



ELSEVIER

Journal of Banking & Finance 21 (1998) 1461–1485

Journal of
BANKING &
FINANCE

The theory of financial intermediation

Franklin Allen, Anthony M. Santomero *

The Wharton School, University of Pennsylvania, Philadelphia, PA 19096, USA

Abstract

Traditional theories of intermediation are based on transaction costs and asymmetric information. They are designed to account for institutions which take deposits or issue insurance policies and channel funds to firms. However, in recent decades there have been significant changes. Although transaction costs and asymmetric information have declined, intermediation has increased. New markets for financial futures and options are mainly markets for intermediaries rather than individuals or firms. These changes are difficult to reconcile with the traditional theories. We discuss the role of intermediation in this new context stressing risk trading and participation costs. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: G2; G1; E5; L2

Keywords: Intermediation; Risk management; Delegated monitoring; Banks; Participation costs

1. Introduction

In this paper we review the state of intermediation theory and attempt to reconcile it with the observed behavior of institutions in modern capital markets. We argue that many current theories of intermediation are too heavily focused on functions of institutions that are no longer crucial in many devel-

* Corresponding author. Tel.: +1 215 898-7625; fax: +1 215 573-8757; e-mail: santo@finance.wharton.upenn.edu.

oped financial systems. They focus on products and services that are of decreasing importance to the intermediaries, while they are unable to account for those activities which have become the central focus of many institutions. In short, we suggest that the literature's emphasis on the role of intermediaries as reducing the frictions of transaction costs and asymmetric information is too strong. The evidence we offer suggests that while these factors may once have been central to the role of intermediaries, they are increasingly less relevant.

We offer in its place a view of intermediaries that centers on two different roles that these firms currently play. They are facilitators of risk transfer and deal with the increasingly complex maze of financial instruments and markets. Risk management has become a key area of intermediary activity, though intermediation theory has offered little to explain why institutions should perform this function. In addition, we argue that the facilitation of participation in the sector is an important service provided by these firms. We suggest that reducing participation costs, which are the costs of learning about effectively using markets as well as participating in them on a day to day basis, play an important role in understanding the changes that have taken place.

The paper proceeds as follows. In Section 2, we offer a review and critique of the usual views of intermediation found in the literature. This critique is supported by data presented in Section 3, which outlines the changes in financial systems that have occurred over the recent past. In Section 4 the current justifications for one of the growth areas of intermediary services, namely risk management, are presented, while Section 5 discusses the risk reduction activities that intermediaries should take. Section 6 then outlines the importance of participation costs as another rationale for intermediation and assisting in risk management. Finally, Section 7 contains concluding remarks.

2. Review and critique of current intermediation theory

In the traditional Arrow–Debreu model of resource allocation, firms and households interact through markets and financial intermediaries play no role. When markets are perfect and complete, the allocation of resources is Pareto efficient and there is no scope for intermediaries to improve welfare. Moreover, the Modigliani–Miller theorem applied in this context asserts that financial structure does not matter: households can construct portfolios which offset any position taken by an intermediary and intermediation cannot create value (see Fama, 1980).

A traditional criticism of this standard market-based theory is that a large number of securities are needed for it to hold except in special cases. However, the development of continuous time techniques for option pricing models and the extension of these ideas to general equilibrium theory have negated this

criticism. Dynamic trading strategies allow markets to be effectively complete even though a limited number of securities exist.

Such an extreme view – that financial markets allow an efficient allocation and intermediaries have no role to play – is clearly at odds with what is observed in practice. Historically, banks and insurance companies have played a central role. This appears to be true in virtually all economies except emerging economies which are at a very early stage. Even here, however, the development of intermediaries tends to lead the development of financial markets themselves (see McKinnon, 1973).

In short, banks have existed since ancient times, taking deposits from households and making loans to economic agents requiring capital. Insurance, and in particular marine insurance, also has a very long history. In contrast, financial markets have only been important recently, and then only in a few countries, primarily the UK and the US. Even there, banks and insurance companies have played a major role in the transformation of savings from the household sector into investments in real assets.

Our understanding of the role or roles played by these intermediaries in the financial sector is found in the many and varied models in the area known as intermediation theory. These theories of intermediation have been built on the models of resource allocation based on perfect and complete markets by suggesting that it is frictions such as transaction costs and asymmetric information that are important in understanding intermediation. Gurley and Shaw (1960) and many subsequent authors have stressed the role of transaction costs. For example, fixed costs of asset evaluation mean that intermediaries have an advantage over individuals because they allow such costs to be shared. Similarly, trading costs mean that intermediaries can more easily be diversified than individuals.

Looking for frictions that relate more to investors' information sets, numerous authors have stressed the role of asymmetric information as an alternative rationalization for the importance of intermediaries. One of the earliest and most cited papers, Leland and Pyle (1977), suggests that an intermediary can signal its informed status by investing its wealth in assets about which it has special knowledge. In another important paper, Diamond (1984) has argued that intermediaries overcome asymmetric information problems by acting as "delegated monitors." Many others followed, expanding on these two contributions and advancing the literature in substantive ways (e.g., see Gale and Hellwig, 1985; Campbell and Kracaw, 1980; Boyd and Prescott, 1986).

Bhattacharya and Thakor (1993) have provided an excellent survey of the current state of the literature on banking, building on an earlier review of the banking literature published in Santomero (1984). Dionne (1991) contains a set of surveys of the literature on insurance. Readers wishing detailed accounts of particular literatures should consult these papers. Our contribution here will not be a duplication of these efforts. Rather, it will attempt to contrast

the traditional view of the role and functions performed by intermediaries with the evolution of these institutions over the last few decades. It is an attempt to confront the literature with a view of the practice to see if the literature adequately addresses the reasons that these institutions exist in the financial markets, and how they perform value added activity.

The reality is that the financial systems in many countries have undergone a dramatic transformation in recent years. Financial markets such as the stock and bond markets have grown in size using nearly any metric, such as the value of companies listed or any other conceivable measure of their importance. At the same time, there has been extensive financial innovation acceleration in the 1970s and 1980s. This includes the introduction of new financial products, such as various mortgage backed securities and other securitized assets, as well as derivative instruments such as swaps and complex options. These have all had a virtual explosion in volume. At the same time, new exchanges for financial futures, options and other derivative securities have appeared and become major markets.

Interestingly, this increase in the breadth and depth of financial markets has been the result of increased use of these instruments by financial intermediaries and firms. They have not been used by households to any significant extent. In fact, the increased size of the financial market has coincided with a dramatic shift away from direct participation by individuals in financial markets towards participation through various kinds of intermediaries.

The importance of different types of intermediary over this same time period has also undergone a significant change. The share of assets held by banks and insurance companies has fallen, while mutual funds and pension funds have dramatically increased in size. New types of intermediary such as non-bank financial firms like GE Capital have emerged which raise money entirely by issuing securities and not at all by taking deposits. In short, traditional intermediaries have declined in importance even as the sector itself has been expanding.

Perhaps in response, but clearly contemporaneously, the activities of traditional institutions such as banks and insurance companies have also changed. Banks which used to take deposits and make loans found that the possibilities for securitizing loans meant that they did not need to keep on their balance sheet all the loans they could originate. At the same time, insurance firms realized that their actuarial function was but a minor part of their asset management capabilities and these firms too innovated and broadened their products and services.

Some of these changes in the volume of financial activity, along with the relative importance of some institutions and the changes in others, can be explained using traditional theories which are based on transaction costs and asymmetric information. But, others cannot. For example, the standard explanation for the existence of mutual funds is that, while diversification is desir-

able, the high costs of trading for individuals makes it expensive for individuals to achieve this. Mutual funds can trade at significantly lower cost and so can achieve diversification much more cheaply. Given this explanation it might be expected that if individuals' trading costs were lowered the share of ownership of mutual funds would fall. However, we have not observed this behavior. Although with the introduction of competition for brokerage fees on the NYSE in the early 1970s trading costs for individuals fell dramatically, the share of assets invested in mutual funds has risen significantly. Likewise, the advent of the technological revolution has substantially reduced the cost of information and reduced information asymmetry. Yet it did not reduce the need for intermediary services and encourage direct lending by households. In fact, the data suggest the opposite. In short, the decline in frictions which were allegedly the market imperfections that led to a need for intermediation services has not reduced the demand for them. Intermediation is growing and prospering even as the frictions decline.

The contrast between theory and reality is perhaps most apparent in the area of risk management. Arguably the most important change in intermediaries' activities that has occurred in the last thirty years is the growth in the importance of risk management activities undertaken by financial intermediaries. As we noted above, the change in the breadth of the markets that are available for hedging risk has not led very many individual or corporate customers to manage their own risk. Rather, it has meant that risk management has now become a central activity of many intermediaries. Most current theories of intermediation have little to say about why risk management should play such an important role in the activities of intermediaries.

In some cases, theories explaining why both financial and non-financial firms should undertake risk management have been added on to our understanding of firm level decision making. However, these descriptions of why they undertake hedging activities are almost an afterthought in the literature. Little is offered as a cogent argument as to why intermediaries should be the ones offering these services, and what value they bring to the activity. In short, the intermediation literature is noticeably quiet as to why these institutions should be engaged in one of their central areas of activity.

An important exception is the work of Merton and Bodie (see, in particular, Merton, 1989, 1993; Merton and Bodie, 1995 for a recent overview). The main theme of this contribution to the literature is to suggest that financial systems should be analyzed in terms of a "functional perspective" rather than an "institutional perspective." A functional perspective is one based on the services provided by the financial system, such as providing a way to transfer economic resources through time. In contrast, an institutional perspective is one where the central focus is on the activities of existing institutions such as banks and insurance companies. The argument in favor of focusing on the functional rather than the institutional perspective is that over long periods of time func-

tions have been much more stable than institutions. This has clearly been a characteristic of the intermediary sector in the recent past. Institutions have come and gone, evolved and changed, but functional needs persist while packaged differently and delivered in substantially different ways.

This constancy of functional needs has led Oldfield and Santomero (1997) to argue that financial services such as origination, distribution, servicing and funding are more stable than either the institutions that provide services or the specific products they offer in order to satisfy customer requirements. The financial services may be packaged differently both across competitive institutions and over time, but the functions are far more stable.

Using this functional approach to the financial sector, the literature that explains its activities can be seen as focusing on one or another function performed by it. The literature on transactions costs can be seen to be rationalizing the role of these institutions in the distribution function. The emphasis on information asymmetries centers on both the origination and the servicing function. To this list Merton (1989) adds another role for the sector. He argues that another central feature of the sector is its ability to distribute risk across different participants. A model is developed where the key value added of intermediaries is that they provide the function of allowing risk to be allocated efficiently at minimum cost.

Merton notes that intermediaries can transact at near zero cost while individuals have high trading costs. This means that intermediaries can create a large number of synthetic assets through dynamic trading strategies. By hedging appropriately, they can create products with very safe payoffs which Merton argues are particularly valuable to some intermediaries' customers. Alternatively they can engineer products with varying degrees of complexity if their customers need such securities. This addition to the list of services provided by the financial sector is quite consistent with its observed activities of late. They have increasingly been focused on the trading of risk and the bundling and unbundling of the risks of various financial contracts. To see this trend in their activities, let us review the recent history.

3. Recent changes in markets and intermediaries

It is widely acknowledged that there has been an unprecedented amount of financial innovation in recent years (see, e.g., Miller, 1986). However, financial innovation has been occurring for many centuries albeit at a slower pace. Allen and Gale (1994a) offer a detailed historical account of financial innovation. They point out that numerous different types of instruments were developed over time but relatively few survived. By the 1930s what might be termed the traditional financial instruments had been developed and had demonstrated some robustness. These instruments are outlined in Table 1.

Table 1
Traditional financial instruments

Issuer	Instrument	Characteristics
Governments	Bonds	A long-term obligation by the firm to make a series of fixed payments
	Notes	An intermediate obligation
	Bills	A short-term obligation
Banks	Deposits	Funds deposited at a bank available on demand or with some delay
	Acceptances	A written promise to pay a given sum at a prespecified date
Firms	Equity	Equityholders are the owners of the firm and are responsible for conducting its affairs
	Bonds	A long-term obligation by the firm
	Convertibles	A bond that can be swapped for equity at a prespecified ratio or vice versa
	Preferred stock	A hybrid security that combines features of debt and equity
	Commercial paper	A short-term debt security issued by firms that can be easily traded
Exchanges	Warrants	A long-term call option on a firm's stock issued by the firm
	Commodity futures	Contracts for future delivery of a commodity

In practically all countries, government securities have been the most important type of instrument traded in financial markets. In the 19th and early 20th century banks and insurance companies played the major role in transforming savings by households into investment in real assets by firms. Banks would take deposits from households and make loans to firms; insurance companies would issue policies and lend the proceeds to firms or invest in security markets. In the US and UK markets for securities issued by firms were significant in terms of the assets outstanding, but in most other countries this was not the case until the post-war era. The financial markets in the US and UK, particularly the equity and bond markets, were predominantly participated in by individuals rather than intermediaries. In the US, in addition to the equity and bond markets, there were also the exchanges in Chicago where commodity futures were traded starting in the mid-19th century.

Markets for traditional instruments have grown substantially in recent years. This is not only in absolute terms but also in relative terms. For example, Fig. 1 shows market capitalization of corporate equity as a percentage of GDP from 1975–1994 for the US. It can be seen that there has been a distinct long term increase in market capitalization relative to GDP.

However, even as financial assets were increasing, another trend in the data has become clear. A notable feature of markets for traditional instruments in the past few decades has been the drop in use by individuals. As Fig. 2 indicates the ownership of corporate equity by individuals in the US has fallen

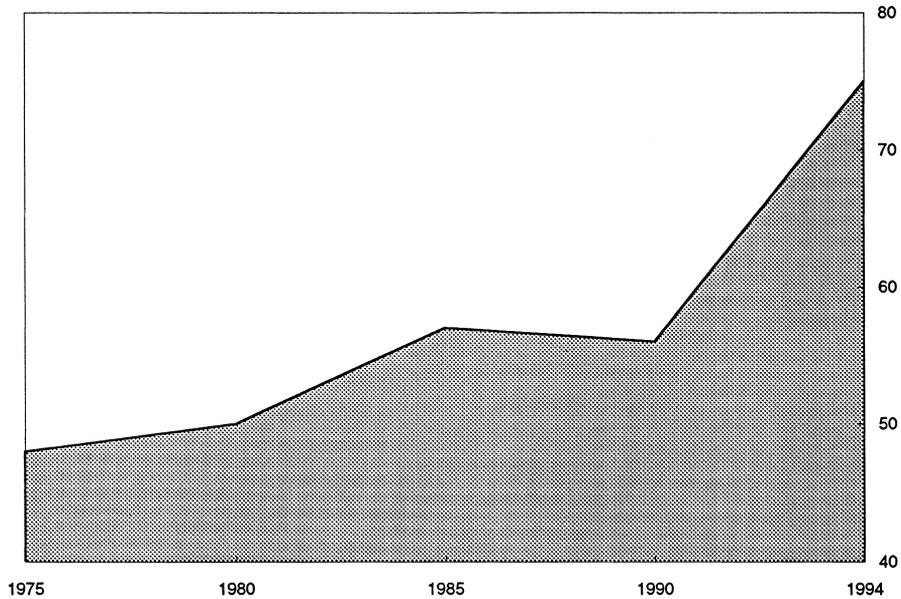


Fig. 1. US market capitalization of corporate equity (% of GDP).

from about 85% in the mid-1960s to around 50% in recent years. At the same time Fig. 3 shows how the share of mutual and closed end funds, pension funds and insurance companies has correspondingly changed. The amount of financial claims held directly by households has clearly fallen dramatically. Intermediation has become significantly more important and has been the predominant source of new financial resources flowing into the capital markets over the past several decades. Fig. 4 shows how the ratio of mutual fund holdings to household equity ownership has risen from about 5% in 1980 to around 25% by 1995.

Starting in the 1960s but primarily in the 1970s and 1980s, the markets themselves have changed significantly. Table 2 shows some of the most important innovations that have occurred (see Allen and Gale, 1994a for a detailed account). Arguably the most successful type of innovation has been the development of various kinds of derivative securities which have been introduced over this period. This includes the financial futures and options listed on exchanges and the new over-the-counter (OTC) instruments such as swaps.

Standardized markets for financial futures and options started with the introduction of foreign currency futures at the International Monetary Market (IMM), (which is part of the Chicago Mercantile Exchange) in 1972. Other types of futures contracts were introduced in the following years. Successful examples were the IMM's ninety-one day T-bill contract and the Chicago Board of Trade's (CBOT) Treasury bond contract which were introduced in 1976 and

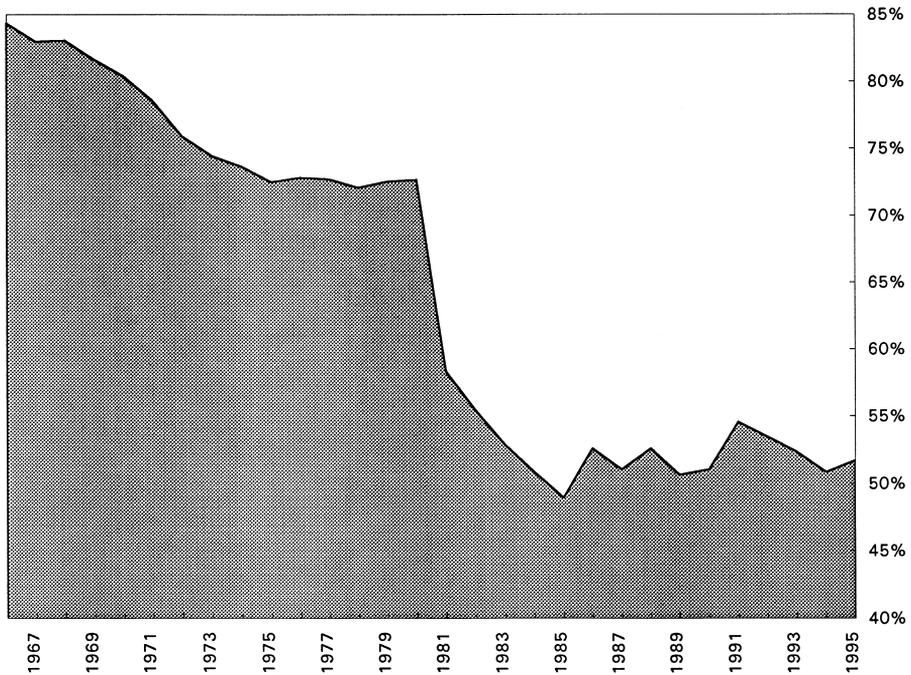


Fig. 2. US – Individual ownership of corporate equity 1966–1995.

1977, respectively. Other countries introduced financial futures markets in the 1980s with the London International Financial Futures Exchange (LIFFE) being started in 1982 and the Tokyo Futures Exchange (TFE) in 1985.

The first standardized options were introduced in 1973 by the Chicago Board Options Exchange (CBOE). The CBOE was immediately successful, and by 1984 it had become the second largest securities market in the world with only the New York Stock Exchange being larger. This success led other US and foreign exchanges to introduce options exchanges. These included the American Stock Exchange, the Philadelphia Stock Exchange, the European Options Exchange in Amsterdam and the London Stock Exchange.

In addition to the development of exchange traded derivatives there has also been a huge increase in the volume of OTC derivatives, particularly swaps. The first swaps were currency swaps and occurred in the 1960s as a way for UK firms to circumvent exchange controls. They involved swapping a stream of payments in one currency for a stream of payments in another currency. The basic techniques developed for currency swaps were then applied in other contexts, most importantly in swapping fixed rate loans for adjustable rate loans.

Table 3 shows the notional amounts outstanding as of 31 March 1995 and the turnover based on April 1995 data for exchange traded derivatives and

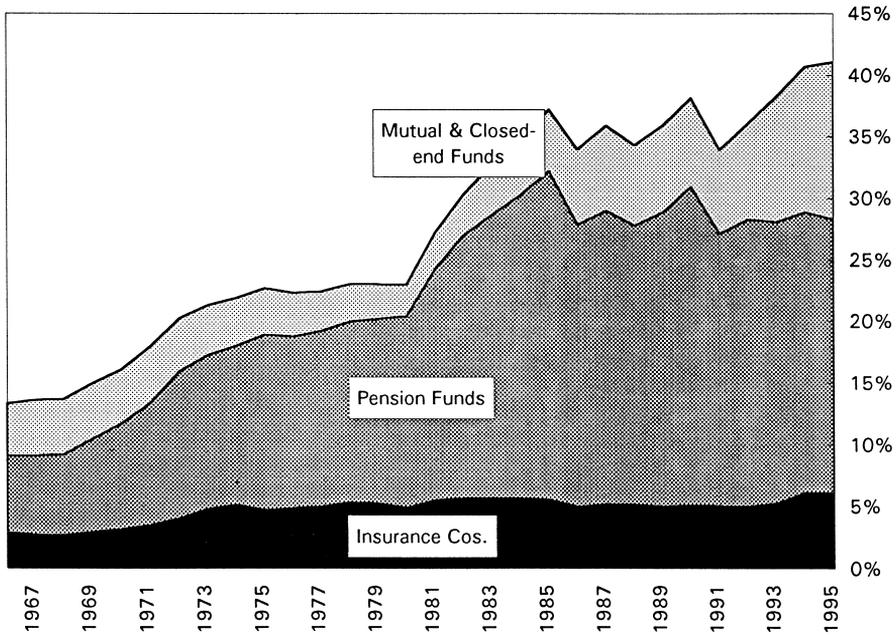


Fig. 3. US – Selected institutional share of corporate equity 1966–1995.

OTC derivatives. The gross market values are also provided for OTC derivatives. The data is given on a global basis. It can be seen that both exchange traded and OTC markets are large in terms of all these measures.

Perhaps the most interesting feature of these markets, however, is who it is that uses them. Fig. 5 shows that financial institutions are the primary users of the OTC markets accounting for 82% of its volume compared to 18% for all other participants. This latter group includes the sum total of all non-financial firms, governments and individuals. These figures indicate that there is a huge amount of derivatives trading by intermediaries. These institutions are buying and selling the unbundled state contingent cash flows associated with financial claims among themselves and on behalf of their clients. In essence they are actively trading risk to and for their clients for risk management purposes.

In addition to derivatives, the other major innovation mentioned in Table 2 that has been successful is securitized loans. As is well known, this market began with the developments in the mortgage market. The market for mortgage-backed securities in the US dates back to the 1950s at least but it was not until the 1970s that it became important in terms of the volume outstanding. The critical development was the introduction of “pass-through” securities by the Government National Mortgage Association (GNMA or Ginnie Mae) in 1970. These allowed shares in a pool of mortgages to be freely traded without

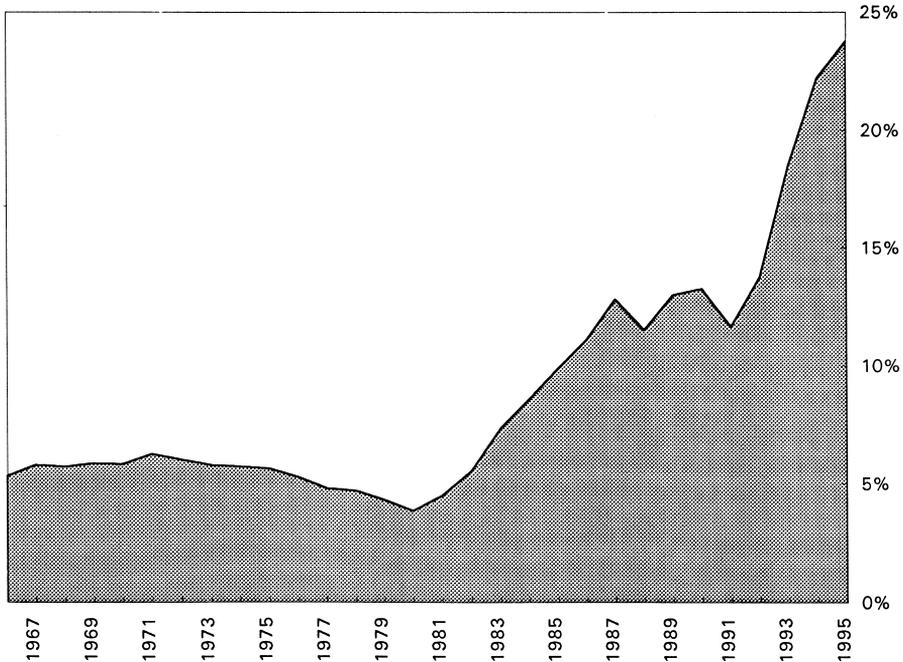


Fig. 4. US: Mutual fund/Householder equity ownership 1966–1995.

Table 2
Recent financial innovations

Main Issuer	Instrument	Characteristics
Exchanges	Financial futures	Contracts for the future delivery of currencies, securities, or an amount of money based on an index
	Options	The right to buy or sell a security on or before a specified date
Banks	Swaps	Transactions in which different streams of income are exchanged
Governments	Securitized loans	Pools of mortgages or other types of loans that are publicly traded
Firms	Floating-rate debt	The interest rate on the debt is based on LIBOR, the T-bill rate or some other index
	Floating-rate preferred	A substitute for money market funds that captures the dividends-received deduction for firms
	Primes and scores	Equity is split into a prime component that has dividends and capital gains up to a stated price and a score component that has capital gains above this
	Synthetics	Securities that allow combinations of assets to be obtained with low transaction costs

Table 3
Global markets for exchange traded and OTC derivatives

	Notional amounts outstanding as of March 31, 1995	Gross market values	Average daily turn- over of notional amounts in April 1995
(US \$ billions)			
A. OTC contracts			
Total OTC Contracts	40 714 (100)	1745 (100)	839 (100)
Foreign exchange	13 153 (32)	1021 (59)	688 (82)
Forwards and swaps	8742 (21)	602 (34)	643 (77)
Currency swaps	1974 (5)	345 (20)	4 (0)
Options	2375 (6)	69 (4)	40 (5)
Interest rates	26 645 (65)	646 (37)	151 (18)
FRAs	4597 (11)	18 (1)	66 (8)
Swaps	18 283 (45)	560 (32)	63 (8)
Options	3548 (9)	60 (3)	21 (3)
B. Exchange-traded derivatives			
Total	16 581 (100)	1136 (100)	
Interest rate contracts	15 674 (95)	1121 (99)	

Figures in parentheses are percentage share.

Source: Bank of Japan, Quarterly Bulletin, May 1996, Tables 2-4, 9.

transfer of title of individual mortgages which was necessary previously. The bank that services the loan, i.e., collects the payments and deals with other administrative aspects, earns a fee for undertaking these tasks.

Other types of securitized loans quickly followed including commercial mortgages, bank loans, automobile loans and credit card receivables. Fig. 6 shows the tremendous growth in just one type of securitized loans, i.e. mortgage pools, that has taken place over the past 30 years particularly since the early 1980s, while Fig. 7 shows the corresponding growth in bank loans and

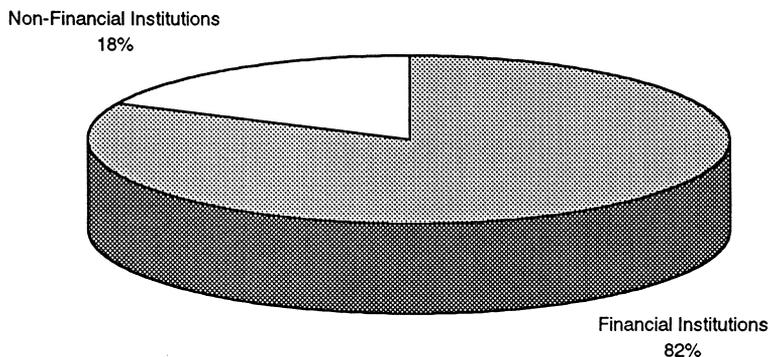


Fig. 5. Notional amounts outstanding of OTC derivatives.

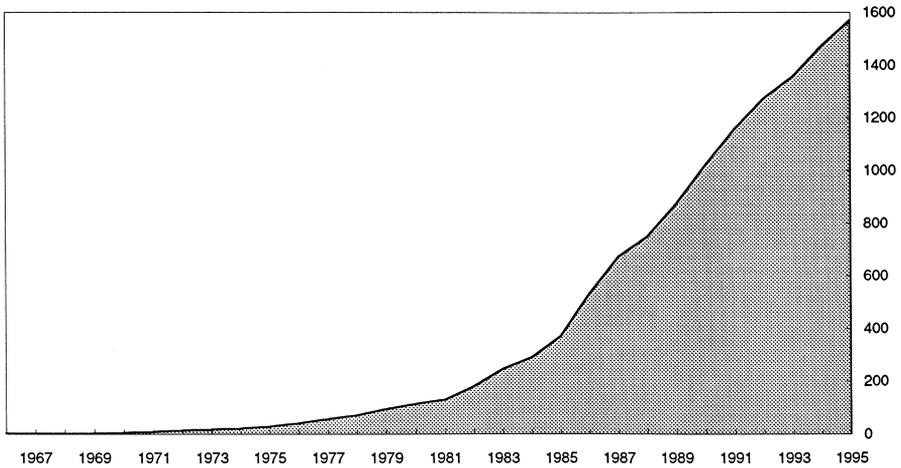


Fig. 6. US – Federal mortgage pools 1966–1995 (\$ billions).

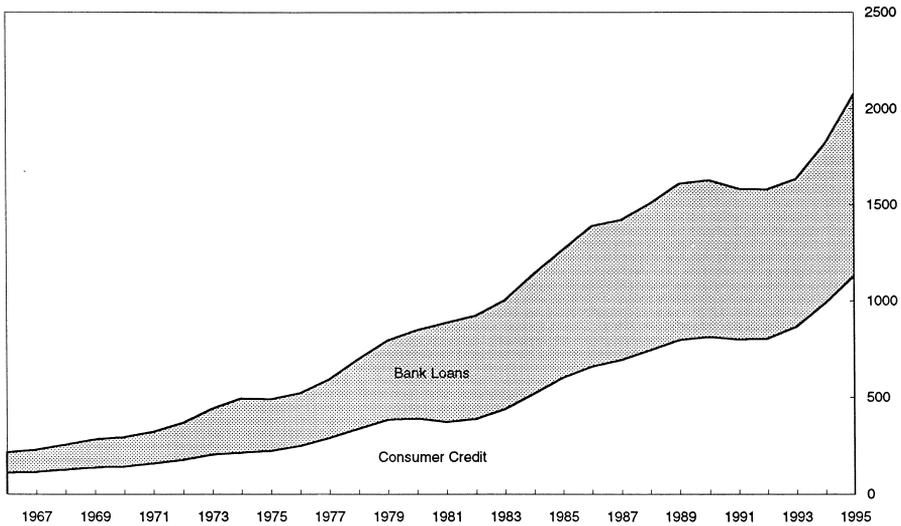


Fig. 7. US – Consumer credit and loans 1966–1995 (\$ billions).

consumer credit. A comparison of the two figures indicates that securitized loans have overtaken both of these categories in terms of volume outstanding. The fact that securitization has become so important in recent years suggests that asymmetric information cannot be that important for the loans that have been securitized. If this were the case, there would be an adverse selection or “lemons” problem with bad risks attempting to securitize more than good risks. As an empirical fact this appears not to be the case.

The final type of innovation mentioned in Table 2 is securities directly issued by firms. In terms of volume issued most of these have been relatively unimportant compared to derivatives or securitized loans. However, the trend is important particularly in light of our previous comments on the changing nature of the information set available to market participants.

Taken as a whole, the evidence presented here indicates that the traditional distinction between financial markets, where securities are issued by firms and directly owned by individuals, and intermediaries, where depositors and policyholders provided funds to banks and insurance companies who lent out these funds, has broken down. Financial markets for equity and debt are becoming increasingly dominated by intermediaries such as mutual and pension funds. The volume of transactions in these markets and those trading more complicated financial claims has become dominated by these same intermediaries, as well as the participants representing the standard institutions, i.e., commercial and investment banks and insurance companies. Indeed, the operations of many large banks and insurance companies have changed dramatically over this period, with trading activity occupying the bulk of their efforts.

The increased use of securitization of loans has exacerbated this trend in that it has altered the lending functions performed by banks. Now much of the asset origination activity is merely the first step to asset sales or complex stripping and repackaging. At the very least such assets are viewed as available for sale.

However, perhaps the most significant trend that is evident in the data is the increased concentration by banks and insurance companies as well as other financial institutions in the business of asset trading and risk shifting. The huge amounts of derivatives outstanding and the turnover suggest that this has become a major, and perhaps the most important, activity for the sector.

4. Current rationales for risk management

As discussed in Section 2, current theories of intermediation focus on transaction costs and asymmetric information. These factors can explain traditional intermediation but are less satisfactory in explaining the developments outlined in the previous section. Moreover, they are unable to satisfactorily explain the huge amount of risk management that is undertaken by intermediaries. In this section we consider the current rationales for the interest in risk management that is evident in the market. While these theories have been described above as somewhat of an add-on to our basic optimization models, it is worthwhile to review our understanding of why customers of intermediaries have a need to trade and manage risk. It is particularly important in light of the fact that the trading of risk appears to have become central to the role of intermediation.

The literature on why firms manage risk at all is usually traced back to 1984. In that year Stulz (1984) first suggested a viable economic reason why a firm's managers, who are presumed to be working on behalf of firm owners, might concern themselves with both expected profit and the distribution of firm returns around their expected value. He provided a rationale for why firm's objective functions may be concave so as to actively want to avoid risk. His contribution is widely cited as the starting point of this burgeoning literature. Since that time a number of alternative theories and explanations have been offered. Recently, Santomero (1995) has presented a useful review of these explanations upon which we shall draw here. They can be divided into four cases.

1. Managerial self-interest.
2. The non-linearity of taxes.
3. The costs of financial distress.
4. The existence of capital market imperfections.

In each case, the economic decision maker is shown to face a non-linear optimization, and this leads it to concern itself with the variability of returns. In the first case the objective function itself is concave, while in the others the effect of some feature of the economic environment is to lead firm managers to behave in a risk averse manner. We begin with a brief description of each theory.

4.1. Managerial self-interest

Stulz (1984) argued that firm managers have limited ability to diversify the significant portion of their personal wealth held in the form of stock in the firm and the capitalization of their earnings from the firm. Therefore, they prefer stability of the firm's earnings to volatility because, other things equal, such stability improves their own utility, at little or no expense to other stakeholders. This argument can be traced back to the literature on agency. In this area, the relationship between firm performance and managerial remuneration is clearly developed in such work as Ross (1973, 1977).

Objections have been offered, however, to this line of reasoning by those arguing that managerial self-interest in diversification need not occur on the firm's balance sheet (see Santomero, 1995 for a discussion of this point). However, the work of Breeden and Viswanathan (1990) and Demarzo and Duffie (1992) makes the managerial self-interest argument more compelling. They point out that observed outcomes may influence owner perception of managerial talent. This would, in turn, favor reduced volatility, or at least the protection of firm specific market value from large negative outcomes that may be found within the distribution of possible returns. For this, if for no other reason, there appears to be ample justification for the assumption that managers will behave in a manner consistent with a concave objective function.

4.2. *The non-linearity of taxes*

Beyond managerial motives, firm level performance and market value may be directly associated with volatility for a number of other reasons. The first is the nature of the tax code, which both historically and internationally is highly non-linear. This point was brought to our attention by Smith and Stulz (1985) and Gennotte and Pyle (1991) and emphasized by others as a key rationale of risk reduction. With a non-proportional tax structure, income smoothing reduces the effective tax rate and, therefore, the tax burden shouldered by the firm. By reducing the effective long term average tax rate, activities which reduce the volatility in reported earnings will enhance shareholder value. Graham and Smith (1996) have conducted a careful examination of the degree of convexity in the current US tax code. They find that on average the tax function is convex primarily because of tax loss carrybacks and carryforwards but in some circumstances it is concave. For firms with a convex tax function the average tax savings from a 5% reduction in volatility of taxable income are about 5% of expected tax liabilities.

Again, there have been objections to this rationale. Reported earnings may be different than economic earnings, for example. Economists have long been suspicious of accounting reasons for economic decisions. However, no matter how tenuous the rationale should not be dismissed out of hand.

4.3. *The costs of financial distress*

The third reason is perhaps the most compelling of the four. Firms may also be concerned about volatility of earnings because low realizations lead to bankruptcy. When bankruptcy is costly the firm will try to avoid it and so will behave as if it had a concave objective function.

This rationale seems the most reasonable one to us, as there is a long literature dating back to Warner (1977) on the costs of bankruptcy. More recently Weiss (1990) has continued to offer evidence of this feature of financial distress.

The cost is, perhaps, more important in regulated industries, however. In these cases, large losses may be associated with license or charter withdrawal and the loss of a monopoly position. This has led some to argue that this rationale offers significant insight into why banks themselves may choose low risk strategies (see, e.g., Marcus, 1984 or Santomero, 1989). In all cases, however, the cost of financial distress must be non-linear and is frequently modeled as discrete, as linear cost functions do not lead to the required behavior.

4.4. *Capital market imperfections*

In a series of papers, Froot et al. (1989, 1993, 1994) accept the basic paradigm of the financial distress model above, but rationalize the cost of bad out-

comes by reference to Myers (1977) debt overhang argument. In their model, external financing is more costly than internally generated funds due to capital market imperfections. These may include discrete transaction costs to obtain external financing, imperfect information as to the riskiness of the investment opportunities present in the firm, or the high cost of potential future bankruptcy.

At the same time, the firm has an investment opportunity set which can be ordered in terms of net present value. The existence of the cost imperfections results in underinvestment in some states, where internally generated funds fall short of the amount of new investment that would be profitable in the absence of these capital market imperfections. Stated another way, the volatility of profitability causes the firm to seek external finance to exploit investment opportunities when profits are low. The cost of such external finance is higher than the internal funds due to the market's higher cost structure associated with the factors mentioned above. This, in turn, reduces optimal investment in low profit states.

The cost of volatility in such a model is the foregone investment in each period that the firm is forced to seek external funds. Recognizing this outcome, the firm embarks upon volatility reducing strategies, which have the effect of reducing the variability of earnings. Hence, risk management is optimal in that it allows the firm to obtain the highest expected shareholder value.

4.5. Discussion

The theories presented are designed to explain why firms, both financial and non-financial, might be interested in undertaking risk management. The discussion above has focused on the benefits from hedging but has not mentioned the costs. These include the direct transaction costs and the agency costs of ensuring managers transact appropriately. These are arguably significant. To begin with the transaction costs of hedging include the costs of trading. Perhaps more importantly, though, they also include the substantial cost of information systems needed to provide the data necessary to decide on the appropriate hedging positions to take. Then, there are the agency costs that such activities bring. These include the problems associated with the opportunities for speculation that participation in derivative and other markets allow. Recent scandals at Metallgesellschaft, Barings and other firms, where billions of dollars were lost are extreme examples of these agency costs. They have been demonstrably substantial.

The plausibility of the four explanations for risk management varies especially if these costs are taken into account. The non-linearity of taxes is perhaps the least plausible. The tax savings indicated by Graham and Smith (1996) seem unlikely to be able to justify the huge amount of risk management absorbed, especially if the costs of these strategies are taken into account.

The arguments based on managerial self-interest and capital market imperfections are more plausible, but again it is not clear that the benefits outweigh

the costs. Shareholders could prohibit managers from managing risks and use the cost savings realized to compensate them for the increased risk. With capital market imperfections the NPVs of the marginal projects that become feasible because of the smoothing of cash flows would need to be substantial to compensate for the risk management costs incurred.

Of the four, perhaps the most plausible rationale is bankruptcy costs. It has been widely argued that these are substantial. They no doubt explain a significant part of the risk management that takes place. However, it is not clear why alternative strategies such as reducing the amount of debt or using instruments which do not precipitate bankruptcy are not superior given the costs associated with risk management.

Nonetheless, the result is to offer a rationale for the demand for both risk management services offered to customers and perhaps risk management at the firm level too. Managers appear to be interested both in expected profitability and the risk, or variability, of reported earnings and market value. The firm is led to treat the variability of earnings as a choice variable that it selects, subject to the usual constraints of optimization (see Smith et al., 1990 for an example of the procedures employed). The banker now has a customer for risk management services.

However, one important characteristic of the data considered in Section 3 is that a majority of trading in financial assets, including both traditional assets and derivatives is undertaken directly by financial institutions. Rationales for why risk management should be offered in this form are not obvious. Specifically, why should institutions be the principal participants in the bulk of financial transactions, what types of transactions should they engage in directly, and what is the core value added of their activity? Theory should be able to explain these things. It is our contention that it is the ability of these institutions to perform value added services in some of these markets most efficiently and knowledgeably that is the key competency of the sector. This determines where they trade and how they add value by reducing participation costs to outsiders who wish the benefit of these markets.

5. Understanding the role of intermediaries in risk trading

By dealing in financial assets, intermediaries are by definition in the financial risk business. By virtue of the fact that they originate, trade, or service financial assets, intermediaries are managing and trading risk. As Merton (1989) noted, a key feature of their franchise is the bundling and unbundling of risks. However, some of the risks inherent in the intermediaries' franchise will not be borne directly by them. Some will be traded or transferred, and others will be eliminated altogether. In fact, it is useful to decompose the risks inherent in financial assets into these three subgroups. This will allow us both to consid-

er which risks belong to each group and how intermediaries deal with each type of risk. Using this approach, risks can be segmented into the following groups:

1. risks that can be eliminated or avoided by business practices;
2. risks that can be transferred to other participants;
3. risks that must be actively managed at the firm level.

In the first of these cases, the charge of the financial intermediary is to engage in actions to reduce the chances of idiosyncratic losses by eliminating risks that are superfluous to the financial transaction's purpose. Common risk avoidance actions, here, are underwriting standards, due diligence procedures, and portfolio diversification. In each case the goal is to rid the financial transaction of risks that are inconsistent with the desired financial characteristics of the asset, or not essential to the financial asset being created. What remains is some portion of systematic risk, and the idiosyncratic risks that are integral to a product's unique business purpose. In both of these cases, risk reduction remains incomplete and could be further enhanced. In the case of systematic risk, any element not required or desired can be minimized by offsetting hedging activity. Whether or not this is done is a decision that can be clearly indicated to asset holders whether they are stockholders of the institution creating the asset and bearing the risk or a buyer obtaining ownership of the traded asset. The idiosyncratic risk also can be virtually eliminated. However, such actions are costly and it is up to the institution to determine the point where the cost of further risk reduction is higher than its value.

Risks can also be eliminated from a financial transaction, or at least substantially reduced, through risk transfer. Other market participants can buy or sell financial claims representing a portion of the state contingent payoffs to diversify or concentrate the risk in their portfolios. This is achieved through separate contracts offsetting certain state contingent payoffs such as swaps, or by the issuance of financial contracts which leaves some of the inherent risk of the transaction with the other party. Adjustable rate lending is a case in point.

To the extent that the financial risks of the asset created or held by the financial firm are understood by the market, they can be sold easily to the open market at their fair market value. If the institution has no comparative advantage in managing the attendant risk, there is no real reason for the firm to absorb and/or manage such risks, rather than transfer them. In essence, there is no value added to risk absorption at the firm level for such asset characteristics.

Then, there is the last class of assets or activities, where the risk inherent in the activity must and should be absorbed at the originating firm level. In these cases, good reasons exist for using further resources to manage firm level risk. These are financial transactions or contractual relationships that have one or more of the following characteristics. First, the equity claimants, or others for whom the institution has a fiduciary interest, may own claims that cannot be traded or hedged easily by the investors themselves. For example, defined benefit pension plan participants can neither trade their claims nor hedge them on an equivalent

after-tax basis. A similar case can be made for policies of mutual insurance companies which are complex bundles of insurance and equity. Second, activities where the nature of the embedded risk may be complex and difficult to reveal to non-firm level interests. This is the case in institutions such as banks, which hold complex, illiquid and proprietary assets. Communication in such cases may be more difficult or expensive than hedging the underlying risk. Moreover, revealing information about customers or clients may give competitors an undue advantage. Third, moral hazard may exist such that it is in the interest of stakeholders to require risk management as part of standard operating procedures. For example, providers of insurance can insist that institutions with insured claims follow business policies that maintain the basis of the insured risk. A fourth reason for risk absorption at an institution is that it is central to the intermediary's business purpose. An index fund invests in an index without hedging systematic risk. A security dealer normally hedges his positions, since profits accrue from order flow. However, if the dealer's purpose is proprietary trading and arbitrage, positions would not be routinely hedged. In all of the above circumstances, risk management activity requires management to monitor the risks and return of its business activities. This is part of the nature of their doing business since it springs directly from their franchise in the financial market.

With legitimate institutional risk management rationales defined and outlined, non-economic or redundant risk taking can also be identified. In short, if the risk absorption plays no fundamental role in the institution's franchise it is best to transfer the risk to the market.

But, who is the market? From our perspective it consists of two different groups, loosely defined as the involved and the uninvolved. The first of these are the market participants of economic theory. They are fully informed at each instant of time and are active participants in the dynamic management of their portfolio of financial assets. Then, there is the second group. These are usually described as uninformed. They are making decisions with limited information on both the nature of the financial claims involved and the most recent information on fair market value. It is to this group that the financial intermediary offers participation services. These may be provided by offering information to the uninformed investor, by investing on their behalf, or by offering a fixed income claim against the intermediary's balance sheet. In any case, the investor gains access to the market through the intermediary's services, which add value to the transaction by reducing the perceived participation costs of the uninformed investor.

6. Intermediation and participation costs

Traditional frictionless theories where intermediaries do not add value and there is no need for intermediaries to manage risk assume all investors are in-

volved and there is full participation in markets. However, there is extensive evidence that full participation is not an assumption which holds in practice. Typical households hold few stocks and participate in only a limited number of financial markets. Rather than full participation there is limited market participation.

Blume et al. (1974) develop a measure of portfolio diversification which takes into account the proportion of each stock held in individuals' portfolios. Based on this measure, they find that the average amount of diversification is equivalent to having an equally weighted portfolio with two stocks. Blume and Friend (1978) provide more detailed evidence of this lack of diversification. They find that a large proportion of investors have only one or two stocks in their portfolios, and very few have more than ten. These results concerning the small number of stocks most investors hold cannot be explained by them holding mutual funds. Fig. 4 indicates that until the 1980s, the value of mutual funds held was only about 5% of the total amount of households' equity.

King and Leape (1984) analyze data from a 1978 survey of 6010 US households with average wealth of almost \$250 000. When assets are categorized into 36 classes, they find that the median number owned is eight. Mankiw and Zeldes (1991) find that only a small proportion of investors own stocks. Of those with liquid assets in excess of \$200 000, only 47.7% hold any stocks.

One plausible explanation of limited market participation is that there are fixed costs of learning about a particular stock or other type of financial instrument. In order to be active in a market, an investor must devote time and effort to learning how the market works, the distribution of asset returns and how to monitor changes through time. Brennan (1975) has shown that with fixed setup costs of this kind it is optimal to invest in a limited number of assets. King and Leape (1984) find empirical evidence that is consistent with this type of model.

In addition to the fixed costs of market participation there are also arguably extensive marginal costs of monitoring markets on a day to day basis. Such monitoring is necessary to see how the expected distribution of payoffs is changing and how portfolios need to be adjusted. To the extent investors are following dynamic trading strategies to create synthetic securities they will need to follow the market on a continuous basis.

The assumption of limited market participation has been used in a number of asset pricing theories including Merton (1987), Hirshleifer (1988), Cuny (1993) and Allen and Gale (1994b). What we shall argue here is that costs of participation are important in understanding intermediation and the changes that have occurred in recent years.

The first change that was focused on in Section 3 was the reduction in the proportion of equity owned directly by households and the corresponding increase in the amount held by intermediaries. The standard story for why stock is held indirectly in these intermediaries is the desirability of diversification and the high trading costs for small portfolios. As pointed out above trading costs

have fallen dramatically since the 1970s which would seem to suggest mutual funds' role should have been reduced. However, as Fig. 4 illustrates the ratio of mutual funds to householder equity ownership has risen from about 5% in 1980 to around 25% by 1995. This change can be explained in terms of participation costs. The value of people's time, particularly that of many professionals, has increased significantly in the last 15 years. Mutual funds have low participation costs and thus are an efficient method to invest for individuals whose costs of direct participation have risen.

The second development outlined in Section 3 was the heavy use of derivatives by financial institutions for risk management. Theory based on the absence of frictions suggests households should be constantly reviewing and altering their portfolios as new information becomes available. If participation costs are taken into account, this is clearly unrealistic. It does suggest, however, that a main role of intermediaries is to create products with relatively stable distributions of returns. This allows investors to monitor their asset holdings on a relatively infrequent basis which is a very valuable characteristic.

One of the striking features of the securities that investors do hold is that many are debt or debt-like. This type of security has low participation costs. Except for the possibility of default, there is no need to monitor these assets through time and the costs of learning about the market are low. To the extent that the probability of default can be lowered, given the cost of participation, the value of the securities will be increased.

A theory of intermediation based on participation costs is thus consistent with the fact that intermediaries trade risk and undertake risk management to such a large extent (see Santomero, 1997; Santomero and Babbel, 1997). By creating products with stable distributions of cash flows they can lower participation costs for their customers. In extreme cases this may involve creating low risk debt, but even with more risky securities the stability of distributions is important in minimizing the costs of revising portfolios through time.

The participation costs rationale is also consistent with firms undertaking risk management. The purchasers of firm stocks will typically be more sophisticated in the sense that they have already incurred the fixed component of participation costs. However, they must also bear some marginal costs to maintain their knowledge of firm performance and its probability of success. To the extent these can be lowered by firms managing risk then value is created. A theory of intermediation based on participation costs can thus explain why firms manage risk but to a lesser extent than financial intermediaries.

7. Concluding remarks

This paper has suggested that theories of intermediation need to reflect and account for the fact that financial systems in many countries have changed sub-

stantially over the past 30 years. Over this period many traditional financial markets have expanded and new markets have come into existence. Transaction costs have fallen and information has become cheaper and more available. However, these changes have not coincided with a reduction in intermediation. In fact, quite the reverse has happened. Intermediaries have become more important in traditional markets and account for a very large majority of the trading in new markets, such as those for various types of derivatives. Standard theories of intermediation based on transaction costs and asymmetric information are difficult to reconcile with the changes that have taken place. We have argued that participation costs are crucial to understanding the current activities of intermediaries and in particular their focus on risk management.

This paper has focused on intermediation theory. The fact that markets have become more dominated by intermediaries also has important implications for asset pricing theory. Current asset pricing theories usually assume investors choose optimal portfolios directly. The fact that there is such extensive intermediation suggests that this approach may miss important features of actual markets. For example, Allen and Gorton (1993) have shown that if intermediaries make investment decisions on behalf of investors markets can be inefficient in the sense that asset prices differ from fundamentals and there are bubbles. In short, given the importance of intermediaries' trading in financial markets asset pricing theories and intermediation theories need to become better integrated.

Acknowledgements

We are grateful to Paul Hoffman for excellent research assistance and to the Wharton Financial Institutions Center for support.

References

- Allen, F., Gale, D., 1994a. *Financial Innovation and Risk Sharing*, MIT Press, Cambridge, MA.
- Allen, F., Gale, D., 1994b. Limited market participation and volatility of asset prices. *American Economic Review* 84, 933–955.
- Allen, F., Gorton, G., 1993. Churning bubbles. *Review of Economic Studies* 60, 813–837.
- Bhattacharya, S., Thakor, A.V., 1993. Contemporary banking theory. *Journal of Financial Intermediation* 3, 2–50.
- Blume, M.E., Crockett, J., Friend, I., 1974. Stock ownership in the United States: characteristics and trends. *Survey of Current Business* 54 (11), 16–40.
- Blume, M.E., Friend, I., 1978. *The changing role of the individual investor: a Twentieth Century Fund report*, Wiley, New York.
- Boyd, J., Prescott, E.C., 1986. Financial intermediary coalitions. *Journal of Economic Theory* 38, 211–232.

- Breeden, D., Viswanathan, S., 1990. Why do firms hedge? An asymmetric information model. Working Paper, Duke University.
- Brennan, M.J., 1975. The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: theory and some empirical results. *Journal of Financial and Quantitative Analysis* 10, 483–496.
- Campbell, T.S., Kracaw, W.A., 1980. Information production, market signaling, and the theory of financial intermediation. *Journal of Finance* 35, 863–882.
- Cuny, C.J., 1993. The role of liquidity in futures market innovations. *Review of Financial Studies* 6, 57–78.
- Demarzo, P., Duffie, D., 1992. Corporate incentives for hedging and hedge accounting. Working paper, Northwestern University.
- Diamond, D., 1984. Financial intermediation and delegated monitoring. *Review of Economic Studies* 51, 393–414.
- Dionne, G., 1991. *Contributions to Insurance Economics*. Kluwer Academic Publishers, Dordrecht.
- Fama, E.F., 1980. Banking in the theory of finance. *Journal of Monetary Economics* 6, 39–58.
- Froot, K., Scharfstein, D., Stein, J., 1989. LDC debt: forgiveness, indexation, and investment incentives. *Journal of Finance* 44, 1335–1350.
- Froot, K., Scharfstein, D., Stein, J., 1993. Risk management: coordinating corporate investment and financing policies. *Journal of Finance* 48, 1629–1658.
- Froot, K., Scharfstein, D., Stein, J., 1994. A framework for risk management. *Journal of Applied Corporate Finance* 7, 22–32.
- Gale, D., Hellwig, M., 1985. Incentive-compatible debt contracts: The one-period problem. *Review of Economic Studies* 52, 647–664.
- Gennotte, G., Pyle, D., 1991. Capital controls and bank risk. *Journal of Banking and Finance* 15, 805–824.
- Graham, J.R., Smith, C.W., 1996. Tax incentives to hedge. Working paper.
- Gurley, J.G., Shaw, E.S., 1960. *Money in a Theory of Finance*. Brookings Institution, Washington, D.C.
- Hirshleifer, D., 1988. Residual risk, trading costs, and commodity futures risk premia. *Review of Financial Studies* 1, 173–193.
- King, M.A., Leape, J.I., 1984. Wealth and portfolio composition: theory and evidence. National Bureau of Economic Research (Cambridge, MA), Working paper No. 1468, September.
- Leland, H.E., Pyle, D.H., 1977. Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance* 32, 371–387.
- Mankiw, N.G., Zeldes, S.P., 1991. The consumption of stockholders and nonstockholders. *Journal of Financial Economics* 29, 97–112.
- Marcus, A., 1984. Deregulation and bank financial policy. *Journal of Banking and Finance* 8, 557–565.
- McKinnon, R., 1973. *Money and Capital Economic Development*. Brookings Institution, Washington, D.C.
- Merton, R.C., 1987. Presidential address: A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42, 483–510.
- 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.
- Merton, R.C., 1993. Operation and regulation in financial intermediation: a functional perspective. In: England, P. (Ed.), *Operation and Regulation of Financial Markets*, The Economic Council, Stockholm.
- Merton, R.C., Bodie, Z., 1995. *A conceptual framework for analyzing the financial environment, The Global Financial System – A Functional Perspective*. Harvard Business School Press.

- Miller, M.H., 1986. Financial innovation: the last twenty years and the next. *Journal of Financial and Quantitative Analysis* 21, 459–471.
- Myers, S.C., 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5, 147–175.
- Oldfield, G., Santomero, A., 1997. The place of risk management in financial institutions. *Sloan Management Review*, Fall.
- Ross, S., 1973. The economic theory of agency: the principal's problem. *American Economic Review* 63, 134–139.
- Ross, S., 1977. The determination of financial structure: the incentive-signaling approach. *Bell Journal of Economics* 8, 23–40.
- Santomero, A., 1984. Modeling the banking firm. *Journal of Money, Credit, and Banking*, Part 2 16, 576–602.
- Santomero, A., 1989. The changing structure of financial institutions: a review essay. *Journal of Monetary Economics* 24, 321–328.
- Santomero, A., 1995. Financial risk management: the whys and hows, *Financial Markets, Institutions, and Investments* 4 (5) 1–14.
- Santomero, A., 1997. Commercial bank risk management: an analysis of the process. *Journal of Financial Services Research*, Sept.
- Santomero, A., Babbel, D., 1997. Risk management by insurers: an analysis of the process. *Journal of Risk and Insurance* 64 (2), 231–270.
- Smith, C., Smithson, C., Wilford, D., 1990. *Strategic Risk Management (Institutional Investor Series in Finance)*. Harper and Row, New York.
- Smith, C.W., Stulz, R., 1985. The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis* 20, 391–406.
- Stulz, R., 1984. Optimal hedging policies. *Journal of Financial and Quantitative Analysis* 19, 127–140.
- Warner, J., 1977. Bankruptcy costs: some evidence. *Journal of Finance* 32, 337–347.
- Weiss, L., 1990. Bankruptcy resolution: direct costs and violation of priority of claims. *Journal of Finance and Economics* 27, 285–314.