The Financing of Research and Development

Bronwyn H. Hall*

January 2002

Abstract

Evidence on the “funding gap” for R&D is surveyed. The focus is on financial market reasons for underinvestment in R&D that persist even in the absence of externality-induced underinvestment. The conclusions are that 1) small and new innovative firms experience high costs of capital that are only partly mitigated by the presence of venture capital; 2) evidence for high costs of R&D capital for large firms is mixed, although these firms do prefer internal funds for financing these investments; 3) there are limits to venture capital as a solution to the funding gap, especially in countries where public equity markets are not highly developed; and 4) further study of governmental seed capital and subsidy programs using quasi-experimental methods is warranted.

Keywords: R&D, financing, liquidity constraints, venture capital

JEL codes: G32, O32, O38

The Financing of Research and Development

Bronwyn H. Hall

I. Introduction

It is a widely held view that research and development (R&D) activities are difficult to finance in a freely competitive market place. Support for this view in the form of economic-theoretic modeling is not difficult to find and probably begins with the classic articles of Nelson (1959) and Arrow (1962), although the idea itself was alluded to by Schumpeter (1942). The argument goes as follows: the primary output of R&D investment is the knowledge of how to make new goods and services, and this knowledge is nonrival: use by one firm does not preclude its use by another. To the extent that knowledge cannot be kept secret, the returns to the investment in it cannot be appropriated by the firm undertaking the investment, and therefore such firms will be reluctant to invest, leading to the underprovision of R&D investment in the economy.

Since the time when this argument was fully articulated by Arrow, it has of course been developed, tested, modified, and extended in many ways. For example, Levin et al (1987) and Mansfield et al (1981) found using survey evidence that imitating a new invention was not costless, but could cost as much as fifty to seventy-five per cent of the cost of the original invention. This fact will mitigate but not eliminate the underinvestment problem. Empirical support for the basic point concerning the positive externalities created by research that was made by Arrow is widespread, mostly in the form of studies that document a social return to R&D that is higher than the private level (Griliches, 1992; Hall, 1996). Recently, a large number of authors led by Romer (1986) have produced models of endogenous macro-economic growth that are built on the

1 See, for example, footnote 1, Chapter VIII of Capitalism, Socialism and Democracy.
increasing returns principle implied by Arrow’s argument that one person’s use of knowledge does not diminish its utility to another (Aghion and Howitt, 1997).

This line of reasoning is already widely used by policymakers to justify such interventions as the intellectual property system, government support of R&D, R&D tax incentives, and the encouragement of research partnerships of various kinds. In general, these incentive programs can be warranted even when the firm or individual undertaking the research is the same as the entity that finances it. However, Arrow’s influential paper also contains another argument, again one which was foreshadowed by Schumpeter and which has been addressed by subsequent researchers in economics and finance: the argument that an additional gap exists between the private rate of return and the cost of capital when the innovation investor and financier are different entities.

This paper concerns itself with this second aspect of the market failure for R&D investment: even if problems associated with incomplete appropriability of the returns to R&D are solved using intellectual property protection, subsidies, or tax incentives, it may still be difficult or costly to finance R&D using capital from sources external to the firm or entrepreneur. That is, there is often a wedge, sometimes large, between the rate of return required by an entrepreneur investing his own funds and that required by external investors. By this argument, unless an inventor is already wealthy, or firms already profitable, some innovations will fail to be provided purely because the cost of external capital is too high, even when they would pass the private returns hurdle if funds were available at a “normal” interest rate.

In the following, I begin by describing some of the unique features of R&D investment. Then I discuss the various theoretical arguments why external finance for R&D might be more expensive than internal finance, going on to review the empirical evidence on the validity of this hypothesis and the solutions that have been developed and adopted by the market and some governments. The paper concludes with a discussion of policy options.
II. Research and development as investment

From the perspective of investment theory, R&D has a number of characteristics that make it different from ordinary investment. First and most importantly, in practice fifty per cent or more of R&D spending is the wages and salaries of highly educated scientists and engineers. Their efforts create an intangible asset, the firm’s knowledge base, from which profits in future years will be generated. To the extent that this knowledge is “tacit” rather than codified, it is embedded in the human capital of the firm’s employees, and is therefore lost if they leave or are fired.

This fact has an important implication for the conduct of R&D investment. Because part of the resource base of the firm itself disappears when such workers leave or are fired, firms tend to smooth their R&D spending over time, in order to avoid having to lay off knowledge workers. This implies that R&D spending at the firm level typically behaves as though it has high adjustment costs (Hall, Griliches, and Hausman, 1986; Lach and Schankerman, 1988), with two consequences, one substantive and one that affects empirical work in this area. First, the equilibrium required rate of return to R&D may be quite high simply to cover the adjustment costs. Second, and related to the first, is that it will be difficult to measure the impact of changes in the costs of capital, because such effects can be weak in the short run due to the sluggish response of R&D to any changes in its cost.

A second important feature of R&D investment is the degree of uncertainty associated with its output. This uncertainty tends to be greatest at the beginning of a research program or project, which implies that an optimal R&D strategy has an options-like character and should not really be analyzed in a static framework. R&D projects with small probabilities of great success in the future may be worth continuing even if they do not pass an expected rate of return test. The uncertainty here can be extreme and not a simple matter of a well-specified distribution with a mean and variance. There is evidence, such as that in Scherer (1998), that the distribution of profits from
innovation sometimes has a Paretoian character where the variance does not exist. When this is the case, standard risk-adjustment methods will not work well.

The natural starting point for the analysis of R&D investment financing is the “neo-classical” marginal profit condition, suitably modified to take the special features of R&D into account. Following the formulation in Hall and Van Reenen (2000), I define the user cost of R&D investment $\rho$ as the pre-tax real rate of return on a marginal investment that is required to earn $r$ after (corporate) tax. The firm invests to the point where the marginal product of R&D capital equals $\rho$:

$$\rho = \frac{(1 - \delta)/(1 - \tau) - A^d}{(r + \delta + MAC)}$$

$\tau$ is the corporate tax rate, $\delta$ is the (economic) depreciation rate, and $MAC$ is the marginal adjustment cost.

In this equation, $A^d$ and $A^t$ are the present discounted value of depreciation allowances and tax credits respectively. In most financial accounting systems, including those used by major OECD economies, R&D is expensed as it is incurred rather than capitalized and depreciated, which means that the lifetime of the investment for accounting purposes is much shorter than the economic life of the asset created and that $A^d$ is simply equal to $\tau$ for tax-paying firms. Many countries have a form of tax credit for R&D, either incremental or otherwise, and this will be reflected in a positive value for $A^t$.

Note that when $A^t$ is zero, the corporate tax rate does not enter into the marginal R&D decision, because of the full deductability of R&D.

---

2 See Hall and Van Reenen (2000) for details. The US has an incremental R&D tax credit with a value for $A^t$ of about 0.13, whereas the UK has no credit at the present time, so $A^t=0$. 

5
The user cost formulation above directs attention to the following determinants of R&D financing:

1. tax treatment such as tax credits, which are clearly amenable to intervention by policy makers.

2. economic depreciation $\delta$, which in the case of R&D is more properly termed obsolescence. This quantity is sensitive to the realized rate of technical change in the industry, which is in turn determined by such things as market structure and the rate of imitation. Thus it is difficult to treat $\delta$ as an invariant parameter in this setting.

3. the marginal costs of adjusting the level of the R&D program.

4. the investor’s required rate of return $r$.

The last item has been the subject of considerable theoretical and empirical interest, on the part of both industrial organization and corporate finance economists. Two broad strands of investigation can be observed: one focuses on the role of asymmetric information and moral hazard in raising the required rate of return above that normally used for conventional investment, and the latter on the requirements of different sources of financing and their differing tax treatments for the rate of return. The next section of the paper discusses these factors.

III. Theoretical background

This section of the paper reviews the reasons that the impact of financial considerations on the investment decision may vary with the type of investment and with the source of funds in more detail. To do this, I distinguish between those factors that arise from various kinds of market failures in this setting and the purely financial (or tax-oriented) considerations that affect the cost of different sources of funds.
One of the implications of the well-known Modigliani-Miller theorem (1958, 1961) is that a firm choosing the optimal levels of investment should be indifferent to its capital structure, and should face the same price for investment and R&D investment on the margin. The last dollar spent on each type of investment should yield the same expected rate of return (after adjustment for nondiversifiable risk). A large literature, both theoretical and empirical, has questioned the bases for this theorem, but it remains a useful starting point.

Reasons why the theorem might fail in practice are several: 1) uncertainty coupled with incomplete markets may make a real options approach to the R&D investment decision more appropriate; 2) the cost of capital may differ by source of funds for non-tax reasons; 3) the cost of capital may differ by source of funds for tax reasons; and 4) the cost of capital may also differ across types of investments (tangible and intangible) for both tax and other reasons.

With respect to R&D investment, economic theory advances a plethora of reasons why there might be a gap between the external and internal costs capital; these can be divided into three main types:

1. Asymmetric information between inventor and investor.
2. Moral hazard on the part of the inventor or arising from the separation of ownership and management.
3. Tax considerations that drive a wedge between external finance and finance by retained earnings.

**Asymmetric information problems**

In the R&D setting, the asymmetric information problem refers to the fact that an inventor frequently has better information about the likelihood of success and the nature of the contemplated innovation project than potential investors. Therefore, the marketplace for financing the
development of innovative ideas looks like the “lemons” market modeled by Akerlof (1970). The lemons' premium for R&D will be higher than that for ordinary investment because investors have more difficulty distinguishing good projects from bad when the projects are long-term R&D investments than when they are more short-term or low-risk projects (Leland and Pyle, 1977). When the level of R&D expenditure is a highly observable signal, as it is under current U.S. and U.K. rules, we might expect that the lemons' problem is somewhat mitigated, but certainly not eliminated.3

In the most extreme version of the lemons model, the market for R&D projects may disappear entirely if the asymmetric information problem is too great. Informal evidence suggests that some potential innovators believe this to be the case in fact. And as will be discussed below, venture capital systems are viewed by some as a solution to this “missing markets” problem.

Reducing information asymmetry via fuller disclosure is of limited effectiveness in this arena, due to the ease of imitation of inventive ideas. Firms are reluctant to reveal their innovative ideas to the marketplace and the fact that there could be a substantial cost to revealing information to their competitors reduces the quality of the signal they can make about a potential project (Bhattacharya and Ritter, 1983; Anton and Yao, 1998). Thus the implication of asymmetric information coupled with the costliness of mitigating the problem is that firms and inventors will face a higher cost of external than internal capital for R&D due to the lemons' premium.

Some empirical support for this proposition exists, mostly in the form of event studies that measure the market response to announcements of new debt or share issues. Both Alam and Walton (1995) and Zantout (1997) find higher abnormal returns to firm shares following new debt issues when the firm is more R&D-intensive. The argument is that the acquisition of new sources of

3 Since 1974, publicly traded firms in the United States have been required to report their total R&D expenditures in their annual reports and 10-K filings with the SEC, under FASB rule No. 2, issued October 1974. In 1989, a new accounting standard, SSAP 13, obligated similar disclosures in the UK. Most continental European countries do not have such a requirement, although they may evolve in that direction due to international harmonization of accounting standards, at least for publicly traded firms.
financing is good news when the firm has an asymmetric information problem because of its R&D strategy. Similarly, Szewcyk, Tsetsekos, and Zantout (1996) find that investment opportunities (as proxied by Tobin’s q) explain R&D-associated abnormal returns, and that these returns are higher when the firm is highly leveraged, implying a higher required rate of return for debt finance in equilibrium.

**Moral hazard problems**

Moral hazard in R&D investing arises in the usual way: modern industrial firms normally have separation of ownership and management. This leads to a principal-agent problem when the goals of the two conflict, which can result in investment strategies that are not share value maximizing. Two possible scenarios may co-exist: one is the usual tendency of managers to spend on activities that benefit them (growing the firm beyond efficient scale, nicer offices, etc.) and the second is a reluctance of risk averse managers to invest in uncertain R&D projects. Agency costs of the first type may be avoided by reducing the amount of free cash flow available to the managers by leveraging the firm, but this in turn forces them to use the higher cost external funds to finance R&D (Jensen and Meckling, 1976). Empirically, there seem to be limits to the use of the leveraging strategy in R&D-intensive sectors. See Hall (1990, 1994) for evidence that the LBO/restructuring wave of the 1980s was almost entirely confined to industries and firms where R&D was of no consequence.

According to the second type of principal-agent conflict, managers are more risk averse than shareholders and avoid R&D projects that will increase the riskiness of the firm. If bankruptcy is a possibility, managers whose opportunity cost is lower than their present earnings and potential bondholders may both wish to avoid variance-increasing projects which shareholders would like to undertake. The argument of the theory is that long-term investments can suffer in this case. The
optimal solution to this type of agency cost would be to increase the long-term incentives faced by the manager rather than reducing free cash flow.

Evidence on the importance of agency costs as they relate to R&D takes several forms. Several researchers have studied the impact of antitakeover amendments (which arguably increase managerial security and willingness to take on risk while reducing managerial discipline) on R&D investment and firm value. Johnston and Rao (1997) find that such amendments are not followed by cuts in R&D, while Pugh, Jahara, and Oswald (1999) find that adoption of an Employee Stock Ownership Plan (ESOP), which is a form of antitakeover protection, is followed by R&D increases. Cho (1992) finds that R&D intensity increases with the share that managerial shareholdings represent of the manager’s wealth and interprets this as incentive pay mitigating agency costs and inducing long term investment.

Some have argued that institutional ownership of the managerial firm can reduce the agency costs due to free-riding by owners that is a feature of the governance of firms with diffuse ownership structure, while others have held that such ownership pays too much attention to short term earnings and therefore discourages long term investments. Institutions such as mutual and pension funds often control somewhat larger blocks of shares than individuals, making monitoring firm and manager behavior a more effective and more rewarding activity for these organizations.

There is some limited evidence that this may indeed be the case. Eng and Shackell (2001) find that firms adopting long term performance plans for their managers do not increase their R&D spending but that institutional ownership is associated with higher R&D; R&D firms tend not to be held by banks and insurance companies. Majumdar and Nagarajan (1997) find that high institutional investor ownership does not lead to short-term behavior on the part of the firm, in particular, it does not lead to cuts in R&D spending. Francis and Smith (1995) find that diffusely held firms are less innovative, implying that monitoring alleviates agency costs and enables investment in innovation.
Although the evidence summarized above is fairly clear and indicates that long term incentives for managers can encourage R&D and that institutional ownership does not necessarily discourage R&D investment, it is fairly silent on the magnitude of these effects, and whether these governance features truly close the agency cost-induced gap between the cost of capital and the return to R&D.

**Capital structure and R&D**

In the view of some observers, the leveraged buyout (LBO) wave of the 1980s in the United States and the United Kingdom arose partly because high real interest rates meant that there were strong pressures to eliminate free cash flow within firms (Blair and Litan, 1990). For firms in industries where R&D is an important form of investment, such pressure should have been reduced by the need for internal funds to undertake such investment and indeed Hall (1993, 1994) and Opler and Titman (1993) find that firms with high R&D intensity were much less likely to do an LBO. Opler and Titman (1994) find that R&D firms that were leveraged suffered more than other firms when facing economic distress, presumably because leverage meant that they were unable to sustain R&D programs in the fact of reduced cash flow.

In related work using data on Israeli firms, Blass and Yosha (2001) report that R&D-intensive firms listed on the United States stock exchanges use highly equity-based sources of financing, whereas those listed only in Israel rely more on bank financing and government funding. The former are more profitable and faster-growing, which suggests that the choice of where to list the shares and whether to finance with new equity is indeed sensitive to the expected rate of return to the R&D being undertaken. That is, investors supplying arms-length finance require higher returns to compensate them for the risk of a “lemon.”
Although leverage may be a useful tool for reducing agency costs in the firm, it is of limited value for R&D-intensive firms. Because the knowledge asset created by R&D investment is intangible, partly embedded in human capital, and ordinarily very specialized to the particular firm in which it resides, the capital structure of R&D-intensive firms customarily exhibits considerably less leverage than that of other firms. Banks and other debtholders prefer to use physical assets to secure loans and are reluctant to lend when the project involves substantial R&D investment rather than investment in plant and equipment. In the words of Williamson (1988), “redeployable” assets (that is, assets whose value in an alternative use is almost as high as in their current use) are more suited to the governance structures associated with debt. Empirical support for this idea is provided by Alderson and Betker (1996), who find that liquidation costs and R&D are positively related across firms. The implication is that the sunk costs associated with R&D investment are higher than that for ordinary investment.

In addition, servicing debt usually requires a stable source of cash flow, which makes it more difficult to find the funds for an R&D investment program that must be sustained at a certain level in order to be productive. For both these reasons, firms are either unable or reluctant to use debt finance for R&D investment, which may raise the cost of capital, depending on the precise tax treatment of debt versus equity.\(^4\) Confirming empirical evidence for the idea that limiting free cash flow in R&D firms is a less desirable method of reducing agency costs is provided by Chung and Wright (1998), who find that financial slack and R&D spending are correlated with the value of growth firms positively, but not correlated with that of other firms.

\(^4\) There is also considerable cross-sectional evidence for the United States that R&D intensity and leverage are negatively correlated across firms. See Friend and Lang (1988), Hall (1992), and Bhagat and Welch (1995).
Taxes and the source of funds

Tax considerations that yield variations in the cost of capital across source of finance have been well articulated by Auerbach (1984) among others. He argued that under the U.S. tax system during most of its history the cost of financing new investment by debt has been less that of financing it by retained earnings, which is in turn less than that of issuing new shares. More explicitly, if $r$ is the risk-adjusted required return to capital, $\tau$ is the corporate tax rate, $\theta$ is the personal tax rate, and $c$ is the capital gains tax rate, we have the following required rates of return for different financing sources:

- **Debt**: $r(1-\tau)$, interest deductible at the corporate level
- **Retained earnings**: $r(1-\theta)/(1-c)$, avoids personal tax on dividends, but capital gains tax
- **New shares**: $r/(1-c)$, eventual capital gains tax

If dividends are taxed, clearly financing with new shares is more expensive than financing with retained earnings. And unless the personal income tax rate is much higher than the sum of the corporate and capital gains rates, the following inequalities will both hold:

$$(1 - \tau) < \frac{1 - \theta}{1 - c} < \frac{1}{1 - c}$$

These inequalities express the facts that interest expense is deductible at the corporate level, while dividend payments are not, and that shareholders normally pay tax at a higher rate on retained earnings that are paid out than on those retained by the firm and invested.\(^5\) It implicitly assumes that the returns from the investment made will be retained by the firm and eventually taxed at the capital gains rate rather than the rate on ordinary income.

\(^5\) A detailed discussion of tax regimes in different countries is beyond the scope of this survey, but it is quite a common in several countries for long term capital gains on funds that remain with a firm for more than one year to be taxed at a lower rate than ordinary income. Of course, even if the tax rates on the two kinds of income are equal, the inequalities will hold. Only in the case where dividends are not taxed at the corporate level (which was formerly the case in the UK) will the ranking given above not hold.
It is also true that the tax treatment of R&D in most OECD economies is very different from that of other kinds of investment: because R&D is expensed as it is incurred, the effective tax rate on R&D assets is lower than that on either plant or equipment, with or without an R&D tax credit in place. This effectively means that the economic depreciation of R&D assets is considerably less than the depreciation allowed for tax purposes -- which is 100 percent -- so that the required rate of return for such investment would be lower. In addition some countries offer a tax credit or subsidy to R&D spending, which can reduce the after tax cost of capital even further.6

The conclusion from this section of the paper is that the presence of either asymmetric information or a principal-agent conflict imply that new debt or equity finance will be relatively more expensive for R&D than for ordinary investment, and that considerations such as lack of collateral further reduce the possibility of debt finance. Together, these arguments suggest an important role for retained earnings in the R&D investment decision, independent of their value as a signal of future profitability. In fact, as has been argued by both Hall (1992) and Himmelberg and Petersen (1994), there is good reason to think that positive cash flow may be more important for R&D than for ordinary investment. The next section reports on a series of empirical tests for this proposition.

IV. Testing for financial constraints

The usual way to examine the empirical relevance of the arguments that R&D investment in established firms can be disadvantaged when internal funds are not available and recourse to external capital markets required is to estimate R&D investment equations and test for the presence of “liquidity” constraints, or excess sensitivity to cash flow shocks. This approach builds on the extensive literature developed for testing ordinary investment equations for liquidity constraints.

---

6 See Hall and Van Reenen (2000) for details.
It suffers from many of the same difficulties as the estimates in the investment literature, plus one additional problem that arises from the tendency of firms to smooth R&D spending over time.

The ideal experiment for identifying the effects of liquidity constraints on investment is to give firms additional cash exogenously, and observe whether they pass it on to shareholders or use it for investment and/or R&D. If they choose the first alternative, either the cost of capital to the firm has not fallen, or it has fallen but they still have no good investment opportunities. If they choose the second, then the firm must have had some unexploited investment opportunities that were not profitable using more costly external finance. A finding that investment is sensitive to cash flow shocks that are not signals of future demand increases would reject the hypothesis that the cost of external funds is the same as the cost of internal funds. However, lack of true experiments of this kind forces researchers to use econometric techniques such as instrumental variables to attempt to control for demand shocks when estimating the investment demand equation, with varying degrees of success.

The methodology for the identification of R&D investment equations is based on a simple supply and demand heuristic, as shown in Figure 1. The curve sloping downward to the right represents the demand for R&D investment funds and the curves sloping upward the supply of funds. Internal funds are available at a constant cost of capital until they are exhausted, at which point it becomes necessary to issue debt or equity in order to finance more investment. When the demand curve cuts the supply curve in the horizontal portion, a shock that increases cash flow (and shifts supply outward) has no effect on the level of investment. However, if the demand curve cuts the supply curve where it is upward sloping, it is possible for a shock to cash flow to shift the supply curve out in such a way as to induce a substantial increase in R&D investment. Figure 2 illustrates
such a case, where the firm shifts from point A to point B in response to a cash flow shock that does not shift the demand curve.

Econometric work that tests the hypothesis that financing constraints matter for R&D investment has largely been done using standard investment equation methodology. Two main approaches can be identified: one uses a neoclassical accelerator model with ad hoc dynamics to allow for the presence of adjustment costs, and the other an Euler equation derived from the forward-looking dynamic program of a profit-maximizing firm that faces adjustment costs for capital.\(^7\)

The accelerator model begins with the marginal product equal to cost condition for capital:

\[
MPK = C
\]

Assuming that the production function for the \(i\)th firm at time \(t\) is Cobb-Douglas and taking logarithms of this relationship yields

\[
k_t = s_t + a_t - c_t
\]

where \(k = \log(\text{R&D capital}), s = \log(\text{output or sales}),\) and \(c = \log(\text{cost of R&D}). a_t\) captures any permanent differences across firms, including differences in the production function.

Lagged adjustment is allowed for by specifying an autoregressive distributed lag (ADL) for the relationship between capital and sales. For example, specifying an ADL(2,2) and approximating \(\Delta k\) by \(R/K-\delta\) yields an estimating equation of the following form:

\[
R/K = f(R(-1)/K(-1), \Delta s, \Delta s(-1), k(-2)-s(-2), \text{time dummies, firm dummies})
\]

The time dummies capture the conventional cost of capital, assumed to be the same for all firms. Firm-specific costs related to financing constraints are included by adding current and lagged values of the cash flow/capital ratio to this equation. Because of the presence of firm dummies, estimation

\(^7\) A detailed consideration of the econometric estimation of these models can be found in Mairesse, Hall, and Mulkay (1999). See also Hall (1981).
is done using first differences of this equation, instrumented by lagged values of the right hand side variables to correct for the potential endogeneity of the contemporaneous values. In principle, this will also control for the potential simultaneity between current investment and the disturbance.

The Euler equation approach begins with the following first order condition for investment in two adjacent periods:

\[ E_{t-1} [MPK_t + (1-\delta)(p_{t+1}+MAC_t) - (1+r)(\alpha_{t-1} \alpha_t^{-1} (p_{t-1}+MAC_{t-1}))] = 0 \]

where \( MAC \) denotes the marginal adjustment costs for capital and \( \alpha_t \) is the shadow value of investment funds in period \( t \), which will be unity if there are no financing constraints. After specifying a Cobb-Douglas production function and quadratic adjustment costs, we obtain the following estimating equation:

\[ \frac{R}{K} = f(\frac{R(-1)}{K(-1)}, \frac{S}{K}, \frac{R}{K}^2, \text{time dummies, firm dummies}) \]

Like the accelerator model, this equation also should be estimated in differenced from with lagged values of the right hand side variables as instruments.

When financial constraints are present, the coefficient of lagged R&D investment in the Euler equation differs from \( (1+r) \) by the term \( (\alpha_{t-1}/\alpha_t) \). The implication is that when the firm changes its financial position (that is, the shadow value of additional funds for investment changes) between one period and the next, it will invest as though it is facing a cost of capital greater than \( r \) (when the shadow value falls between periods) or less than \( r \) (when the shadow value rises between periods). Clearly this is a very difficult test to perform because \( (\alpha_{t-1}/\alpha_t) \) is not constant across firms or across time periods, so it cannot be treated as a parameter.

Three solutions are possible: the first is to model \( (\alpha_{t-1}/\alpha_t) \) as a function of proxies for changes in financial position, such as dividend behavior, new share issues, or new debt issues. The second is more ad hoc: recall that this term also multiplies the price \( p_t \) of R&D capital to create a firm-specific cost of capital. Most researchers simply include the cash flow to capital ratio in the
model to proxy for the firm-specific cost of capital and test whether it enters in the presence of time dummies that are the same for all firms. This method assumes that all firms face the same R&D price (cost of capital), except for the cash flow effect.

The third possibility is to stratify firms in some way that is related to the level of cash constraints that they face (for example, dividend-paying and non-dividend paying firms) estimate separate investment equations for each group, and test whether the coefficients are equal. This last was the method used by Fazzari, Hubbard, and Petersen (1988) in the paper that originated this literature. Note that this approach does not rely on the full Euler equation derivation given above, but uses a version of the neoclassical accelerator model (the first model given above).

During the past few years, various versions of the methodologies described above have been applied to data on the R&D investment of U.S., U.K., French, German, Irish, and Japanese firms. The firms examined are typically the largest and most important manufacturing firms in their economy. For example, Hall (1992) found a large positive elasticity between R&D and cash flow, using an accelerator-type model and a very large sample of U.S. manufacturing firms. The estimation methodology here controlled for both firm effects and simultaneity. Similarly and using some of the same data, Himmelberg and Petersen (1994) looked at a panel of 179 U.S. small firms in high-tech industries and find an economically large and statistically significant relationship between R&D investment and internal finance.

Harhoff (1998) found weak but significant cash flow effects on R&D for both small and large German firms, although Euler equation estimates for R&D investment were uninformative due to the smoothness of R&D and the small sample size. Combining limited survey evidence with his regression results, he concludes that R&D investment in small German firms may be constrained by the availability of finance. Bond, Harhoff, and Van Reenen (1999) find significant differences between the cash flow impacts on R&D and investment for large manufacturing firms in the United
Kingdom and Germany. German firms in their sample are insensitive to cash flow shocks, whereas the investment of non-R&D-doing UK firms does respond. Cash flow helps to predict whether a UK firm does R&D, but not the level of that R&D. They interpret their findings to mean that financial constraints are important for British firms, but that those which do R&D are a self-selected group that face fewer constraints. This is consistent with the view that the desire of firms to smooth R&D over time combines with the relatively high cost of financing it to reduce R&D well below the level that would obtain in a frictionless world.

Mulkay, Hall, and Mairesse (2001) perform a similar exercise using large French and U.S. manufacturing firms, finding that cash flow impacts are much larger in the U.S. than in France, both for R&D and for ordinary investment. Except for the well-known fact that R&D exhibits higher serial correlation than investment (presumably because of higher adjustment costs), differences in behavior are between countries, not between investment types. This result is consistent with evidence reported in Hall, Mairesse, Branstetter, and Crepon (1999) for the U.S., France, and Japan during an earlier time period, which basically finds that R&D and investment on the one hand, and sales and cash flow on the other, are simultaneously determined in the United States (neither one “Granger-causes” the other, whereas in the other countries, there is little feedback from sales and cash flows to the two investments. Using a nonstructural R&D investment equation together with data for the US, UK, Canada, Europe, and Japan, Bhagat and Welch (1995) found similar results for the 1985-1990 period, with stock returns predicting changes in R&D more strongly for the US and UK firms.

Recently, Bougheas, Goerg, and Strobl (2001) examined the effects of liquidity constraints on R&D investment using firm-level data for manufacturing firms in Ireland and also found evidence that R&D investment in these firms is financially constrained, in line with the previous studies of US and UK firms.
Brown (1997) argues that existing tests of the impact of capital market imperfections on innovative firms cannot distinguish between two possibilities: 1) capital markets are perfect and different factors drive the firm's different types of expenditure or 2) capital markets are imperfect and different types of expenditure react differently to a common factor (shocks to the supply of internal finance). He then compares the sensitivity of investment to cash flow for innovative and non-innovative firms. The results support the hypothesis that capital markets are imperfect, finding that the investment of innovative firms is more sensitive to cash flow.

The conclusions from this body of empirical work are several: first, there is solid evidence that debt is a disfavored source of finance for R&D investment; second, the “Anglo-Saxon” economies, with their thick and highly developed stock markets and relatively transparent ownership structures, typically exhibit more sensitivity and responsiveness of R&D to cash flow than continental economies; third, and much more speculatively, this greater responsiveness may arise because they are financially constrained, in the sense that they view external sources of finance as much more costly than internal, and therefore require a considerably higher rate of return to investments done on the margin when they are tapping these sources. However, it is perhaps equally likely that this responsiveness occurs because firms are more sensitive to demand signals in thick financial equity markets; a definitive explanation of the “excess sensitivity” result awaits further research. In addition to these results, the evidence from Germany and some other countries suggests that small firms are more likely to face this difficulty than large established firms (not surprisingly, if the source of the problem is a “lemons” premium).

---

8 It is also true that much of the literature here has tended to downplay the role of measurement error in drawing conclusions from the results. Measurement error in Tobin’s q, cash flow, or output is likely to be sizable and will ensure that all variables will enter any specification of the R&D investment equation significantly, regardless of whether they truly belong or not. Instrumental variables estimation is a partial solution, but only if all the errors are serially uncorrelated, which is unlikely.
From a policy perspective, these results point to another reason why it may be socially beneficial to offer tax incentives to companies in order to reduce the cost of capital they face for R&D investment, especially to small and new firms. Many governments, including those in the United States and the United Kingdom, currently have such programs. Such a policy approach simply observes that the cost of capital is relatively high for R&D and tries to close the gap via a tax subsidy. However, there is an alternative approach relying on the private sector that attempts to close the financing gap by reducing the degree of asymmetric information and moral hazard rather than simply subsidizing the investment. I turn to this topic in the next section.

V. Small Firms, Startup Finance, and Venture Capital

As should be apparent from much of the preceding discussion, any problems associated with financing investments in new technology will be most apparent for new entrants and startup firms. For this reason, many governments already provide some of form of assistance for such firms, and in many countries, especially the United States, there exists a private sector “venture capital” industry that is focused on solving the problem of financing innovation for new and young firms. This section of the paper reviews what we know about these alternative funding mechanisms, beginning with a brief look at government funding for startups and then discussing the venture capital solution.

**Government funding for startup firms**

Examples of such programs are the U.S. Small Business Investment Company (SBIC) and Small Business Innovation Research (SBIR) programs. Together, these programs disbursed $2.4 billion in 1995, more than 60% of the amount from venture capital in that year (Lerner 1998a). In Germany, more than 800 federal and state government financing program have been established for
new firms in the recent past (OECD 1995). In 1980, the Swedish established the first of a series of
investment companies (along with instituting a series of measures such as reduced capital gains taxes
to encourage private investments in startups), partly on the United States model. By 1987, the
government share of venture capital funding was 43 percent (Karaomerliolu and Jacobsson 1999).
Recently, the UK has instituted a series of government programs under the Enterprise Fund
umbrella which allocate funds to small and medium-sized firms in high technology and certain
regions, as well as guaranteeing some loans to small businesses (Bank of England 2001). There are
also programs at the European level.

A limited amount of evidence, most of it U.S.-based, exists as to the effectiveness and
“additionality” of these programs. In most cases, evaluating the success of the programs is difficult
due to the lack of a “control” group of similar firms that do not receive funding.9 Therefore most of
the available studies are based on retrospective survey data provided by the recipients; few attempt
to address the question of performance under the counterfactual seriously. A notable exception is
the study by Lerner (1999), who looks at 1435 SBIR awardees and a matched sample of firms that
did not receive awards, over a ten-year post-award period. Because most of the firms are privately
held, he is unable to analyze the resulting valuation or profitability of the firms, but he does find that
firms receiving SBIR grants grow significantly faster than the others after receipt of the grant. He
attributes some of this effect to “quality certification” by the government that enables the firm to
raise funds from private sources as well.10

9 See Jaffe (this issue) for a review of methodologies for evaluation such government programs. For a complete review
of the SBIR program, including some case studies, see the National Research Council (1998).
10 Also see Spivack (2001) for further studies of such programs, including European studies, and David, Hall, and Toole
(2000) and Klette, Moen, and Griliches Klette (2000) for surveys of the evaluation of government R&D programs in
general.
Venture capital

Many observers view the rise of the venture capital (VC) industry, especially that in the United States, a “free market” solution to the problems of financing innovation. In fact, many of the European programs described above have as some of their goals the provision of seed capital and the encouragement of a venture capital industry that addresses the needs of high technology startups. Table 1 shows why this has been of some concern to European policymakers: the amount of venture capital available to firms in the United States and Europe was roughly comparable in 1996, but the relative allocation to new firms (seed money and startups) in Europe was much less, below 10% of the funds as opposed to 27%. A correspondingly greater amount was used to finance buyouts of various kinds.

In the United States, the VC industry consists of fairly specialized pools of funds (usually from private investors) that are managed and invested in companies by individuals knowledgeable about the industry in which they are investing. In principle, the idea is that the lemons premium is reduced because the investment managers are better informed and moral hazard is minimized because a higher level of monitoring than that used in conventional arm’s length investments is the norm. But the story is more complex than that: the combination of high uncertainty, asymmetric information, and the fact that R&D investment typically does not yield results instantaneously not only implies option-like behavior for the investment decision but also has implications for the form of the VC contract and the choice of decision maker. That is, there are situations in which it is optimal for the investor (VC) to have the right to shut down a project and there are other situations in which optimal performance is achieved when the innovator has control.

A number of studies have documented the characteristics and performance of the VC industry in the United States. The most detailed look at the actual operation of the industry is that by Kaplan and Stromberg (2000), who examine 200 Venture Capital contracts and compare their
provisions to the predictions of the economic theory of financial contracting under uncertainty. They find that the contracts often provide for separate allocation of cash flow rights, control rights, voting rights, board positions, and liquidation rights, and that the rights are frequently contingent on performance measures. If performance is poor, the VCs often gain full control of the firm. Provisions such as delayed vesting are often included to mitigate hold-up by the entrepreneur as suggested by Anand and Galetovic (2000).

Kaplan and Stromberg conclude that these contracts are most consistent with the predictions of Aghion and Bolton (1992) and Dewatripont and Tirole (1994), all of whom study the incomplete contracts that arise when cash flows can be observed but not verified in sufficient detail to be used for contract enforcement. Put simply, the modal VC contract is a complex debt-equity hybrid (and in fact, frequently contains convertible preferred securities and other such instruments) that looks more like debt when the firm does poorly (giving control to the investor) and more like equity when the firm does well (by handing control to the entrepreneur, which is incentive-compatible).

In a series of papers, Lerner (1992, 1995) studied a sample of VC-financed startups in detail, highlighting the important role that investing and monitoring experience has in this industry. He found that the amount of funds provided and the share of equity retained by the managers are sensitive to the experience and ability of the capital providers and the maturity of the firm being funded. VCs do increase the value of the firms they fund, especially when they are experienced investors. Firms backed by seasoned VC financiers are more likely to successfully time the market when they go public, and to employ the most reputable underwriters.

At a macro-economic level, VC funding tends to be pro-cyclical but it is difficult to disentangle whether the supply of funding causes growth or productivity growth encourages funding (Kortum and Lerner 2000; Gompers and Lerner 1999a,b; Ueda 2001). The problem here is very
similar to the identification problem for R&D investment in general: because of feedback effects, there is a chicken-egg simultaneity in the relationship. Some evidence (Majewski 1997) exists that new and/or small biotechnology firms turn to other sources of funding in downturns, but that such placements are typically less successful (Lerner and Tsai 2000), due to the misallocation of control rights (when the startup firm is in a weak bargaining position, control tends to be allocated to the more powerful corporate partner, but this has negative consequences for incentives).

The limited evidence from Europe on the performance of VC-funded firms tends to confirm that from the U.S. Engel (2001) compares a matched sample of German firms founded between 1991 and 1998 and finds that the VC-backed firms grew faster than the non-VC-backed firms. Lumme et al (1993) compare the financing and growth of small UK and Finnish firms. This approach permits a comparison between a financial market-based and a bank-centered economy, and indeed, they find that small UK firms rely more on equity and less on loan finance and grow faster than small Finnish firms. Further evidence on small UK high technology firms is provided by Moore (1993), who looks at 300 such firms, finding that the availability and cost of finance is the most important constraint facing these firms, but that they are affected only marginally more than other types of small firms. That is, the financing “gap” in the UK may be more related to size than to R&D intensity.

For Japan, Hamao, Packer, and Ritter (1998) find that the long run performance of VC-backed Initial Public Offerings (IPOs) are no better than that for other IPOs, unlike Lerner’s evidence for the United States. However, many VCs in Japan are subsidiaries of major securities firms rather than specialists as in the United States. Only these VCs have low returns, whereas those that are independent have returns more similar to the US. They attribute the low returns to conflicts of interest between the VC subsidiary and the securities firm that owns it, which affects the price at
which the IPO is offered. This result highlights the importance of the institutions in which the venture capital industry is embedded for the creation of entrepreneurial incentives.

Black and Gilson (1997) and Rajan and Zingales (2001) take the institutional argument further. Both pairs of authors emphasize the contrast between arms’ length market-based financial systems (e.g., the US and the UK) and bank-centered capital market systems (e.g., much of continental Europe and Japan), and view venture capital as combining the strengths of the two systems, in that it provides both the strong incentives for the manager-entrepreneur characteristic of the stock market system and the monitoring by an informed investor characteristic of the bank-centered system. They emphasize the importance of an active stock market, especially for newer and younger firms, in order to provide an exit strategy for VC investors, and allow them to move on to financing new startups. Thus having a VC industry that contributes to innovation and growth requires the existence of an active IPO (Initial Public Offering) market to permit successful entrepreneurs to regain control of their firms (and incidentally to provide powerful incentives for undertaking the startup in the first place) and also to ensure that the VCs themselves are able to use their expertise to help to establish new endeavors.

VI. Conclusions

Based on the literature surveyed here, what do we know about the costs of financing R&D investments and the possibility that some kind of market failure exists in this area? Several main points emerge:

1. There is fairly clear evidence, based on theory, surveys, and empirical estimation, that small and startup firms in R&D-intensive industries face a higher cost of capital than their larger competitors and than firms in other industries. In addition to compelling theoretical arguments and empirical evidence, the mere existence of the VC industry and the fact that it is concentrated
precisely where these startups are most active suggests that this is so. In spite of considerable entry into the VC industry, returns remain high, which does suggest a high required rate of return in equilibrium (Upside 2001).

2. The evidence for a financing gap for large and established R&D firms is harder to establish. It is certainly the case that these firms prefer to use internally generated funds for financing investment, but less clear that there is an argument for intervention, beyond the favorable tax treatment that currently exists in many countries.\textsuperscript{11}

3. The VC solution to the problem of financing innovation has its limits: First, it does tend to focus only on a few sectors at a time, and to make investment with a minimum size that is too large for startups in some fields. Second, good performance of the VC sector requires a thick market in small and new firm stocks (such as NASDAQ or EASDAQ) in order to provide an exit strategy for early stage investors.

4. The effectiveness of government incubators, seed funding, loan guarantees, and other such policies for funding R&D deserves further study, ideally in an experimental or quasi-experimental setting. In particular, studying the cross-country variation in the performance of such programs would be desirable, because the outcomes may depend to a great extent on institutional factors that are difficult to control for using data from within a single country.

\textsuperscript{11} It is important to remind the reader of the premise of this paper: I am focusing only on the financing gap arguments for favorable treatment of R&D and ignoring (for the present) the arguments based on R&D spillovers and externalities. There is good reason to believe that the latter is a much more important consideration for large established firms, especially if we wish those firms to undertake basic research that is close to industry but with unknown applications (the Bell Labs model).
References


Figure 1
Unconstrained Firm

Figure 2
Constrained Firm
Table 1
Venture Capital Disbursements by Stage of Financing (1996)

<table>
<thead>
<tr>
<th></th>
<th>United States</th>
<th>Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total VC disbursements</td>
<td>9,420.6</td>
<td>8,572.0</td>
</tr>
<tr>
<td>(millions $1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share seed and startups</td>
<td>27.1%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Share for expansion</td>
<td>41.6%</td>
<td>39.3%</td>
</tr>
<tr>
<td>Share other (incl. buyouts)</td>
<td>31.3%</td>
<td>54.2%</td>
</tr>
</tbody>
</table>

Source: Rausch (1998) and author’s calculations.