

The Changing Nature of Debt and Equity: A Financial Perspective

*Franklin Allen**

Historically, corporations have mainly financed their activities with two securities, debt and equity. The stockholders have responsibility for the operation of the firm through the election of the board of directors; the dividends they receive in return for their subscription of capital are not guaranteed and are paid at the discretion of the board of directors. In contrast, debtholders are promised a particular rate of return; they have no rights of control unless payments by the firm are omitted, in which case they have the right to foreclose on assets or, in some cases, force bankruptcy. Dewing (1934, pp. 236-37) ascribes these differences in rights between debtholders and equityholders to the historical distinction in Anglo-Saxon law between debtors and creditors.

As a result of the importance of debt and equity, the focus of inquiry into firms' choice of capital structure has traditionally been "What is the optimal debt-equity ratio?" Modigliani and Miller (1958) and subsequent authors¹ showed that if capital markets are perfect and complete and no taxes are in effect, a firm's debt-equity ratio has no effect on its value because investors' opportunity sets are not affected by its capital structure. If a corporate income tax is in effect, with interest deductibility, Modigliani and Miller (1963) used the same logic to show firms should use entirely debt finance since this allows corporate taxes to be avoided.

*Associate Professor of Finance and Economics, the Wharton School, University of Pennsylvania. The author is grateful to Jaime Zender and to his discussants, Oliver Hart and Robert Merton, for helpful comments. Financial support from the National Science Foundation (Grant No. SES-8813719) is acknowledged.

¹ See, for example, Hellwig (1981) and the references therein.

This prediction of the theory did not square well with empirical evidence; despite interest deductibility and a corporate tax rate of almost 50 percent at that time, firms typically used only moderate amounts of debt. This led a number of authors² to point to the capital market imperfection of bankruptcy and liquidation costs. They suggested that a firm balances these costs against the tax advantage of debt and it is this trade-off that determines the optimal debt-equity ratio.

The trade-off theory has been criticized on a number of grounds. Evidence on the direct costs of bankruptcy, such as lawyers' fees, suggested they were small (Warner 1977). Direct measurement of the indirect costs of bankruptcy, such as the difficulties of running a firm while it is in bankruptcy court, are difficult to obtain; proponents of the trade-off theory suggest they are significant while detractors suggest they are small relative to the tax advantage of debt. It is widely agreed that liquidation costs, which are the costs of breaking up a firm and selling it off piecemeal, are sufficiently large to explain firms' observed debt ratios if included with bankruptcy costs. However, Haugen and Senbet (1978) argued that liquidation costs should not be included with bankruptcy costs since liquidation was not implied by bankruptcy; if the firm was worth more as a going concern it would not be liquidated. In addition, they argued that if bankruptcy was costly it could be avoided by firms' buying back their debt just before it became due. These arguments depend on perfect markets; a number of recent papers have investigated why bankruptcy and liquidation may be linked and why bankruptcy may be difficult to avoid by repurchasing securities when markets are imperfect.³

The deficiencies of the trade-off theory resulted in the development of a number of alternative theories. Miller (1977) pointed to the importance of personal taxes. He argued that personal taxes on equity were lower than on debt and presented a model where this personal tax disadvantage of debt entirely offset its corporate tax advantage so that in equilibrium each firm was indifferent between the use of equity and debt. De Angelo and Masulis (1980) and subsequent authors⁴ developed this model to allow for bankruptcy costs and other factors; in this case again a trade-off exists between the use of debt and equity and firms have an interior optimal capital structure.

Some of the alternative theories that did not rely on the inclusion of personal taxes were based on asymmetric information. Agency theories

² See, for example, Kim (1978) and the references therein.

³ See, for example, Titman (1984); Allen (1987); Webb (1987); Giammarino (1989); and Mooradian (1989).

⁴ See Kim (1989) for a survey of this literature.

started from the premise that managers' actions could not be fully contractually specified because they were unobservable and would be influenced by capital structure choices (Jensen and Meckling 1976; Myers 1977; and Green 1984). Signaling theories were based on the idea that firms' capital structure choices could convey information about their prospects to investors (Ross 1977; Myers and Majluf 1984; and Brennan and Kraus 1987). More recently, it has been suggested that imperfectly competitive markets for outputs and inputs and opportunities for product innovation may influence firms' choice of capital structure.⁵

The deficiencies of these theories in explaining the use of debt and equity by firms are well documented by Myers (1984). He gives the following succinct summary of the literature (p. 575): " 'How do firms choose their capital structures?' . . . the answer is 'We don't know.' "

Financial Innovation

The notion that firms finance their activities with debt and equity is a simplification; corporations have issued securities other than standard debt and equity for many centuries. Dewing (1934, p. 135) recounts that multiple classes of stock with certain preferences or disabilities were issued by some of the first English companies in the middle of the sixteenth century. He also gives examples (pp. 377-78) of a number of English firms that issued convertible securities in the seventeenth and eighteenth centuries.

In the United States, corporations also have a long history of use of securities other than debt and equity. Since the late 1880s, firms have issued significant amounts of preferred stock. This form of stock combines many of the features of equity with those of debt; in particular, a level of payments is specified, as with debt, but unlike debtholders, investors in preferred stock cannot force bankruptcy if the firm omits these payments. Firms have also issued income bonds at various times since 1848. Like preferred stock, income bonds have a number of features of debt and equity. Unlike preferred stock, the specified payments are not at the discretion of the board of directors but depend on the level of accounting earnings. If they are omitted, however, the securityholders cannot force bankruptcy. Still other types of securities such as convertible bonds and warrants have also been issued by corporations for many decades. Dewing (1934) gives a full account of the early history of these securities.

Financial innovation is, therefore, not a recent phenomenon. How-

⁵ See Ravid (1988) for a survey and also Baldwin (1983a, 1983b, and 1988).

ever, Miller (1986) suggests that financial innovation has proceeded at a particularly fast pace during the last twenty years. Not only have corporations started to issue new securities such as zero coupon bonds and adjustable rate bonds, but also entirely new markets such as the Chicago Board Options Exchange have been established.⁶

Miller argues that much of this recent innovation is in response to features of the tax code and to regulation. A classic example of innovation in response to the tax code is zero coupon bonds. Before the Tax Equity and Fiscal Responsibility Act of 1982 (TEFRA), the tax liability on zero coupon bonds was allocated on a straight line basis; that is, the annual interest deduction was the amount to be repaid at the due date less the issue price, divided by the number of years until repayment. This rule ignored the effect of compounding of interest and created an opportunity for corporations to avoid taxes by issuing long-term zero coupon bonds to tax-exempt investors. When interest rates were high in the early 1980s, the potential tax benefits from this type of security became large and corporations issued a large amount of these bonds. Although TEFRA closed this loophole, the market for zero coupon bonds continued but now was mainly supplied by investment banks "stripping" government securities into principal and interest (Kane-masu, Litzenberger, and Rolfo 1986).

An alternative rationale for financial innovation, stressed by Van Horne (1985), is that new securities may make markets more complete in the sense that they increase opportunities for risk sharing between investors. In a categorization of the primary factors responsible for the introduction of sixty-eight new types of security, Finnerty (1988) lists tax and regulatory advantages in twenty-seven cases and risk reallocation in fifty-three cases. (More than one factor is possible for each type of security.)

In addition to taxes and regulation and risk reallocation, another important class of security innovation has resulted from attempts by incumbent managements to discourage takeovers. Examples of these "poison pill" defenses are preferred stock plans, flip-over plans, back-end plans and voting plans. The securities associated with these plans all have the common feature that on the occurrence of a takeover attempt not approved by the board of directors, certain rights accrue to the securityholders. For example, target shareholders may be given the right to buy the stock of the bidder at a substantial discount on completion of the takeover.⁷

⁶ For a full account of recent innovation see Finnerty (1988).

⁷ See Malatesta and Walking (1988) for a more complete description of actual poison pills.

Tufano (1988) has constructed a data base of fifty-eight financial innovations introduced by investment banks between 1974 and 1987. These innovations, often bonds, equities or preferred stocks with novel features, can cost substantial amounts to develop. Tufano finds that the banks that create these products almost immediately face competition from rivals offering imitative products. During the brief period of monopoly before imitation, originators do not charge high prices to recoup their development costs. Moreover, once the imitative products appear, they charge a lower rather than a higher price than the imitators. The main difference between the originating bank and imitators is that the originating bank obtains a larger share of the market. Tufano gives a number of reasons why market share may allow originators to recoup the costs of developing the products. Sunk costs may be involved in entering the underwriting business. These may deter entry and allow positive profits; price competition may be limited by the type of noncooperative collusion considered by the threat of reverting to the single-period equilibrium. Another possibility is that the bank may make profits on related business so that it can recapture the costs in this way.

The fact that debt and equity are not the only securities that firms use to finance their activities, and the constant introduction of new forms of securities, suggest that a more fundamental issue than "What is the optimal debt-equity ratio?" is "What are the optimal securities that should be issued?" Many recent studies of capital structure have taken this perspective. These studies provide some insight into the changing nature of debt and equity.

This literature has two branches. The first has been concerned with trying to identify the circumstances in which debt and equity are optimal. This will be considered in the next section of this paper. The second branch has been concerned with the optimal securities that a firm should issue. The succeeding section considers this, followed by a summary and conclusions.

When Are Debt and Equity Optimal?

A number of papers have identified situations where debt contracts are optimal. Townsend (1979) considers the optimal contract between a risk-averse agent and a risk-neutral principal. In one version of the model, the agent requires funds at the beginning of the period to produce a random income at the end. The principal can observe the realization of the agent's income only if bankruptcy is declared and the agent's income is transferred to the principal. This bankruptcy process is costly. Among the class of deterministic strategies, where the principal

observes the agent's income with probability either one or zero, Townsend shows that debt is an optimal contract. This requires the agent to pay a constant amount to the principal; if the agent's income is insufficient to pay this amount, then bankruptcy is declared and the agent's income is transferred to the principal.

This basic idea has been used by a number of authors to consider the role of debt contracts in various contexts. For example, Diamond (1984) used a similar framework to explain the use of debt contracts by financial intermediaries such as banks. Gale and Hellwig (1985) consider the case where the agent's investment is mutually observable in order to show that underinvestment can occur.

Allocation of Cash Flows

An important issue is whether this type of analysis can be applied to corporate securities. If the agent is interpreted to be the insiders that operate the firm, and the principal the outside investors that supply capital, then the optimal security for the firm to issue is debt. The question is whether equityholders correspond to the insiders that run the firm or the outside investors. For privately held firms, the equityholders correspond to the insiders that run the firm. For publicly traded corporations, however, most equityholders are outside investors with access to the same information as bondholders; in this case it is not immediately evident that Townsend's type of analysis can be used to justify the existence of debt and equity.

Williams (1989) develops a model to consider this issue. He assumes markets are complete in the sense that everybody is effectively risk neutral with respect to aggregate-state prices. However, asymmetric information about the earnings of individual firms in any particular period can only be observed by the managers or insiders that run the firms; as a result securities cannot be made contingent on earnings in a manner similar to Townsend's type of analysis. In addition, Williams introduces "ex ante monitoring," such as accounting controls, which prevents the managers from simply expropriating a firm's assets. It is also assumed that an agency problem exists between managers and outside investors. It is shown that it is optimal for the firm to issue debt or stock or both to outside investors, with the precise mix of securities depending on the nature of the agency problem.

An important issue is how general the assumptions of the model are and, in particular, the circumstances in which markets are complete in the sense that managers are effectively risk neutral with respect to aggregate-state prices. One possibility is that the managers are risk neutral; if they are risk averse, the fact that they cannot trade securities

state by state that are contingent on the firm's earnings will presumably prevent markets from being effectively complete.

In addition to the applicability of this type of analysis to corporations, another issue to be considered is the assumption by Townsend that strategies are deterministic, so that income is observed by declaring bankruptcy with probability one or zero. Mookherjee and Png (1989) show that if random strategies are possible, then the optimal contract involves randomization. To see why it is possible to do better with random strategies, consider the optimal deterministic contract, which is a debt contract. Suppose that the agent is now made to announce his income, and bankruptcy occurs with probability one whenever the announced income is less than the required payment. During bankruptcy, the true value of the agent's income is revealed. By rewarding the agent when he has correctly announced his income level, it is possible to provide a strict incentive to tell the truth. This means it is no longer necessary to force bankruptcy all the time. Since the agent is risk averse and the principal is risk neutral, this change allows a Pareto improvement. The important issue here is whether randomization is possible. If a device exists that both parties know is truly random, then Townsend's type of analysis is unable to provide a rationale for debt contracts, but if such randomization devices do not exist, it can.

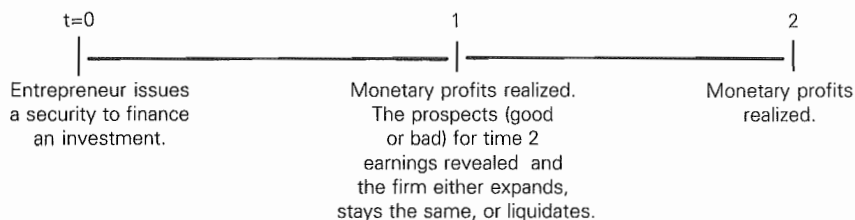
Allocation of Control Rights

The papers considered above are primarily concerned with the allocation of cash flows. In a recent paper, Aghion and Bolton (1988) take a different approach by looking at the allocation of control rights among different securityholders in closely held firms. They consider a model with the sequence of events shown in Figure 1. An entrepreneur has insufficient resources of his own to finance a project he wishes to undertake. The project involves an outlay at time 0 and yields revenues at time 1 and time 2. The entrepreneur can finance the investment by issuing securities at time 0 to an outside investor who receives a portion of the firm's profits at time 1 and time 2. Both the entrepreneur and the investor are assumed to be risk neutral, so that risk-sharing issues are not considered in the model.

At time 1, the firm's monetary profits and its prospects for future earnings, which can be either good or bad, are determined. After receiving this information, the party in control of the firm decides on which of three possible courses of action to undertake: expand the firm, continue as before, or liquidate. If the time 1 prospects for future earnings are good, expansion leads to the highest expected profits, continuing as before the next highest, and liquidating the least. If the

Figure 1

The Sequence of Events in the Aghion and Bolton (1988) Model



prospects are bad, the reverse is true. The private costs to the entrepreneur of the three courses are different, with liquidation being the most costly, expansion the next most costly, and keeping operations the same the least costly. The magnitudes of the expected monetary profits and private costs to the entrepreneur are such that in the first-best world where all states can be contracted on, it is optimal for the firm to continue operations as before in the state where prospects are good, and liquidate in the state where prospects are bad.

The critical assumption that Aghion and Bolton make is that contracting possibilities are incomplete. In particular, the earnings prospects cannot be contracted upon; the only variable that can be contracted on is monetary profits. This creates two problems. The first occurs if the entrepreneur uses securities that cede control of the firm to the investor and the good state is realized. In this case, the investor would like the firm to expand since this maximizes expected monetary profits. However, this is not optimal since it imposes large costs on the entrepreneur; when these costs are taken into account, continuing the current level of operations is optimal.

The second problem occurs if the entrepreneur retains control. Now if prospects are good the efficient action of continuing operations will be chosen; however, if prospects are bad the entrepreneur may not have the correct incentives to liquidate. The entrepreneur bears high private

costs with liquidation; unless he also receives a high proportion of the monetary profits so that most of the marginal benefits of liquidation are obtained, it will not be worth doing. The problem is that since it is not possible to distinguish between states where earnings prospects are good and states where they are bad, it is also necessary to give the entrepreneur most of the monetary profits in the good state. The overall payoff to the investor is then insufficient to make financing the project worthwhile. Hence, a drawback also exists if the entrepreneur retains control.

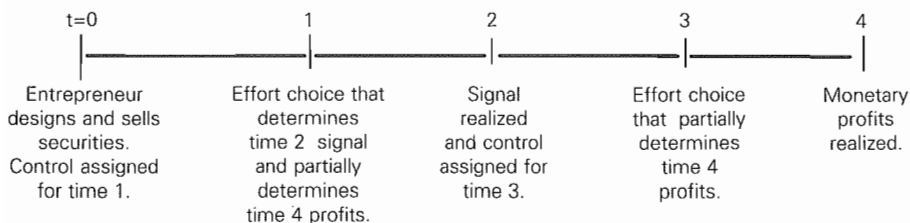
These arguments imply that giving the control entirely to either the investor or the entrepreneur may mean the first-best contract cannot be implemented: if the investor has control, the entrepreneur may be forced to expand, which has high private costs; but if the entrepreneur has control, he may be unwilling to liquidate because of the high private costs associated with that. Ideally, what is required is a mechanism that grants control to the entrepreneur when earnings prospects are good and to the investor when they are bad. Aghion and Bolton argue that the use of debt by the entrepreneur and the institution of bankruptcy can achieve this outcome, if monetary profits and the prospects for future earnings are positively correlated at time 1: for example, in the case where they are perfectly correlated, when earnings prospects are good, monetary profits are high and the entrepreneur retains control. When earnings prospects are bad, monetary profits are low, and if the level of debt issued initially has been correctly chosen, the firm will go bankrupt and control will be transferred to the outside investor.

Zender (1989) also develops a model based on the allocation of control rights where the use of debt and equity is optimal for closely held firms. Once again, all agents are risk neutral so that risk-sharing considerations are not considered in the model. The sequence of events is illustrated in Figure 2. At time 0, an entrepreneur designs and sells securities to two identical investors to finance a project. Individually, neither investor has the funds to finance the project so both must contribute money if the project is to be undertaken. The investor who is assigned control then hires a manager who undertakes an effort choice at time 1. No agency problem exists between the manager and the investor, so the manager acts as the investor specifies. The time 1 effort choice determines the level of a signal at time 2 and partially determines the level of profits at time 4. In addition to the signal that is observed at time 2, control is allocated for time 3. At time 3, the party in control again specifies an effort choice for the manager. This, together with the effort choice at time 1, determines the expected monetary profits realized at time 4.

The problem in the model is to provide the correct incentives for the effort choices at times 1 and 3. A single investor with sufficient funds to

Figure 2

The Sequence of Events in the Zender (1989) Model



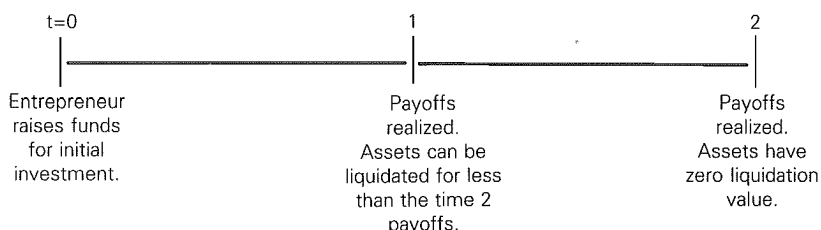
finance the entire project would obtain the full marginal benefits of the effort choices and so would be prepared to undertake the efficient level. However, because neither investor has sufficient funds to finance the entire project, the securities must be such that both have a chance of obtaining part of the time 4 payoffs. This means that the investor in control does not get the full marginal benefit of the effort choice at times 1 and 3.

Zender shows that the optimal contract involves making control at time 3 and the allocation of payoffs at time 4 contingent on the time 2 signal. If a good signal is observed at time 2, the investor in control at time 1 remains in control and retains the residual of the payoffs less a constant amount at time 4. If a bad signal is observed, then control is switched to the second investor who obtains the payoffs at time 4. This optimal contract is interpreted as the investor in control initially having equity and the other investor having debt; it ensures that the investor who is delegated control is made the residual claimant and so has incentives to make the proper decisions.

Another paper that is related to Aghion and Bolton (1988) is Hart and Moore (1989). They also consider a model of an entrepreneur who wishes to raise funds to undertake a project when contracting possibilities are incomplete. The focus of their analysis, however, is the problem of providing an incentive for the entrepreneur to repay the borrowed funds. It is the ability of the creditor to seize the entrepreneur's assets that provides this incentive.

Figure 3

The Sequence of Events in the Hart and Moore (1989) Model



The sequence of events in the simplest version of their model is shown in Figure 3. A risk-neutral entrepreneur raises funds from a risk-neutral outside investor to purchase assets that can realize payoffs at times 1 and 2. If the entrepreneur does not fulfill the contract at time 1, the outside investor can renegotiate or can seize some proportion of the assets and liquidate them. Liquidation is socially inefficient, however, because the liquidation value of the assets at time 1 is less than the present value of the time 2 payoffs. Although both the entrepreneur and the outside investor have symmetric information, third parties such as the courts cannot observe the asset payoffs so these cannot be contracted upon. The entrepreneur can appropriate the cash flows from the assets for his own use, so the problem is to design a contract that provides incentives for the entrepreneur to repay the loan.

It is shown that the optimal contract is a debt contract and the incentives to repay are provided by the threat of liquidation. Since the present value of the time 2 payoffs of the assets is above their liquidation value, the entrepreneur will always want to hold on to as high a proportion of the assets as possible and will be prepared to pay up to the assets' present value. In low payoff states, the entrepreneur will have insufficient cash to make the required payment; the outside investor therefore renegotiates the loan and liquidates a certain proportion of the assets to make the payment up to the required amount. Although this liquidation is inefficient relative to an ideal world, it is necessary because the entrepreneur cannot commit to pay any of the time 2 payoffs to the

investor. The threat of liquidation also ensures that the entrepreneur pays the required amount in high output states.

One interesting implication of the analysis is that reducing the amount borrowed is not always desirable. If the time 1 payoffs or time 2 liquidation values are uncertain, it may be better for the entrepreneur to borrow more than strictly the initial cost of the assets. This allows him to make a higher payment in low output states at time 1, so that a smaller proportion of the assets is liquidated.

A version of the model where the assets pay off at time 3 is also considered. It is shown that the use of short-term or long-term debt depends on when information arrives and the pattern of payoffs. Short-term debt gives the outside investor a high degree of control early on since the entrepreneur has to renew the loan. This has the advantage that the size of the debt can be kept low, which avoids the inefficiencies associated with liquidation. However, it has the disadvantage that the outside investor may liquidate projects early on even though this is inefficient from a social point of view. For example, if information arrives at time 1 that a project will have high time 2 payoffs and low time 3 payoffs, the outside investor may force liquidation at time 1, anticipating that it will not be possible to extract any payment at time 2. This type of inefficiency can be avoided with long-term debt.

The papers by Aghion and Bolton (1988), Zender (1989), and Hart and Moore (1989) provide rationales for the use of debt and equity by closely held firms. Their analyses raise at least two issues that remain to be fully resolved. The first is which results depend on risk neutrality and which are robust to the introduction of risk aversion. The second is that it is not immediately evident how this type of theory can be applied to justify the use of debt and equity by large corporations. The problem is how to identify the interests of managers with those of outside equity-holders, given the latter are in a similar position to outside bondholders. These are important topics for future research.

Allocation of Voting Rights

Another strand of the literature has considered the question of control in terms of the way in which voting rights should be assigned to securities. The aspect of equity that has been of particular concern is the use of one vote per share and majority voting as the decision rule. A number of papers have identified the circumstances where these provisions are optimal.

Grossman and Hart (1988) argue that the voting structure of securities is important primarily because of its impact on the market for corporate control. When securities are widely held, a free-rider problem

exists: individual shareholders do not have an incentive to carefully monitor management and vote them out when they perform badly. Monitoring of management is likely to be important when a single individual or group has a large enough ownership share to make the free-rider problem insignificant. A prime example of the type of situation where this occurs is the case of a takeover bid. Grossman and Hart therefore consider a model where the allocation of voting rights and dividends to securities is determined by its effect on allowing rivals to obtain control from an incumbent management.

Initially, the firm is owned by an entrepreneur who wishes to draw up a corporate charter that maximizes the value of the firm. Grossman and Hart are interested in schemes that are privately optimal for the entrepreneur. A number of different classes of shares can be created and the share of votes and the share of dividends accruing to each can be varied. The entrepreneur anticipates that these securities will be widely held and that the firm will be run by an incumbent management. At some date in the future, a rival team, which may or may not be able to manage the firm better than the incumbent team, may attempt to acquire control by bidding for the securities to which control rights are attached. The incumbent team makes a counteroffer and holders of the securities decide which offer to accept.

The critical assumption of the model is that management teams can obtain private benefits from controlling the firm; the optimal allocation of voting rights and dividends depends on the absolute and relative sizes of the private benefits accruing to the incumbent management team and the rival team. If private benefits are negligible, then the allocation of control is unimportant and one share, one vote is as good as any other allocation.

Grossman and Hart first consider the case where all securities of a particular class must be treated equally, so that the whole class must be purchased if the votes of that class are necessary for control. Suppose that the private benefits of control are one-sided; for example, suppose the incumbent team has no private benefits of control but the rival team does. In this case one share, one vote is optimal because it maximizes the amount the rival must pay to obtain control. If a firm has a voting structure that allows the rival to obtain control by buying securities with only a small proportion of dividends attached, then he can obtain control and the associated benefits it provides to him at a small price. This may even be worth doing when the rival cannot generate as high a dividend stream as the incumbent. In order to make sure the rival pays as much as possible for control and its associated private benefits, and in particular at least as much as the value of the dividend stream provided by the incumbent, votes must be spread as widely as possible. This

implies one share, one vote. A similar argument holds if the incumbent team has one-sided private benefits of control.

If private benefits are two-sided so that both teams value control, one share, one vote is no longer optimal. The reason is that by separating votes from dividends it is possible to get the incumbent and rival to compete for control and pay for the associated private benefits they obtain. Grossman and Hart argue that this case is of little interest empirically for large publicly owned corporations, since the extent to which management can extract benefits is limited by corporate law, which gives a corporation's directors a fiduciary duty to all shareholders. It then follows that their theory is consistent with the widespread use of one share, one vote among publicly owned corporations.

Finally, Grossman and Hart consider the case where it is not necessary to treat all holders of a particular class of securities equally; it is only necessary for the rival to obtain the proportion of votes specified in the charter to obtain control. This prespecified proportion is assumed to be between 50 and 100 percent. Ignoring the case where both incumbent and rival have private benefits of control for the reasons mentioned above, the analysis of the optimal proportion is similar to before. The main difference occurs when the incumbent has one-sided benefits of control. In this case, it is optimal to set the proportion at the lowest value of 50 percent, since this minimizes the chance of the incumbent team maintaining control. Their paper thus provides some rationale for the use of a single class of equity with control requiring a majority of the votes.

Harris and Raviv (1988a) also consider the optimal allocation of voting rights and dividends to securities. Although the details differ somewhat, the framework is similar. One of the main differences between the papers is in the focus of the analysis. Grossman and Hart consider arrangements that are privately optimal as far as the original entrepreneur who designs the charter is concerned; they do not consider a criterion of social optimality, which includes the private benefits accruing to the incumbent and rival management teams. In contrast, Harris and Raviv do explicitly distinguish between private and social optimality.

Harris and Raviv show that one-share, one-vote majority rule is socially optimal since it ensures that the management team that generates the greatest total amount (including payouts to shareholders and private benefits to managers) controls the firm. This is because the arrangement allows the team that can pay the most to gain control; any deviation gives an advantage to the incumbent or rival that may allow them to gain control even though they generate a lower total amount. The arrangement that is privately optimal for the original owner involves issuing two extreme classes of security, one with all the voting

rights and one with all the dividends. The reason this is optimal is that it allows the securityholders to extract as much of the benefits of control from the management teams as possible because it forces them to compete for them. Thus, in general, the rules that are privately and socially optimal are not the same.

Grossman and Hart obtain one-share, one-vote majority rule as privately optimal, whereas Harris and Raviv obtain issuing extreme securities as privately optimal, because of differences in their assumptions. Among other things, the two papers are concerned with different parameters for the benefits of control the incumbent and rival can capture. Grossman and Hart argue that the case where both have benefits of control is of little empirical interest, whereas Harris and Raviv do not make this restriction. If the private benefits for both incumbent and rival are high, concentrating votes among a small class of equity is optimal in the Grossman and Hart model.

In the cases where only the rival or only the incumbent obtains benefits of control, both one-share, one-vote majority rule and extreme securities are optimal arrangements in the Harris and Raviv model. Extreme securities are optimal in these circumstances in their model but not in Grossman and Hart's, because Harris and Raviv assume that each investor can construct an optimal portfolio containing both of the extreme securities and that each investor's tender decision can be pivotal. This means that investors take into account the effect of tendering their votes on the value of their nonvoting shares. In contrast, Grossman and Hart assume each investor ignores any effects his actions may have on the outcome of the tender.

These differences between the assumptions and results of the two papers raise a number of issues. The private optimality of firms issuing equity with one share, one vote apparently depends critically on the assumption that the private benefits of control of the incumbent and the rival are asymmetric. If both have significant benefits, then concentration of votes appears to be (privately) desirable. If this type of theory is to explain the predominance of one share, one vote, it is necessary to provide some theoretical or empirical justification for why asymmetric private benefits of control is a plausible assumption. A priori, one might expect private benefits would be symmetric, since the limitations on the amounts managers can capture are set by corporate law and other factors that are the same for both incumbents and rivals. The main private benefit that can differ is, perhaps, the psychic satisfaction of control. An important question empirically is, therefore, how much this does differ between incumbents and rivals. Another issue is the best way to model shareholders' decisions; in particular, in close contests do they in practice regard themselves as unimportant in influencing the outcome, or pivotal?

Blair, Golbe and Gerard (1989) consider a model similar to that of Harris and Raviv (1988a) in that they are concerned with social optimality and both the rival and incumbent have private benefits of control, but these authors obtain rather different results. They are able to show that, in the absence of taxes, one-share, one-vote majority rule and extreme securities that unbundle voting rights and cash flows are equivalent and both lead to social optimality. In contrast, Harris and Raviv show that only one-share, one-vote majority rule is socially optimal; extreme securities can lead to suboptimal outcomes. The reason for this difference is that Blair, Golbe and Gerard assume the rival and incumbent bid simultaneously, whereas Harris and Raviv assume they bid sequentially. Again, this difference in approaches and its effect on the results raises the question of which is the most appropriate way of modeling the situation.

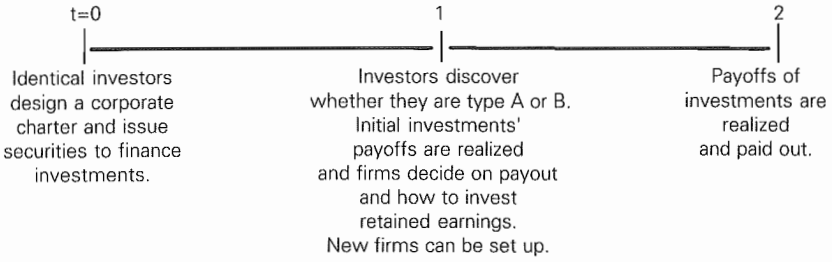
The main concern of Blair, Golbe and Gerard is to consider the effect of capital gains taxes on the allocation of voting rights and cash flows. If capital gains taxes are in effect, then welfare is improved if extreme securities are used. This is because a lock-in effect means capital gains taxes may prevent a superior rival from winning if there is one-share, one-vote majority rule; tax liabilities may be higher when the rival wins than when the incumbent wins. Allowing separate trading of votes alleviates this effect.

Taking the security structure of voting equity and debt as exogenous, Harris and Raviv (1988b) stress the importance of capital structure for takeover contests, because high leverage allows a controlling interest to be acquired for a low outlay. Harris and Raviv (1989) combine this idea with the approaches in Grossman and Hart (1988) and Harris and Raviv (1988a) by considering the allocation of voting rights and cash flows when the firm is not restricted to issuing just equity. They use a similar model to that of Grossman and Hart. In particular they focus on privately optimal securities, only the incumbent (or the rival) is assumed to have private benefits of control, and each investor ignores any effect his actions may have on the outcome of the tender.

The problem of the entrepreneur who owns the firm initially is to design securities that prevent the incumbent management that has private benefits from maintaining control when a superior rival appears. This means that the cost of resisting takeovers must be maximized. As in the papers focusing only on equity, one share, one vote among voting securities is an important component of this, since it means that control cannot be acquired cheaply by the party with private benefits. In addition, they show that nonvoting risky securities should not be sold to outside investors; if nonvoting securities are sold to outside investors, they should be risk-free debt. The reason is again that these maximize the cost of obtaining control and so tend to favor the superior rival.

Figure 4

The Sequence of Events in the Bagwell and Judd (1989) Model



The private optimality of one share, one vote in Harris and Raviv (1989) again appears to depend on the assumption of asymmetric benefits of control between the incumbent and rival. If both had private benefits of control, extreme securities of some sort might be optimal as in Harris and Raviv (1988a). An interesting issue is whether debt and equity remain optimal in this case.

The models to analyze the design of equity that have been considered above are all concerned with the effect of voting when an incumbent management team is challenged by a rival team. Bagwell and Judd (1989) take a different approach by considering the optimality of majority rule where control is concerned with payout and investment decisions.

The sequence of events in their model is shown in Figure 4. Initially all investors are identical; they design corporate charters and issue securities to finance firms' investments. At time 1 investors discover whether they are type A or B. Type As value consumption at time 1 and time 2 and require a minimum level of consumption at time 1. Type Bs only value consumption at time 2 and are less risk averse than type As at that time. Just after investors' types are discovered, firms decide on how much of the cash generated by the initial investment to pay out to shareholders and whether to invest the retained earnings in a safe or a risky project. If investors have any cash remaining at time 1 they can invest it in new firms. At time 2, the final payoffs from firms' investments are realized and paid out to shareholders.

A crucial feature of Bagwell and Judd's model is the existence of transaction costs for trading securities at time 1 after investors have discovered their type. The particular cost that is modeled is the capital gains tax. In the absence of this cost, investors would simply reallocate their portfolios. Type A investors would choose firms that pay out their required consumption at time 1 and invest in relatively safe projects and type B investors would choose firms that invest all their time 1 earnings in risky projects. When this type of rebalancing is prohibitively costly, each firm will have shareholders with different views about its optimal policy and control will be important. For example, suppose there is majority rule and type As are just in the majority. In this case they will prefer dividends to share repurchase even though the former strategy involves a higher tax burden, because this allows them to maintain control and implement the investment choice they prefer.

Bagwell and Judd show that the optimal decision rule in the corporate charter depends on the level of these transaction costs for rebalancing at time 1. For small transaction costs, majority rule is optimal because investors can rebalance at low cost and not much shareholder diversity is found among firms. However, for transaction costs that are so high that no rebalancing occurs, majority rule is not optimal. In this case the corporate charter should specify that the firm's policy is chosen to maximize a welfare function where the weights assigned to each type correspond to their representation in the firm at time 1. This maximizes investors' welfare initially since they only know the probability of being a particular type.

Bagwell and Judd's model illustrates that control may be important in situations other than takeovers. They focus on a particular situation of this type. One issue is in what other circumstances control matters. Another is how important empirically each of these possible scenarios is in influencing the design of corporate charters.

Overall, the papers considered in this section indicate circumstances do exist where debt and equity are optimal. However, these circumstances appear to be rather special relative to the wide set of circumstances in which debt and equity are used in practice. Thus the contribution of the literature to date is to provide some insights into why debt and equity are used, rather than a single comprehensive theory. The literature has also succeeded in identifying a number of important issues and has provided paradigms within which to consider these issues.

What Are the Optimal Securities?

As mentioned above, the circumstances so far identified where debt and equity are optimal are fairly restricted. In particular, most of the

papers mentioned require that the firm or its investors or both be risk neutral. Since it has traditionally been argued that one of the main roles of the stock market is to allow risk to be shared, this assumption is fairly restrictive. Moreover, the long history and extent of financial innovation suggest that firms' financing needs are not satisfied by debt and equity.

Rather than ask "What are the circumstances where debt and equity are optimal?" another branch of the literature has been concerned with the question "What are the optimal securities to issue?" The Modigliani and Miller result, that capital structure is irrelevant when markets are complete, suggests that the form of securities issued is also irrelevant in these circumstances. In order to develop a theory of optimal securities, it is necessary that markets be incomplete. One possible reason for incompleteness that is often suggested is transaction costs. Allen and Gale (1988; 1989) have considered the implications of the transaction costs of issuing securities.

Allen and Gale (1988) develop a simple model of financial innovation with two dates and a finite set of states of nature. Information is symmetric; the state is unknown to everybody at the first date and revealed to all at the second. A single good exists at both dates, along with a finite number of investor and firm types with a continuum of each type. Instead of assuming that firms are restricted to issuing debt and equity, however, Allen and Gale assume that firms choose the securities that they issue and this determines the transaction costs they incur. This means the market structure is endogenous and it is possible to consider the theoretical issues raised by financial innovation.

The equilibrium concept used is based on that of Hart (1979) and is essentially Walrasian. Markets are perfectly competitive since there is a continuum of firms and consumers. Prices are quoted to both firms and investors for every possible security. This includes all those securities that are issued in equilibrium as in Hart's model. It also includes all those securities that could be issued but in equilibrium are not (that is, demand and supply are both zero). This contrasts with Hart's approach where markets for these unissued securities are closed to investors and prices are quoted only to firms.

The first result obtained is that under standard assumptions equilibrium exists provided short sales are not possible. If securities can be costlessly sold short, then equilibrium may not exist because short-sellers are effectively able to expand the supply of firms' securities more cheaply than firms can. For example, suppose a firm can issue two securities rather than one for some additional cost. In order for the firm to be willing to do this, its gross value with two securities must be larger than with one to allow it to recoup this additional cost. However, if costless short sales are possible this implies an arbitrage opportunity is available, since by going short in a two-security firm and long in a

one-security firm, an investor can earn the difference between the two. An equilibrium where all firms issue one security may not be feasible because at the prevailing prices issuing two securities may be profitable. Thus equilibrium may not exist unless short sales are ruled out.

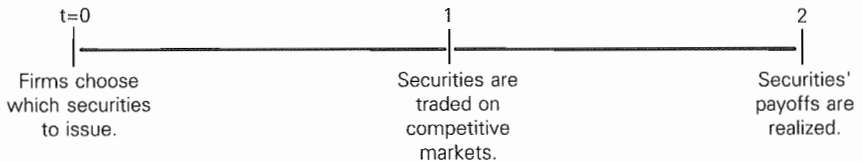
The short sales constraint means that with incomplete markets distinct types of investor value securities differently on the margin. The price of a security, whether issued or unissued, is determined by the group that values it most. In equilibrium, the firm issues the securities that maximize its value and sells them to the groups or clienteles that value them the most.

The second result obtained is that every equilibrium is constrained efficient. In other words, a planner subject to the same transaction costs for issuing securities and able to make transfers between investors at the first date cannot make everybody better off than in the market allocation. This result arises because of the assumption that the prices of unissued securities are quoted to both firms and investors. If prices are only quoted to firms, then inefficient equilibria may exist because of a pecuniary externality. To see this, suppose there are two types of firm, each of which produces output in one state only. Investors have Cobb-Douglas utility functions so that consumption in one state will not have value unless consumption is positive in the other. If markets for unissued securities are closed to investors, an equilibrium exists where the firms do not issue any securities because the price quoted to them for all securities is zero. This cannot be an equilibrium if prices are quoted to investors as well, because at zero prices they would demand securities that allow them to consume in both states.

A third result is that debt and equity are not optimal but that the optimal securities do have a particularly simple form. To see this, suppose there are two types of investor, one type of firm and two states. When firms issue only equity, the more risk-averse investors have a lower marginal utility of consumption in the high-output state than the less risk-averse investors; in the low-output state, the reverse is true. If a firm issues debt and levered equity, the more risk-averse group will pay a premium for the debt since it allows them to smooth consumption; the levered equity will be held by the less risk-averse group since they value consumption most in the high-output state. This split is not optimal, however, because the debt allocates payoffs in the good state to the more risk-averse group that values consumption the least. The firm could obtain more for its securities by allocating all the payoffs in the good state to the security that is held by the less risk-averse group, which values consumption most in this state. In general, it can be seen that optimal securities involve allocating all the firm's output in a particular state to the security held by the group that values consumption most in that state.

Figure 5

The Sequence of Events in the Allen and Gale (1989) Model



The critical assumption for all these results is the one ruling out short sales. In practice, short selling of corporate securities is costly and only a limited amount is undertaken (Pollack 1986). This suggests that in some circumstances it may be appropriate to rule out short sales. However, markets for stock options and index futures may represent a low-cost substitute for short sales. This suggests that the case of unlimited short sales is also of interest. In addition, the fact that unlimited short sales is a crucial assumption of many models in financial economics means this case is important theoretically.

Allen and Gale (1989) develop a model where unlimited short sales are possible. The main differences between this model and the one in Allen and Gale (1988) are that the number of agents is finite and the sequence of events is as shown in Figure 5. Firms first choose the securities to issue, these securities are then traded on competitive markets, and finally the securities' payoffs are realized. When choosing securities initially, firms play a noncooperative game; they take into account the effect of their actions on the equilibrium of the securities market at the next stage.

In contrast to the model of Allen and Gale (1988), firms are not price-takers; if a firm issues a new security it changes the security market equilibrium. Nevertheless it can be shown that if short sales are ruled out, then as the number of agents approaches infinity the equilibrium is essentially equivalent to that in Allen and Gale (1988); each firm's actions have a negligible impact on the equilibrium at the second stage.

If short sales are *not* ruled out, the equilibrium of the model may

differ significantly from that in Allen and Gale (1988). Even if the value of a two-security firm is the same as that of an identical one-security firm, so that no arbitrage opportunity exists in the second-stage equilibrium, a firm may nevertheless have an incentive to issue a costly security initially. A new security may increase the value of the firm in the second-stage equilibrium relative to the equilibrium that would occur if no innovation were made. Thus there can be an *ex ante* incentive to innovate even when there is no *ex post* incentive. This is true even as the number of agents approaches infinity. Now a single firm can affect the security-market equilibrium even though it is negligible, because the existence of short sales means that the open interest in the security may be large.

The fact that firms are no longer price-takers ensures that the existence of equilibrium is not a problem even though short sales are possible. However, the equilibrium is no longer constrained efficient. An example is given of too little innovation; the change in firm value across security-market equilibria is such that firms fail to issue a security even though everybody could be made better off if such a security were issued and appropriate initial transfers were made. An example is also given of too much innovation; in this case firms issue securities even though everybody could be made better off if fewer securities were issued. In the context of this model, therefore, the endogenous incomplete market structure that arises from profit-maximizing behavior is not necessarily efficient. Another aspect of this result is that the equilibrium with short sales ruled out may be superior to the equilibrium where short sales are not ruled out. For a given set of securities, allowing short sales improves possibilities for risk sharing. However, allowing short sales reduces the incentives to innovate, so overall risk-sharing opportunities may be reduced.

As far as the form of optimal securities is concerned, an example is given where debt and equity are optimal. This example is clearly a special case, however, and in general the optimal securities have a complex form that cannot be characterized simply.

This section has considered models of financial innovation where corporations issue securities. However, in addition to corporations a number of other types of institution such as futures and options exchanges issue securities. Duffie and Jackson (1989 and the references therein) consider innovation by futures exchanges; Allen and Gale (1990) consider innovation by options exchanges. The implications of incomplete markets for the design of government securities are considered in Gale (1989).

As with the literature on the optimality of debt and equity, the literature on optimal securities is still at a very early stage. The results in Allen and Gale (1988), showing that optimal securities involve allocating

all the firm's payoffs in a particular state to the security held by the group that values consumption the most, provide some insight into the option-like form of many new securities. However, the literature to date does not provide much insight into the actual path of most financial innovations. Its main contribution is again in identifying the theoretical issues and in providing models to analyze these issues.

Summary and Conclusions

The traditional approach to understanding firms' choice of capital structure has been to consider firms' optimal debt-equity ratios. This approach has not been very successful in terms of providing an understanding of the capital structures firms choose in practice. The introduction of many new securities in recent years suggests the alternative approach of considering the optimal form of securities that firms should issue. The literature based on this approach has been the subject of this paper.

The first branch of this literature has considered the circumstances in which debt and equity are optimal. A number of situations where debt is optimal have been identified. These typically involve a principal-agent relationship where an investor (the principal) lends money to an entrepreneur (the agent) to allow him to undertake an investment project. A debt contract is optimal in these models because it ensures that the entrepreneur takes a particular action. Although these theories are suggestive of why a public corporation may want to issue debt and equity, they cannot be directly applied in this case. Williams (1989) has extended this type of analysis to public corporations by assuming ex ante monitoring that prevents managers from expropriating firms' assets.

The assumptions of all these models are fairly restrictive. It is usually critical that either one or both parties is risk neutral and/or the earnings from the investment or actions of the entrepreneur are difficult for the outside investor to observe and so cannot be contracted upon. If earnings or anything else related to the management's performance can be observed at all, and the management is risk averse, the results of Holmstrom (1979) suggest that the optimal payments to the bondholder should be conditioned on this information. In practice, even though typically the parties are risk averse and some information on earnings is available, payments on debt contracts are fixed and do not vary with the available information. An exception is provided by income bonds but these are rarely used.

Another part of the literature has looked at the question of why public corporations typically have equity securities with one vote per

share and majority rule. Most of these papers are concerned with the effect of voting on the market for corporate control. Again, the circumstances where these results hold are rather special. Moreover, they critically depend on the magnitude of the private benefits of control and the distribution of these between incumbents and rivals.

Overall, the literature on the optimality of debt and equity suggests that the circumstances in which these commonly used securities are the best are fairly restrictive. This is difficult to reconcile with the fact that debt and equity are so widely used. However, the literature has identified a number of important issues and identified ways to think about these issues. A similar argument can be made concerning the literature on the form of optimal securities.

The results to date do suggest a number of important questions to be investigated in future research. Debt and equity have been used in numerous diverse situations. Why is it that they are so robust? What are the incentives for firms to issue securities other than debt and equity, and what are the general principles underlying the design of these securities? Finally, even though the securities that are issued may be optimal privately, the results of Harris and Raviv (1988a) and Allen and Gale (1989) suggest that no particular reason exists to believe that they are optimal from a social point of view. In other words, as far as the issue of securities is concerned, it is not immediately obvious that the "invisible hand" operates and ensures that market structure is efficient. A critical issue is, therefore, under what circumstances the market structure that arises is socially desirable and under what circumstances government intervention is justified.

The papers considered above all assume discrete time. The use of continuous time models to price derivative securities has not been discussed. As Hakansson (1979) has pointed out in the context of option securities, these models rely on the fact that dynamic trading strategies make markets effectively complete. This makes the analysis of financial innovation using continuous time techniques difficult. However, Merton (1989) has made progress in this direction by considering a world where individuals face transaction costs but intermediaries do not so that continuous time techniques can still be used. The relationship between financial innovation and dynamic trading strategies is an important topic for future research.

In conclusion, the theoretical literature has just begun to look at the question "What are the optimal securities for firms to issue?" Recent research has shed some light on the changing nature of debt and equity by identifying some of the important issues in this area.

References

- Aghion, Philippe and Patrick Bolton. 1988. "An 'Incomplete Contract' Approach to Bankruptcy and the Financial Structure of the Firm." Technical Report 536, IMSSS, Stanford University.
- Allen, Franklin. 1987. "Capital Structure and Imperfect Competition in Product Markets." Working paper 11-87, Rodney L. White Center, University of Pennsylvania.
- Allen, Franklin and Douglas Gale. 1988. "Optimal Security Design." *Review of Financial Studies*, vol. 1, pp. 229-263.
- . 1989. "Arbitrage, Short Sales and Financial Innovation." Working paper 10-89, Rodney L. White Center, University of Pennsylvania.
- . 1990. "Incomplete Markets and Incentives to Set Up an Options Exchange." *Geneva Papers on Risk and Insurance*, special issue on "The Allocation of Risk with Incomplete Asset Markets," edited by Herakles Polemarchakis, vol. 15, pp. 15-44.
- Bagwell, Laurie Simon and Kenneth L. Judd. 1989. "Transactions Costs and Corporate Control." Working paper 67, Kellogg Graduate School of Management, Northwestern University.
- Baldwin, Carliss Y. 1983a. "Productivity and Labor Unions: An Application of the Theory of Self-Enforcing Contracts." *Journal of Business*, vol. 56, pp. 155-185.
- . 1983b. "Innovation and the Vertical Structure of Industry." Working paper, Harvard Business School.
- . 1988. "Time Inconsistency in Capital Budgeting." Working paper, Harvard Business School.
- Blair, Douglas H., Devra L. Golbe and James M. Gerard. 1989. "Unbundling the Voting Rights and Profit Claims of Common Shares." *Journal of Political Economy*, vol. 97, pp. 420-443.
- Brennan, Michael and Alan Kraus. 1987. "Efficient Financing under Asymmetric Information." *Journal of Finance*, vol. 42, pp. 1225-1243.
- DeAngelo, Harry and Ronald W. Masulis. 1980. "Optimal Capital Structure under Corporate and Personal Taxation." *Journal of Financial Economics*, vol. 8, pp. 3-30.
- Dewing, Arthur S. 1934. *A Study of Corporation Securities: Their Nature and Uses in Finance*. Ronald Press Company: New York.
- Diamond, Douglas W. 1984. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies*, vol. 51, pp. 393-414.
- Duffie, Darrell and Matthew O. Jackson. 1989. "Optimal Innovation of Futures Contracts." *Review of Financial Studies*, vol. 2, pp. 275-296.
- Finnerty, John D. 1988. "Financial Engineering in Corporate Finance: An Overview." *Financial Management*, vol. 17, pp. 14-33.
- Gale, Douglas. 1989. "Notes on Efficient Public Debt." Working paper, University of Pittsburgh.
- Gale, Douglas and Martin Hellwig. 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem." *Review of Economic Studies*, vol. 52, pp. 647-663.
- Giammarino, Ronald M. 1989. "The Resolution of Financial Distress." *Review of Financial Studies*, vol. 2, pp. 25-47.
- Green, Richard C. 1984. "Investment Incentives, Debt and Warrants." *Journal of Financial Economics*, vol. 13, pp. 115-136.
- Grossman, Sanford J. and Oliver D. Hart. 1988. "One Share/One Vote and the Market for Corporate Control." *Journal of Financial Economics*, vol. 20, pp. 175-202.
- Hakansson, Nils H. 1979. "The Fantastic World of Finance: Progress and the Free Lunch." *Journal of Financial and Quantitative Analysis*, vol. 14, pp. 717-734.
- Harris, Milton and Artur Raviv. 1988a. "Corporate Governance: Voting Rights and Majority Rule." *Journal of Financial Economics*, vol. 20, pp. 203-235.
- . 1988b. "Corporate Control Contests and Capital Structure." *Journal of Financial Economics*, vol. 20, pp. 55-86.
- . 1989. "The Design of Securities." Working paper, Graduate School of Business, University of Chicago.
- Hart, Oliver D. 1979. "On Shareholder Unanimity in Large Stock Market Economies." *Econometrica*, vol. 47, pp. 1057-1084.

- Hart, Oliver D. and John Moore. 1989. "Default and Renegotiation: A Dynamic Model of Debt." Working paper 89-069, Harvard Business School.
- Haugen, Robert A. and Lemma W. Senbet. 1978. "The Insignificance of Bankruptcy Costs to the Theory of Optimal Capital Structure." *Journal of Finance*, vol. 33, pp. 383-393.
- Hellwig, Martin F. 1981. "Bankruptcy, Limited Liability and the Modigliani-Miller Theorem." *The American Economic Review*, vol. 71, pp. 155-170.
- Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics*, vol. 10, pp. 74-91.
- Jensen, Michael C. and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, vol. 3, pp. 305-360.
- Kanemasu, Hiromitsu, Robert H. Litzenberger and Jacques Rolfo. 1986. "Financial Innovation in an Incomplete Market: An Empirical Study of Stripped Government Securities." Working paper 26-86, Rodney L. White Center, University of Pennsylvania.
- Kim, E. Han. 1978. "A Mean-Variance Theory of Optimal Capital Structure and Optimal Debt Capacity." *Journal of Finance*, vol. 33, pp. 45-63.
- . 1989. "Optimal Capital Structure in Miller's Equilibrium." In *Financial Markets and Incomplete Information: Frontiers of Modern Financial Theory*, Volume 2, edited by Sudipto Bhattacharya and George M. Constantinides. Totowa, N.J.: Rowman and Littlefield.
- Malatesta, Paul H. and Ralph A. Walkling. 1988. "Poison Pill Securities: Stockholder Wealth, Profitability and Ownership Structure." *Journal of Financial Economics*, vol. 20, pp. 347-376.
- Merton, Robert C. 1989. "On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance." *Geneva Papers on Risk and Insurance*, vol. 14, pp. 225-261.
- Miller, Merton H. 1977. "Debt and Taxes." *Journal of Finance*, vol. 32, May, pp. 261-275.
- . 1986. "Financial Innovation: The Last Twenty Years and the Next." *Journal of Financial and Quantitative Analysis*, vol. 21, pp. 459-471.
- Modigliani, Franco and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *The American Economic Review*, vol. 48, pp. 261-297.
- . 1963. "Corporate Income Taxes and the Cost of Capital: A Correction." *The American Economic Review*, vol. 53, pp. 433-443.
- Mookherjee, Dilip and Ivan Png. 1989. "Optimal Auditing, Insurance and Redistribution." *Quarterly Journal of Economics*, vol. 104, pp. 399-416.
- Mooradian, Robert M. 1989. "Recapitalizations and the Free-Rider Problem: Bankruptcy and the Theory of Capital Structure." Working paper 89-3, Graduate School of Business, University of Florida.
- Myers, Stewart C. 1977. "Determinants of Corporate Borrowing." *Journal of Financial Economics*, vol. 5, November, pp. 147-175.
- . 1984. "The Capital Structure Puzzle." *Journal of Finance*, vol. 39, pp. 575-592.
- Myers, Stewart C. and Nicholas S. Majluf. 1984. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics*, vol. 13, pp. 187-221.
- Pollack, Irving M. 1986. "Short-Sale Regulation of NASDAQ Securities." NASDAQ pamphlet.
- Ravid, S. Abraham. 1988. "On Interactions of Production and Financial Decisions." *Financial Management*, vol. 17, pp. 87-99.
- Ross, Stephen A. 1977. "The Determination of Financial Structure: The Incentive Signaling Approach." *Bell Journal of Economics*, vol. 8, pp. 23-40.
- Titman, Sheridan. 1984. "The Effect of Capital Structure on a Firm's Liquidation Decision." *Journal of Financial Economics*, vol. 13, pp. 137-152.
- Townsend, Robert M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory*, vol. 22, pp. 265-293.
- Tufano, Peter. 1988. "Financial Innovation and First-Mover Advantages: An Empirical Analysis." Working paper, Harvard Business School.

- Van Horne, James C. 1985. "Of Financial Innovation and Excesses." *Journal of Finance*, vol. 40, pp. 621-631.
- Warner, Jerold B. 1977. "Bankruptcy Costs: Some Evidence." *Journal of Finance*, vol. 32, pp. 337-347.
- Webb, David C. 1987. "The Importance of Incomplete Information in Explaining the Existence of Costly Bankruptcy." *Economica*, vol. 54, pp. 279-288.
- Williams, Joseph. 1989. "Ex-Ante Monitoring, Ex-Post Asymmetry, and Optimal Securities." Working paper, University of British Columbia.
- Zender, Jaime F. 1989. "Optimal Financial Instruments." Working paper, University of Utah.

Discussion

*Oliver D. Hart**

Discussing a survey of the literature is never an easy task. In this case the task is made harder by the fact that (a) this survey is very good; (b) (perhaps because of this) I agree with a large proportion of what Franklin Allen says. In fact, probably the only disagreement we have concerns the assessment of the strengths and weaknesses of a recent body of work that tries to analyze the optimality of debt and equity arrangements. Allen believes that this work relies on rather restrictive assumptions—risk neutrality of management, unobservability of corporate earnings—and that current models can explain the existence of debt, but not of equity as well. In contrast, I believe that the results of these models are a bit more robust than Allen suggests. I will devote the rest of my discussion to explaining why I think this is so.

My main point can in fact be summarized very easily: it is that debt and equity are not just two out of a huge universe of potential financial instruments, with the puzzle being why firms select them, rather than the others, so much of the time. Rather I believe that debt and equity have some quite special features, which make their selection (or the selection of some variants) less surprising.

My starting point is the idea that in a world of transaction costs where individuals can contract at best imperfectly over the uncertain future, asset ownership matters. In particular, the owner of an asset has residual rights of control over that asset; that is, the right to decide how the asset should be used in any way not inconsistent with a prior contract, custom, or any law (Grossman and Hart 1986). For example,

*Professor of Economics, Massachusetts Institute of Technology.

the owner of a machine can typically decide who has access to it, the owner of a production line can decide at what speed it runs, and so forth.

As far as I can see, equity is just another name for ownership: the equityholders of a firm collectively have the right to determine how the firm's assets should be used—in particular who should manage them, or, to be more precise (through the choice of a board of directors), who should oversee their management. Ownership is like any other economic good: there is an efficient allocation of it. It is not difficult to see, for example, that if a single agent 1 takes actions that cause an asset A to be productive (agent 1 might learn how to operate it, say), then it will often be efficient for 1 to own A. In particular, if some other agent 2 owns A, then 2's ability to exclude 1 from access to A may diminish 1's return from his asset-specific activities, thus causing him to underinvest in these. On the other hand, if several agents take actions, each of which has a positive payoff in conjunction with A, then some form of shared ownership (for example, a partnership) may be optimal, since this will allow each agent to receive at least some return on his activities (Hart and Moore 1988).

Sometimes a single-ownership arrangement, even though it may be efficient, will not be feasible. For example, suppose agent 1 is the only agent who causes asset A to be productive—and therefore on efficiency grounds 1 should own A. It may be the case that 1 does not have the funds to buy A (asset A may initially be owned by somebody else, or A may not have been constructed yet). Under these conditions the best arrangement may be for agent 1 to borrow the funds necessary to buy A. What this means is that 1 owns and controls the asset unless he (or she) fails to repay his loan: in this event, the creditor has the right to seize the asset, that is, the residual rights of control shift to the creditor.¹ The advantage of such an arrangement is that agent 1 still has an incentive to take actions that increase the asset's productivity since he benefits from these in non-default states, but at the same time agent 1 is given a strong incentive to repay his loan. In particular, such an arrangement may be superior to one in which the creditor owns the asset from the beginning and employs agent 1 to work with it. (Under these conditions, agent 1 will on average get a smaller fraction of any productivity increase.)

The situation just described is one where a liquidity-constrained "manager" raises funds from a single investor-creditor. However, in many cases, several investors will finance the asset's purchase. The manager could again borrow from each of them. Now, however, it is

¹ This is under the assumption that the loan is secured by the asset.

necessary to specify more carefully what happens in a default state. Obviously, it is impossible for residual control rights over the asset to shift to each creditor. Instead, each creditor could be given the right to foreclose on some fraction of the asset (if it is divisible) or the right to sell the whole asset and divide up the proceeds on a pro rata basis (if it is not).

An alternative to borrowing from several investors would be for the manager to issue equity in the asset, that is, to give each investor control rights or votes in the asset from the beginning. Again a variety of arrangements is possible, depending on the power each equityholder has to act alone. In the case of a divisible asset, one could suppose that each equityholder has the right to liquidate his pro rata share of the whole asset at any time (open-ended mutual funds are set up in this way). In the case of an indivisible asset, however, it is more natural to suppose that some degree of consensus is required before the equityholders can act; for example, that it takes a majority to unseat management (or the board of directors). The latter is, of course, the standard practice in public corporations.

Many other arrangements are possible in addition to pure debt and pure equity. One obvious one is for the firm to issue combinations of debt and equity. Another is for the firm to issue various hybrid securities. For example, the firm could promise to hand investors either a fixed amount of money or a certain fraction of equity at a future date, with the choice being up to the firm (preferred shares are defined in this way); or the firm could issue debt that can be converted into equity at the investor's discretion (convertible debt); or it could issue options—that is, sell investors the right to purchase future equity in the firm.

Such arrangements all involve possibly contingent exchanges of equity (current or future) for cash (current or future). In this sense they can be regarded as variants of debt and equity; or to the extent that debt is itself just a particular type of contingent equity claim (the creditor gets either cash or an ownership claim in the future depending on whether the debtor defaults), as variants of equity itself.

So far I have spoken entirely about the votes or control rights that accrue to equity. Of course, each equityholder also has the right to his pro rata share of any dividends the firm pays. However, to the extent that dividends are at management's discretion—as they typically are—this right almost goes hand in hand with the residual right of control over assets, where assets now include any cash the firm disburses (over and above cash that has been committed to others).

Nothing that I have said so far about the special features of ownership, and various claims written on it, seems to depend on particular assumptions about manager or investor attitudes towards risk. Nor do I think the ideas would be much changed if some aspect of

the firm's performance, such as earnings or profits, could be verified. Under these conditions, the firm's payment to creditors could be chosen to be some function of performance. However, this simply gives rise to contingent debt; or, if default provides investors with the right to acquire equity rather than the right to foreclose on assets, to contingent preferred shares. Conceptually, I see little difference between contingent debt or preferred shares and their noncontingent counterparts.

In particular, it is worth stressing that the verifiability of earnings and profits does not eliminate the rationale for equity. It is true that with earnings and profits verifiable, management can be put on some sort of incentive scheme. Given that the firm has assets, however, it will still matter who has control over these. For example, regardless of the incentive scheme, a situation where outside investors are the equityholders and can fire management (or the board of directors), perhaps with compensation, and hire a new management team to operate the assets, is very different from a situation where management owns the equity and can veto any management change. Moreover, the incentive scheme itself can be thought of as part of the firm's financial structure; it is just a contingent debt from the firm to management (the firm owes management a sum of money as a function of the profit level realized).

Franklin Allen might accept much of the above, but argue that even though the theory survives the extension to verifiable profits or earnings, the real question is why we do not see in practice the types of contingent debt or equity that such an extension suggests we should. This is similar to the question of why (arguably) we do not see the highly nonlinear incentive schemes that principal-agent theory predicts, and I don't have a good answer to it.² I can, however, give *an* answer. It may be the case that those variables that would impart useful information about managerial performance—and on which we would like financial structure to be contingent—are costly to observe and verify, while those variables that can be verified more easily are not that informative. Earnings or "profits" seem to be examples of relatively uninformative variables since they are open to a good deal of manipulation by management. Allen argues in his conclusion that, by a result of Holmstrom, even a variable that is only slightly informative should (almost always) be included in an optimal incentive scheme or as part of an optimal financial structure. However, as far as I know, this is true only if management's actions are unidimensional, which is not the case if management can, say, divide its time between leisure, making profit, and manipulating the accounts. Under these conditions, making the

² Not all principal-agent models predict complex schemes, however. See, for example, Holmstrom and Milgrom (1987).

manager's salary (or financial structure) sensitive to earnings may be positively undesirable since it may cause him to devote too much time to manipulation. In addition, Holmstrom's result is no longer valid if observation of additional variables is costly. Thus, it may not be so surprising after all that most debt is not contingent.

In conclusion, I believe that recent work on optimal financial structure has more robust foundations than might be apparent at first sight. We are still a long way from a theory that can explain the simultaneous use of debt, equity, preferred shares, and the like, but I believe that the control rights approach does offer some hope for the development of such a theory in the future.

References

- Grossman, Sanford J. and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, vol. 94, no. 4, pp. 691-719.
- Hart, Oliver D. and John Moore. 1988. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, forthcoming.
- Holmstrom, Bengt and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica*, vol. 55, no. 2, pp. 303-328.

Discussion

*Robert C. Merton**

I agree with Oliver Hart that Franklin Allen's survey article covers the literature rather well. Hart, as I hoped he would, has discussed in detail some of the specific theoretical papers to which Allen refers. I would, however, briefly mention two relevant theories that were not explicitly discussed by either Hart or Allen. One is the work of Oliver Williamson on so-called transaction cost economics, as perhaps an alternative to agency theory, wherein the unit of analysis is the "transaction" rather than the "firm." The other is some work that my colleague Carliss Baldwin has done over the years, which I (not she, perhaps) would characterize as the "strategic capital" theory. This second theory may be illustrated by two examples. First, a dominant firm in an industry may not have as much debt as, say, tax theory would suggest, because by having more equity capital, the firm can credibly threaten to cut prices and suffer losses in order to deter competitors. Second, consider a firm bargaining with another strong entity such as a union. By carrying substantial debt, the firm can reinforce its bargaining position by threatening to shut down, should the union insist on too great a wage settlement or on employment contracts that, in conjunction with the debt, are too burdensome. Perhaps names of some companies will come to mind that, in the past year or so, have assumed considerable debt, suffered huge losses, and now appear to be using the threat of bankruptcy to renegotiate employee labor contracts.

Before discussing the debt-equity question, I would like to reinforce

*George Fisher Baker Professor of Business Administration, Harvard University Graduate School of Business Administration.

Allen's observation that many of the financial innovations of the last two decades, although heralded as novel, were not entirely new. Consider, for example, exchange-traded stock options and futures contracts. In the seventeenth century, options and contracts resembling futures accounted for the bulk of transactions on the Amsterdam stock exchange, which at the time was the financial center of the western world. Moreover, from the accounts given by Joseph de la Vega (1688), it appears that the concerns raised about these contracts in Amsterdam at that time (for example, insider trading, manipulation, excessive speculation and price volatility, and default risks) are much the same as those expressed about options and futures trading today. It is perhaps not surprising that as we resurrect ancient financial instruments, we revisit ancient financial problems.

My contribution to this session is essentially a negative report. I would suggest that promising explanations of why nonfinancial business firms issue the variety and mix of debt and equity instruments that they do are probably not to be found on the "demand-side" for these securities. That is, other than for financial intermediaries, the firm's choice of capital structure should not rest on the tastes of investors and theories of investor risk-sharing. As I have discussed elsewhere (Merton 1989; 1990, Chapters 14–16), finance theory in the context of well-developed financial markets would largely rule out nonfinancial business firms issuing a variety of financial instruments solely to satisfy the risk-sharing needs of investors. Financial intermediaries and specialized "zero-supply" markets (for example, futures and options) can meet these needs more efficiently. Thus, such a priori reasoning implies that the issuance of multiple types of financial instruments by business firms is not driven by the demand side of the market for these securities. Hence, one should look to the supply side (for example, corporate tax, regulation, and agency issues) for explanations of this issuing activity.

In further elaboration on this point, business firms need not issue instruments other than equity in order for investors to achieve an efficient blend of risk and returns in their portfolios. For example, if a business firm issues only shares, a separate entity like a mutual fund, a financial intermediary, or a holding company could acquire these shares as assets and finance their acquisition by offering investors various claims with payoffs contingent on these shares. Such partitioning of the payoffs could, for example, emulate the payoffs to "junk" bonds. Collateralized mortgage obligations and collateralized bond obligations are examples of such intermediation.

By using intermediation in this fashion to separate a business's operations from its financial structure, the system permits investors to acquire the instruments they demand while insulating the operations of the businesses from any defaults on these instruments. This separation

essentially permits "no-fault default." For instance, suppose an intermediary issues two securities against the shares of a firm. The first promises to pay a fixed amount in, say, two years; the second is a residual claim against the value of the shares. In the case of a default, the intermediary turns over the assets of the firm to the holders of the first type of security. Thus, the first security has essentially the same structure as junk debt, but without the prospect of bankruptcy costs. This arrangement allows for all the demand-driven risk-sharing entailed by junk debt, but in the case of a default, the company's operating management is not influenced a whit. I would suspect that most policy-makers who are worried about default, especially on junk debt, are not worried about some investors failing to be paid in full; what they are really worried about is a failure of the financial system and broad disruption of the real sectors of the economy, due to the effects of widespread default on business operations.

The separation of financial structure from business operations is not just hypothetical. This separation already functionally exists in the options market. The expiration unexercised of a call option on a stock is equivalent to a no-fault default. I have not seen these option expirations trigger any response from managements of the underlying firms or anyone else. Another example of this separation is the Americus Trust structure, which funds a trust with shares of a company's stock (for example, DuPont) and finances the share purchases by issuing two types of securities called "primes" and "scores." The prime is entitled to all dividends on the stock and a fixed payment at maturity of the trust. The score receives the residual value, if any, at maturity of the trust. The prime security is functionally equivalent to a junk bond on a highly levered firm with operations the same as the underlying stock.

Investors' behavior also suggests that it is very unlikely that theories of risk-sharing are crucial for explaining the financial strategies and capital structures of businesses. I think that the evidence shows that given a choice, investors are really more concerned with controlling the risk of their portfolios than with controlling the risk of the individual securities that constitute their portfolios. Option contracts on individual stocks, for example, initially were traded very actively, but after options on portfolios of stocks were introduced, the volume and use of individual options dropped dramatically.

Not only does the separation of the business operations from its financial structure make possible no-fault default, it also reduces the moral hazards that arise when management is beholden to only a subset of the firm's investors (that is, its stockholders as distinct from its creditors). Management, when it is disconnected from the firm's financial structure, has no particular reason to choose a dividend policy or any other financing or investment strategy that benefits one group of the

firm's liability holders, equityholders, while injuring another group of its liability holders, such as the firm's creditors. Consequently, the separation of the business from its financial instruments seems to be compelling in almost every respect if the only role for such instruments is the efficient sharing of risk among investors.

I believe, however, that an adequate theory of business behavior must consider liabilities that do not appear on business balance sheets as well as those formal financial liabilities that do. Accordingly, not all liabilities can be managed by an intermediary distinct from the business. For example, a firm's obligations to customers, suppliers, and employees entail risk-sharing. If customers have long-term, unfulfilled contracts with a company, these contracts become liabilities just like debt. I would suspect, however, that this type of liability would be less significant for companies with relatively generic and nondurable products. If you buy Kleenex tissues and Kimberly-Clark goes out of business, you are not worried that you will fail to obtain facial tissues elsewhere. But if you invest heavily in software that works only on one kind of computer, then discontinuance of warranty and hardware support for that computer can be very costly.

Finance theory already provides powerful tools for analyzing the complex financial instruments to which Allen and Hart allude. For example, the tool of contingent claims analysis (see Merton 1990, Chapters 11–14) permits one to convert each of the complex securities issued by firms into an equivalent portfolio of unlevered equity and short-term, default-free debt. By making this conversion for each security, one can thereby compare the risk profiles and functional roles of seemingly quite different types of financial instruments.

Theory can also help us distinguish important policy issues from those that are not. For example, most observers would agree that the levering of corporate America during the past decade has caused both corporate debt and equity returns to become more volatile. This observation has often been used to conclude that such levering has increased the riskiness of American firms. Theory tells us that the riskiness of both debt and equity increases with a corporation's leverage, but paradoxically the riskiness of the whole firm's value (that is, the sum of debt plus equity) may not change at all and may even decline. Consequently, when the returns on equities and debt are described as being more volatile than before, one cannot infer anything about firms as a whole being more risky. Similarly, one can ask whether institutional investor portfolios of debt and equity have become more risky. Debt and equity returns, taken separately, are almost surely more risky today than in the past. But, debt returns today are less volatile than equity returns in the past. And, institutional investors now hold more corporate debt relative to equity than they did. Thus, it is entirely possible that their total

equity-cum-corporate-debt portfolios are no more risky today than they were five or ten years ago. Essentially, these large investors are, in total, holding the same underlying assets that they did before, but their separate financial claims are just packaged differently.

In closing, a word on future empirical work: During the past generation, finance has relied on large sample-size, generic data bases (for example, stock and bond market prices) as the core for testing hypotheses. However, I think that in the future many of the issues surrounding business financial behavior and risk sharing will be resolved using smaller sample-size, specialized data bases gathered through good field work. For example, as we all know, there is a signalling theory for explaining dividend policies of firms. Although stock market data support this theory to a degree, these data support alternative theories as well. We might resolve this issue in large part by interviewing one hundred or more boards of directors to discover how much time they spend on setting the dividend and what factors they considered in making their decisions. More generally, such field work might help distinguish among the alternative theories of optimal security design discussed in Allen's paper.

References

- Merton, Robert C. 1989. "On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance." *The Geneva Papers on Risk and Insurance*, vol. 14, July, pp. 225-262.
- . 1990. *Continuous-Time Finance*. Oxford: Basil Blackwell.
- Vega, de la, J.P. 1688. *Confusión de Confusiones*. English translation by H. Kellenbenz, 1957. Publication No. 13, The Kress Library Series of Publications, The Kress Library of Business and Economics, Harvard University. Cambridge, Massachusetts: Harvard University Press.