Innovations in Financial Services, Relationships, and Risk Sharing

Franklin Allen • Douglas Gale
Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
Economics Department, New York University, New York, New York 10003

Relationships between intermediaries and their customers have become increasingly important in recent years. This paper argues that the need for costly ex ante information acquisition and analysis is a major barrier to the participation of investors and firms in sophisticated markets. Long-term relationships between intermediaries and their customers, in which intermediaries provide implicit insurance to customers, can be an effective substitute for costly ex ante investigation. In this way, intermediaries allow firms and investors to reap the benefits of financial markets. Relationships are easiest to sustain when the ongoing benefits to both parties are high. As a result, competition may lower the benefits that can be obtained from relationships.
(*Markets*; *Intermediaries*; *Unforeseen Contingencies*; *Misunderstandings*; *Implicit Contracts*)

1. Introduction

The financial services industry has undergone a dramatic transformation in recent decades. What might be termed the *traditional paradigm* for intermediation is illustrated in Figure 1. Banks and insurance companies convey funds from low wealth households to firms. High wealth households and large companies (with the help of investment banks) mostly use financial markets directly. The primary role of intermediaries is perceived to be reducing transaction costs and providing information. Markets and intermediaries are *alternative* ways of channelling funds.

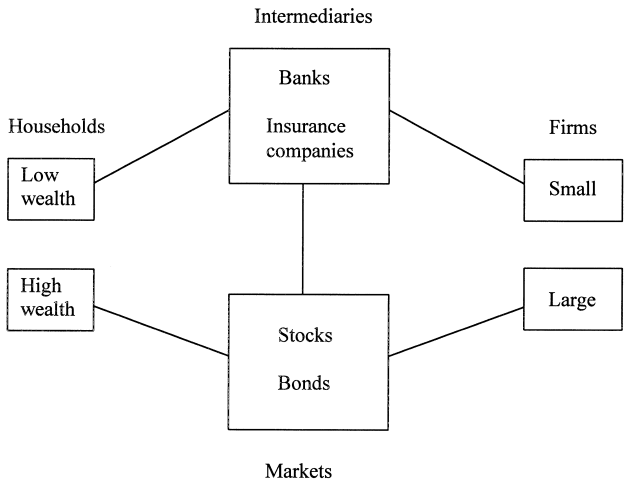
One aspect of the transformation has been the increase in importance of traditional financial markets in the U.S., such as those for equity and debt. For example, the market capitalization of corporate equity in the U.S. has risen steadily as a percentage of GDP, from around 50 percent in 1975 to nearly 75 percent in 1994 (OECD-Financial Market Trends, #62, November 1995). Another aspect is that the range of markets available has widened with the development of liquid options, financial futures, and other derivative markets.

Despite a significant fall in the direct costs of trading after the deregulation of stock commissions in 1975, the increased availability of information about corporations and access to price data and standard valuation models through services such as Bloomberg, this expansion in the importance of financial markets is not because of higher participation by individuals or firms. In fact, the share of individual ownership of corporate equity in the U.S. has fallen during the period 1967–1995. There was a particularly sharp change in the early 1980s, when individual ownership fell from over 75 percent to around 50 percent in only a few years (Board of Governors of the Federal Reserve System—Flow of Funds Accounts).

The change has occurred because intermediaries are using markets more extensively than before. Over the period 1967–1995, the share of pension funds' ownership of equity in the U.S. has risen from less than 10 percent to over 20 percent. In the same period, the share of mutual funds' ownership of equity in the U.S. has grown from around five percent to nearly 25 percent (Board of Governors of the Federal Reserve System—Flow of Funds Accounts). In derivatives

0025-1909/99/4509/1239$05.00
Copyright © 1999, Institute for Operations Research and the Management Sciences

Figure 1 The Traditional Paradigm



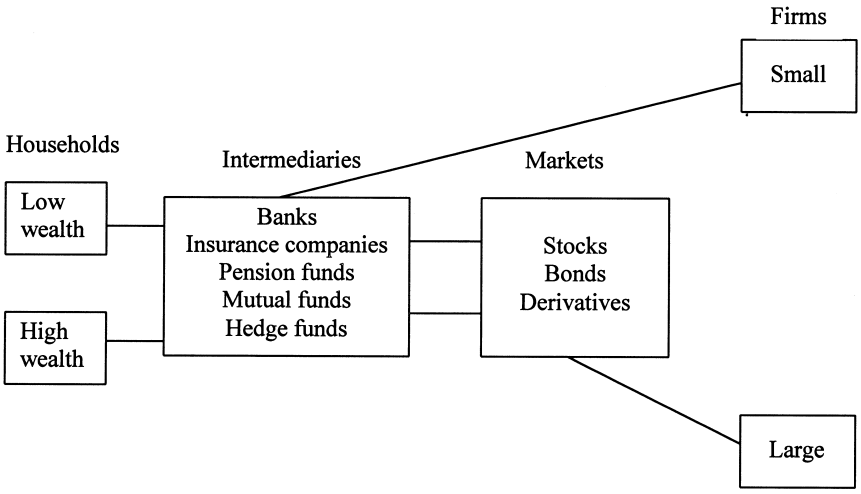
markets, intermediaries play an even more significant role. As of 1995, financial institutions accounted for 82 percent of the notional amounts of OTC derivatives outstanding while nonfinancial institutions accounted for the remaining 18 percent (Bank for International Settlements—Central Bank Survey of Derivatives Market Activity, 1995).

Figure 2 illustrates what might be termed the *emerging paradigm* for intermediation that these changes are leading to. Most households increasingly deal with intermediaries, such as pension and mutual funds that invest in markets on their behalf. Even among the very wealthy, the use of private banking services and hedge funds, where advisors make investment decisions on behalf of their clients, has become increasingly common. Small firms deal with banks and other entities, such as limited partnerships, providing venture capital and other forms of private equity. Only the very largest firms (with the help of investment banks) deal directly in financial markets.

Allen and Santomero (1997) have suggested that the focus of intermediaries' activities has shifted away from reducing transaction costs and providing information. Instead, risk management has become an increasingly important activity. Intermediaries are using the wide range of markets available to them to transfer, transform, and redistribute risk. Understanding the practice and role of risk management has become the center of a growing literature (see Allen and Santomero 1997 and the references therein). Figure 2 illustrates that in the emerging paradigm of intermediation the increased importance of risk management is only one part of the change in the financial services industry that has occurred. Another part is the increase in the interaction of intermediaries with households and firms.

To understand these changes in the financial services industry we have to revise the traditional view of the comparative advantages of financial markets and financial intermediaries. The traditional paradigm

Figure 2 The Emerging Paradigm



stresses three differences between financial markets and financial intermediaries.

• First, markets offer the opportunity to trade standard securities, whereas intermediaries can tailor securities to the needs of the individual investor or firm. A market needs a certain amount of depth (liquidity) to attract traders. As a result, there exist viable markets for a relatively small fraction of the set of all conceivable securities.[1] On the other hand, intermediaries can design securities for specific customers and specific needs. Investment banks such as Goldman Sachs will provide a wide range of securities for a price.[2] In short, there is an inherent trade-off between liquidity and customization.[3]

• Markets are more competitive and offer services at lower cost. This may result from greater concentration in the intermediary sector or from the existence of hold-up problems and lock-in effects. The intermediary may have information about the customer or about the financial product that allows it to extract surplus. Switching costs and lemons problems may lead to lock-in effects, which the intermediary can use to extract rents (Sharpe 1990, Rajan 1992). In addition, markets may actually be cheaper because they can tap a larger pool of resources and reduce the cost of raising funds. The junk bond market was successful initially because it allowed a significant reduction in the cost of funds compared to bank finance.[4]

• Traditionally, the role of intermediaries is explained in terms of the existence of increasing returns to scale in executing transactions, monitoring investments, and acquiring information. Because there are fixed costs involved in each of these activities, a specialist institution can reduce the per unit expense for investors. Intermediaries may also be able to lower the costs of its customers by providing information. Here the advantage comes from increasing returns to scale. Once the intermediary has paid the fixed cost of acquiring information, it can provide it to all its customers at nearly zero marginal cost (Diamond 1984).[5] An alternative is to package risk in such a way that it is easier to evaluate than the securities provided by the market. Markets can also do this to some extent (cf. Boot and Thakor 1993, Duffie and DeMarzo 1995), but since different individuals have different information available, it may be impossible for markets to duplicate this function in the form of standard securities.

The development of more sophisticated markets has provided a new and somewhat different role for intermediaries.

• As new markets have opened up and made new hedging opportunities available, they have increased the expertise necessary to devise strategies and make effective use of these opportunities. So while tangible transaction costs such as fees and the cost of observing prices may have fallen, some types of information costs have increased. To evaluate a complex security, a complex portfolio, or a complex strategy requires

[1] Even though markets provide a limited number of securities, it may be argued that it is possible to synthesize a large number of different securities by adopting dynamic trading strategies involving a small number of actively traded securities; but this requires a lot of expertise, which a small investor or firm may not have. An intermediary can provide the security directly, saving the investor or firm the trouble of figuring out the trading strategy (Merton 1989).

[2] Even when the product being offered by the intermediary is "vanilla-flavored," the intermediary may still be able to offer services tailored to the needs of the customer that a market cannot. For example, in the U.K. and other European countries, the relationship with the bank manager is important, because even small depositors and borrowers expect the privileges of relationship banking, although the products being offered are not very sophisticated.

[3] A good illustration is provided by a comparison of futures which offer the liquidity that is possible in an exchange market and forwards which offer greater opportunities for customization. The exchanges have recognized the value of customization by offering a broader menu of instruments such as LEAP and FLEX options on the CBOE, that attempt to afford greater customization and liquidity.

[4] A related issue is that the investor or firm may be the victim of moral hazard because it lacks information about what the intermediary is doing. In that case, too, intermediaries may turn out to be costly.

[5] An example is provided by information intermediaries such as Reuters, Bloomberg, or the internet which can provide a platform that enables information vendors to reap the advantages of economies of scale in information distribution. As a result of these economies, "standard" financial information, such as stock and bond prices, can be distributed more cheaply to market participants.

more than just knowing "the facts" about a firm's balance sheet. It requires financial expertise that an ordinary investor usually does not possess. Even large firms may lack this expertise. Intermediaries assume the role of advisors, who bridge the gap between the investors' lack of knowledge and the expertise required to get the most out of sophisticated markets.

Thus, intermediaries play a crucial role in allowing firms and investors to participate in financial markets and, at the same time, ensure that financial markets have enough depth to survive. In this sense, financial markets and financial intermediaries have a symbiotic relationship. Each is necessary to the other. Without intermediaries, the informational barriers to participation would prevent investors from reaping the benefits of the new markets and the markets themselves might not survive. At the same time, financial markets reduce costs for the intermediaries and their customers and allow them to hedge risks more effectively than they previously could.

In this paper, we focus on one aspect of the emerging paradigm and consider the role of intermediaries in reducing the risk to investors and firms of operating in the new and highly sophisticated markets. In an Arrow-Debreu world with complete markets and complete information, there would be no role for intermediaries, but in the world we live in there are numerous reasons why the risk-sharing opportunities offered by markets are incomplete:

• *Complexity*. It may be very expensive to write complex contracts, so the market provides only simple contracts, which is what most investors want.

• *Liquidity*. In order to maintain a viable market, the volume of trade must be large enough to allow the market makers to cover their fixed costs and to ensure sufficient market depth to avoid excessive price volatility. This may not be possible for very exotic securities, with the result that we tend to observe a limited set of securities (Pagano 1989, Allen and Gale 1994).

• *Legal uncertainty*. One of the barriers to the introduction of new securities is the uncertainty about how the legal system will treat them. As a result, there is a preference for securities on which there is a settled body of case law. This necessarily limits the selection of contracts on offer.

• *Gains from standardization*. Trading securities requires general knowledge about classes of securities, stocks, bonds, options, futures, etc., as well as specific knowledge about individual securities, mean return, variance, beta, etc. As a result, information costs are reduced by dealing in standard securities and this may also discourage the introduction of new securities (Gale 1992).

As a benchmark, a simple example with incomplete markets is studied in §2. A customer can hedge an idiosyncratic risk in the market, but the risk-sharing opportunity is incomplete. This means that there is the potential for an intermediary to increase the customer's welfare by offering a supplementary risk-sharing contract. It is assumed that the intermediary can write an *explicit contract*; that is, a written contract that can be enforced by third parties such as the courts. It is also assumed that the contract is *complete*; that is, it is the optimal contract under the assumption that there is complete information and no transaction costs. In this case, the intermediary is able to increase welfare because we assume that it can do something that the market cannot do, namely, provide an explicit, complete risk-sharing contract. This provides a role for the intermediary, but it does not explain why the intermediary has an advantage over the market. Why can the intermediary provide a risk-sharing contract at lower cost than the market? The subsequent sections attempt to provide an explanation.

In §3, we take the argument a step further by assuming that writing explicit contracts is costly. The market provides a limited set of securities (and hence a limited set of hedging opportunities) because it is too costly to provide a broader set of securities. However, intermediaries have an advantage over the market, because they can offer risk-sharing through *implicit contracts*; that is, contracts that are unwritten and hence unenforceable by the courts. The problem with implicit contracts, of course, is that they must be *self-enforcing*. The intermediary must have an incentive to carry out the terms of the implicit contract. In the static model studied in §2, the intermediary would have no incentive to keep the terms of an implicit contract. If the investor suffers a loss in some state of nature, it would not be rational for the intermediary to

make good the loss. However, the intermediary may have an incentive to abide by the terms of the implicit contract in the context of a *long-term relationship*, rather than renege and put an end to the relationship. In §3, we show how an implicit contract can be made self-enforcing in a long-term relationship. The degree of risk sharing that can be sustained by an implicit contract will be limited by the value of the ongoing relationship. If the amount of money transferred between the parties ex post is greater than the value of the relationship, the party that is required to pay will dissolve the relationship rather than abide by the contract. So the need for the implicit contract to be self-enforcing implies that the contract will be incomplete. This has the further implication that an increase in competition, which reduces ex post payoffs to the intermediary, will reduce the completeness of the implicit contract.

In §4 we turn to the central question of the paper, which is how an intermediary can reduce investors' participation costs and in particular their cost of acquiring information. The problem faced by the investor in the market is not so much one of "risk" as one of "uncertainty." The investor simply does not know what he is getting into. He lacks the expertise needed to evaluate the security, derivative, or risk-sharing contract that the intermediary wants him to trade. The intermediary can reduce the uncertainty faced by the investor by offering implicit insurance against "unforeseen contingencies" or "misunderstandings."

We model uncertainty by assuming that the outcome of a security depends on two different types of states of nature. We distinguish between *salient states*, which correspond to the kinds of risk that everyone, including investors and firms, can understand, and *obscure states*, which correspond to uncertainty about the nature of the security that can only be eliminated by acquiring costly expertise. Writing a complete contingent contract does not solve this problem, because including more contingencies simply increases the number of obscure states and makes it harder for the customer to evaluate the contract.

We model the cost of becoming informed by a fixed cost of *evaluating* the security in each state. Investors take an attitude of extreme risk aversion toward obscure states and hence toward any security or contract that contains contingencies based on obscure states. The intermediary can reduce participation costs by providing an implicit agreement to insure the customer. If things turn out badly, there is the possibility of compensation. In effect, the intermediary, by providing insurance against "unforeseen contingencies" (obscure states), reduces the security's contingency on obscure states and hence reduces the need for the investor to have information in the first place.

The implications of our analysis for improving the effectiveness of relationships in the financial services industry are discussed in §5. Since it is crucial, for relationships to be successful, that parties have some surplus at stake, attempts at increasing the scope of relationships should focus on combining profitable lines of business. Section 6 contains concluding remarks.

Many of the arguments developed in this paper may have applications to a wider range of economic phenomena. In many principal-agent relationships the principal employs an agent precisely because the principal lacks relevant knowledge and expertise. For example, informational problems that are similar to those of an investor trading a complex security may arise in the relationship between a doctor and patient, a lawyer and client, a computer manufacturer and computer user. However, we believe that the informational problems arising from financial transactions are serious, so it is particularly important to consider this issue in a financial setting. Financial transactions are often concerned with risk sharing and the outcomes are typically unpredictable. A firm or investor buying a complex security faces a wide range of possible outcomes and the intermediary cannot insure against all of them. For example, if there is a change in interest rates this may affect most of the derivatives sold by the firm. Indeed, it is rarely optimal for one party to bear all of the risk. Because each party has a lot of private information, for example, information about their idiosyncratic risks, it is hard to be clear about what they expect as a result of a complex financial transaction. So financial arrangements seem to be

more opaque and to offer more scope for problems than other kinds of transactions.

The paper is related to a number of strands of the literature. Merton (1989) justified the use of continuous time techniques which assume no frictions by distinguishing between intermediaries and individual investors and firms. He suggested intermediaries could trade at almost zero marginal costs in markets. These intermediaries could rebundle payoff streams for individual investors who would pay a price for this service. He focused on intermediaries' activities in markets whereas we focus on their interaction with unsophisticated investors. The existing literature on the importance of relationships and implicit contracts for intermediaries has a rather different focus from ours. Sharpe (1990) and Rajan (1992) consider the generation of information during relationships between banks and borrowers and the possibility for ex post exploitation of the monopoly rents these contracts lead to. They show how ex ante competition limits the potential amount that can be extracted and the effect this has on loan rates and choices of financing. In contrast, in our model there is not a problem with opportunistic behavior. Instead, relationships are beneficial because they can provide implicit insurance. Boot and Thakor (1996) consider the effect of competition on the composition of bank lending between transaction loans, which are like capital market funding, and relationship loans, which are like traditional lending. They are able to show the surprising result that relationship lending increases with competition. Their focus is thus on the composition of banks' activities whereas ours is on the interaction between explicit and implicit contracts.

2. Explicit Risk-Sharing Contracts

As pointed out above, one advantage of intermediaries is that they can tailor security design to the individual customer. In effect, they make markets less incomplete. In this section, we take for granted the incompleteness of risk sharing provided by the market and investigate the potential for an intermediary to provide a tailor-made insurance contract. We analyze this benchmark case as a simple risk-sharing problem, in which it is assumed that the intermediary and the customer are symmetrically informed and can write an explicit, enforceable contract. In later sections, we relax these assumptions, one by one.

Suppose that the customer, which could be a firm or an investor, wants to hedge the risk of a random income. Risk preferences are represented by von Neumann-Morgenstern utility functions $u(x)$ for the customer and $v(y)$ for the intermediary. The customer receives income $w(s)$ in state s. There is a market in which a security f is traded. This security pays a net return $f(s)$ in state s, where the net return includes the price of the security as well as any commissions and fees. Without loss of generality, we can scale the security's payoffs so that the customer wants to trade exactly one unit in equilibrium. Then, if it trades the security f on the market, its net income will be $f(s) + w(s)$ in state s and its expected utility will be

$$E[u(f + w)] = \sum_s p(s)u(f(s) + w(s)),$$

where $p(s)$ is the probability of state s.

Now suppose the customer went instead to the intermediary and wanted to hedge the risk w. We ignore other income for the intermediary, so if the intermediary wants to sell a security g to the customer, its income is simply $-g(s)$ in state s. Without further information, the expected utilities of the customer and intermediary, conditional on trading the security g, will be $E[u(g(s) + w(s))]$ and $E[v(-g(s))]$, respectively. The particular security chosen by the intermediary has two functions, to share risks efficiently and to provide the maximum payoff to the intermediary. If the intermediary has all the bargaining power, it maximizes its own payoff subject to the constraint that the customer can always go to the market:

$$\begin{array}{ll} \max_g & E[v(-g(s))] \\ \text{s.t.} & E[u(g(s) + w(s))] \geq E[u(f(s) + w(s))]. \end{array}$$

In other contexts, the customer may have some bargaining power. For example, the customer may be able to extract informational rents if the intermediary does not know its reservation utility. In that case, the surplus will be shared in a less extreme way. In any case, the contract chosen will be efficient.

The bottom line is that the intermediary can im-

prove risk sharing by tailoring the security to the customer's needs, but may or may not share the gains with the customer.

This may not be the whole story, however. In the first place, the intermediary may have higher costs than the market. For example, if there is a fixed cost $c > 0$ of writing a tailor-made security (i.e., a complete contract), then the intermediary's payoff is $E[v(-g(s) - c)]$ and if the intermediary's reservation utility is $v(0)$, then even if the customer has all of the bargaining power, the most that it can obtain is

$$\begin{aligned} \max \quad & E[u(g(s) + w(s))] \\ \text{s.t.} \quad & E[v(-g(s) - c)] \geq v(0), \end{aligned}$$

which may be less than $E[u(f(s) + w(s))]$. Even if there is positive surplus, the intermediary needs something less than full bargaining power to make the firm no better off than it would be dealing in the market. A similar issue arises if the intermediary's cost of funds is greater than the market's. That will also appear as a cost that must be subtracted from g in order to define the intermediary's payoff.

The essential issue in this situation is the trade-off between the costs and benefits of markets and intermediaries. For some investors and firms, intermediaries may be superior while for others markets are superior. The existence of both intermediaries and markets allows all agents to choose the least-cost method of lending or borrowing funds.

3. Self-Enforcing Implicit Contracts

In analyzing the risk-sharing problem, we have assumed that the customer and the intermediary write an explicit and enforceable contract. However, the costs that make complete markets an implausible assumption may also make complete explicit contracts unrealistically expensive too. One way of getting around this problem is to assume that the intermediary and the customer do not have an explicit (written) contract, but rather have an implicit (unwritten) understanding about what might happen if a "bad outcome" occurs.

This solution raises another problem, however. If the implicit contract between the intermediary and the customer is unwritten and hence unenforceable by a third party, why should either the customer or the intermediary believe the other will keep the terms of the implicit contract? In a one-shot contracting problem, unenforceability may prevent any insurance from being offered by the intermediary. In the context of a long-term relationship, however, it may be possible to sustain risk sharing through a self-enforcing implicit contract, in which the value of the ongoing relationship to both parties provides incentives to abide by the terms of the agreement, rather than renege and end the relationship. However, the amount of insurance that can be offered in a self-enforcing way will be limited by the value of the ongoing relationship. We study this next.

3.1. Sustainable Relationships

The model in §2 is extended to allow for the possibility of long-term relationships and implicit contracts. As before, there are two parties to a risk-sharing contract, a customer and an intermediary. There is a sequence of dates $t = 1, 2, \cdots$ and a new risk-sharing opportunity at each date. The states $\{s_t\}_{t=1}^{\infty}$ are assumed to be i.i.d. random variables with probability density $p(s_t)$. The customer's random income is $w(s_t)$ if the state s_t is observed at date t. The written contract is a function $f(s_t)$, where $f(s_t)$ is the amount of money paid by the intermediary to the customer in state s_t. In the absence of any other transfers, the expected utility of the customer at date t is $E[u(w(s_t) + f(s_t))]$ and the expected utility of the intermediary, ignoring other risks, is $E[v(-f(s_t))]$. The contract f is taken as given, that is, it has been designed by the intermediary and represents the best hedge that the intermediary could write for the customer. The question now is how much better can the intermediary and the customer do by entering into an *implicit* risk-sharing arrangement?

Let $h(s_t)$ be the transfer that is paid by the intermediary to the customer in state s_t as a result of this implicit agreement. In order to make the implicit contract self-enforcing, there must be some threat of punishment for either party if it reneges on the agreement. Sometimes concern for "reputation" is sufficient to sustain the relationship. If the intermediary reneges on its part of the bargain, it will lose its reputation and the possibility of future business. In the same way, a customer that reneges on its part of the bargain will

find it difficult to get an intermediary to offer implicit insurance in the future. Without modeling this part of the story formally, we assume that if either party reneges on its part of the implicit contract, it will be impossible to find another partner. On the other hand, we assume that the customer can always resort to the formal contract as a security level. For this reason, even if the customer is restricted to dealing with its current intermediary, the customer has a level of expected utility below which the implicit contract cannot push it in any state of nature.

The participation constraints that both the customer and the intermediary face are similar. For ease of exposition, we will illustrate the operation of these constraints by focusing on those faced by the customer and assume in what follows that the intermediary can commit itself to the implicit agreement. Let

$$u^* = E[u(w(s_t) + f(s_t) + h(s_t))]$$

be the ex ante expected utility of the customer in each period when there is an implicit risk-sharing agreement. Let

$$u^{**} = E[u(w(s_t) + f(s_t))]$$

be the corresponding value without the implicit agreement. Then the participation constraint for the customer is

$$u(w(s_t) + f(s_t) + h(s_t)) + \frac{u^*}{r} \geq u(w(s_t) + f(s_t)) + \frac{u^{**}}{r},$$

where r is the customer's discount rate. The left-hand side of the inequality is the expected utility of continuing the implicit agreement while the right-hand side is the expected utility of deviating this period and ending it. This inequality puts a lower bound on the value of $h(s_t)$ that is a function of s_t and u^*; call it $H(s_t, u^*, u^{**})$. Assuming that competition among intermediaries leads them to maximize the expected utility of the customer subject to some period opportunity cost v^{**} (we assume for simplicity that each intermediary can deal with at most one customer), the equilibrium implicit contract has to solve the following problem:

$$\begin{aligned} \max \quad & u^* = E[u(w(s_t) + f(s_t) + h(s_t))] \\ \text{s.t.} \quad & h(s_t) \geq H(s_t, u^*, u^{**}) \quad \forall\, s_t, \\ & E[v(-f(s_t) - h(s_t))] \geq v^{**}. \end{aligned}$$

This is an equilibrium problem, since the participation constraint $h(s_t) \geq H(s_t, u^*, u^{**})$ depends on the expected utility that is available from continuing the relationship, which in turn is the value of the maximization problem.

The first-order conditions that must be satisfied by the solution to this problem are

$$u'(s_t + f(s_t) + h(s_t)) = \lambda v'(-f(s_t) - h(s_t)) \quad \text{if } h(s_t) > H(s_t, u^*, u^{**}),$$

$$u'(s_t + f(s_t) + h(s_t)) \geq \lambda v'(-f(s_t) - h(s_t)) \quad \text{if } h(s_t) = H(s_t, u^*, u^{**}),$$

where λ is the Lagrange multiplier on the intermediary's participation constraint.

The first condition applies in states where the optimal insurance payment is consistent with the customer's participation constraint. The second condition applies if the risk-sharing transfer is constrained by the future surplus obtained from the relationship.

3.2. The Effect of Competition on Relationships

The value of the implicit insurance contract is revealed by the fact that $u^* > u^{**}$ whenever $f(s_t)$ is not itself an optimal contract. However, there may be other constraints on the contract. The more opportunities either party has to renege, the tighter the ex post participation constraint will be and the lower the value of u^* relative to u^{**}. Possibilities for reneging reduce the future surplus and hence the level of transfer the implicit agreement can involve.

In the preceding section, we assumed that the only alternative to complying with the terms of an implicit contract was for the customer to renege and accept the loss of implicit insurance. Neither the customer nor the intermediary was allowed to abandon the relationship and take up with another party. This could be equilibrium behavior if every other customer and intermediary were willing to punish a customer or intermediary who reneged on such an agreement, but it is not entirely clear (because we have not explicitly modelled it) that this would be supported by an

equilibrium. In any case, it seems likely that in most explicit formulations the more competitors there are on each side of the market and the more anonymous the market becomes, the less likely it is that any customer or intermediary will be punished for reneging on an implicit agreement. The more outside options each party has available, the tighter the (ex post) participation constraints become and the lower the (ex ante) expected utility from the implicit contract.

Suppose, to take an extreme case, that customers can switch intermediaries whenever it suits them. This requires in particular that the written contract f does not include any penalties for terminating the relationship or any bonding devices such as payment of fees and commissions to the intermediary in advance. Then the ex post participation constraint takes the form

$$u(w(s_t) + f(s_t) + h(s_t)) + \frac{u^*}{r} \geq u(w(s_t) + f(s_t)) + \frac{u^*}{r}$$

or $h(s_t) \geq 0$. In other words, the implicit agreement can never require the customer to make a transfer to the intermediary. This is a tighter constraint, as long as $u^* > u^{**}$, and so it reduces the expected utility that can be achieved for the customer. Increased competition, in the form of easier switching between intermediaries, makes it harder to enforce agreements and causes a real efficiency cost in terms of reduced risk sharing. Another way of saying the same thing is that the set of self-enforcing contracts is smaller and excludes the contracts that offer better risk sharing.

We can take this argument one step further and consider the effect of competition from financial markets. Since intermediaries can duplicate anything markets can offer, the introduction of a market does not introduce any new opportunities for risk sharing. Furthermore, since intermediaries can provide at least as good products as financial markets, there is no tightening of the ex post participation constraint. What markets may do is to change the distribution of surplus between the customer and the intermediary. To make this very clear, consider the case where there is initially either a monopolistic intermediary or a collusive cartel of intermediaries that can extract all the surplus from the customer. In that case, the implicit contract solves the problem:

$$\begin{aligned} \max \quad & v^* = E[v(-f(s_t) - h(s_t))] \\ \text{s.t.} \quad & h(s_t) \geq H(s_t, u^*, u^{**}) \quad \forall\, s_t, \\ & u^* = E[u(w(s_t) + f(s_t) + h(s_t))] \geq u^{**} \\ & \quad = E[u(w(s_t))]. \end{aligned}$$

Now suppose that a financial market opens up, in which the security k is traded competitively. This security may be identical to f except for the fees and commissions that are incorporated in the definition of the security and are assumed to be lower than those charged by the intermediary. Then the customer's reservation utility changes from $u^{**} = E[u(w(s_t))]$ to $u^{***} = E[u(w(s_t) + k(s_t))]$. This change has two effects. First, it makes the customer better off. Second, it changes the ex post participation constraint. The first is a distributional change and comes at the expense of the monopolistic intermediaries. The second has efficiency repercussions, because it restricts the risk-sharing properties of the implicit contract being offered by the intermediaries. So there is a trade-off between efficiency and distribution. A reduction in monopoly rents may lead to a reduction in risk sharing.

4. Reducing the Costs of Market Participation

4.1. Insurance Against Unforeseen Contingencies

One of the things that discourages small investors and firms from making use of financial markets is the information that is needed to operate successfully in volatile and sophisticated markets. Quite apart from the risk attached to the underlying variables that are being hedged, there is uncertainty about the exact nature of the securities being traded, about the behavior of the markets, the dealing process, etc. Unsophisticated investors and firms do not know what they are getting into.

Intuitively, dealing with a trusted advisor or intermediary seems to avoid some of this risk. The intermediary can give advice on the choice of securities; but beyond this it can offer some assurance that there will be no unpleasant surprises. If something unex-

pected does occur, the intermediary may be willing to bear the cost or, at least, share the cost with the customer. In short, dealing with an intermediary provides insurance against nonspecific risks. Were an agent to operate directly in a financial market, however, there would be no such insurance. The agent gets what the security specifies, no more and no less. For this reason, investors and firms may feel much more "comfortable" dealing with an intermediary. This is not an appeal to irrationality. The fact that one has a relationship with the intermediary provides insurance.

It is not easy to model the kind of uncertainty that the customer faces in trying to make decisions in an unfamiliar environment. Here we try to develop a model that distinguishes this uncertainty from ordinary risk associated with asset returns. One of the costs of going to the market is the cost of becoming informed about the securities traded and about the market mechanism itself. Suppose that the customer does not know what the security f is; that is, he is uncertain not just about the state of nature s that will eventually be revealed, but also about the value of the security in state s. In this case, the intermediary may be able to offer the customer insurance about the nature of the security by promising to compensate the customer for any "unpleasant surprises."

This has the additional advantage that the intermediary is taking advantage of the market's lower costs by piggybacking on the standard security f. So the customer is trading f indirectly through the intermediary and getting an additional security $h = g - f$ that has two functions, one being to customize the standard security f to the customer's individual needs, and the second being to relieve the customer of the need to obtain extra information about the nature of the security f.

How are we to interpret the transfers $h = g - f$? They could take the form of cash payments but need not. They could take the form of some readjustment of the terms of the contract f. Or, if the customer and the intermediary do business repeatedly, h may take the form of an adjustment in the terms of a subsequent transaction. If the customer gets "burned" on one deal, it may receive favorable terms on the next deal as compensation. Or the intermediary may provide services in lieu of direct compensation, for example, it could arrange a loan if the customer is facing liquidity problems because the last deal turned out badly. Or the intermediary may pay for the cost of some action to make the customer whole after it suffers some damage because of an unanticipated outcome from a previous transaction. On occasion the customer may require compensation in cash. In some cases this may involve legal action, but this does not have to be the norm.

4.2. Risk, Uncertainty, and Implicit Insurance

To clarify some of the ideas introduced above, it will be necessary to have a slightly richer model. Suppose that f is the security traded in the market and that the outcome of f depends on two types of contingencies or states. The first type of state we call *salient* and the second we call *obscure*. The salient states consist of contingencies the individual investor or firm is likely to think of and understand when evaluating a security; the obscure states consist of those which are not consciously brought to mind at the time a decision is made. Ex post, both kinds of states are observable. In fact, we may consider both to be salient ex post because it will be clear that they have affected the outcome in a particular way. Let S denote the finite set of salient states. For each state $s \in S$ the finite set of obscure states is $T(s)$. The probability of a salient state s is $p(s)$ and the conditional probability of an obscure state t given s is $\pi(t|s)$. It is important to note that the division of contingencies into salient and obscure states is subjective, that is, it depends on the firm or individual in question and may vary from one to another.

For simplicity, we assume that the customer's income risk depends only on the salient states and that the customer has well-defined probabilities attached to these states. The customer behaves like a von Neumann-Morgenstern expected utility maximizer with respect to states $s \in S$ where $T(s)$ contains one element so there is no obscurity. On the other hand, when $T(s)$ contains more than one element the obscure states represent a source of uncertainty in the Keynesian sense, precisely because they are not being consciously scrutinized. It is less clear how the cus-

tomer should respond to this uncertainty. We assume that for obscure states the customer attaches a very low utility denoted $\underline{u}$. There are a number of formal ways this can be justified. For simplicity and the sake of illustrating the general idea, suppose that the customer acts like an infinitely risk-averse individual when confronted with this kind of uncertainty; that is, if $f(s, t)$ is the payoff to the security in the combined state (s, t), then the customer acts as if the outcome were $f_0(s) \equiv \inf_{t \in T(s)}\{f(s, t)\}$ in state s and proceeds to calculate the expected utility of the security f_0 in the usual way. The utility of f_0 in state s is $\underline{u}(s)$. This rather pessimistic way of evaluating the security is not likely to make the customer very keen to trade the security. In many cases, the optimum response may be to choose the quantity $\theta = 0$. Note that infinite risk aversion does not lead the customer to sell the security short: When $\theta < 0$ the customer's pessimism will take the form of replacing $f(s)$ with $f_0(s) \equiv \sup_{t \in T}\{f(s, t)\}$. Hence, unlike standard expected utility theory the model is consistent with the fact that a large proportion of potential investors do not participate in standard financial markets at all (cf. Dow and Werlang 1992). For example, Mankiw and Zeldes (1991) find that only a small proportion of investors participate in the stock market; in their sample of those with liquid assets in excess of $100,000 only 47.7 percent held stocks.

The existence of obscure states and the uncertainty to which they give rise is a barrier to the participation of unsophisticated individuals in financial markets. But why should securities depend on obscure states in the first place? There are at least two reasons. The first is that it may not be possible to provide securities that avoid obscure states altogether. Salience and obscurity are subjective: What is salient to one individual may not be salient to another because of a difference in background or expertise. Second, the incompleteness of the set of securities exacerbates the problem by reducing the set of securities from which the individual has to choose. The formulas that can be written down in a legally enforceable contract may inevitably involve contingencies that are obscure to some individuals; the case law that is likely to determine how contracts are enforced in practice may be obscure to all but a few lawyers specializing in an area, etc.

This is where the intermediary can provide a valuable service. It can insure the individual or firm against obscure contingencies, thus reducing or eliminating the uncertainty that lowers the value of the security. Note that this does not require the intermediary to write a contract that is in some sense more detailed or complete than the security offered in the market. The insurance can be implicit in the relationship between the intermediary and the customer. What the intermediary is offering the customer is a substitute contract that pays the customer the certain amount $f_1(s)$ in state s rather than the uncertain amount $f(s, t)$. For example, $f_1(s)$ may be the mean of $f(s, t)$ according to the intermediary's beliefs about the distribution of t. This security f_1 can do a perfectly adequate job of hedging against the risk in $y(s)$, which depends only on the salient state s. Of course, the legally binding contract is f; the difference between f and f_1 is something that the customer cannot quite understand, but that does not matter since only f_1 needs to be taken into account in the outcome. The important point is that the implicit contract is based on states that the customer and the intermediary can both fully distinguish. The extent to which customers can rely on the implicit contract will depend on the profitability of the ongoing relationship versus the benefit of reneging as before. The difference is that now customers are unaware of the benefit of reneging. In making their decision customers must rely on the information available to them, such as previous experience on reneging with contracts of this type.

In the next subsection, we extend this simple view of implicit insurance to allow for information acquisition.

4.3. The Operation of Markets with Ex Ante Information

Consider an (implicit) contracting problem between an intermediary and a customer. As before, suppose that the intermediary is designing a risk-sharing arrangement with the customer. The intermediary has the expertise to identify both salient and obscure states. The customer can only identify salient states. This has the important implication that explicit contracts cannot be used to provide insurance to the unsophisticated against

obscure states. The customer has no idea how to interpret such contracts and will simply assign a low utility to them. Implicit contracts must therefore be used to insure against outcomes in obscure states.

Consider the use of securities traded in the market first. As usual, the unsophisticated customer faces a random income $w(s)$ in state s so if it trades the security f its net income will be $f(s, t) + w(s)$ in state (s, t). We ignore other income for the intermediary in the usual way, so its income is simply $-f(s, t)$ in state (s, t). With no further information, the expected utilities of the customer and intermediary, conditional on trading the randomly chosen security f, will be

$$E[u(f(s) + w(s))] \quad \text{and} \quad E[v(-f(s))],$$

respectively. There are two sources of uncertainty here, of course, one arising from the uncertainty about the salient state and the other arising from uncertainty about the obscure state t. The uncertainty about the salient state of nature will be resolved at some time in the future and we shall have no more to say about it here. On the other hand, we assume that for some cost, information is available to customers about the obscure states t for each salient state s. One way of capturing the possibility of reducing uncertainty about the nature of $f(s, t)$ is to assume that, for some fixed cost $c > 0$ and for some state s, the customer can observe the true value of $f(s, t)$ for all states t. If the number of salient states is large it will never be optimal to pay this cost and learn the entire security $f(s, t)$; but, depending on the circumstances, it may pay to learn in some states. For each state s where information is acquired it is possible to calculate expected utility in the standard way. For other states where the cost c is not incurred, the utility assigned is $\underline{u}(s)$.

The problem of evaluating f can be regarded as a search problem. If the security is evaluated in a subset of states $\Phi \subset S$ then the expected utility of the contract to the customer is

$$U(f|\Phi) = \sum_{s\in\Phi} \sum_{t\in T(s)} p(s)\pi(t|s)u(w(s) + f(s, t) - \#\Phi c) + \sum_{s\in S\setminus\Phi} p(s)\underline{u}(s),$$

where $\#\Phi$ denotes the number of elements of Φ. The value of information comes from the possibility of making the trade conditional on the information, in which case the customer gets the maximum of $U(f|\Phi)$ and $E[u(w(s) - \#\Phi c)]$. The optimal choice of Φ maximizes

$$E[\max\{U(f|\Phi), E[u(w(s) - \#\Phi c)]\}].$$

Investing in information only makes sense if it sometimes changes the decision to buy the security. If it does not, there is no gain and the customer would be better off not paying the cost of evaluating the contract. For this reason, there is the usual well-known nonconvexity in the objective function. If there is a large number of individually insignificant states of nature, information about any one state is unlikely to make a difference to the decision whether to accept the contract or not. Therefore, a small amount of information, for example, information about the contract's payoff in a single state, may have no value. Clearly, an intermediate amount of information may affect the decision and therefore has a positive value. So there must be increasing returns to information over some range. On the other hand, when a large amount of information has been collected, getting information about an additional state is unlikely to have much effect on the decision, so the marginal value of information is decreasing beyond some point.

For $c = 0$ complete contingent contracts will be optimal. Both sides will understand exactly how the contract will work in each state and all possibilities for risk sharing will be exploited by the contract. In that case relationships will have no role to play.

For moderate c, the security will be evaluated at states which are likely to occur but not at other states. Ex post there will be "surprises" in the sense that for those states at which the security was not evaluated, the utility obtained will be different from that expected. The optimal expenditure on evaluating how the contract will work involves trading off the transaction costs against the risk of surprises.

For sufficiently high c, the security will never be evaluated. In this case the basic underlying risk (salient states) will be compounded by the uncertainty about how the contract will work (obscure states). Surprises will occur frequently.

Depending on the level of c and the cost of surprises in terms of the risk, there may be no trade at all if

$$E\pi[u(f(s) + w(s))] + E\pi[v(-f(s))]$$
$$< E[u(w(s)] + v(0).$$

This might be true in spite of the fact that the security does have some genuine risk-sharing benefits, simply because of the customers' uncertainty about the actual outcome in some states and the costs of eliminating this uncertainty.

4.4. Intermediaries and Ex Ante Versus Ex Post Security Evaluation

The essential problem with using contracts in the framework developed above is that the costs of evaluation are borne ex ante. This means that the security is evaluated in many states even though ex post only one will occur. This raises the obvious possibility of improving the allocation of resources by dispensing with ex ante evaluation and having some type of insurance should a surprise occur. The difficulty, of course, is how this type of insurance can be provided. Explicit contracts based on obscure information cannot be evaluated by unsophisticated investors. When confronted with contracts which attempt to condition on obscure states they will simply assign $\underline{u}$ unless they incur the costs of evaluation. This is where the implicit insurance allowed by relationships is important. Customers may be willing to forego ex ante evaluation because they understand that if there is a problem the intermediary will eliminate or share the risk of any surprises. As a result they may be willing to trade even though explicit contracts alone would not provide any surplus. The implicit ex post insurance allowed by relationships means that the ex ante costs of complex explicit contracts can be avoided.

To illustrate this in terms of the model developed, suppose the explicit written contract provides a payoff $f(s, t)$. The implicit contract is such that the net payment to the customer is

$$f_1(s) = f(s, t) - h(s, t).$$

In other words, the customer is provided with a payoff that can be understood ex ante. There is no need to undertake complex evaluation ex ante. The implicit insurance provided by the intermediary means customers do not have to incur the costs of investigating obscure possibilities. By insuring against these possibilities the intermediary can potentially provide a huge savings in costs and/or increase the funds people are willing to invest in markets.

There are various ways in which implicit insurance contracts can be executed. After an unexpected outcome, the aggrieved party may engage in bargaining with the counter-party, threatening to take his business elsewhere or to create unfavorable publicity until the intermediary makes amends. In some cases, a legal action may be brought against the intermediary. Whatever form the "renegotiation" takes, it is part of the implicit contract that is played out over time. We can imagine that all of this is anticipated by the parties when they write the explicit contract. The real contract is the anticipated mapping from states of nature to payoffs, not the written contract, and it is the real contract that has to be evaluated and enforced. Formally, this is equivalent to an explicit contract with the same contingent payoffs, but the crucial difference is the cost of writing and enforcing the contract. The long-term relationship with its implicit insurance is just a cheaper way of achieving a particular pattern of state-contingent payoffs.

5. Strategies for Implementing Relationships

The previous sections have shown how the ex ante costs of explicit contracts can sometimes be avoided by the development of long lasting relationships and implicit insurance between intermediaries and firms or consumers. A question of some interest to providers of financial services is how relationships can best be structured to maximize these benefits from implicit insurance.

In the U.S. in recent years, a number of intermediaries have tried to construct "financial supermarkets" where consumers can obtain a whole range of financial services from the same providers. For example, in the 1980s Sears bought Dean Witter and a number of other specialist financial service firms. At one point it was offering a full range of services including FDIC insured deposits, consumer loans, credit cards, mortgage banking, commercial lending, mutual funds, brokerage services for securities, and insurance. American Express combined with Shearson and Lehman and a number of other firms and offered a similar range of financial services. Ultimately, these strategies were unsuccessful in the sense that the

groups were broken up and returned to competing separately. Prudential thought that insurance agents and stock brokers could sell each other's products and merged with Bache. However, they do not appear to have been particularly successful at implementing this idea. A recent and extreme example of the trend to provide consumers with a full range of financial products from one firm is the combination of Citibank and the Travelers Group to form Citigroup. In addition to attempts to extend relationships between intermediaries and consumers there have also been attempts to extend relationships between intermediaries and firms. One example of this is the recent creation of Section 20 subsidiaries to perform investment banking services by many large commercial banks. The merger between Banker's Trust and Alex Brown in 1997 is another. Citigroup may benefit from the combination of investment banking and commercial banking allowed by links between Salomon Smith Barney within Travelers and Citibank.

In many European countries such as Germany it has long been the case that universal banks have provided a whole array of services to both consumers and firms. On the consumer side, in addition to standard banking accounts and loans, insurance and other types of financial service are offered. On the firm side, loans and underwriting services are often provided. Relationships are wider in scope and much more important than in the U.S.

The model of relationships developed above can be used to shed some light on when strategies of extending the scope of relationships are likely to be successful. The result of §3 that increased competition reduces the amount of implicit insurance that is possible with relationships indicates that it is essentially the available future surplus that is important. The more future surplus at stake the higher the transfers that are incentive compatible and hence the more implicit insurance there can be. By extending the scope of relationships and putting a greater amount of future surplus at stake, the more valuable the relationships can be made. For example, suppose a bank is providing loans to consumers and a mutual fund is providing investment products. Both have long-term relationships with their customers. The extent of risk sharing through implicit contracts that can be achieved by both firms is limited by the future stream of profits each will separately expect to earn in the future. To the extent the current outcomes in each, and hence the need for transfers under the implicit insurance arrangement, are independent, there will be a gain to combining the two and forming a more extensive relationship with customers. Now the transfer to provide the implicit insurance that can take place in each division will depend on the future profits in the combined entity. In states where both divisions need transfers simultaneously it cannot be worse to have the two combined and in states where only one division requires a transfer a strict improvement in the implicit insurance can be obtained.

This argument is consistent with the observation that extensive relationships involved in universal banking have been successful in Europe while financial supermarkets have been unsuccessful in the U.S. In Europe there is not much competition in financial services while in the U.S. there typically is. Combining relationships will only be successful to the extent that large future streams of profits are put at risk, which happens in Europe but not in the U.S.

In order for intermediaries in the U.S. to be successful when creating broader relationships, the objective should be to put together profitable services and not just give consumers a wide range of products. The financial supermarkets that have been constructed have really just given consumers convenience through "one stop shopping"; they have not provided more implicit insurance because each individual line is fairly competitive. For consumers, one stop shopping for financial services does not appear to be very valuable and thus the mergers have been unsuccessful. The results above suggest that combining commercial banking and investment banking in the U.S. may be more successful. The reason is that there are areas of both which are highly profitable and hence extended implicit insurance should be possible.

6. Concluding Remarks

Financial markets have become increasingly important over the last few decades. However, they have mostly become markets for intermediaries rather than markets for individuals and firms. We have argued

that one of the main reasons for this change is that increased participation costs in markets have meant that individuals and firms have withdrawn from markets and rely increasingly on intermediaries. If complex complete contracts were used between intermediaries and their customers there would be no gain because of the extensive ex ante costs of evaluating how the contracts will operate. Instead, customers can rely on the implicit insurance made possible by long-term relationships. Households and firms know that if there is a surprise, the intermediary will share the risk. Such risk sharing is only possible if both parties will benefit from the relationship in the future. This means that competition by intermediaries may be undesirable if it reduces future profits and hence the amount of risk sharing that can occur. It also suggests that when choosing strategies to maximize the benefits of relationships firms should be concerned with offering a range of profitable services rather than simply expanding the number of services available.

Throughout the discussion of explicit and implicit contracts, we took the explicit written contracts as exogenously given. This is obviously a gross oversimplification. There are complex issues to be discussed about how these contracts are written and how they interact with the implicit contract. Only one issue will be mentioned here, but it illustrates a wider class of issues. There are gains to locking in the customer in order to provide better insurance to both the customer and the intermediary. One way of doing this is by getting the customer to post a bond by paying fees and commissions up front. These can be thought of as an "insurance premium" that the customer will forfeit if it switches to another intermediary or reneges on the implicit contract and reverts to the written contract. However, the customer may have difficulty paying this premium up front because it lacks liquidity. In that case, the written contract can easily be amended to provide a stream of payments to the intermediary. Since this stream of payments is noncontingent, there is no problem writing it into the contract and since the written contract is enforceable it serves as a bonding device. Further, since the payments are spread over time it solves the firm's liquidity problem. However, this change in the contract makes the ex ante value of the written contract lower and so reduces the value of the relationship and may adversely affect the ex post participation constraint. So the problem of contract and security design cannot be separated from the problem of providing optimal implicit insurance. We do not pursue this issue here since it raises the question of how to write down the problem in a satisfactory way, for example, to specify the constraints on what contracts can be written and what contracts cannot. This is an important problem for future research.[6]

[6] We thank Tony Santomero, Charles Kahn and other participants at the Wharton Financial Institution's Conference on *Performance of Financial Institutions* in May 1997, Stavros Zenios (the editor), and particularly an anonymous referee for helpful comments. Financial support from the National Science Foundation and the Wharton Financial Institutions Center is gratefully acknowledged.

References

Allen, F., A. M. Santomero. 1997. The theory of financial intermediation. *J. Banking and Finance* **21** 1461–1486.

——, D. Gale. 1994. Limited market participation and volatility of asset prices. *Amer. Econom. Rev.* **84** 933–955.

Boot, A. W. A., A. V. Thakor. 1993. Security design. *J. Finance* **48** 1349–1378.

——, ——. 1996. Can relationship banking survive competition? Mitsui Life Financial Research Center Working Paper 96-6, University of Michigan, Ann Arbor, MI.

Diamond, D. W. 1984. Financial intermediation and delegated monitoring. *Rev. Econom. Stud.* **51** 393–414.

Dow, James, Sergio Ribeiro da Costa Werlang. 1992. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica* **60** 197–204.

Duffie, D., P. M. DeMarzo. 1995. A liquidity-based model of market design. Working Paper, Stanford University, Stanford, CA.

Gale, D. 1992. Standard securities. *Rev. Econom. Stud.* **59** 731–755.

Mankiw, N. G., S. P. Zeldes. 1991. The consumption of stockholders and non-stockholders. *J. Financial Econom.* **29** 97–112.

Merton, R. C. 1989. On the application of the continuous-time theory of finance to financial intermediation and insurance. *Geneva Papers on Risk and Insurance Theory* **14** 225–261.

Pagano, M. 1989. Endogenous market thinness and stock price volatility. *Rev. Econom. Stud.* **56** 269–288.

Rajan, R. 1992. Insiders and outsiders: the choice between informed and arm's length debt. *J. Finance* **47** 1367–1400.

Sharpe, S. 1990. Asymmetric information, bank lending and implicit contracts: a stylized model of customer relationships. *J. Finance* **55** 1069–1087.

Accepted by Stavros A. Zenios; received May 1997. *This paper has been with the authors* 2 *months for* 2 *revisions.*