Stephen A. Ross
Sterling Professor of Economics and Finance
President of the American Finance Association
1988

THE JOURNAL OF FINANCE • VOL. XLIV, NO. 3 • JULY 1989

Institutional Markets, Financial Marketing, and Financial Innovation

STEPHEN A. ROSS*

ABSTRACT

Firms and institutions are monitored and controlled through a complex set of implicit and explicit contractual relations. Because of these agency theoretic relations, institutional behavior in financial markets is not a simple reflection of the preference structures of individuals. Institutional preferences give rise to a demand for new financial instruments and innovations, even when the returns on these instruments are "spanned" in the sense of complete pricing. The innovations can be thought of as solving moral hazard problems. An agency theoretic example serves to illustrate the demand, supply, and financial marketing of stripped securities. In short, institutions matter.

THE PAST TWO DECADES have witnessed an explosion of financial innovations. The changes are too varied and diffuse to permit an easy taxonomy, but, broadly, they seem to fall into two classes. First, there are the new securities and their markets. Beginning with puts and calls we have moved on to Euromarkets, mortgage pass throughs and CMO's, strips, straps and swaps, putable convertibles, primes and scores and S&P futures, and all of the other instruments that inhibit the financial zoo. Second, there are the dynamic trading strategies that make heavy use of the new instruments. Basket and program trading, the asset pricing theories that now characterize different types of money management, and the various incarnations of portfolio insurance have changed financial decision making as dramatically as the available array of instruments has changed.[1]

Along with these changes has come an enormous increase in the demand for the services of financial economists. We are called upon not only to value the new instruments and the new strategies but to design them as well. Like engineers who use physics, financial engineers use the techniques of modern finance to build the financial equivalent of bridges and airplanes. Unfortunately, though, when asked to explain why all of this is happening and why it takes the forms that it does, our answers are much more timid. Not uncommonly we fall back on old canards such as "spanning." Why do we have so many new fixed income instruments? The most common explanation is that markets are incomplete and that these instruments permit us to complete them. Indeed, in the original reports to the SEC that attended the setting up of the CBOT options markets, the social value of options was largely described in such terms.

* Yale School of Organization and Management. I am grateful to Joel Demski, Phil Dybvig, Bengt Holmstrom, Jon Ingersoll, Offer Kella, and Richard Roll for helpful comments. Any errors are my own.

[1] See Miller (1986) for a more exhaustive list of financial innovations as well as an exciting analysis of the recent history of financial innovation. An interesting database of financial innovations is provided by Tufano (1988).

This is an awkward explanation and points to an embarrassing gap in the science. After all, our success at valuing these instruments comes entirely from treating them as derivative securities that are already spanned by the underlying securities on which they are written, stocks and a riskless asset in the case of a simple option on a stock. While it is trivially true that markets can never be fully complete and that, as a consequence, any option contributes a new security in some exact, nearly tautological sense, pricing the new security within a few percent of its trading value never seems to be a problem. In effect, markets are "price complete." What, then, is the role of the new innovations in such essentially complete markets?

There seem to me to be two dominant features at work in prompting the wave of financial innovation. First, our financial markets have become institutional markets in that institutions are often the most significant traders and holders of active positions. Financial institutions range from transparent through translucent to opaque. Most mutual funds are transparent institutions. Insurance companies and savings and loan organizations are opaque. In between are translucent institutions such as pension funds. Transparent institutions are pass throughs that trivially and rather directly reflect retail forces. Opaque institutions are clouded and veiled from the view of participants and, at best, are only weak reflections of those forces.

The unconstrained individual is the retail player of our models. When impacted by taxes and constrained by regulations and agency considerations that make borrowing costly or limited, the individual becomes our simplest and most transparent institution. Only slightly less transparent, many open end mutual funds are really just efficient vehicles for individuals to hold securities. The control and agency problems seem quite manageable and relatively innocuous. At any moment of time the participants in the fund can be fully informed about the assets of the fund. There is no substantial equity investment in the fund by those who manage it, and their principal compensation is a percentage of the assets under management. While there is no limit to human ingenuity and while there certainly have been conflicts and even lawsuits dealing with such institutions, relative to the more opaque ones there is far less room for a divergence in interests between management and the participants.

By contrast, opaque institutions are those for which monitoring, bonding, and control costs, i.e., agency problems, are most severe. Information is asymmetric between the participants and the institution, and the need for control and contractual structures to permit them to function is so critical that they essentially define these institutions. A depositor in an S&L has almost no knowledge of the particular loans the institution is making. Are they local or national, secured or not? Little of this is visible to the participant. The same is true for the insurance company that invests premiums prior to paying claims, and the situation is even worse for the manufacturing companies whose financial activities are sidelines to their main businesses.[2]

[2] The monitoring and bonding requirements of institutions are often so well established that they become part of a regulatory environment through the workings of political economy. For our purposes it doesn't matter whether monitoring and agency restrictions arise from the natural needs of the market relations between institutions and participants or whether they are imposed exogenously by

It is by no means clear that the demand and supply for financial assets by opaque institutions simply reflect retail forces. In the prevailing equilibrium models of securities markets, demand comes from the individuals solving portfolio optimization problems. However, when we take account of the intervening contractual relations under which opaque institutions operate, it seems heroic to think that they mirror these forces. Furthermore, given the significance of opaque institutions in the market, it seems unlikely that retail forces and individual portfolio optimization will be dominant in price formation. The only hope for this position is that somehow or other these institutions in aggregate will act like individuals.

Of course, to the extent that credit restrictions and the like prevent individuals from behaving as the unconstrained players of the simple optimization models and to the extent that regulations restrict their actions, they, too, can be a focus for a different description of supply and demand at even the retail level. In this case, though, it is often large institutions that force innovation. For example, the options markets began as retail markets. Institutions saw options trading as a way to develop retail interest where the alternative was to try to solve the agency (and regulatory) problems associated with extending credit to retail clients. Instead of having a client take a beta of 2 position in a beta 1 stock by leveraging, options provided a limited liability approach.[3]

To view the financial markets as institutional markets is not to say that they are solely markets where institutions operate but, rather, that institutions are major and significant forces. To summarize, institutions are financial market players whose activities are governed by agency relations. Institutional markets are markets where such players are important. The second element in the analysis of financial innovation is the role of marketing.

If we think of markets as the perfect and frictionless markets of neoclassical finance, then there is really no role for marketing per se. Selling a financial instrument in a frictionless market is, by definition, costless. However, of course, in practice the less familiar and the more esoteric the instrument, the more costly it is to sell it. Here is where the religious view of spanning gives way to a more practical perspective. Assuming that the states of nature are exogenously specified, a security is defined by its payoffs in the different states. To the extent that any uncertainty remains about such payoffs, new states are generated and they are not yet spanned. However, how important are such states empirically? Such uncertainty may be diversifiable or "nearly" spanned. Most interestingly, it is the role of marketing to leave as little nonspanned uncertainty as is efficient. In the parlance of economics, marketing "explains" the payoffs to an exactitude

---

the government. Although, presumably in the latter case an argument can be made that regulation serves as an efficient device for establishing contracts that the private sector would have written anyway or of solving free rider problems in contract design and enforcement that would be too costly or incentive incompatible for the private sector to implement.

[3] Similarly, government intervention brings other constraints on activities that cause deviations from unconstrained maximization. Taxes, for example, as Miller (1986) has argued, are a source of much of the motivation for financial innovation. To the extent to which the government induces payment, taxes will introduce agency-like constraints. We will not deal explictly with taxes and Miller is certainly correct that much financial innovation is the effort to deal with an ever changing tax code, but what we say will be consistent with such tax induced innovation.

where the marginal cost of further explanation equals the marginal benefit from the proposed transaction.

This approach retains the view of markets as complete enough to permit new instruments to be priced with considerable precision. However, when a new instrument or a new financial strategy is introduced, it recognizes that there is a cost to marketing it. If the new instrument is successful and long lived it can go through a life cycle from innovation and initial marketing to establishment and, ultimately, to displacement. In the beginning it is hand tailored to the needs of particular institutions or retail clients. In its mature phase, marketing costs diminish and the instrument becomes a standard commodity with only a slight bid/ask spread and a tad of market impact to deter from its movement in an efficient and nearly frictionless markets.

This paper combines these familiar pieces of analysis, agency theory and marketing, into a less familiar construct, but one within which financial innovation is natural. Demand and supply is a function of players whose actions are constrained by agency type considerations and, explicitly, by contracts and regulations that reflect such considerations. (These players could be individuals subject, for example, to credit restrictions.) Within this model, the marketing of new instruments is governed by the cost structure of the marketing networks and distribution channels.

The resulting analysis complements the work of both Miller (1986) and Merton (1988). Miller makes a compelling argument for the centrality of taxes and regulations in the recent wave of innovation. Merton provides a detailed examination of the production function underlying the creation of new derivative securities, paying careful attention to the role of transactions costs as the raison d'être for the existence of financial intermediaries. By contrast with Miller, we consider taxes and regulations as part of a larger framework within which institutions are controlled and monitored. Since we ignore production costs and focus instead on the role of marketing costs, the paper provides a converse to Merton's work.

In our analysis, financial innovation arises as a natural adjunct of the supply and demand of agency constrained participants, and marketing costs help to shape the form of the new institutional features. From this perspective, a major role of investment banks is to run the marketing networks. Section I outlines a very simple agency model which is consistent with the above description of institutions. Section II outlines a simple marketing network. Here we derive the cost structure that characterizes financial marketing. Section III puts the agency and marketing models together to examine a particular financial innovation, fixed income pooling and the stripping of the components. The CMO (Collateralized Mortgage Obligation), in which a pool of mortgages is formed and then stripped into bonds and a residual component, and Treasury bond strips are canonical examples of this kind of innovation. The final section briefly concludes the paper with a summary and some possibilities for research.

I. Institutions and Institutional Markets

Consider a stylized model of an opaque financial institution, perhaps an S&L or an insurance company. The institution is assumed to serve its stockholers who

have contributed a capital infusion of S and who wish to maximize the value of their holdings, i.e., the total wealth of the institution minus the amount that will be paid to the participants (depositors or claimants). We will ignore any possible divergence between the desires of the stockholers and those of the incumbent management. For simplicity, we assume that the world lasts for one period. Letting W_T denote the terminal value of the institution, the goal of the stockholders is to

$$\max E\{\rho[W_T - B_T]\},$$

where B_T is the claim that will be paid to the participants and ρ is the state price density, i.e., the state price per unit of probability. (In this approach to pricing the value, $E(\rho x)$, of any security with payoffs of x is just the sum of the state payoffs of x weighted by the prices of pure contingent claims. See Dybvig and Ross (1987) for a descrpition of this approach to pricing.)

Letting B denote the contribution of the participants, in a complete market the institution solves the following agency problem:

$$\max E\{\rho[W_T - B(W_T)]\}$$

subject to

$$E\{\rho W_T\} = B + S$$

and

$$E\{U(BW_T))\} \geq U(Br),$$

where $U(\cdot)$ is the participants' representative utility function and r is the riskless interest factor (one plus the interest rate).[4]

Since the market is complete, the first constraint permits the institution to obtain any pattern of contingent payoffs in their budget set. The second constraint is the familiar one that the participants be at least as well off as they could be on their own.

In the familiar case where $\rho = 1/r$ the objective function is risk neutral and the solution is to fully insure the participants and set

$$B(x) = Br.$$

In the more general problem we have set up, the solution is to set the marginal rates of substitution for the participants across the states of nature equal to the ratios of the state space prices. If the participants are sufficiently risk averse, though, and if the state space densities are sufficiently close to a constant function, the solution will not differ much from the risk-neutral case, and we will take the risk-neutral promise as the constraint on the institution.

There are at least two important difficulties with this analysis, though. First, even if we assume that the investments must be limited liability, then there is no assurance that the terminal wealth, w_T, will be sufficient to pay the promised commitment, Br, in all states of nature. Second, and more important, the promise

[4] For a description of the basic agency problem, see Ross (1973) or Holmstrom (1979). For a study of such problems in complete market settings see Dybvig and Spatt (1985). Demski, Sappington, and Spiller (1988) study the agency problem with limited liability constraints.

itself is not incentive compatible, which is simply to say that, having obtained the funds from the participants, the equity holders in the institution have an incentive to take on risky investments whose shortfall under Br is borne by the participants.

This is, of course, just a version of the familiar conflict between stockholders and bondholers in corporate finance. The more opaque the institution, the more costly it is to monitor and the more difficult it is to resolve the conflict. To the extent to which the participants have competitive alternatives, the institution must bear the monitoring costs and set up a system that will ensure that the participants receive their commitment. The solution to the problem, under the assumption that the commitment must be honored, provides some interesting clues as to the form this monitoring must take.

To further simplify the analysis we will assume that there are only two possible states of nature, state 1 and state 2. Figure 1 illustrates the analysis. Analytically, the problem is to choose some point—perhaps reflecting the particulars of the cost structure of the institution—on the budget line,

$$\pi_1 \rho_1 W_1 + \pi_2 \rho_2 W_2 = B + S$$

(π_1 and π_2 are the state probabilities) and satisfying

$$\min\{W_1, W_2\} \geq Br.$$

If the firm can be perfectly monitored, then the participants can commit it to accept, say, the point A. However, if the firm can be completely and costlessly monitored, it is not at all clear what function it serves in this complete market. An opaque firm generally has access to risky investments which the participants cannot even observe, let alone analyze. We will refer to such investments as low grade or "L" for short. By contrast, the firm can also invest in high grade

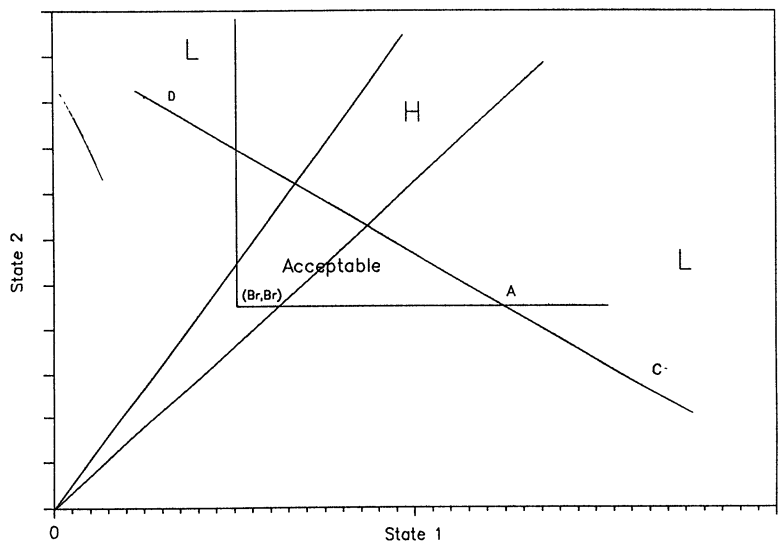

Figure 1. Institutional choice sets.

securities, "H" for short, which are quite visible and trade in competitive securities markets. We will also assume that, perhaps through a third party auditor or monitor, the net amount invested in such securities can be observed and controlled. The two ranges of securities types are, rather arbitrarily, divided into two cones in Figure 1.

Notice that points in L in Figure 1 can be decomposed into limited liability investments in both the H and the L regions. For example, point A, which is in L but is just to the right of the point (Br, Br) and is acceptable to the participants, can be decomposed into

$$\begin{aligned} A &= (A_1, A_2) \\ &= (A_2, A_2) + (A_1 - A_2, 0) \\ &= (Br, Br) + (A_1 - Br, 0), \end{aligned}$$

i.e., a riskless asset and a low grade asset.

This decomposition suggests several ways in which such an acceptable choice could be enforced. For example, a reserve requirement of B invested in H (actually somewhat more if risky assets in H are available) would assure that, whatever was done with the remaining assets, the resulting W_T would exceed Br in all states. Equivalently, there could be an upper bound on the amount invested in low grade assets or a lower bound on the amount invested in high grade assets. Government regulation is an alternative to an outside private sector monitoring agent or to the participants' monitoring this themselves. Government regulation is pervasive for opaque financial institutions, and the lack of incentive compatibility is the most straightforward rationale for such regulation; without it the private sector would be doing the same thing (which is not to take a stand for or against such regulation but simply to point out that the issue is less one of constraints on private sector actions and more one of the most cost-efficient approach to contract writing and enforcement).

The right triangle lying between the ordinates with origin at (Br, Br) and the budget constraint will be called the acceptable set (see Figure 1). From a point such as C the institution must move into the acceptable set to stay in business, and, without being specific about the loss function, we will assume that the firm will be willing to move along the budget line, perhaps with a price concession of ε so as to move into the acceptable set. (This would be illustrated by a parallel budget line lying ε below the line in Figure 1.)

It is interesting to go a bit further with our analysis. After all, there is little here to justify the existence of the institution since there is nothing preventing individuals from investing directly in the market. Suppose, though, that the institution can make investments in a class of low grade, high return (value) investments and suppose, too, that they have some monopoly power over the participants and can acquire funds from them at a rate $r_B < r$.

With these conditions, the firm seeks

$$\max\ W_H + W_L v_L - B(r_B/r)$$

subject to

$$W_H + W_L = B + S$$

and

$$W_H r + W_L m \geq B r_B,$$

where W_H and W_L are the respective investments in high and low grade assets, v_L is the value of the low grade investments per dollar invested ($v_L > 1$), and m is the minimum gross value in the worst possible state of the allowable class of low grade investments. (We are ignoring the competitive feasibility condition that the value of the stockholders' claim must at least equal their initial investment, S).

To prevent the firm from growing arbitrarily we also assume that there is an upper limit both to B, the amount contributed by participants, and to the supply of high return low grade projects. Alternatively, we could make the usual neoclassical assumptions of a rising supply curve of participant capital and a diminishing schedule of net present value projects.

For any choice of wealth, we can restate the agency constraint defining the acceptable region as a limitation on the proportion of firm value placed in low grade assets, ω_L,

$$\omega_L \leq [r - b r_B]/[r - m],$$

where b is the proportion of value contributed by the participants, B/W. Equivalently, this translates into a lower limit on the proportion invested in high grade assets,

$$\omega_H \geq [b r_B - m]/[r - m].$$

At any moment of time the institution can find itself inside or outside of its acceptable region. Firms at points like C or D wish to sell off low grade assets to reenter the acceptable region. Of course, in this simple example, C and D could just swap to become acceptable. Absent this happy coincidence, though, finding desirable sales and purchases is the marketing problem of the next section. If there were no constraints on the participants' contributions, B, or on the supply of positive (bounded above zero) net present value projects, the firm could move into the acceptable region by an infusion of equity capital, S, but moral hazard considerations might rule this out. We will assume that the institutions must sell off some low grade assets to move into the acceptable region. Since these assets are, to some degree, unique to the firm they are probably neither well or easily understood by the market. This characteristic will play an important role in our marketing analysis.

The above analysis could also be used to characterize the demand and supply for options by individuals. To the extent to which individuals are not permitted to borrow against security holdings, there will be a range of permissible holdings. For example, endowed with a risky security with payoffs at point A on the budget line, and in a complete market but one in which only the risky security and a riskless bond are traded, the segment of the budget line between point A and the risk-free point on the budget line is the set of holdings available to the individual without borrowing (or short selling). To achieve an equity exposure beyond the point A, the individual must borrow funds, but doing so implies the need to monitor the uses of those funds and to assure that they are, in this example, in

fact invested in the stock. The option is a cheaper way of accomplishing this since it will lie to the right of A and is a limited liability instrument. The option reduces the cost of monitoring the individual compared with the neoclassically arbitrage equivalent transaction of borrowing and purchasing the security, and it also reduces the cost of adjusting the position over time in an intertemporal setting. Once the option is in place it has a value set by arbitrage, but that does not mean that the innovation was not an improvement in efficiency, nor does it preclude slight deviations in ex post values from the hypothetical arbitrage prices prior to its introduction.

To the extent to which an institution "packages" the whole trade for the individual, both arranging for the loan and purchasing the security and adjusting the position over time, the institution has, effectively, created the option. Doing so obviates the need to check on the individual collateral since the position as a whole has limited liability. The question of why this doesn't occur but has, instead, been supplanted by an options market is more complex and beyond the scope of this paper, but we can identify two important specific elements. First, there are economies of scale in having a market in which individual risk characteristics of the securities themselves don't have to be evaluated by each institution. Second, the existence of a market aids in the revelation of information, such as volatility, although going too far down this road threatens the assumption of complete markets.

II. Marketing

Buying and selling securities that are well graded, uniform, and that trade in well functioning securities markets with relatively large volumes incurs almost no marketing costs. High grade securities, by their very nature, are uniform and can trade in such markets, but this is not the case for tailored, low grade securities.

An institution that wishes to change its risk exposure by disposing of or acquiring a nontraded (and unsynthesizable) exotic low grade instrument has a tailoring problem. Consider the problem of an institution that wishes to sell off some of its low grade assets and, as in Figure 1, wishes to move along its budget line in the direction of the acceptable set. In ideal circumstances this institution knows of another institution in precisely the opposite circumstances that wishes to move along its budget line in precisely the opposite direction and to precisely the same extent. As in all such matters, there is a question of price, but in our nearly complete markets we are assuming that price considerations are secondary. In effect, we are saying that at a sufficiently small price concession it will always be possible to make a move. (A less generous interpretation of our analysis would be that we are simply ignoring price elasticity).

Generally, though, the institution will not have any idea who its opposite number is and it will not even have an exact opposite. A package of buyers and sellers will have to be put together to match at some reasonable price. This packaging is the marketing role of the investment banker. If the seller is large, then a package of purchasers must be put together, and, conversely, if the buyer is large a package of sellers must be put together. Not uncommonly, both sides will be packaged.

The logic of this lies in the cost structure of the marketing network. Suppose an institution wishes to sell a fixed amount (in value terms), say "s," of a particular low grade security. Assume that the asset must be described and explained to be understood and that such explanations are costly. For one thing, there is the fixed cost of simply getting the venture off the ground and putting together a presentation. This includes, among other things, the investment first made in understanding the problem and the legal costs associated with validating the proposed transaction. Investment banks and law firms specialize in such activities. An investment bank that has observed the divergent characteristics of its clients has a comparative advantage at this initial phase of identifying the asset to be sold and configuring the asset and the sales campaign to reach the market. Furthermore, to the extent to which such analysis requires the use of specialized human capital, there is a cost advantage in having it in an investment banking institution that can apply it over many different projects rather than in the originating institution that may see such opportunities only sporadically. In addition, the selling price must also cover the moral hazard costs arising from the inability to perfectly assess the asset as well as the adverse selection costs involved in any asset purchase.

To sell a security, the investment bank must contact a selection of potential purchasers. Assume that it does so sequentially, stopping when it has lined up purchasers for all of s.[5] For simplicity, assume that each sales call costs a unit amount "c" independent of whether or not the call is successful. As Figure 1 suggests, a number of these calls are destined to be unsuccessful simply because they will be made to institutions that are on the same side of the market as the seller or are already in their acceptable sets. In a more enriched picture than that of Figure 1, institutions may wish to move, but not in the direction of the purchase.

For the moment we will assume that the distribution of purchases, $F(\cdot)$, by prospective buyers, is independent and identically distributed across sales calls. This assumes away the possibility that the marketing agent learns about the distribution from the calls. It also assumes away the important feature that they will sequence the calls in their network in the order of the most likely ones first.

We can now readily derive the expected cost of selling s. As a formal matter the problem is the renewal problem of stochastic processes. Let s_i denote the amount sold by call i. The expected cost, $C(s)$, is given by

$$C(s) = cE(\text{number of calls to sell } s).$$

The solution to this problem (under appropriate regularity conditions) is known from renewal theory to be given by

$$C(s) = c[M(s) + 1],$$

[5] Of course, it may have misjudged the market and may discover that there is inadequate demand at the quoted prices. Assuming that prices cannot be moved (generally for the same agency related institutional reasons that brought about the supply in the first place, e.g., the loss of control that comes from marking down assets that have declined in value), the deal may just "fall through".

where $M(s)$ satisfies the renewal equation,

$$M(s) = F(s) + \int_0^s M(s - y)\, dF(y).$$

While the exact shape of $C(s)$ is a function of the distribution of purchases, $F(\cdot)$, we can make some general observations. Since there is a positive probability that a sales call will produce no sales, $F(0) > 0$. It follows that the expected cost of selling any small $\varepsilon > 0$ will be greater than c since at least one call must be made. In fact, for small ε the expected cost will be on the order of

$$C(\varepsilon) \approx c/[1 - F(0^+)] > 0.$$

A number of further results on the shape of the cost function follow from the renewal theorem. (See Karlin and Taylor (1975).) First, we know that $C(s)$ is bounded,

$$c(s/\mu) < C(s) < c(s + k)/\mu_k,$$

where μ is the average sale (including the no sales) and for arbitrary $k > 0$, μ_k is the truncated average sale, i.e., the average sale assuming that sales greater than k are only alloted k. We also know that as $s \to \infty$,

$$C(s) \to c(s/\mu)$$

and that at large sales levels the marginal expected cost of a sale of amount b approaches

$$C(s + b) - C(s) \to (c/\mu)b.$$

Figure 2 summarizes the behavior of the typical expected cost function, the average expected cost, $C(s)/s$, and the marginal expected cost. The fixed cost of selling any positive amount causes the average cost function to decline initially, and the asymptotic behavior makes it reasonable, but not inevitable, that the average cost is globally declining. This average cost function is the basis for much of the marketing behavior we find in financial markets.

If the marketing agent learns from previous calls, then the marginal cost of the next sale can decline, both because the sales presentation becomes more effective and because of the improved ability to order prospects so that the more likely ones can be visited sooner. Countering this tendency is the fact that marketing sampling is done without replacement. That is to say, once a call has been made it is removed from the remaining pool. When the investment banker knows the characteristics of the market, this will tend to complicate the cost function. For example, suppose that the investment banker knows that there is one institution that is far more likely to take a large amount than are the others; then minimizing selling costs will lead to this being the first call. As a consequence, the average expected cost may first increase as the banker has to delve into less likely portions of the market. Knowledge of the market is a valuable piece of capital; it cuts down marketing costs.

If the average cost is declining, then it is efficient to bundle sales or purchases

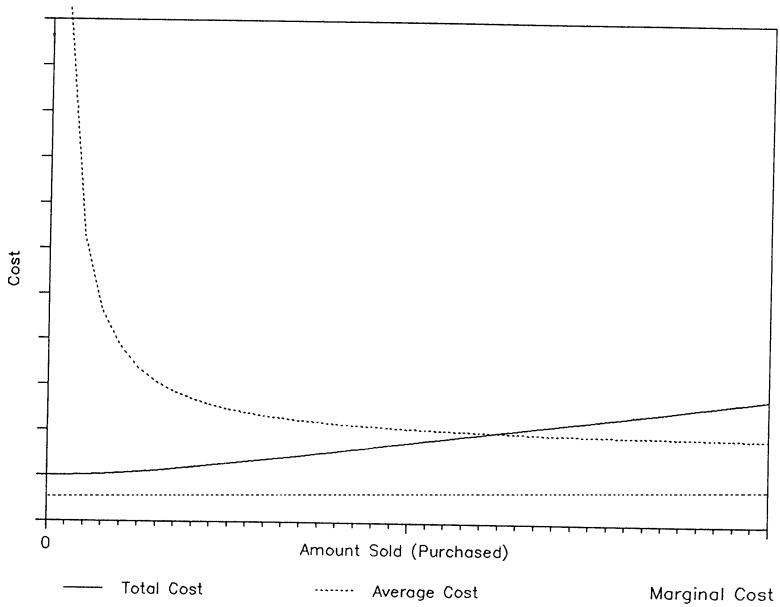

Figure 2. Marketing cost curves for a financial innovation.

into large packages. This is easy to verify analytically. For any amounts, a and b, the cost of selling the combination, $a + b$, is

$$\begin{aligned} C(a + b) &= (a + b)AC(a + b) \\ &\leq aAC(a) + bAC(b) \\ &= C(a) + C(b), \end{aligned}$$

where $AC(\cdot)$ denotes the average cost function. Intuitively, this just means that the fixed cost of marketing should be spread over large efforts.

The lumpy nature of both sales and securities gives rise to yet another, albeit probably small, combinatorial force for combining positions. In the above derivation, typically when enough calls have been made to cover the desired sale, the total of desired purchases will exceed the amount offered for sale. Additionally, prior to having fully sold out, the shortage may be so small as to not justify further calling. It is reasonable to assume that there are costs associated with failures to perfectly match sales with desired purchases. Pooling the desired sales of several institutions makes it more likely that the sum of the desired purchases will be closer to the sum of the sales amounts (certainly as a proportion of the total sale). This lowers the cost of a mismatch and augments the pooling effect from the original development of the cost function.

The analysis changes little, at least in principle, for high grade securities, except that the fixed costs of marketing are not necessary when such securities trade in organized markets. The characteristics of such securities are well understood by many participants and the resulting per unit sales costs, c, are much lower than those for the low grade securities we have considered here. In

such markets, the sequential sales model gives way to a simultaneous one. From the perspective of the model of the institution we have used, one could imagine that the reserve requirement leads to an ongoing excess demand and supply for the riskless asset. For example, if the investment banks can cheaply monitor changes in total participant cash inflows and outflows to institutions, then they can cheaply ascertain which would be willing to buy and which to sell high grade securities. Notice, though, that even this information would not enable them to tell which firms would be buyers and which would sell low grade securities.

Thus, the distinction is not so much between high and low grade securities as it is between marketed and nonmarketed securities. Stocks are low grade securities but they trade in well functioning competitive markets. For the equity of large companies there is great depth in the market, i.e., many participants willing to buy and sell at small price differentials because the securities are well studied and understood. Financial futures are recent examples of low grade—which is to say, high risk—innovations which have evolved into efficient, low cost, well traded commodities. The point we are making is that this evolution itself is costly and its ultimate success depends upon the ability to standardize the instrument to the point where the resulting market is thick enough to sustain a sufficient volume of trade to justify the initial costs. The tradeoff is interesting and unexplored; the more the commodity is standardized the less useful it is to individual institutions, but the lower the transactions costs.

III. Innovation

We now have the ingredients in place to see how a particular innovation can occur. To start things out, suppose that some institutions have been moved out of their acceptable sets by an unanticipated event that was not incorporated in the original acceptable set. Perhaps there has been a regulatory shift or a tax change has moved the boundaries. Merton Miller (1986) refers to the latter as the "grain of sand in the oyster." Such shifts can be the seeds of financial innovation, but it is difficult to understand how the financial innovation can remain even when regulations or tax rules have shifted in such a way as to eliminate that incentive.

The ongoing survival of an innovation—the source of the demand and the supply—depends on the institutional agency structure. The private sector equilibrium involves a contractual structure to deal with agency problems not dissimilar to the tax code or government regulations. A tax or regulatory shift may well provide the incentive to overcome the fixed cost of innovation, but, alternatively, there could simply be a shock to participants' cash flows or to investment realizations (e.g., real estate or third world loans underperform). However, it is the institutional structure of contracts and incentives that permits such innovations to continue, be it the Japanese tax rules for insurance companies, government regulations on S&L's, or the private placement lender's insistence on loan priority.

Whatever the initial shock, assume that a number of institutions now find themselves at points like C or D in Figure 1. The investment banker may not

know the exact situation of individual institutions, and, strictly speaking, pooling securities involves the same sort of marketing search as selling, but often the change will be in a common direction for the institutions in a particular industry. The proposed sales of such institutions can then be combined into a single pool. As we have seen, in the presence of high fixed marketing costs, this will be superior to selling each institutions' low grade asset individually.

We can now show that it pays to crack the security into a high grade and a stripped off low grade component. Since high grade securities, for example, riskless bonds, trade in a well organized and frictionless financial market, the unit marketing and distribution costs, c_H, are lower than those in the market for low grade securities. Assuming that the average cost of selling the low grade security declines and asymptotes to its marginal cost, c_L/μ (see the renewal theorem cited above), this implies that $c_H < c_L/\mu$. More generally, we will assume that the marginal cost of selling these ungraded L securities is higher than that of selling high grade ones; i.e., for any a and b,

$$C(a + b) - C(b) > c_H a.$$

Consider selling the low grade security, L, plotted in Figure 3. (The sale of this security and the purchase of high grade bonds will enable an institution to move from a point like C in Figure 1 into the acceptable set.) Simply selling a total dollar amount, s, of such securities as low grade securities will entail marketing costs of

$$C(s)$$

where

$$s = \pi_1 \rho_1 L_1 + \pi_2 \rho_2 L_2.$$

As an alternative, consider stripping off the high grade portion of these low grade securities and leaving the low grade security labeled as the strip in Figure 3. The value of the strip is given by

$$E\{\rho L\} = \pi_1 \rho_1 [L_1 - L_2],$$

and the cost of marketing it is

$$C(\pi_1 \rho_1 [L_1 - L_2])).$$

Adding the cost of marketing the high grade portion of the asset left after the strip, we obtain a total marketing cost of

$$C(\pi_1 \rho_1 [L_1 - L_2]) + c_H L_2 / r,$$

where r is the riskless interest factor,

$$r^{-1} = \pi_1 \rho_1 + \pi_2 \rho_2.$$

Since the marginal high grade cost is less than the low grade cost,

$$\begin{aligned} C(\pi_1 \rho_1 [L_1 - L_2]) + c_H L_2 / r &= C(\pi_1 \rho_1 [L_1 - L_2]) + (\pi_1 \rho_1 + \pi_2 \rho_2) c_H L_2 \\ &< C(\pi_1 \rho_1 L_1 + \pi_2 \rho_2 L_2) \\ &\equiv C(s). \end{aligned}$$

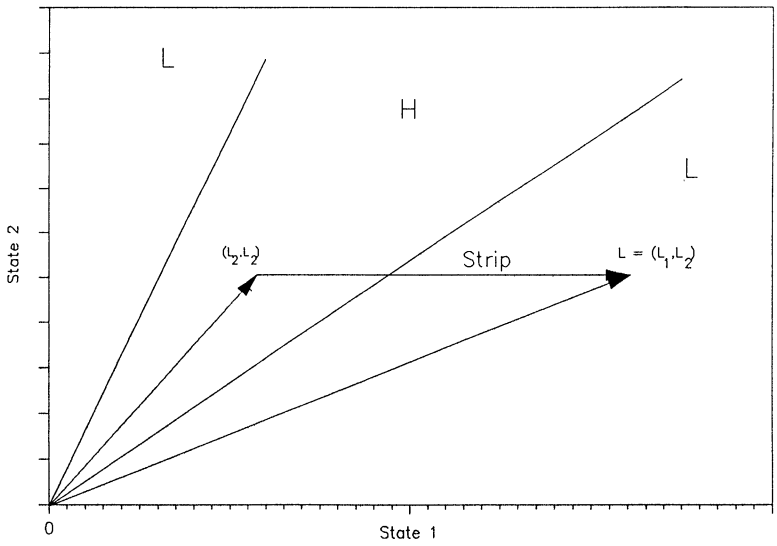

Figure 3. Security stripping.

Summarizing our findings, selling the strip and the high grade security is cheaper than simply selling the original low grade security. This is precisely what happens when a Treasury bond is stripped into its component parts or when a mortgage pool is decomposed into an *IO* (interest only) and a *PO* (principal only).

IV. Summary and Conclusion

What we have constructed is intended as a broad outline of a theory of financial markets and financial innovations. It does not purport to be a theory of all financial innovation. For example, while institutions are major players in the markets for options on individuals stocks, that is not how these markets began.[6] However, many of our same arguments can be used to explain such retail derivatives; they permit individuals to hold levered positions in securities by using limited liability instruments with lowered monitoring requirements. Similarly, our approach to taxation suggests a more interesting class of restrictions than we usually capture in the mechanical analyses of marginal tax rates.

Less conventionally still, institutional markets and financial marketing are central to understanding financial innovation and are almost entirely unexplored. Institutional markets may well have pricing theories that differ from those of the retail markets of our current theories. It is possible, though, to discern some

[6] Leaving aside the role of put/call parity in permitting institutions to write calls to a public that was more interested in buying calls than in writing them.

outlines. Most likely, some similar form of the APT will continue to hold since the absence of arbitrage should remain necessary, but the CAPM and its sister, the consumption beta theory, will have greater difficulties. Insofar as equilibrium theories derive their power from preference restrictions, the likely inapplicability of such restrictions in institutional markets weakens their appeal. The resolution of these matters is the next item on the research agenda.

We have also modeled the marketing cost structure, but marketing opens up a host of strategic issues which we have not considered. Surely, the possibility exists for differentiating products that have not become standardized commodities. One interesting example is provided by the variety of stock market indexes used for performance measurement. Currently financial services firms compete with one another by offering their own differentiated indexes. It is striking how little modern finance currently has to say about such fascinating behavior. It also illustrates that there should be much to gain from an integration of marketing analysis and finance.

REFERENCES

Demski, J. S., Sappington, D. E. H., and P. T. Spiller, 1988, Incentive schemes with multiple agents and bankruptcy constraints, *Journal of Economic Theory* 44, 156–167.

Dybvig, P. H. and S. A. Ross, 1987, Arbitrage, in J. Eatwell, M. Milgate, and P. Newman, eds.: *New Palgrave, A Dictionary of Economics*. (MacMillan Press, Ltd., London).

——— and C. Spatt, 1985, Agency and the market for portfolio managers: The principle of preference similarity, Working Paper #86, Yale School of Management.

Holmstrom, B., 1979, Moral hazard and observability, *Bell Journal of Economics* 12, 74–91.

Karlin, S. and Taylor, H. M., 1975, *A First Course in Stochastic Processes*, 2nd ed. (Academic Press, New York).

Merton, R. C., 1988, On the application of the continuous-time theory of finance to financial intermediation and insurance, Geneva Association Lecture, Centre Hec-Isa, France.

Miller, M., 1986, Financial innovation: The last twenty years and the next, *Journal of Financial and Quantitative Analysis* 21, 459–471.

Ross, S. A., 1973, The economic theory of agency: The principal's problem, *American Economic Review* 63, 134–139.

Tufano, P., 1988, Financial innovation and first-mover advantages: An empirical analysis, Unpublished manuscript, Harvard University.