Liquidity, Banks, and Markets

Douglas W. Diamond

University of Chicago

This paper examines the roles of markets and banks when both are active, characterizing the effects of financial market development on the structure and market share of banks. Banks lower the cost of giving investors rapid access to their capital and improve the liquidity of markets by diverting demand for liquidity from markets. Increased participation in markets causes the banking sector to shrink, primarily through reduced holdings of long-term assets. In addition, increased participation leads to longer-maturity real and financial assets and a smaller gap between the maturity of financial and real assets.

I. Introduction

Financial markets and banks are competing mechanisms that provide investors with liquidity by providing access to their capital, at good terms, on short notice. This paper examines the impact of banks on the liquidity provided to investors and, in addition, on the liquidity provided by markets. Markets can provide too little liquidity when some potential investors are not continuously available for trade. If there is this limited participation in the market, banks lower the cost of giving investors rapid access to their capital. Banks hold assets to finance demand deposits offered to those who deposit, and

I gratefully acknowledge helpful comments from Patrick Bolton, John Boyd, Jerry Caprio, Mike Dotsey, Phil Dybvig, Gary Gorton, Lars Hansen (the editor), Bengt Holmström, Pete Kyle, Jeff Lacker, Matthew Rothman, Bruce Smith, Lars Stole, Jean Tirole, Neil Wallace, John Weinberg, an anonymous referee, and seminar participants at Chicago, Duke, Massachusetts Institute of Technology, the Federal Reserve Bank of Richmond, and the World Bank. This paper was presented at the Nobel Symposium on Law and Finance in Stockholm in August 1995.

[*Journal of Political Economy,* 1997, vol. 105, no. 5]

they divert some demand for liquidity away from markets. This paper characterizes the effects of increased participation in financial markets on the structure of banks, the maturity structure of real and financial assets, and the fraction of capital invested through banks.

Investors are concerned about the return they can obtain on short notice because they are uncertain when they will need their funds. The activities of banks provide liquid investment opportunities through two channels. First, bank deposits offer an option to obtain funds on short notice at a lower opportunity cost than is available with markets alone. Second, banks improve the liquidity of markets. The liquidity of markets is enhanced because long-term assets can be sold before maturity at higher prices than would prevail without banks.

The model can most clearly be applied to effects of financial development in developing economies. The optimal financial mechanism, which includes a banking sector, adapts as market participation increases. The maturity of both financial assets and real investments increases when market participation increases. In addition, the scale of the banking sector falls (more capital is invested in marketable assets), and the maturity of assets held by banks falls. Some evidence consistent with these predictions is discussed in Section V.

The model also resolves some theoretical issues about the roles of banks and markets in Diamond and Dybvig's (1983) model. The model differs from the Diamond-Dybvig model by adding a financial market with limited participation and endogenizing the liquidity of assets. The Diamond-Dybvig model, where asset liquidity is not linked to the operations of markets, has been interpreted as inconsistent with active markets. Jacklin (1987) introduces a secondary market in which bank deposits trade for other financial assets and shows that this implies that banks are not important.[1] Haubrich and King (1990), von Thadden (1997), and Hellwig (1994) also present models that question the liquidity role of banks when there is a financial market. Wallace (1988) argues that the Diamond-Dybvig model can usefully be interpreted as a model in which there are no financial markets because investors are physically separated, and none participates in financial markets. This paper examines the roles of banks and markets when there is a financial market with limited participation. Such a market has an impact on bank activities, but banks remain important. Banks and markets coexist and influence each other's activities.

Limited market participation is introduced by assumption. I argue

[1] The Jacklin result drops the Diamond-Dybvig assumption that consumption is observable and that those who withdraw use the proceeds for consumption.

below that the most plausible interpretation is that participation is limited because of private information about the value of assets. The model has limited participation, but no explicit private information about the value of assets: assets are riskless. Private information possessed by others can lead investors to avoid trading in markets (Akerlof 1970). It is possible to interpret increases in participation in markets as equivalent to better access to information by less informed investors. A full analysis of the effects of changes in information structure on markets and banks is beyond the scope of this paper.

A Sketch of the Model and Its Results

The key elements of the analysis are a demand for liquidity by investors and costs of obtaining liquidity. There is a demand for liquidity because investors are uncertain about the date on which they need their funds, as in Bryant (1980) and Diamond and Dybvig (1983). Liquidity is costly for an individual to obtain because short-term, self-liquidating real assets have a lower rate of return than long-term assets (which are not self-liquidating). Either trade in a market or a bank contract that offers an option to withdraw is needed to provide liquidity without investing primarily in short-term assets.

Investors in the model are ex ante identical and risk averse. Some will turn out to need all of their wealth for consumption soon thereafter, whereas the others will be in no rush to liquidate. How do markets perform without banks when some investors need to liquidate? Those who must consume their wealth immediately will sell their assets and will participate in markets. The other investors, who do not need to liquidate immediately, have no need to participate in financial markets. A fraction is assumed not to participate, but this is best interpreted that a fraction will turn out to have a high opportunity cost of participating. Plausibly, this cost is the expense of obtaining information to value the assets offered for sale. Limited participation has the following implications for markets. First, assets offered for sale in the market will not attract bids from all possible buyers, which will depress the selling price. Consequently, anticipation of a low resale price will depress the investment in long-term assets that mature after an investor might need liquidity. Investors will increase their investment in shorter-term assets that are self-liquidating. Second, investors who turn out not to need immediate liquidity (potential buyers), but do not participate in the market for buying assets offered for sale, will reinvest the proceeds from their shorter-term assets to obtain a lower return than they could obtain in the market. Excessive investment in self-liquidating assets is doubly inefficient: it yields a low immediate return that is sometimes reinvested in low-return assets. In short, limited participation in markets causes overinvestment in short-term real assets that are valued for their ability to self-liquidate.

Banks (or other financial institutions) can substitute for illiquid markets. Banks economize on liquid assets by avoiding the possibility that a nonparticipating potential buyer holds excessive liquidity. An individual will turn out to need liquidity on a single date. A bank is a coalition of individuals that has predictable liquidity needs on each date. When financial markets by themselves perform poorly, the bank as a coalition can manage the maturity of its assets to avoid the need to trade, while offering the option to withdraw on short notice to the members of its coalition. This holding of assets based on the anticipated timing of liquidity needed by a bank is often called *asset management of liquidity*. Besides avoiding the need to use markets, this asset management diverts some demand for liquidity from markets, which can improve the performance of markets.

It is also possible that, besides managing asset maturity better than an individual, financial institutions can participate in markets for investors. Given the informational motivation for limited participation, economies of scale or scope could allow institutions to have better access to information. Liquidity management based on superior access to markets is closely related to the other traditional approach to bank liquidity management: *liability management of liquidity*. Both roles for financial institutions are examined.

At the outset, I impose no constraints on financial institutions' participation in markets and determine how much of their participation is required. Later, I analyze the effects of restricted market participation by financial institutions. In the interest of simplicity, no costs of operating financial institutions are introduced. The existence of variable costs, however, would lead to the smallest scale of the banking industry that implements a given set of liquid consumption opportunities. Analyzing the smallest feasible scale of banks allows the model to examine the effects of increased market participation on the scale, structure, and activities of the banking sector.

Related models that use limited participation to understand financial markets are Merton (1987), which examines the effect on the relative prices of risky assets, and Allen and Gale (1994), which examines the volatility of asset prices. The most closely related study relating banks to the demand for liquidity is Holmström and Tirole (1996). It studies the impact of managerial moral hazard in firms on the ability of markets to allocate liquidity efficiently. The ex ante amount of liquidity that a firm should have available to it can differ from what spot markets would provide, as in Diamond (1991, 1993). Uncertainty about the amount of liquidity a firm needs is similar to the uncertain timing of consumption in this paper. Also related is Gorton and Pennacchi (1990), which examines the ability of intermediaries or firms to create riskless (and thus easy to value) securities when private information causes problems in risky asset markets. Bhattacharya and Gale (1987) examines the effects of a market for bank reserves on the liquid asset holdings of banks. These papers do not have limited participation in markets, but their results have a related focus.

Section II describes the model and characterizes the total amount of liquidity optimally created by the combination of the financial markets and the banking system. Section III describes the implications of optimal liquidity creation for the scale of the banking industry, the contracts the banking system offers, the assets that banks fund with those deposits, and the maturity structure of financial and real assets. Section IV discusses ways in which the analysis might be generalized and gives some more general interpretations of the existing results. Section V concludes the paper.

II. The Model

There are three dates, 0, 1 and 2. All investors are small: there is a continuum of investors. Each is endowed with one unit of date 0 capital. Investors find out their need for liquidity and their participation in the market on date 1, and each investor's type is his or her private information. As of date 0, all investors are identical, but each is uncertain on which date he or she will need to consume and need liquidity. There are three types of agents as of date 1: type 1, type 2A, and type 2B. Type 1 investors will need liquidity at date 1: they need to consume at date 1 and place no value on date 2 consumption. Types 2A and 2B do not need liquidity on date 1 and place no value on date 1 consumption. The only difference between types 2A and 2B is their participation in a secondary market for assets on date 1. Type 2A agents are active in the secondary market, and type 2B agents are not. As of date 0, an investor is of type τ on date 1 with probability q_τ. Assume that $q_1 > 0$, and some investors will need to consume at date 1. Whenever $q_{2B} > 0$, there is limited participation in markets.

Define $c_{t\tau}$ as the consumption on date t of a type τ investor. Investors are risk averse, and the date on which they prefer to consume depends on their type. The form of the utility function of investor j, who consumes c_1 at date 1 and c_2 at date 2, is

$$u_j(c_1,\ c_2) = \begin{cases} U(c_1) & \text{if } j \text{ is of type 1} \\ U(c_2) & \text{if } j \text{ is of type 2A or 2B,} \end{cases}$$

where U: $R_{++} \rightarrow R$ is twice continuously differentiable, increasing, and strictly concave and satisfies the Inada conditions $U'(0) = \infty$ and $U'(\infty) = 0$. Also, the relative risk aversion coefficient $-cU''(c)/U'(c) \geq 1$ everywhere. Investors maximize expected utility. These preferences are identical to those assumed in Diamond-Dybvig, except here types 2A and 2B are distinguished. Diamond-Dybvig allows no secondary market, which essentially assumes that there are only investor types 1 and 2B.

The date 0 objective function of each investor is given by

$$\max_{c_{11}, c_{22A}, c_{22B}} \quad \Psi = q_1 U(c_{11}) + q_{2A} U(c_{22A}) + q_{2B} U(c_{22B}),$$

subject to resource and incentive constraints specified below.

There are two real assets. The first is a short-term asset that yields a one-period-ahead cash flow of $R \leq 1$ per unit invested (with constant returns to scale). The second real asset is a long-term asset that yields a two-period-ahead cash flow of $X > R^2$ (with constant returns to scale) and nothing in one period. This implies that liquidity is difficult to obtain, because repeated investment in real assets that pay quickly is less profitable than long-term investments. In addition, limited participation limits the ability of markets to allocate available liquidity to its best use. Table 1 summarizes the timing of events.

In the Diamond-Dybvig model, there are no secondary markets, but long-term assets can be physically liquidated for a return that weakly exceeds the return on short-term assets, implying that all investment should be long-term.[2] In the current model, long-term assets cannot be physically liquidated. This implies a nontrivial decision on how to allocate investment between short- and long-term assets.

The Performance of Markets When All Assets Are Held Directly

Suppose that there are no banks. This means that all investors must hold assets directly, with the possibility of trade between those who turn out to be of type 1 or 2A. Each investor has one unit to invest at date 0. Let each investor put α into one-period assets and $1 - \alpha$ into two-period assets. Before any trade, each investor holds date 1 claims of αR and date 2 claims of $(1 - \alpha)X$. Type 2B investors have

[2] Diamond-Dybvig also assumes that the return on one-period assets is $R = 1$ and that date 2 utility is discounted by a factor $\rho \leq 1$. The minor differences here of $R \leq 1$ and $\rho = 1$ are not significant.

TABLE 1

Timing of Events

Date 0	Date 1	Date 2
Identical investors receive endowment of one to allocate between one-period assets, two-period assets, and bank deposit claims, if available	Investors learn their type and choose their withdrawal from bank, and some may trade	
	Type 1 investors sell any date 2 claims held directly or withdrawn from the bank and consume all wealth	Type 1 investors do nothing
	Type 2A investors sell any date 1 claims held directly or withdrawn from the bank to buy date 2 claims or fund new one-period investment	Type 2A investors consume all wealth
	Type 2B investors take any date 1 claims held directly or withdrawn from the bank to fund new one-period investment	Type 2B investors consume all wealth

no access to the market and take their maturing date 1 claim of αR and invest it in a new one-period asset, yielding αR^2 of date 2 consumption, in addition to their original date 2 holding of $(1 - \alpha)X$. Trade at date 1 would lead the type 1 investors to trade their date 2 claims for date 1 claims of type 2A investors. Markets clear when the date 1 value of date 2 claims offered by type 1 investors is equal to the date 1 claims offered by type 2A investors. Let the date 1 price of a unit claim on date 2 consumption be b_1. Investors are price takers because each is small. Trade with all assets held directly leads to the outcome described in lemma 1.

Lemma 1. When all assets are held directly, the date 1 price of a unit claim on date 2 consumption, b_1, is less than or equal to R/X. Investors place a fraction greater than or equal to q_1 of date 0 wealth in one-period assets ($\alpha \geq q_1$), and type 1 consumption, c_{11}, is less than or equal to R. With limited participation ($q_{2B} > 0$), all inequalities are strict ($b_1 < R/X$, $\alpha > q_1$, and $c_{11} < R$). With full participation, the price is $b_1 = R/X$, the fraction of short-term investment is $\alpha = q_1$, and type 1 consumption is $c_{11} = R$.

Proof. Each investor on date 0 allocates an endowment of one, choosing to put α into short-term assets and $1 - \alpha$ into long-term assets to maximize expected utility. Expected utility is

$$\Phi = q_1 U(c_{11}) + q_{2A} U(c_{22A}) + q_{2B} U(c_{22B}),$$

where consumption of each type is given by $c_{11} = \alpha R + (1 - \alpha) b_1 X$, $c_{22A} = \alpha(R/b_1) + (1 - \alpha) X$, and $c_{22B} = \alpha R^2 + (1 - \alpha) X$. The first-order condition for an interior optimum with $\alpha \in (0, 1)$ is $d\Phi/d\alpha = 0$, or

$$q_1 U'(c_{11})(R - b_1 X) + q_{2A} U'(c_{22A})\left(\frac{R}{b_1} - X\right)$$
$$+ q_{2B} U'(c_{22B})(R^2 - X) = 0.$$

The date 1 price b_1 will be positive (and c_{11} will be positive) only if there is some date 0 investment in short-term assets (only if $\alpha > 0$). There is short-term investment as long as $q_{2B} < 1$, because otherwise $U'(c_{11}) = \infty$ and $R/b_1 = \infty$. If investors are to choose to invest anything in short-term assets at date 0, the date 1 price of a unit claim on date 2 consumption, b_1, must be less than or equal to R/X. Otherwise, decreasing α will increase the consumption of all types (increasing expected utility). The date 1 price, b_1, is therefore less than or equal to R/X. The price is strictly less than R/X when $q_{2B} > 0$ (limited participation) because c_{22B} is decreasing in α. In addition, one can show that $b_1 < 1/R$, and trading one unit of a date 1 claim in the market will buy more than R units of date 2 consumption. It follows because $X > R^2$, implying $R/X < R/R^2 = 1/R$, which implies $b_1 < 1/R$. As a result, only type 2B investors will reinvest in short-term investments at date 1. In aggregate, type 1 investors trade all their date 2 claims for all the date 1 claims of type 2A investors. This leads to a market-clearing price of $b_1 = q_{2A}\alpha R/q_1(1 - \alpha)X$. In summary, this discussion implies that the consumption levels are given by

$$c_{11} = \frac{q_1 + q_{2A}}{q_1}\alpha R,$$

$$c_{22A} = \frac{q_1 + q_{2A}}{q_{2A}}(1 - \alpha)X,$$

$$c_{22B} = \alpha R^2 + (1 - \alpha)X.$$

The date 1 price of a claim of one unit maturing on date 2 is

$$b_1 = \frac{q_{2A}\alpha R}{q_1(1 - \alpha)X}.$$

To show that $\alpha \geq q_1$, with strict inequality when there is limited participation ($q_{2B} > 0$), substitute $\alpha = q_1 > 0$ into the equation for the price b_1, yielding

$$b_1 = \frac{q_{2A} q_1 R}{q_1(1 - q_1)X} = \frac{q_{2A} R}{(q_{2A} + q_{2B})X}.$$

The price $b_1 < R/X$ if $q_{2B} > 0$, implying that expected utility is increasing in α. When there is full participation and $q_{2B} = 0$, $\alpha = q_1$ is the equilibrium value because expected utility is strictly increasing in α for all lower values and strictly decreasing in α for all higher values of α. Q.E.D.

Lemma 1 characterizes the liquidity that financial markets provide and the real investment decisions implied by this level of market liquidity. Markets with limited participation lead to lower liquidity than full-participation markets because the secondary market price, b_1, is lower, and the consumption of those who turn out to need liquidity, c_{11}, is lower than when all investors participate in the market. In addition, there is less investment in high-return illiquid real assets when there is limited participation. One can show that increased participation increases all these quantities toward the full-participation values. Lemma 1 is primarily used as a benchmark with which to compare the outcomes when banks are formed. When there is limited participation, investors put more capital into short-term liquid assets and obtain lower date 1 consumption than when banks are formed. The next subsection characterizes the optimal allocations when investors form banks.

Banks and Optimal Mechanisms

Financial institutions such as banks can improve access to liquidity in two ways. First, by centralizing the holding of liquid assets, the institution reduces the opportunity cost of excess date 1 liquidity held by investors who do not participate in the market. Second, financial institutions possess some ability to cross-subsidize investors. Investors who need to consume unexpectedly at date 1 (type 1) can receive higher returns at the expense of those who cannot trade in the market (type 2B).

To characterize the role of intermediaries and markets in providing the optimal amount of liquidity, I solve for the optimal set of incentive-compatible consumption opportunities that a coalition of investors can choose at date 0, and later determine how they are related to markets and intermediaries. The standard method for characterizing the optimal consumption is to examine direct mechanisms in which each investor reveals his or her type and is given type-contingent consumption on each date, subject to the constraint that each investor is willing to make an honest report. It turns out that the report of an investor's type corresponds to a choice of which withdrawal option to select from those offered by a bank.

The optimal financial mechanism solves the date 0 maximization problem described above:

$$\max_{c_{11}, c_{22A}, c_{22B}} \Psi = q_1 U(c_{11}) + q_{2A} U(c_{22A}) + q_{2B} U(c_{22B}),$$

subject to resource and incentive constraints. As of date 0, one unit of date 1 consumption costs $1/R$ and one unit of date 2 consumption costs $1/X$. There is one unit of endowment per capita on date 1, and as a result, the resource constraint is given by

$$\frac{q_1 c_{11}}{R} + \frac{q_{2A} c_{22A} + q_{2B} c_{22B}}{X} \leq 1.$$

Consumption on the "wrong" date (by a type who assigns no value to consumption on that date) is never optimal, and at the optimum, $c_{21} = c_{12A} = c_{12B} = 0$. There are several incentive constraints as well, which may not be binding. If only the resource constraint is imposed, the first-order condition for optimal consumption levels is given by

$$U'(c_{11}) = U'(c_{22A}) \frac{X}{R} = U'(c_{22B}) \frac{X}{R}.$$

This equates the ratio of marginal utility of consumption of the two periods with the marginal rate of transformation across periods and equates the date 2 consumption of types 2A and 2B: $c_{22B} = c_{22A}$. Investors who are sufficiently risk averse would choose cross-subsidization to allow them to hold liquid claims with high one-period returns (high $c_{11} > R$) at the expense of lower two-period returns (lower $c_{22A} = c_{22B} < X$). This allows increased consumption when they have low consumption, due to forced liquidation of assets, and is financed by reduced consumption when they have high consumption. Consider the base case without cross-subsidization. This occurs when each investor receives consumption equal to the value of investing all of his or her date 0 endowment in a real asset that matures on the date on which he or she needs to consume: R at date 1 or X at date 2. A bank could achieve this because it knows the timing of the aggregate consumption of all depositors. If there is no cross-subsidization, then $c_{11} = R$ and $c_{22A} = c_{22B} = X$. The first-order condition holds with no cross-subsidization if and only if each investor's relative risk aversion equals one, and $U(c) = \log(c)$. For risk aversion greater than one, the empirically relevant case, $U'(R) > U'(X) \times (X/R)$, implying that liquidity is increased: there is cross-subsidization with $c_{11} > R$ and $c_{22A} = c_{22B} < X$.[3]

The cross-subsidized optimal consumption levels may not be incentive-compatible. Each investor's type is private information, and in addition types 1 and 2A can trade anonymously at date 1. The incentive constraints that trade implies are analyzed in the next subsection. It turns out that there is scope for a beneficial subsidy to type 1 agents.

Date 1 Incentive Constraints Imposed by a Limited Participation Market

On date 1, an agent who joins a financial mechanism at date 0 will be given a choice of claims on date 1 and date 2 consumption. Type 1 and 2A agents can trade anonymously at date 1 and privately consume the proceeds from those trades. As a result, they can choose claims on both dates' consumption without wasting goods. Let $W_{t\tau}$ denote the pretrade date 1 holding of date t claims by a type τ investor. After choosing these claims, type 1 and 2A investors have the ability to trade at date 1. Let b_1 denote the date 1 price of a claim on one unit of date 2 consumption. Type 2B investors cannot trade, but if a type 2B investor has a claim maturing on date 1, he or she can earn a return R per unit by initiating a new one-period asset at date 1. Consumption of each type is given by

$$c_{11} = W_{11} + W_{21} b_1,$$

$$c_{22A} = \frac{W_{12A}}{b_1} + W_{22A},$$

$$c_{22B} = W_{12B} R + W_{22B}.$$

The date 1 market value of the claims selected by type 1 agents is c_{11}, and the date 2 market value of claims selected by type 2A agents is c_{22A}. The price, b_1, on date 1 of claims on a unit of date 2 consumption must be c_{11}/c_{22A}; otherwise $c_{11} \neq b_1 c_{22A}$, and the claims intended for one type will have a higher market value than the other. If the market values differ, the ability to trade at date 1 implies that it is not incentive-compatible for one type to select the correct claims.

[3] If risk aversion is below one, then optimal risk sharing leads to cross-subsidization to reduce liquidity ($c_{11} < R$ and $c_{22B} > X$) if type is observable and there is no possibility of trade.

The date 1 incentive constraint for types 1 and 2A not to prefer the claims intended for the other is $b_1 = c_{11}/c_{22\text{A}}$: the rate of return from taking the claims intended for one's type must equal the rate of return from taking the claims intended for the other type and then trading them. If trade allows a return higher than $c_{22\text{A}}/c_{11}$, type 2A investors will take the withdrawal intended for type 1. If trade offers a return lower than $c_{22\text{A}}/c_{11}$, type 1 investors will take the withdrawal intended for type 2A.

A type 2B investor does not have access to the financial market. A type 2B investor can use the proceeds of date 1 claims to invest in new short-term investments that mature on date 2. He or she must not have incentives to take the wrong claims for the purpose of initiating a new one-period investment at date 1. It is feasible to give types 1 and 2A claims only on date 1 consumption, which a type 2A uses to buy date 2 claims (supplied by the bank) in the financial market. As a result, a type 2B can be induced to choose the proper claim as long as $c_{22\text{B}} \geq c_{11}R$. This constraint is not binding. Finally, the date 1 incentive constraint that type 2A investors not prefer the date 2 claims intended for type 2B investors is $c_{22\text{A}} \geq c_{22\text{B}}$. Lemma 2 summarizes the date 1 incentive constraints in a limited participation market.

LEMMA 2. The date 1 incentive-compatibility constraints are satisfied if and only if $b_1 = c_{11}/c_{22\text{A}}$, $c_{22\text{A}} \geq c_{22\text{B}}$, and $c_{22\text{B}} \geq c_{11}R$.

Proof. See the Appendix.

In addition to these constraints that each type be willing to withdraw the proper amount, there are date 0 constraints. These constraints are imposed by the possibility that investors invest directly at date 1, or that they join a competing bank. These constraints are described in the next subsection.

Date 0 Incentive Constraints: Competing Banks and Voluntary Deposits

Individuals can form alternative mechanisms (''competing banks'') at date 0, realizing that the members of all banks who turn out to be of type 1 or 2A will be able to trade in the same anonymous market at date 1. This ability imposes coalition incentive-compatibility constraints, because trade in the market allows investors to form coalitions at date 1. For investors to choose to join a financial coalition (deposit in the bank), each must get type-contingent consumption as desirable as can be obtained from joining another coalition (bank) or from investing directly. Because there are no costs of establishing banks, any individual asset holdings can be replicated by those of a competing bank. The ability to form competing banks imposes coalition incentive constraints at date 1. The importance of competing banks in models of this type was first identified by von Thadden (1997).

The constraint imposed by date 0 coalition formation is that the return from trading assets must equal the physical returns offered by the real assets. If these returns are not equal, a competing bank coalition can offer claims that allow its members who participate in markets to consume more. Real assets allow a date 0 cost of $1/R$ per unit of date 1 consumption and $1/X$ per unit of date 2 consumption. If the market price on date 1 of one unit of date 2 consumption is not R/X, then offering the option for depositors to withdraw and trade at that price will allow higher consumption. If some depositors can trade, then the date 0 marginal rate of transformation of date 1 to date 2 consumption must be equal to the ratio of prices that will prevail on date 1 of claims on unit claims maturing on those dates. Lemma 3 describes the incentive constraints.

LEMMA 3. A dominating competing bank coalition exists unless $b_1 = R/X$ and prices are in line with marginal productivity. Combined with the date 1 incentive constraints in lemma 1, this implies that the date 1 and date 0 incentive constraints are $c_{22A} = c_{11}(X/R)$, $c_{22B} \geq c_{11}R$, and $c_{22A} \geq c_{22B}$. The resource constraint then reduces to

$$(1 - q_1 - q_{2A})c_{22B} = \left[1 - (q_1 + q_{2A})\frac{c_{11}}{R}\right]X.$$

Proof. See the Appendix.

Combining lemma 2 and lemma 3 implies that the subsidy provided to type 1 investors must also be provided to type 2A investors. The ratio of consumptions c_{11}/c_{22A} must be constant. If not, one type can get a higher market value of proceeds by choosing the wrong withdrawal, or a competing bank can be set up that offers its depositors better returns. Cross-subsidy of both types 1 and 2A is possible, with the subsidy of liquidity provided by type 2B investors who cannot trade.

The banking system creates more liquidity than there would be without a banking system or secondary markets. The banking system also makes the secondary market more liquid: secondary markets will offer the amount of liquidity implied by the short-term physical return on capital. I assume that interbank deposits are identifiable as such (if only by their size). This prevents a competing bank from obtaining the liquidity subsidy provided by another bank by simply investing directly in the one-period deposits of the bank.[4] Some ben-

[4] If interbank deposits could not be identified, an argument similar to the one above shows that not only must $b_1 = R/X$, but also $c_{11} = R$ and $c_{22A} = c_{22B} = X$.

efits of liquidity creation can be focused on the individual bank's depositors. However, the effect of banks' liquidity creation on market liquidity is available to all competitors because of free entry into trades in the anonymous secondary market.

The condition for banks to create more liquidity than secondary markets is that risk aversion exceed one (so cross-subsidy is valuable) and that not too many investors participate in the secondary market (so much of the subsidy goes to type 1 investors who need liquidity). Proposition 1 states this result.

PROPOSITION 1. If the coefficient of relative risk aversion is above one and a sufficient fraction of investors do not participate, $q_{2B} > \hat{q}_{2B} > 0$, then banks provide more liquidity than the secondary market and set $c_{11} > R$ and $c_{22B} < X$. If the coefficient of relative risk aversion is less than or equal to one, the banks increase the liquidity of the market but provide no more liquidity than the market and set $c_{11} = R$ and $c_{22A} = c_{22B} = X$.

The proof is in the Appendix, but it is useful here to describe the key first-order condition. If relative risk aversion is greater than one, the first-order condition for optimal consumption is

$$q_1 U'(c_{11}) \leq \{q_1 U'(c_{22B}) + q_{2A}[U'(c_{22B}) - U'(c_{22A})]\} \frac{X}{R},$$

with equality whenever $c_{11} > R$, which occurs for q_{2A} not too large. If relative risk aversion is less than or equal to one, $c_{11} = R$ and $c_{22A} = c_{22B} = X$, because $c_{22A} = c_{11}(X/R) \geq c_{22B}$ is binding and there is no potential for a cross-subsidy provided to type 1 investors (with none desirable if risk aversion equals one).

The next proposition shows the effect of increased secondary market liquidity (increased q_{2A}) on the amount of liquidity created by banks.

PROPOSITION 2. Increasing individual participation in the secondary market (increasing q_{2A} by reducing q_{2B}) weakly reduces the liquidity that banks create relative to secondary markets ($c_{11} - R$) and reduces c_{11} (strictly if $c_{11} > R$), the short-term return available to investors.

Proof. See the Appendix.

Increased market participation reduces the cross-subsidy that banks provide to short-term holders, because it increases the fraction of the benefit that goes to those who profit from trading (type 2A investors who have high consumption) rather than to those who need liquidity for consumption (type 1 investors who have low consumption). The consumption of those who do not need liquidity and do not participate in the secondary market can increase or decrease (c_{22B} can rise or fall), but $c_{22B} - c_1$ rises as participation increases: there is less risk sharing between those who turn out to need liquidity and those who do not participate in secondary markets.

The Diamond-Dybvig and Jacklin Models

The Diamond-Dybvig model characterizes the optimal amount of liquidity cross-subsidization to provide when the option for an increased short–holding period return is financed by a lower long–holding period return. Jacklin (1987), Haubrich and King (1990), von Thadden (1997), and Hellwig (1994) examine the effects of competitive financial markets and reach largely negative conclusions about the viability of bank liquidity created by cross-subsidy. The results in the two propositions show that these results follow not from the existence of a financial market, but from a market in which all investors participate continuously. Increased participation in markets reduces the subsidy that short-term holders receive from long-term holders of bank deposits. In addition, even when banks provide no cross-subsidy, banks are important and changes in market participation have interesting effects.

The polar cases of propositions deliver the Diamond-Dybvig and Jacklin (1987) models. Proposition 1 delivers the Diamond-Dybvig result when $q_{2A} = 0$. If trade between agents is impossible ($q_{2A} = 0$ and no type 2A's exist), then investors with relative risk aversion above one choose cross-subsidization to allow them to hold liquid claims with high one-period returns (high $c_{11} > R$) at the expense of lower two-period returns (lower $c_{22B} < X$). Banks must offer a demand deposit in which each depositor is offered a choice between c_{11} at date 1 and c_{22B} at date 2 because no trade is possible, and because $c_{11} > Rc_{22B}$, all investors will self-select.

Proposition 1 delivers the Jacklin (1987) result when $q_{2B} = 0$. Jacklin shows that the ability of banking mechanisms to cross-subsidize investors is eliminated when there exists a full-participation secondary market for bank assets. In this case, cross-subsidization is impossible, and $c_{11} = R$ and $c_{22A} = X$ is the only feasible compatible consumption pair (there are no type 2B investors). The reason is the date 0 constraint: investors will prefer not to join the bank at date 0 if $c_{11}/c_{22A} \neq R/X$, because they then would prefer to trade in the market (no competing banks are required when there is full participation). In this case, $c_{22A} = c_{11}(X/R)$ combines with the resource constraint to yield $c_{11} = R$ and $c_{22A} = X$.

With a full-participation secondary market, there is not only no scope for cross-subsidization but no beneficial role for banks. If each investor holds a fraction q_1 in short-term assets and $q_{2A} = 1 - q_1$ in long-term assets, then $c_{11} = R$ and $c_{22A} = X$, without using banks. Jacklin's result has been interpreted as meaning that banks and markets cannot coexist, and if liquidity is enhanced ($c_{11} > R$), either markets or the direct holding of assets must be prohibited.

When there is a limited-participation market, there is a role for banks even without cross-subsidization. Unless q_{2A} is very high, there is cross-subsidization. The existence of a limited-participation market reduces but does not eliminate cross-subsidization. In addition, there is an interaction between the amount of cross-subsidization, the scale of the banking sector, and the degree of market participation. This is explored in the next section.

III. Direct Holdings and Bank Claims: The Scale of the Banking Sector

The contracts that banks write with investors influence the performance of financial markets. If there is limited participation in markets and all claims are held directly, financial markets will provide too little liquidity. Investing a fraction of wealth through banks diverts some demand for liquidity away from markets. This allows the limited supply of liquidity in the market to be better matched with demand. Not all the financial claims need to be held by banks at date 0 for banks and markets to provide increased liquidity to investors. If there are variable costs associated with running wealth through intermediaries, the scale of the banking sector is the minimum needed to implement the desired amount of liquidity. The scope for direct holdings arises because the optimal mechanism leads to a set of tradable claims held by investors before trade at date 1 that can be separated into two components: one is a holding that is identical for all investors (and thus not type-specific), and the other is a type-specific choice selected from the options offered by the bank at date 1. The total claim, $W_{t\tau}$, on date t consumption held by investor type τ on date 1 (before any trade) is decomposed into two parts: $W_{t\tau} = d_t + w_{t\tau}$, where d_t is the component that is not type-specific and $w_{t\tau}$ is type-specific. Each investor can directly hold assets that constitute the claims that are not type-specific. The claim $w_{t\tau}$ is the claim on date t consumption withdrawn by a type τ investor at date $t \in \{1, 2\}$.

The holding of short-term liquidity must be centralized to avoid inefficient reinvestment in short-term assets at date 1 by type 2B investors. This requires that no date 1 claims be held by type 2B investors ($W_{12B} = 0$), which requires that individuals hold no short-term assets directly ($d_1 = 0$). All directly held claims are long-term. If investors hold a fraction β of their date 0 wealth as bank claims, they invest $1 - \beta$ in long-term claims, and this gives each a direct holding of date 2 claims of $d_2 = (1 - \beta)X$. A lower bound on the date 0 scale of banks is the amount of assets needed to finance short-term investment. This lower bound on date 0 investment in banks is

$$\beta \geq \frac{q_1 c_{11}}{R} \equiv \beta_1$$

(scale of banks when they hold just all the short-term assets), because total date 1 consumption is c_{11} by a fraction q_1 of investors. If this minimal fraction of assets were invested in banks, the banking system would hold only short-term assets. All long-term assets would be held directly, and each investor would hold date 2 claims of $(1 - \beta_1)X$.

Incentive-Compatible Bank Withdrawals

Too large a direct holding of long-term assets can be inconsistent with self-selection of the proper type-specific withdrawals at date 1. The problem arises when banks cross-subsidize short-term holders, because the high short-term return (c_{11} in excess of R) is not reflected in the market prices of directly held assets. Positive holdings of date 2 claims by all types tighten the incentive constraint that type 2B investors choose to leave their funds in the bank until date 2. This constraint is loosest when all claims selected by types 1 and 2A are date 1 claims, because trade allows them a higher return at date 1 than is available to nonparticipating type 2B investors. When investors all directly hold a fraction $1 - \beta$ of date 2 claims, the bank claims selected by both types 1 and 2A are $w_{11} = w_{12A} = c_{11} - (1 - \beta)R$ and $w_{21} = w_{22A} = 0$. The value of the date 2 bank withdrawal intended for type 2B investors is $w_{22B} = c_{22B} - (1 - \beta)X$. It is incentive-compatible for type 2B investors to choose w_{22B} instead of taking w_{22} and investing it in a short-term investment at date 1 only if $w_{22B} > Rw_{11}$, which is equivalent to

$$\beta \geq 1 - \frac{c_{22B} - Rc_{11}}{X - R^2} \equiv \beta_{IC}.$$

This incentive constraint need not be binding because β_{IC} can be less than $\beta_1 = q_1 c_{11}/R$. For sufficiently low risk aversion or for sufficiently high participation ($q_{2A} \to 1 - q_1$), cross-subsidization is low, and $\beta_{IC} \to 0$ (and is less than β_1), because $c_{22B} - Rc_{11} \to X - R^2$. In this case, the banking system does not hold long-term assets. For sufficiently high risk aversion and sufficiently low market participation, $\beta_{IC} > \beta_1$, and the banking system holds a fraction $\beta_{IC} - \beta_1$ of long-term assets. Increasing market participation (q_{2A} increases for fixed q_1) reduces β_{IC}, reducing the banks' holdings of long-term assets, eventually to zero as $q_{2A} \rightarrow 1 - q_1$.

Market Clearing and Bank Trades

Up to this point, no constraint has been imposed on bank trades in markets. Given total pretrade holdings W_{tt}, individual traders' supply of short-term claims to the market is $q_{2A} W_{12A}$, and the date 1 value of the date 2 claims offered by individuals is $q_1 W_{21}(R/X)$. For individuals, the date 1 excess supply of date 1 claims is $q_{2A} W_{12A} - (R/X) q_1 W_{21} \equiv m_1$. Market clearing implies that the bank buys (with date 2 claims) date 1 claims of m_1 on date 1 (sells date 1 claims of $-m_1$ when $m_1 < 0$). Setting a small value of β can require $m_1 < 0$, which requires that the bank sell short-term claims to buy existing long-term assets in the market at date 1. Factors such as limited information that constrain type 2B agents' ability to value existing long-term assets might plausibly also prevent banks from valuing those assets and require that $m_1 \geq 0$. The impact of requiring $m_1 \geq 0$ is analyzed in the next subsection. Note that a date 1 market is required whenever there is cross-subsidization ($c_{22A} > c_{22B}$), because some date 1 claims must be selected by type 2A agents to allow type 2A investors to choose high consumption not available to nontrading type 2B investors.

Scale of the Banking Sector and Banks' Ability to Trade Assets

To determine the link between bank trades in the market and the implied scale of the banking sector, begin with the benchmark in which banks do not trade in the financial market and $m_1 = 0$. If the bank makes no trades, it must hold sufficient assets to provide type 1 agents with consumption c_{11}, plus provide enough date 2 assets to provide the excess of type 2B's consumption over that obtained from their direct holdings of assets. If the bank holds more assets, it must sell some date 2 assets at date 1 or give date 2 claims to types 1 and 2A for them to trade. If the bank holds fewer assets, it must buy some date 2 claims in the market at date 1. When $m_1 = 0$, the date 2 assets held directly by type 1 and 2A investors will finance the consumption of type 2A investors, and those held directly by type 2B investors will finance the part of their own consumption that does not come from bank deposits.

All investors will choose the same direct holding on date 0, when their liquidity need and type are unknown. Without trading with the bank, consumption of type 2A investors can come from holdings of date 2 claims by type 1 and 2A agents. The date 2 value of date 2 claims held by types 1 and 2A must be $q_{2A}c_{22A}$. Date 0 direct holdings of long-term claims must equal $q_{2A}c_{22A}/(q_1 + q_{2A})X$, because a fraction $q_1 + q_{2A}$ of direct claims are held by types 1 and 2A. Because type 2A consumption is $c_{22A} = c_{11}X/R$, the date 0 value of direct holding when banks do not trade is $1 - \beta = q_{2A}c_{11}X/(q_1 + q_{2A})R$. This implies that the balance of date 0 capital is invested by the bank, and if $m_1 = 0$, the scale of the banking sector is

$$\beta = 1 - \frac{q_{2A}c_{11}X}{(q_1 + q_{2A})R} \equiv \beta_{MC}$$

(scale of banks with no trade by banks). The minimum scale of banking when there is no bank trade weakly exceeds the minimum scale q_1c_{11}/R of holding only short-term assets, because the market-clearing condition requires that the bank hold some long-term assets when $c_{11} > R$.

The scale of the banking sector at date 0 is $\beta_{MC} + (m_1/R)$. If the bank cannot value others' long-term assets and $m_1 \geq 0$, then the minimum scale of the banking industry is $\max\{\beta_{MC}, \beta_{IC}\}$. The value of β_{MC} decreases with increased market participation, and β_{MC} goes from one to q_1 as q_{2A} goes from zero to $1 - q_1$. If $\beta_{MC} < \beta_{IC}$, then the bank holds more long-term assets at date 0 than needed and sells them to type 2A depositors at date 1. An alternative explanation of this is that the bank must raise some deposits at date 1 from type 2A investors by paying market rates of interest.

In summary, if the bank participates fully in markets, $\beta = \max\{\beta_{IC}, q_1c_{11}/R\}$. If the banks face the market participation constraint equal to that of type 2B investors, the scale of banking is

$$\beta = \max\left\{\beta_{IC}, \frac{q_1c_{11}}{R}, \frac{q_{2A}c_{11}X}{(q_1 + q_{2A})R}\right\}.$$

Whether or not the bank can participate in markets on behalf of investors, the banking sector shrinks as financial market participation increases. There are other interesting interpretations of this result. Because the banking system issues liabilities with an option to withdraw at date 1 and the remaining assets held by investors are long-term, the scale of the banking system measures the proportion of financial assets that are short-term. Whenever the scale of the banking sector exceeds that implied by the minimal liquidity needs of the economy (its scale exceeds $\beta_1 = q_1c_{11}/R$), the banks hold long-term assets as well. An increase in the amount of long-term assets held by banks is an increase in the mismatch between the maturities of real and financial assets in the economy. Proposition 3 summarizes the results in this section.

Proposition 3. The scale of the banking sector, the fraction of financial assets that are short-term, the fraction of real assets that are short-term, and the gap between the maturities of financial and real assets all decrease as direct market participation increases (q_{2A} increases and q_{2B} decreases).

IV. Alternative Interpretations and Possible Extensions

Several of the assumptions of the model can be generalized without qualitatively changing the results and implications. This section describes alternative assumptions and alternative interpretations of what has been assumed. Some open questions for future work are also identified.

The key exogenous variable in the model is the amount of direct participation in financial markets. Exploration of the factors that limit participation, and the impact of these factors on banks and markets, is beyond the scope of this paper. It is useful to outline some additional motivation for the specification used here. A motivation for limited participation based on asymmetry of information follows. Suppose that evaluating existing assets offered for resale in the market is costly, and they cannot be distinguished from less valuable assets. Investors with low costs will trade in the market for existing assets, and those with high costs will not (see Akerlof 1970). One could imagine that there is a continuous cost of information and that the informed are those whose cost is below a given level. For simplicity imagine that there are informed investors who get information at no cost and uninformed investors who get it only at a prohibitive cost. Only the informed investors (type 2A) and those who are in need of liquidity (type 1) are active in the secondary market. Increased disclosure of public information or a reduction in the amount of private information would increase the fraction of traders active in the secondary market, by transferring some traders from the uninformed group to the informed group. As a result, one can interpret an increase in direct participation as a reduction in the cost of acquiring information.

Explicit study of private information would also be useful in motivating the limits to arbitrage by competing banks. The current model assumes that a bank cannot borrow at date 1 from the type 2B customers of another bank by issuing deposits backed by the bank's existing long-term assets. The information-based motivation is that retaining a claim in an existing bank that does not buy others' long-term assets gives investors the unconditional mean return from those assets. Because of the potential for adverse selection, this exceeds the anticipated return from switching to another bank that chooses to raise additional funds from uninformed type 2B investors rather than informed 2A investors in the market. The alternative assumption that the other bank can always borrow from type 2B investors at date 1 would imply that having assets held by a bank allows full participation in the market for all claims that are derivatives of the underlying real assets in the economy. This would yield the same consumption levels as a full-participation market in the underlying assets. There would be no cross-subsidization, but banks would still be needed to hold assets. It is possible that some of these issues could be clarified by embedding the analysis in an overlapping generations model, such as that in Qi (1994).

The model can easily accommodate costs of financial intermediation (e.g., a proportional cost of banks' holding assets), but the analysis is much more complicated with very few new insights. The effects of these costs can be seen by comparing the case in which there are no intermediation costs, analyzed here, with that in which the costs are so large that all assets are held directly. When intermediation costs are so high that all assets are held directly, each investor invests more in short-term assets. Yet the consumption of those who need liquidity (c_{11}) is lower, and the secondary market price of long-term assets (b_1) is lower than it is when banks face low intermediation costs. Reduced intermediation costs make financial markets more liquid and lower the opportunity cost of liquidity. In addition, in the case in which banks can participate in the market, the reduction of intermediation costs also increases the volume of trade in the financial market. Reduced bank costs can increase the volume of trade even when banks cannot access the secondary market, because of the increased holding of long-term assets by individuals due to the higher secondary market prices of long-term assets. This suggests that improvements in banking, through reduced costs or less oppressive regulation, will be conducive to the liquidity of financial markets and to financial market development. Improvements in access to financial markets (increased disclosure and transparency) that make the market more liquid will diminish the role of banks but will also reduce banks' costs if the improvements provide increased bank access to the market. This two-way causality suggests that empirical study of the roles of banks and markets must use structural information to disentangle the effects. The line of empirical research started by Demirgüç-Kunt and Levine (1996) on banks, markets, and development has documented that banks and markets tend to develop together. Future work should attempt to disentangle the conflicting effects of banks and markets on each other.

Finally, this analysis abstracts from important problems with enforcement of property rights over collateral and other bankruptcy/enforcement issues that are also present in many developing countries. Explicit analysis of information and incentives could allow these issues to be integrated into the analysis of participation in markets.

V. Conclusion

With limited participation in markets, the banking system creates liquidity in two ways. First, banks fill the liquidity gap in markets by diverting demand for liquidity from markets. This improves the market's liquidity, increasing the price of illiquid assets above what it is when all assets are held directly. Second, if investors are sufficiently risk averse and enough do not participate in markets, bank deposits provide higher short-term returns than the market. The short-term assets held by individuals are bank liabilities, and the short-term real ''reserve assets'' are held only by banks. In this case, banks provide a cross-subsidy to those who withdraw early that is financed by those who hold bank claims for many periods.

The model can be most clearly applied to the understanding of financial markets and institutions in developing economies. Limited participation in secondary markets implies that the maturity structure of financial claims will adjust to fill the gap by allowing individuals to hold self-liquidating financial claims. As the financial markets develop, one should expect to see increased use of longer-term claims such as long-term debt or equity. The analysis implies that there will be a small supply of long-term direct claims in economies in which few participate in financial markets. The banking system will have a large role in the allocation of capital and the provision of liquidity.

More participation in markets leads to less cross-subsidization of short-term returns by banks, a smaller banking sector, and a longer average maturity of financial assets. More participation also leads to longer-maturity physical investment and a smaller gap between the maturity of financial assets and physical investments. In addition, as more liquid markets force the banking system to shrink, the banks' holdings of long-term assets (term loans) will shrink more rapidly than their holdings of shorter-term loans. The empirical study by

Hoshi, Kashyap, and Scharfstein (1990) documents the effects of market development on banks and their structure. Regulatory changes opened access to the Japanese bond market. The effects were broadly in line with the implications of this model. Banks' market share was reduced, and their holdings of long-term assets fell more rapidly than their holdings of short-term assets.

The analysis also has implications for the effect of development of the banking sector on financial markets. Adding banks, or reducing their costs of operation, makes liquidity cheaper to obtain, and this makes markets more liquid. Because investors then choose to hold more long-term assets, the development of a banking system will lead to increased turnover and volume in financial markets. These links between banks and markets are worthy of further study. The current model of the link between liquidity provided by financial institutions and liquidity provided by markets is quite rudimentary, but I hope that further study of this link will provide more insight into these issues in financial structure and development.

Appendix

Proof of Lemma 2

Let $W_{t\tau}$ denote the pretrade date 1 holding of date t claims by a type τ investor. After choosing these claims, type 1 and 2A investors have the ability to trade anonymously at date 1 and to privately consume the proceeds from those trades. Let b_1 denote the date 1 price of a claim on one unit of date 2 consumption. A type 2B investor does not have access to the financial market. As a result, the only way that a type 2B investor can convert date 1 consumption into date 2 consumption is to use the proceeds of date 1 claims and invest in new short-term investments that pay off on date 2. This implies that the final consumption levels of each type of agent are given by

$$c_{11} = W_{11} + W_{21} b_1,$$

$$c_{22\mathrm{A}} = \frac{W_{12\mathrm{A}}}{b_1} + W_{22\mathrm{A}},$$

$$c_{22\mathrm{B}} = W_{12\mathrm{B}} R + W_{22\mathrm{B}}.$$

The type-contingent consumption offered on date 1 is incentive-compatible if and only if no investor prefers the consumption implied by the claims $W_{t\tau}$ intended for another type of investor. Let $c_{t\tau}^{\bar{\tau}}$ denote the consumption on date t of a type τ investor who misrepresents himself or herself as a type $\bar{\tau}$ investor, choosing the claims $W_{1\bar{\tau}}$, $W_{2\bar{\tau}}$ and trading at the market price b_1 if of type 1 or 2A. With this definition and the definitions of individual consumption, $c_{t\tau}$, given above, the following are the date 1 constraints on incentive-compatible consumption (IC τ, $\bar{\tau}$):

$$c_{11} = W_{11} + b_1 W_{21} \geq W_{12A} + b_1 W_{22A} \equiv c_{11}^{2A} = c_{22A} b_1, \qquad \text{(IC 1, 2A)}$$

$$c_{11} = W_{11} + b_1 W_{21} \geq W_{12B} + b_1 W_{22B} \equiv c_{11}^{2B} = c_{22B} b_1, \qquad \text{(IC 1, 2B)}$$

$$c_{22A} = \frac{W_{12A}}{b_1} + W_{22A} \geq \frac{W_{11}}{b_1} + W_{21} \equiv c_{22A}^{1} = \frac{c_{11}}{b_1}, \qquad \text{(IC 2A, 1)}$$

$$c_{22A} = \frac{W_{12A}}{b_1} + W_{22A} \geq \frac{W_{12B}}{b_1} + W_{22B} \equiv c_{22A}^{2B} \geq c_{22B}, \qquad \text{(IC 2A, 2B)}$$

$$c_{22B} = W_{12B} R + W_{22B} \geq W_{11} R + W_{21} \equiv c_{22B}^{1}, \qquad \text{(IC 2B, 1)}$$

$$c_{22B} = W_{12B} R + W_{22B} \geq W_{12A} R + W_{22A} \equiv c_{22B}^{2A}. \qquad \text{(IC 2B, 2A)}$$

The constraints (IC 1, 2A) and (IC 2A, 1) together imply that $c_{11} = c_{22A} b_1$ and $b_1 = c_{11}/c_{22A}$. If the relative price of date 2 consumption in terms of date 1 consumption were not c_{11}/c_{22A}, either type 1 or type 2A would prefer to take and then sell the claim withdrawn by the other type of investor, because the date 1 market value of the claims would differ. The market value of the amount withdrawn by type 1 investors must equal that of type 2A investors; otherwise both will take the one with higher market value and trade to get higher consumption on the desired date.

This implies that types 1 and 2A are indifferent about taking the claims intended for either of the two types. One feasible allocation is $W_{11} = W_{12A} = c_{11}$, $W_{21} = W_{22A} = W_{12A} = 0$, and $W_{22B} = c_{22B}$. In this case, the consumption of type 2B's must exceed $c_{11} R$, which is not binding. The date 1 market clears as follows. All wealth is deposited in the bank at date 0. Type 1 depositors take and consume W_{11}, and type 2A's take W_{12A} and sell in the market to the bank in exchange for $c_{22A} = W_{12A}(X/R)$. In this allocation, the bank holds all assets at date 0 and is the only seller of date 2 claims at date 1, when type 2A agents are the only buyers. The type 2B's do not withdraw at date 1, but take $W_{22B} = c_{22B}$ at date 2. The bank need not hold all assets, and individuals can both buy and sell assets at date 1. This has implications for the equilibrium scale of the banking sector when banks and markets coexist. These implications are developed in Section III. Q.E.D.

Proof of Lemma 3

Suppose that at date 0, a competing bank contract can be proposed by "bank II." Bank II accepts deposits at date 0 and offers date 1 and date 2 type-contingent payments and a portfolio policy. A contract offered by bank I is date 0 coalition incentive-compatible if no dominating contract can be proposed on date 0 by bank II. A contract offered by bank II can offer its members claims on date 1 and date 2 consumption that its type 1 or type 2A members can use to trade on date 1 in the anonymous market that includes members of bank I. The constraint imposed by the possibility of trade after withdrawing from an individual bank is that the price $b_1 = c_{11}/c_{22A}$. The constraint that a competing bank not propose a dominating contract is $b_1 = R/X$. If this does not hold, then a competing bank can give tradable claims to its members that lead to superior consumption opportunities for its members who can trade at price b_1.

Suppose that bank I proposes a contract that, if no competing contract were proposed, would lead to type-contingent consumptions (c^{I}_{11}, $c^{\mathrm{I}}_{22\mathrm{A}}$, and $c^{\mathrm{I}}_{22\mathrm{B}}$), with $c^{\mathrm{I}}_{11}/c^{\mathrm{I}}_{22\mathrm{A}} > R/X$. If no competing contract is proposed, then b_1, the date 1 price of date 2 claims, will be high: $b^{\mathrm{I}}_1 = c^{\mathrm{I}}_{11}/c^{\mathrm{I}}_{22\mathrm{A}} > R/X$. This allows bank II to propose a dominating contract. Suppose that bank II proposes a contract that gives the same $c_{22\mathrm{B}}$ as bank I ($c^{\mathrm{II}}_{22\mathrm{B}} = c^{\mathrm{I}}_{22\mathrm{B}}$) but invests more of the remaining capital in long-term assets (and less in short-term) to give types 1 and 2A tradable claims slightly biased toward date 2 consumption. Investing one unit more in long-term assets and one fewer in short-term claims allows an R decrease in date 1 and an X-unit increase in date 2 claims. Choose $\epsilon > 0$ such that

$$W^{\mathrm{II}}_{11} = W^{\mathrm{II}}_{12\mathrm{A}} = W^{\mathrm{II}}_{1} = \frac{q_1 c^{\mathrm{I}}_{11}}{q_1 + q_{2\mathrm{A}}} - \epsilon,$$

$$W^{\mathrm{II}}_{21} = W^{\mathrm{II}}_{22\mathrm{A}} = W^{\mathrm{II}}_{2} = \frac{q_{2\mathrm{A}} c^{\mathrm{I}}_{22\mathrm{A}}}{q_1 + q_{2\mathrm{A}}} + \frac{\epsilon X}{R}$$

such that

$$\frac{R}{X} \leq \frac{q_1 W^{\mathrm{II}}_1}{q_{2\mathrm{A}} W^{\mathrm{II}}_2} < \frac{c^{\mathrm{I}}_{11}}{c^{\mathrm{I}}_{22\mathrm{A}}}.^{5}$$

This implies that

$$\begin{aligned} c^{\mathrm{II}}_{11} &= \frac{q_1 c^{\mathrm{I}}_{11}}{q_1 + q_{2\mathrm{A}}} - \epsilon + \left(\frac{q_{2\mathrm{A}} c^{\mathrm{I}}_{22\mathrm{A}}}{q_1 + q_{2\mathrm{A}}} + \frac{\epsilon X}{R}\right) b_1 \\ &= \frac{q_1 c^{\mathrm{I}}_{11}}{q_1 + q_{2\mathrm{A}}} - \epsilon + \left(\frac{q_{2\mathrm{A}} c^{\mathrm{I}}_{22\mathrm{A}}}{q_1 + q_{2\mathrm{A}}} + \frac{\epsilon X}{R}\right) \frac{c^{\mathrm{I}}_{11}}{c^{\mathrm{I}}_{22\mathrm{A}}} \\ &= c^{\mathrm{I}}_{11} + \epsilon\left(\frac{c^{\mathrm{I}}_{11} X}{c^{\mathrm{I}}_{22\mathrm{A}} R} - 1\right) > c^{\mathrm{I}}_{11}. \end{aligned}$$

Similarly, for type 2A agents,

$$\begin{aligned} c^{\mathrm{II}}_{22\mathrm{A}} &= \left(\frac{q_1 c^{\mathrm{I}}_{11}}{q_1 + q_{2\mathrm{A}}} - \epsilon\right) \frac{1}{b_1} + \frac{q_{2\mathrm{A}} c^{\mathrm{I}}_{22\mathrm{A}}}{q_1 + q_{2\mathrm{A}}} + \frac{\epsilon X}{R} \\ &= \left(\frac{q_1 c^{\mathrm{I}}_{11}}{q_1 + q_{2\mathrm{A}}} - \epsilon\right) \frac{c^{\mathrm{I}}_{22\mathrm{A}}}{c^{\mathrm{I}}_{11}} + \frac{q_{2\mathrm{A}} c^{\mathrm{I}}_{22\mathrm{A}}}{q_1 + q_{2\mathrm{A}}} + \frac{\epsilon X}{R} \\ &= c^{\mathrm{I}}_{22\mathrm{A}} + \epsilon\left(\frac{c^{\mathrm{I}}_{11} X}{c^{\mathrm{I}}_{22\mathrm{A}} R} - 1\right) > c^{\mathrm{I}}_{22\mathrm{A}}. \end{aligned}$$

Trade with members of bank I at price b^{I}_1 would allow members of bank II to get date 1 consumption at date 0 cost $(b^{\mathrm{I}}_1 X)^{-1} < 1/R$, which is less

[5] As an alternative to increasing W^{II}_2, bank II could directly sell long-term assets at date 1.

than the actual date 0 cost of date 1 consumption. If the price ratio, b_1, of date 1 to date 2 consumption is not in line with marginal productivity, R/X, a competing bank can offer a dominating contract. A symmetric argument rules out $b_1 < R/X$. If, and only if, $b_1 = R/X$, is there no dominating contract possible for a competing bank.

Proof of Propositions 1 and 2

When we substitute in the resource constraint, the objective function, Φ, becomes

$$\Phi = q_1 U(c_{11}) + q_{2A} U\left(\frac{c_{11} X}{R}\right) + q_{2B} U\left(\frac{[1 - q_1(c_{11}/R) - q_{2A}(c_{11}/R)]X}{q_{2B}}\right),$$

$$\Phi'(c_{11}) = q_1 U'(c_{11}) + q_{2A} U'\left(\frac{c_{11} X}{R}\right)\frac{X}{R} - (q_1 + q_{2A}) U'\left(\frac{[1 - q_1(c_{11}/R) - q_{2A}(c_{11}/R)]X}{q_{2B}}\right)\frac{X}{R}.$$

At $c_{11} \leq R$, the resource constraint implies that

$$c_{22B} = \frac{[1 - q_1(c_{11}/R) - q_{2A}(c_{11}/R)]X}{q_{2B}} \geq X$$

and

$$c_{22A} \leq X.$$

The function $U(c)$ is more risk averse than $\log(c)$, and $U'(c) > ZU'(cZ)$ for $Z > 1$, implying that $U'(c_{11}) > U'(c_{11}(X/R))(X/R)$. Risk aversion implies that

$$U'\left(\frac{c_{11} X}{R}\right)\frac{X}{R} \geq U'\left(\frac{[1 - q_1(c_{11}/R) - q_{2A}(c_{11}/R)]X}{q_{2B}}\right)\frac{X}{R}.$$

These two results imply that $\Phi'(c_{11}) > 0$ for $c_{11} \leq R$. Because $\Phi(c_{11})$ is continuous but not differentiable at $c_{11} = R$, the optimal value of $c_{11} \geq R$.

The function Φ is concave:

$$\Phi''(c_{11}) = q_1 U''(c_{11}) + q_{2A} U''\left(c_{11}\frac{X}{R}\right)\frac{X^2}{R^2} + \frac{(q_1 + q_{2A})^2}{q_{2B}} \times U''\left(\frac{[1 - q_1(c_{11}/R) - q_{2A}(c_{11}/R)]X}{q_{2B}}\right)\frac{X^2}{R^2} < 0,$$

because $U''(c) < 0$.

Proof That the Right Derivative at $c_{11} = R$ *Is Negative if* q_{2B} *Is Small*

Set $c_{11} = R + \epsilon$ for $\epsilon > 0$. From $q_{2B} = 1 - q_1 - q_{2A}$,

$$\begin{aligned}\Phi'(R+\epsilon) &= q_1 U'(R+\epsilon) + q_{2A} U'\left(\frac{(R+\epsilon)X}{R}\right)\frac{X}{R} \\ &\quad - (q_1 + q_{2A})U'\left(\frac{\{1-(q_1+q_{2A})[(R+\epsilon)/R]\}X}{1-q_1-q_{2A}}\right)\frac{X}{R} \\ &= q_1 U'(R+\epsilon) + q_{2A}U'\left(\left(1+\frac{\epsilon}{R}\right)X\right)X \\ &\quad - (q_1+q_{2A})U'\left(\frac{\{1-(q_1+q_{2A})[1+(\epsilon/R)]\}X}{1-q_1-q_{2A}}\right)\frac{X}{R} \\ &= q_1\left[U'(R+\epsilon) - U'\left(X - \frac{(q_1+q_{2A})\epsilon X}{1-q_1-q_{2A}}\right)\frac{X}{R}\right] \\ &\quad + q_{2A}\frac{X}{R}\left[U'\left(X+\frac{\epsilon X}{R}\right)X - U'\left(X - \frac{(q_1+q_{2A})\epsilon X}{1-q_1-q_{2A}}\right)\right].\end{aligned}$$

For any fixed $\epsilon > 0$ (and for all smaller values of ϵ), one can choose $q_{2B} = 1 - q_1 - q_{2A} > 0$ such that

$$q_{2A}\frac{X}{R}\left[U'\left(X+\frac{\epsilon X}{R}\right) - U'\left(X - \frac{(q_1+q_{2A})\epsilon X}{1-q_1-q_{2A}}\right)\right]$$

is arbitrarily negative; in particular, is less than

$$-q_1\left[U'(R+\epsilon) - U'\left(X - \frac{(q_1+q_{2A})\epsilon X}{1-q_1-q_{2A}}\right)\frac{X}{R}\right],$$

implying that the right derivative at $c_{11} = R$ is negative if $q_{2B} > 0$ is sufficiently small.

By a similar argument, if q_{2A} is sufficiently small, then there exists $\epsilon > 0$ such that $\Phi'(R+\epsilon) > 0$; because $\Phi(c_{11})$ is concave, the right derivative at $c_{11} = R$ is positive, and the optimal value of c_{11} exceeds R.

As to proposition 2, increasing liquidity implies that more type 2 agents participate and that q_{2B} decreases as q_{2A} increases. If the solution is not at the kink at $c_{11} = R$, then $c_{11} > R$, and $c_{22A} > c_{22B}$. If $c_{11} > R$, then $c_{22A} = c_{11}X/R > c_{22B}$, from concavity of $U(\cdot)$. We have $U'(c_{11})(X/R) > U'(c_{22B})(X/R)$, and as a result, $\Phi'(c_{11})$ is strictly decreasing in q_{2A}.

The expression

$$\frac{\partial \Phi'(c_{11})}{\partial q_{2A}} = \frac{X}{R}\left[U'(c_{22A}) - U'(c_{22B}) - \frac{q_1+q_{2A}}{q_{2B}}(c_{22B}-c_{22A})U''(c_{22B})\right]$$

is less than zero because at the optimum $c_{22A} > c_{22B}$, and concavity of $U(\cdot)$ implies both $U'(c_{22A}) < U'(c_{22B})$ and $U''(c_{22B}) < 0$.

Combined with the previous result that $\Phi''(c_{11}) < 0$, this implies that

$$\frac{\partial c_{11}}{\partial q_{2A}} = -\frac{\partial \Phi'(c_{11})}{\partial q_{2A}} \Phi''(c_{11}) < 0,$$

and the optimal value of c_{11} is decreasing in q_{2A}. This proves proposition 2.

References

Akerlof, George A. ''The Market for 'Lemons': Quality Uncertainty and the Market Mechanism.'' *Q.J.E.* 84 (August 1970): 488–500.

Allen, Franklin, and Gale, Douglas. ''Limited Market Participation and Volatility of Asset Prices.'' *A.E.R.* 84 (September 1994): 933–55.

Bhattacharya, Sudipto, and Gale, Douglas M. ''Preference Shocks, Liquidity, and Central Bank Policy.'' In *New Approaches to Monetary Economics,* edited by William A. Barnett and Kenneth J. Singleton. New York: Cambridge Univ. Press, 1987.

Bryant, John. ''A Model of Reserves, Bank Runs, and Deposit Insurance.'' *J. Banking and Finance* 4 (December 1980): 335–44.

Demirgüç-Kunt, Asli, and Levine, Ross. ''Stock Market Development and Financial Intermediaries: Stylized Facts.'' *World Bank Econ. Rev.* 10 (May 1996): 291–321.

Diamond, Douglas W. ''Debt Maturity Structure and Liquidity Risk.'' *Q.J.E.* 106 (August 1991): 709–37.

———. ''Seniority and Maturity of Debt Contracts.'' *J. Financial Econ.* 33 (June 1993): 341–68.

Diamond, Douglas W., and Dybvig, Philip H. ''Bank Runs, Deposit Insurance, and Liquidity.'' *J.P.E.* 91 (June 1983): 401–19.

Gorton, Gary, and Pennacchi, George. ''Financial Intermediaries and Liquidity Creation.'' *J. Finance* 45 (March 1990): 49–71.

Haubrich, Joseph G., and King, Robert G. ''Banking and Insurance.'' *J. Monetary Econ.* 26 (December 1990): 361–86.

Hellwig, Martin. ''Liquidity Provision, Banking, and the Allocation of Interest Rate Risk.'' *European Econ. Rev.* 38 (August 1994): 1363–89.

Holmström, Bengt, and Tirole, Jean. ''Private and Public Supply of Liquidity.'' Working paper. Cambridge: Massachusetts Inst. Tech., Dept. Econ., 1996.

Hoshi, Takeo; Kashyap, Anil K.; and Scharfstein, David S. ''Bank Monitoring and Investment: Evidence from the Changing Structure of Japanese Corporate Banking Relationships.'' In *Asymmetric Information, Corporate Finance, and Investment,* edited by R. Glenn Hubbard. Chicago: Univ. Chicago Press (for NBER), 1990.

Jacklin, Charles J. ''Demand Deposits, Trading Restrictions, and Risk Sharing.'' In *Contractual Arrangements for Intertemporal Trade,* edited by Edward C. Prescott and Neil Wallace. Minneapolis: Univ. Minnesota Press, 1987.

Merton, Robert C. ''A Simple Model of Capital Market Equilibrium with Incomplete Information.'' *J. Finance* 42 (July 1987): 483–510.

Qi, Jianping. ''Bank Liquidity and Stability in an Overlapping Generation Model.'' *Rev. Financial Studies* 7 (Summer 1994): 389–417.

von Thadden, Ernst-Ludwig. ''The Term Structure of Investment and the Banks' Insurance Function.'' *European Econ. Rev.* 41, no. 3 (1997).

Wallace, Neil. ''Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously.'' *Fed. Reserve Bank Minneapolis Q. Rev.* 12 (Fall 1988): 3–16.