepreneur's value function will contain a stopping rule allowing the entrepreneur to time the 'switch' optimally. Closure is also socially optimal rather than the result of financial market imperfections.

However, the role of uncertainty in the model does suggest that the censorious attitudes to small business failure embodied in the UK's bankruptcy laws should be toned down: if chance plays such a potentially important role in the occurrence and timing of failure, and is by definition unanticipated and outside the control of the entrepreneur, a sympathetic understanding of unfortunate ('unlucky') individuals is more sensible response than social opprobrium.

Acknowledgements

I should like to thank Dennis Glycopantis, Pat McCloughan, Simon Parker and seminar participants at the Royal Economic Society conference, St. Andrews, Scotland, and the University of Edinburgh, together with the referees of this journal and the journal editor for helpful comments. I absolve all these individuals from misdemeanours for which I may be responsible.

Notes

[1] Bruderl, et al., (1992) found that for German startups, 24% went out of business in the first 2 years, 37% in the first 5 years of trading; Cressy (1996) found for UK startups that 45% died in the first 2.5 years of trading and 80% in the first 6 years; Mata (1992) found that for Portuguese startups 20% died in the first year and 50% in the first 4 years. The average size of the startups in these samples is very small, bring always less than five employees.

[2] Cressy (1993) for example, records that some 45% of businesses die in the first 2.5 years of trading, implying an annual failure rate of some 18%. This rate reduces dramatically to about 11% per annum for firms of 5 years of age.

[3] Startup firms are notoriously risk-prone measured along almost any dimension: in particular number of products, customers and suppliers are typically very small, making the firm highly vulnerable to changes in the environment.

[4] Kihlstrom and Laffont (1979) are an exception to this general rule. Their theory of entrepreneurship is based on uncertainty and differences in risk aversion amongst potential entrepreneurs.

[5] Jovanovic (1982) regards entrepreneurship itself as a learning experiment in which the entrepreneur continuously updates information on his ability during the course of the business 'experiment'. Her ability is fixed and, entrepreneurship is regarded as a process of discovery of its true value over time. We note that (a) Jovanovic's model does not allow the entrepreneur to influence the rate at which he learns (there is no learning by doing – see Ericson and Pakes, 1995), (b) there is no risk aversion informing the entrepreneur's decision on her growth rate, (c) learning has no *cost* to the entrepreneur, and (d) there is no initial dispersion of entrepreneurial ability – see Frank (1986). In the model presented below the entrepreneur by contrast does influence her growth rate (via market positioning) and this choice will partly depend on her attitudes towards risk. Her growth rate will also be influenced by her endowment of management human capital (which may vary deterministically over time) Entrepreneurs differ in their cost of achieving growth where this cost is logically measured in terms of the *disutility of the extra risk it involves*. This in turn has a *dynamic* effect, changing the *whole distribution* of income and failure risk through time as it influences the growth trajectory and firms' risk parameters. In line with the empirical facts (Cressy, 1996) we also allow for the dispersion of human capital at the point of entry into entrepreneurship.

[6] Firm-specific risk matters to an entrepreneur since (s)he typically has a highly undiversified portfolio of assets.

[7] See also Parker (1996) for an application of a Merton portfolio approach to the self-employment vs. employee decision.

[8] A little known fact about small business is that most businesses do not fail as a result of bankruptcy, rather they close voluntarily. Even in a deeply recessionary environment such as the UK experienced in the early 90s, the proportion of bankrupt/insolvent closures rose to only 20% of the total (Cressy, 1993).

[9] Empirical studies that have refuted Gibrat's law have found that firm growth and survival vary inversely and directly respectively with the age of the firm (i e. time trading). We control for human capital as a function of age when imposing Gibrat's law and thus the assumption is weaker than might otherwise be thought.

[10] We have from equation (1) $dx/x = \mu_x dt + \sigma_x \sqrt{dt}\varepsilon$ where $E(\varepsilon) = 0$, $V(\varepsilon) = 1$. Taking the expectation we get $E(dx/x) = \mu_x dt$ and so by rearrangement $\frac{E(dx/x)}{dt} = \mu_x$. Likewise, taking the variance, $V(dx/x) = \sigma_x^2(\sqrt{dt})^2 \;\; V(\varepsilon) = \sigma_x^2 dt$, and by rearrangement $V\frac{(dx/x)}{dt} = \sigma_x^2$.

[11] The assumption that the systematic component of firm growth is zero is consistent with a number of studies that have shown the typical small firm's growth rate to be close to zero. For example, Storey et al., (1987) and Watson (1990).

[12] Businessmen often argue that they eliminate as much risk as possible and leave therefore only a 'calculated' risk to their business. See Delmar (l995) for a description of such entrepreneurial psychology from Swedish case studies.

[13] Cressy (l996) shows that managerial human capital is a concave function of the average age of the team running the business with a maximum in the mid-50s.

[14] The empirical evidence shows that the great majority of small businesses do not fail through bankruptcy, contrary to common belief. This is because the majority do not borrow from their bank.

[15] Equation (6) is the Bellman equation of dynamic programming. See Dixit and Pindyck (1994, p. 105).

[16] It is important to recognise that most business closures do not involve bankruptcies or insolvencies, nor do they on average result in trade sales. See Cressy (1996).

[17] We assume in equation (12) thaI V′ > 0 and V″ < 0. This will be proved later.

[18] We shall see later that this is equivalent to assuming that the growth of expected utility is positive.

[19] These results rely on the derivative of θ^* with respect to any parameter z which takes the form:

$$d\theta^*/dz = X^{\gamma-1}[YX_z + X\log X.Y_z] \quad \text{where } Y, X^{Y-1} > 0.$$

[20] The exception is the parameter μ which yields indeterminate results for $\mu = 0$.

[21] We assume in this derivation that all profits are distributed to shareholders. This implies there is no carry-forward from previous periods in the form of reserves.

[22] Available from the author on request.

[23] It is clear from Corollary 4.1 that we need the expected growth in utility (μ_v) to be less than the rate of interest to ensure convergence.

[24] $g(t|.)$ is strictly not a density if μ is not restricted in sign. However, to save space in what follows we shall refer to it as the failure densify.

[25] See Cox and Miller (1965).

[26] Cressy (1996) in a UK empirical study shows that human and financial capital are positively correlated and that once human capital is controlled for, financial capital becomes irrelevant in the survival equation.

References

Aitcheson, J. and J. A. C. Brown, 1957, *The Lognormal Distribution*, Cambridge: Cambridge University Press.

Audretsch, D., 1991, 'New Firm Survival and the Technological Regime', *Review of Economics and Statistics* **68**(3), 520–526.

Audretsch, David B., 1994, 'Business Survival and the Decision to Exit', *Journal of the Economics of Businesss* **1**(1), 125–138.

Audretsch, David B. and Talat Mahmood, 1995, 'New Firm Survival: New Results using a Hazard Function' *Review of Economics and Statistics* **77**(1) (February), 97–103.

Agarwal, Rashree and David B. Audretsch, 2001, 'Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival', *Journal of Industrial Economics* **XLIX**(1), (March), 21–44.

Bates, Timothy, 1990, 'Entrepreneur Human Capital Inputs and Small Business Longevity', *Review of Economics and Statistics* **LXXII**(4), 551–559.

Bruderl, Joseph, Peter Preisendorfer and Rolf Ziegler, 1992, 'Survival Chances of Newly Founded Business Organisations, *American Sociological Review* **57**(2) (April), 227–242.

Cox, D. R. and H. D. Miller, 1965, *The Theory of Stochastic Processes*, London Chapman and Hall.

Cressy, Robert C., 1983, 'Goodwill, Intertemporal Price Dependence and the Repurchase Decision', *The Economic Journal* 93 (December), 847–861.

Cressy, Robert C., 1993, *The Startup Tracking Exercise: Third Year Report*, prepared for National Westminster Bank of Great Britain, November.

Cressy, Robert, 1996, 'Are Business Startups Debt-rationed?', *The Economic Journal* **106**(438) (September), 1253–1270

Cressy, Robert C., 1999, 'Small business failure: failure to fund or failure to learn by doing?', in Acs, Z. Carlsson and Charlie Karlsson (eds) *Entrepreneurship, SMEs and the Macro Economy,* 1977, Cambridge: *Cambridge University Press.*

Delmar, F., 1995, 'Risk management of the entrepreneur', in Bo Green (ed.), *Risk Behaviour and Risk Management*, Proceedings of the First International Stockholm seminar on Risk Behaviour and Risk Management, June 12–14.

Dixit, A vinash and Robert S. Pindyck, 1994, *Investment under uncertainty*, Princeton, NJ: Princeton University Press.

Ericson, Richard and Arie1 Pakes, 1995, 'Markov-Perfect Industry Dynamics: A Framework for Empirical Work', *Review of Economic Studies* **62**, 53–82.

Evans, David and Boyan Jovanovic, 1989, 'An Estimated Model of Entrepreneurial Choice Under Liquidity Constraint', *Journal of Political Economy* **97**(4), 808–827.

Folks, J. L. and R. S. Chhikara, 1978, 'The Inverse Gaussian Distribution And Its Statistical Application – A Review (with Discussion)', *Journal of the Royal Statistical Society B* **40**, 263–289.

Frank, Murray Z., 1986. 'An Intertemporal Model of Industrial Exit', *Quarterly Journal of Economics* **103** (May), 333–344.

Ganguly, P., 1985, *UK small business statistics and international comparisons*. Small Business Research Trust, Harper Row, London.

Gimeno, J., Timothy B. Folta, Arnold C. Cooper, and Carolyn Woo, 1997, 'Survival of the Fittest? Capital and the persistence of underperforming firms', *Administrative Science Quarterly* **42**, 750–783.

Hart, P. and S. Prais, 1956, 'The Analysis of Business Concentration', *Journal of the Royal Statistical Society Series A* **119**, 150–91

Holtz-Eakin, D., D. Joulfaian, and H. S. Rosen, 1994a, 'Sticking it Out: Entrepreneurial Survival and Liquidity Constraints', *Journal of Political Economy* **102**(11), 53–75.

Holtz-Eakin, D., D. Joulfaian, and H. S. Rosen, 1994b, 'Entrepreneurial Decisions and Liquidity Constraints', *Rand Journal of Economics* **25**(2), Summer, 34–347.

Ingersoll, Jonathan E., 1987, *Theory of Financial Decision Making*, New York: Rowman and Littlefield.

Jovanovic, Boyan, 1982, 'Selection and the Evolution of Industry', *Econometrica* **50**(3) (May), 649–670.

Kihlstrom, R. E. and Jean Jacques Laffont, 1979, 'A general Equilibrium Theory of Firm Formation Based on Risk Aversion', *Journal of Political Economy* **87**, 719–748.

Knight, F., 1965, Risk, *Uncertainty and Profit*, New York: Sentry Press.

Mata, Jose, 1994, 'Life duration of New Firms', *Journal of Industrial Economics* DLII(3) (September), 227–245.

Merton, Robert, 1969, 'Lifetime Portfolio Selection Under Uncertainty: The Continuous Time Case' *Review of Economics and Statistics* **51**, 247–257.

Ozga, S. A., 1960, 'Imperfect markets through lack of knowledge', *Quarterly Journal of Economics* **74**, 29–52.

Parker, Simon C., 1996, 'A Time Series Model Of Self-Employment Under Uncertainty', *Economica* **63**, 459–475.

Prais, S J, 1983, *The Evolution of Giant Firms in Britain*, Cambridge University Press.

Stigler, George J., 1961, 'The Economics of Information', *Journal of Political Economy* **69**, 213–225.

Storey, David, Kevin Keasey, Robert Watson and Pooran Wynarczyck, 1987, The *Performance of Small Firms*, Croom Helm, Beckenham, Kent.

Watson, Robert, 1990, 'Employment Change, Profits and Directors' Remuneration in Small and Closely-held UK Companies', *Scottish Journal of Political Economy* **37**(3) (August), 259–274.