# Financial Regulation in a Changing Environment*

Douglas Gale
Department of Economics
New York University
269 Mercer Street
New York, NY 10003
douglas.gale@nyu.edu

January 3, 2002

## 1 The flight from theory

Financial crises have become a popular academic subject since the recent events in Asia, Russia, and elsewhere. Of course, financial crises are nothing new. They are part of the long and colorful history of the development of the financial system. They are also an important part of the history of central banking. Originally established to help governments finance wars, central banks later took on the prevention and control of financial crises as one of their central functions. The Bank of England perfected the technique in the nineteenth century. The Federal Reserve System, founded in the early

*I am indebted to the participants in the Conference on "Framing Financial Structure in an Information Environment" at Queen's University for the lively discussion following this address and, as always, to thank Franklin Allen for his advice and comments. The financial support of the C.V. Starr Center for Applied Economics at NYU, the National Science Foundation under Grant No. SES 0095109, and the conference sponsors, is gratefully acknnowledged.

twentieth century, was a slow learner and only mastered the technique in the 1930s.[1]

For the most part, the development of central banking and financial regulation has been an essentially empirical process, a matter of trial and error driven by the exigencies of history, rather than formal theory. An episode that illustrates the character of this process is the Great Depression in the US. The financial collapse in the US was widespread and deeply disruptive. It led to changes that shape our current regulatory framework. Investment and commercial banking were segregated by the Glass-Steagall Act. The SEC was established to regulate financial markets. The Federal Reserve Board revised its operating procedures in the light of its failure to prevent the financial collapse. The FDIC and FSLIC were set up to provide deposit insurance to banks and savings and loan institutions.

Looking back, there is no sign of formal theory guiding these changes. Everyone seems to have agreed the experience was terrible; so terrible that it must never be allowed to happen again. But why was this set of institutions and rules adopted? And why are many of them still with us today? The mindset of the 1930s continues to influence thinking about policy. According to this mindset, the financial system is extremely fragile and the purpose of prudential regulation is to prevent financial crisis at all costs. In addition, policy making continues to be an empirical exercise, with little attention to theoretical reasoning.

The Basel Accords, which impose capital adequacy requirements on the banking systems of the signatory countries around the world, are a case in point. Practitioners have become experts at the details of a highly complex system for which there is no widely agreed rationale based in economic theory. What is the optimal capital structure? What market failure necessitates the imposition of capital adequacy requirements? Why can't the market be left to determine the appropriate level of capital? We do not find answers to these questions in the theoretical literature.

It is not my intention to pass judgment on the practical value of any of the innovations mentioned above, but simply to point out that this empirical procedure is unusual. Indeed, the area of financial regulation is somewhat unique in the extent to which the empirical developments have so far outstripped theory. In most areas of economics, when regulation becomes an

[1] For a more detailed discussion of the development of central banking, see Chapter 2 of Allen and Gale (2000e).

issue, economists have tried to identify some specific market failure that justifies intervention. Sometimes they have gone further to derive the optimal form of regulation. But there is no theory of optimal financial regulation.

In these remarks, I want to use some simple theoretical models to address the rationale for capital adequacy requirements. The models are too simple to provide convincing answers. The purpose is to clarify issues, offer some insights, and perhaps provide a useful baseline for further reflection.

## 2 Insights from a simple model of risk sharing

In a series of related papers (Allen and Gale, 1998, 2000a, b, c, d), Franklin Allen and I describe a model that integrates intermediation and capital markets in a way that proves useful for the analysis of asset-price volatility, liquidity provision, financial crises, and related issues.

The model can be briefly described as follows. There are two types of assets in the economy, short-term assets that yield an immediate but low return and long-term assets that yield a higher but delayed return. Risk averse individuals want to invest to provide for future consumption. However, they are uncertain about their preferences regarding the timing of consumption. If they invest in the long-term asset, they earn a high return, but it may not be available when they want to consume it. If they invest in the short-term asset, they have the certainty that it will be available when they want it, but they have to forego the higher return of the long-term asset. In short, there is a trade-off between liquidity and rate of return.

Banks are modeled as institutions that provide an optimal combination of liquidity and return.[2] Banks take deposits from consumers and invest them in a portfolio of long- and short-term assets. In exchange, the bank gives the individual a deposit contract, that is, an option to withdraw from the bank. The amount withdrawn depends on the date at which the option is exercised, but for a given date, liquidity is guaranteed. By pooling independent risks, the bank is able to provide a better combination of liquidity and return than an individual could achieve on his own. The aggregate demand for liquidity is less volatile than individual risks, so the bank can guarantee the same degree of liquidity while investing a smaller fraction of the portfolio in short-term

[2] In this respect we are simply following Diamond and Dybvig (1983) and a host of other writers, e.g., Chari and Jagannathan (1988), Jacklin and Bhattacharya (1988), Postlewaite and Vives (1987), and Wallace (1988, 1990).

assets, thus giving the depositor the benefit of the higher returns from the long-term assets.

Bank behavior can be represented as the solution of an optimal contracting problem. Banks compete for customers by offering combinations of a portfolio and a deposit contract. Free entry into the banking sector guarantees that banks will earn zero profit in equilibrium and will offer the combination of portfolio and contract that maximizes the depositor's expected utility.

Risk can taken the form of shocks to asset returns or the demand for liquidity. These shocks provide a role for financial markets. Specifically, we introduce markets for securities that allow banks to insure against aggregate shocks. We also introduce markets on which banks can buy and sell long-term assets in order to obtain or provide liquidity.

The introduction of these two types of markets has important implications for the welfare properties of the model. First, the existence of markets on which assets can be liquidated ensures that bankruptcy involves no inefficiency ex post. Firesale prices transfer value to the buyer but do not constitute a deadweight loss. Secondly, ex ante risk sharing is optimal if there is a complete set of Arrow securities for insuring against aggregate shocks.

For a long time, policy makers have taken it as axiomatic that crises are best avoided. By contrast, in the present framework, with complete markets, a laisser faire financial system achieves the constrained-efficient allocation of risk and resources. When banks are restricted to using non-contingent deposit contracts, default introduces a degree of contingency that may be desirable from the point of view of optimal risk sharing. Far from being best avoided, financial crises may actually be *necessary* in order to achieve constrained efficiency. Avoiding default is costly. It either requires holding a very safe and liquid portfolio and earning lower returns, or it requires reducing the liquidity promised to the depositors. In any case, the bank optimally weighs the costs and benefits and chooses the efficient level of default in equilibrium.

The important point is that avoidance of crises should not be taken as axiomatic. If regulation is required to minimize or obviate the costs of financial crises, it needs to be justified by a microeconomic welfare analysis based on standard assumptions. Furthermore, the form of the intervention should be derived from microeconomic principles. After all, financial institutions and financial markets exist to facilitate the efficient allocation of risks and resources. A policy that aims to prevent financial crises has an impact on the normal functioning of the financial system. Any government intervention may impose deadweight costs by distorting the normal functioning of the financial system. One of the advantages of a microeconomic analysis of financial crises is that it clarifies the costs associated with these distortions.

The model described so far has no role for capital. Banks are like mutual companies, operated for the benefit of their depositors, with no investment provided by the entrepreneurs who set them up. We can add capital to the model by assuming the existence of a class of risk neutral investors who are willing invest in the bank in return for an equity share. These investors are assumed to have a fixed opportunity cost of capital, determined by the best investment returns available to them outside the banking sector. We assume this return is at least as great as the return on the long-term asset. These investors can also speculate on the short- and long-term assets, for example, holding the short-term asset in order to buy up the long-term asset at a firesale price in the event of a default. This kind of speculation provides liquidity. It is superfluous in the case of complete Arrow securities, but essential in equilibrium with incomplete markets.

## 2.1 Complete markets

As a benchmark, consider first the case in which there is a complete set of Arrow securities. An Arrow security is a contingent claim that pays one unit if a specified event occurs and nothing otherwise. In practice, we observe a rich array of securities traded in financial markets. These securities allow for the hedging of many different risks and the risk sharing possibilities are further expanded by the use of derivatives and synthetic derivatives generated by dynamic trading strategies. If the set of available securities, derivatives and synthetics is rich enough, it may be possible to span all possible risk sharing opportunities by constructing an appropriate portfolio. This is equivalent to having a complete set of Arrow securities. In the present context, we might think of the Arrow securities as performing the function of a complex array of credit derivatives. The first insight we gain from this model is the following:

- If there exist complete markets for insuring aggregate risks, such as shocks to aggregate liquidity demands or asset returns, then there is no value (to the bank) in having capital at all.

The only function of capital is to improve cross-sectional risk sharing between investors (shareholders) and depositors in the bank, but this can be achieved just as well (even better, as we shall see) using Arrow securities. Thus, in spite of the incompleteness of risk sharing and the welfare loss caused by the use of non-contingent debt contracts, capital provides no benefits in terms of economic efficiency or improved risk sharing.

A corollary is a "Modigliani-Miller" type result:

- Capital structure is irrelevant because Arrow securities can be used to undo any changes in the debt to equity ratio. Any capital ratio is optimal (for the bank), including zero.

Unlike the classical Modigliani-Miller theorem, this result derives from the bank's ability to undo the effects of a change in capital structure, rather than from the investors' ability to undo it.

It might be thought that, even if the individual bank finds it optimal to have no capital or is indifferent between different capital requirements, it makes a difference to the welfare of the economic system to require a particular level of capital. Even this is not true, however. An equilibrium with complete Arrow security markets is constrained efficient:

- As long as banks are restricted to using demand deposits (rather than contingent incentive-compatible contracts), no intervention by the banking authorities can improve economic welfare.

## 2.2 Incomplete markets

In order to provide an opportunity for welfare-improving intervention, some kind of friction or market failure must be introduced. Here I assume that markets for liquidity services are not complete. Specifically, there are no markets for Arrow securities.

In this case, there is a role for capital in promoting improved risk sharing. If banks use non-contingent liabilities to finance investment in risky assets, there is a risk of bankruptcy in bad states where asset returns are low. Even if bankruptcy involves no deadweight costs ex post, depositors end up bearing risk and the allocation of this risk may be suboptimal in the absence of Arrow securities. By using capital to finance investment, the bank increases the total value of its portfolio in each state. The depositors (debt holders) receive all the value in bad states, where the bank is bankrupt, and the shareholders receive the excess returns (total value minus debt) in good states. From the point of view of depositors, there has been a shift in returns from the good states to the bad states, equalizing consumption across states and improving risk sharing.

How far can this process go? It depends on the cost of capital, that is, the difference between the return on external investments and the bank's portfolio. As long as the returns on assets held by the bank are equal to the best returns available to the investors elsewhere in the economy, an increase in capital does not impose any cost on the bank. The increased investment provides a return high enough to cover the opportunity cost of the capital. However, as the bank's portfolio increases in value, it becomes possible to pay the depositors a fixed return equal to the average return on bank assets without risk of default. This is just what an efficient allocation of risk requires: there is no risk of default and the depositors' consumption is equalized in all states. In this case, capital provides the same services as complete markets.

- If the opportunity cost of capital is the same as the return on bank assets, the optimal capital structure provides the same allocation of risk and resources as the complete Arrow securities.

However, as every CEO knows, capital is expensive. If the opportunity cost of funds is greater than the return on bank assets, increasing capital will impose a real cost on the bank depositors.[3] They will have to give up part of the return on their investments in order to compensate the shareholders for the lower average return of bank assets. This tradeoff between cost of capital and improved risk sharing will limit the extent to which it is optimal to share risk between shareholders and depositors.

[3] The assumption that $\rho$, the return on outside investments, is greater than $\bar{R}$, the return on bank investments, suggests the existence of an arbitrage opportunity. For example, why don't risk neutral investors borrow from the bank, invest the funds to get the return $\rho$ and pay a return higher than $\bar{R}$ to the banks? One reason may be risk: the assets that yield a return $\rho$ may be much riskier than the bank's assets. Although the investors are risk neutral, their consumption is non-negative, so they cannot insure the banks against risk if all their wealth is invested in the risky asset. Another reason might be liquidity: the assets that yield a return $\rho$ may not yield returns for a long time, perhaps long after the bank needs the returns to pay its depositors. Of course, it might be optimal for the bank to invest a small amount in these higher-yielding assets. The failure to do so could be explained by regulatory constraints or asymmetric information.

- If the cost of capital is high, the optimal level of capital may be positive, but it will not guarantee complete (first-best) risk sharing. For a sufficiently high cost of capital, the optimal (for the bank) level of capital is zero.

It is worth stopping to ask why first-best risk sharing cannot be achieved between risk neutral equity holders and risk-averse debt holders. Technically, the reason is limited liability. If the shareholder's liability is limited to their investment, so is the depositors' insurance in the worst states. Because the cost of capital is positive, the liability constraint may be binding and risk sharing will be less than complete. If the equity holders and debt holders could write a complete contingent contract, they would replicate the effect of complete Arrow securities, but this would require payments from the equity holders to the debt holders in the worst states. In fact, with complete contingent contracts, the optimal level capital would be independent of the insurance provided and would be optimally set to zero if the cost of capital is positive.

This suggests an interesting way that risk sharing could be improved: multiple liability has been discussed recently by Macey and O'Hara (2000). Double or multiple liability was common in the United States until the introduction of deposit insurance in the 1930s. Absent collection and liquidity costs, multiple liability provides a way of increasing effective capital without increasing capital costs. For example, with double liability, the debt holders receive the same insurance indemnity in bad states with half the capital cost. In the limit, no capital is required, only liability.

- If shareholder liability is a multiple $m$ of their investment, first-best risk sharing is achieved in the limit as $m$ diverges to infinity.

This is like the situation in Lloyd's of London, where names invest their capital in various ways to get the highest rate of return and simultaneously use it as collateral to underwrite insurance contracts. There are two drawbacks to this solution, illiquidity and collection costs.

If the shareholders' liability is limited to the capital invested in the bank's portfolio, the receiver can easily dispose of those assets, assuming no fraud on the part of the bank's management. If the shareholders' liability extends to assets they own, it may be very costly for the receiver to pursue the shareholders and enforce their liability. Again, Lloyd's of London provides a useful illustration. These costs limit the effectiveness of multiple liability as a source of inexpensive insurance.

Illiquidity is another problem. If the best returns achieved by shareholders outside the bank are generated by illiquid investments, the shareholders may have difficulty meeting their liability to the bank when the bank defaults. Another way of putting the same point is that part of the cost of capital is the need to maintain a certain portion of the bank's portfolio in liquid investments.

What can government intervention accomplish? Because markets are incomplete, the equilibrium allocation may not be Pareto-efficient. However, the bank chooses its portfolio and deposit contract to maximize the welfare of the depositors, taking as given the prices in the market. So there is a market failure only if banks are facing the "wrong" prices. There are two possible ways in which prudential regulation can improve economic welfare. First, it could execute intertemporal trades that banks, investors, and depositors cannot achieve, effectively replacing missing markets. Secondly, it could alter the allocation of resources in a way that changes prices and causes economic decision makers to change their own intertemporal decisions.

The first kind of intervention is not as interesting as the second. If regulatory authorities can replace missing markets, there is an obvious welfare gain; but it is not obvious what technological advantage the authorities have over the market when it comes to executing intertemporal trades. For example, if there are missing markets because transaction costs are high, the regulatory authorities will be subject to the same transaction costs. It may be unrealistic to assume that they have a technological advantage in this activity. One advantage that government's do have over the market is their ability to tax and it may be that the government's ability to tax allows it to provide some forms of insurance that the market cannot or will not provide. Deposit insurance may be a case in point.[4] At the same time, it is well known that deposit insurance creates an offsetting moral hazard problem. In any case, even if the government has a superior technology available, it can be argued that markets efficiency should be defined relative to the available technology.

The second possibility is more interesting. We say that an equilibrium

[4]The case for deposit insurance is complicated by asymmetric information. On the one hand, private insurance markets have well known failings because of adverse selection, so there may be a need for intervention. On the other hand, deposit insurance is widely believed to create a moral hazard problem (risk shifting). In the present model, these issues of asymmetric information are ignored.

is *constrained efficient* if it is impossible to make every agent better off (or some better off and no one worse off), by changing the allocation of goods and services at the first date, while relying on the existing (incomplete) markets at the second and subsequent dates. Constrained inefficiency does imply that markets have failed to produce the most efficient allocation possible relative to that technology. The pecuniary externality created by intervention does not require a superior transaction technology, just a manipulation of agents' incentives by changing prices. This is the idea that lies behind a famous result of Geanakoplos and Polemarchakis (1986), who show that in a model of perfectly competitive general equilibrium with incomplete markets, the equilibrium allocation is generically constrained inefficient. A welfare-improving intervention does not require the regulator to make intertemporal trades that are impossible for the market. The regulator only needs to affect the allocation of resources at a single point in time and leave it to the market to respond intertemporally to the changed incentives.

In the present model, there is no welfare-relevant pecuniary externality. Asset prices at the second date are uniquely determined by the opportunity cost of capital and the investors' first-order conditions for an optimal portfolio. An increase in the required capital adequacy ratio or reserve ratio may change the bank's portfolio, but it will not change asset prices. The future asset prices are determined by the risk neutral investors' willingness to hold the short asset: in equilibrium, the investors must earn their opportunity cost and in this example that requirement uniquely determines the capital gain they make in bad states. Since the bank is already maximizing the expected utility of the depositors taking prices as given, there is no feasible welfare improvement.

- Equilibrium is constrained efficient. There is no role for using capital adequacy requirements to improve economic welfare.

This result holds for this particular (non-generic) example, but it makes the point that without a welfare-relevant pecuniary externality, intervention cannot be justified. If investors have different risk preferences (for example, allowing for positive risk aversion) and if there are more states of nature, the general-equilibrium effects of any intervention become more complicated. The Geanakoplos-Polemarchakis theorem suggests that there will typically be *some* intervention that can make everyone better off, but it does not identify the nature of the intervention and in general it is very hard to say what the intervention will look like. Even in simple examples, the general equilibrium effects of a regulatory intervention can contradict our intuition about the policy's likely impact (cf. Allen and Gale, 2000d). Without a theory of optimal policy, intervention is a shot in the dark.

# 3 Asymmetric information and capital structure

The risk-sharing example makes the point that, under certain circumstances, capital adequacy requirements are, at best, unnecessary and, at worst, harmful. Banks, left to themselves, will choose the optimal capital structure. If regulation forces them to increase capital ratios, the result will be a reduction of economic welfare. Of course, this example ignores a number of other ways in which capital structure may influence bank behavior and economic welfare, particularly those associated with asymmetric information (moral hazard, adverse selection).

In the presence of moral hazard, debt finance is associated with risk shifting. Banks are financed by debt-like liabilities (deposits) and this may produce an incentive to take excessive risk. Capital, like collateral, counteracts this tendency, because it increases the shareholders' sensitivity to downside risk. Macey and O'Hara (2000) remark on the temptation for shareholders to expropriate debtholders (depositors) in their proposal for multiple liability.

Of course, this argument assumes that the bank is being run for the benefit of the shareholders. The recent literature on corporate governance has emphasized the agency problem that exists between managers and shareholders. Managers may derive private benefits from decisions that are in conflict with the interests of shareholders. Getting managers to to maximize the shareholders' objectives is a non-trivial task. The mere existence of shareholders with capital at risk may not prevent risk shifting behavior.

Shareholders with capital at risk will have an incentive to monitor the managers, and this may have some effect on their risk taking behavior, but depositors have the same incentive in the absence of deposit insurance, as Calomiris and Kahn (1991) point out. Monitoring suffers from a free-rider problem whenever the numbers of shareholders or depositors is large. For this reason, monitoring by regulatory authorities may provide better discipline.

In any case, the assumption that adequate capital is necessary to prevent excessive risk taking does not by itself provide an argument for capital adequacy requirements. The bank can internalize this agency cost and adopt the optimal capital structure without any assistance from the regulator. In the absence of a pecuniary externality, there is no reason to think that the privately optimal capital structure is not socially optimal.

The same argument can be made in connection with other determinants of optimal capital structure. If there are deadweight losses from bankruptcy, for example because illiquid markets imply that an orderly liquidation is difficult to achieve or because there is loss of charter value or assets cannot be managed as efficiently by other banks, these costs should be internalized in the bank's choice of the optimal capital structure. Only if there is a pecuniary externality and markets are incomplete will there be an argument for regulation.

It is not clear that any of these considerations are actually the ones that motivate the regulators who set capital adequacy requirements. But whatever the motivation, the onus seems to be on the regulator to identify the pecuniary externality so that we can assess the importance of the market failure and the effectiveness of capital adequacy requirements as a solution. Financial fragility, the idea that one bank failure may trigger others and bring down the whole financial system, would be an example of a pecuniary externality on a very large scale. Perhaps this is what motivates the system of capital adequacy requirements. If so, we need better models of financial fragility before we can provide a theoretical basis for the current system.

# 4 Monitoring and survival strategies in a volatile environment

Another function of capital is to make continuous monitoring unnecessary. Imagine a world in which the following assumptions are satisfied:

- Monitoring is continuous;
- Portfolio is marked to market;
- Portfolio value changes continuously;
- Markets are perfectly liquid.

In this world, banks would never make losses without the forbearance of the regulator. When the net worth of the bank reaches zero, the assets and liabilities will be liquidated and the bank closed without loss to depositors. There is no need for capital to act as a buffer for creditors (depositors) since the creditors are always paid in full. Similarly, there is no incentive for risk shifting. The problem of moral hazard is resolved by continuous monitoring. Now, of course, these are strong assumptions. If monitoring is not continuous, if asset returns are not continuous, or if asset markets are illiquid, there may be deadweight losses associated with bankruptcy. But, again, that is not necessarily a market failure that can be rectified by setting high capital adequacy requirements. The bank will internalize these costs and choose the optimal capital structure to maximize shareholder value. Again, in the absence of a pecuniary externality, the private optimum will be the social optimum.

# 5 Concluding remarks

I began by noting the lack of theory in the practice of financial regulation. These notes have raised questions rather than answered them, so this lacuna remains to be filled. I close with the economist's familiar lament that further research is needed. Our understanding of financial institutions and their interaction with financial markets is growing and we are near the point where we will have the tools to address questions of policy in a more systematic and theoretically sophisticated way. The questions before us are fascinating in their own right and deserve to be addressed by the brightest minds in the field.

# 6 Appendix

To illustrate some of the points made in Section 2, I use a variation of the model found in Allen and Gale (2001). The main difference is that bank capital can be provided by risk neutral investors.

*Dates.* There are three dates $t = 0, 1, 2$ and a single good at each date. The good is used for consumption or investment.
*Assets.* There are two assets, a short-term asset (the *short asset*) and a long-term asset (the *long asset*).

- The short asset is represented by a storage technology: one unit of the good invested at date $t$ yields one unit at date $t+1$, for $t = 0, 1$.
- The long asset takes two periods to mature and is more productive than the short asset: one unit invested at date 0 produces a random return $\tilde{R}$ at date 2. The long asset is more productive than the short asset: $E[\tilde{R}] > 1$.

*Consumers.* There is a continuum of ex ante identical consumers, whose measure is normalized to unity. Each consumer has an endowment consisting of one unit of the good at date 0 and nothing at subsequent dates. Ex post, there are two types of consumers, *early consumers*, who consume at date 1, and *late consumers*, who consume at date 2. The probability of being an early consumer is denoted by $0 < \lambda < 1$ and consumption at date $t = 1, 2$ is denoted by $c_t$. The consumer's ex ante utility is

$$\lambda U(c_1) + (1 - \lambda) U(c_2).$$

We adopt the usual "law of large numbers" convention and assume that the fraction of early consumers is identically equal to the probability $\lambda$. The period utility function $U : \mathbf{R}_+ \to \mathbf{R}$ is twice continuously differentiable and satisfies the usual neoclassical properties, $U'(c) > 0, U''(c) < 0$, and $\lim_{c \searrow 0} U'(c) = \infty$.

*Investors.* There is a continuum of risk neutral investors who have a large endowment at date 0 and maximize expected consumption at date 2. They can invest directly in the short and long asset and they can also hold equity in financial institutions. The investors are assumed to have access to investment opportunities that are not available to the banks. The expected return on these investments (measured in terms of consumption at date 2) is $\rho \geq E[\tilde{R}]$.

*Uncertainty.* There are two aggregate states of nature $H$ and $L$. The return to the long asset $\tilde{R}$ is a function of the state of nature:

$$\tilde{R} = \begin{cases} R_H & \text{w.pr. } 1 - \varepsilon \\ R_L & \text{w.pr. } \varepsilon \end{cases}$$

where $0 < R_L < R_H$.

*Information.* All uncertainty is resolved at date 1. The true state $H$ or $L$ is revealed and each consumer learns his ex post type, i.e., whether he is an early consumer or a late consumer. Note that knowledge of the true return $\tilde{R}$ is available one period before the return itself is available.

*Banking.* A bank is a cooperative enterprise that provides insurance to consumers. At date 0 consumers deposit their initial endowments in a bank, which offers them a deposit contract promising $d_t$ units of consumption if they withdraw at date $t = 1, 2$. The bank holds a portfolio $(x, y)$ consisting of $x$ units of the long asset and $y$ units of the short asset. The bank can also obtain $e$ units of capital from risk-neutral investors, in exchange for a claim on the bank's profits.

## 6.1 Equilibrium with Arrow securities

Suppose that there are Arrow securities for the two states $H$ and $L$ at date 0 and a capital market at date 1. Then markets are complete with respect to aggregate uncertainty. We assume that banks and investors can participate in these markets, but that consumers cannot. In this case, the risk neutral speculators will absorb all the risk and the banks can pay depositors the expected return $\bar{R} = E[\tilde{R}]$ on the long asset. The bank's decision problem becomes

$$\begin{array}{ll} \max & \lambda U(d_1) + (1-\lambda)U(d_2) \\ \text{s.t.} & d_1 \leq d_2 \\ & \lambda d_1 = y \\ & (1-\lambda)d_2 = \bar{R}(1-y). \end{array}$$

The first constraint is the incentive constraint: it ensures that late consumers are willing to wait until date 2 to withdraw instead of imitating the early consumers, withdrawing at date 1, and saving the withdrawal until date 2. The second constraint is the budget constraint at date 1: the total consumption provided to depositors is $\lambda d_1$ and the efficient way to provide this consumption is by holding an equal amount of the short asset. The third constraint is the budget constraint at date 2: the total consumption at date 2 is $(1-\lambda)d_2$ and the most efficient way to provide this consumption is by holding $(1-y)$ units of the long asset, which produce on average $\bar{R}(1-y)$ units of consumption at date 2. Substituting from the budget constraints into the objective function and ignoring the incentive constraint, the decision problem is equivalent to the maximization of

$$\lambda U\left(\frac{y}{\lambda}\right)+(1-\lambda)U\left(\frac{\bar{R}(1-y)}{(1-\lambda)}\right).$$

The first-order condition for a maximum is

$$U'\left(\frac{y}{\lambda}\right)=U'\left(\frac{\bar{R}(1-y)}{(1-\lambda)}\right)\bar{R}.$$

Note that this implies $d_1 < d_2$ so the incentive constraint is satisfied.

The presence of Arrow securities means that capital is not needed. Any allocation of consumption that can be achieved by a bank with positive capital can also be achieved by a bank with zero capital and vice versa.

## 6.2 Equilibrium without Arrow securities

In the absence of Arrow securities, there is a role for capital. Equity holders share risk with the depositors, taking a higher return in good states and lower (or no) return in bad states. Limited liability prevents achievement of the first best: equity holders cannot be asked to make a positive payment in bad states. Additional liquidity can be provided by capital markets. Speculators hold the short asset in order to purchase the long asset at date 1. This liquidity provision is inefficient, because the speculators make a profit by paying low prices in the bad state (cf. Allen and Gale (1998)).

Here I sketch the structure of an equilibrium to illustrate the role of capital. If there is no default in equilibrium at date 1, the representative bank offers a deposit contract $(d_1, d_2)$, early consumers at date 1 receive the promised payment $d_1$ and the late consumers are the residual claimants at date 2. The bank must pay the late consumers $d_2$ if possible, and the liquidated value of the portfolio otherwise. Without loss of generality we can put $c_L \leq c_H = d_2$. In one case, there is no default at date 2 and the late consumers receive $d_2$ in both states. In the other case there is default in state $L$ (only) and consumers receive $d_2$ in state $H$ and the liquidated value of the portfolio in state $L$. In the first case, the bank's decision problem (DP) can be written as

$$
\begin{array}{ll}
\max & \lambda U(d_1) + (1-\lambda)U(d_2) \\
\text{s.t.} & x + y \leq 1 + e \\
& \lambda d_1 + (1-\lambda)p_H d_2 + p_H \pi_H \leq y + p_H R_H x \\
& \lambda d_1 + (1-\lambda)p_L d_2 + p_L \pi_L \leq y + p_L R_L x \\
& d_1 \leq d_2 \\
& (1-\varepsilon)\pi_H + \varepsilon \pi_L \geq \rho e,
\end{array}
$$

where $\pi_s$ is profits in state $s$. Note that profits are assumed to be paid at date 2. The first constraint is the budget constraint at date 0: the investment in assets is bounded by the depositors' endowment and the capital provided by investors. The second and third constraints are the date-1 budget constraints corresponding to states $H$ and $L$ respectively: the left hand side is the present value of depositors' consumption and profits and the right hand side is the value of the bank's portfolio. The fourth constraint is the incentive constraint: late consumers have no incentive to imitate early consumers. The final constraint ensures that investors earn the rate $\rho$ on the capital invested in the bank.

In this case, the demand for consumption at date 1 is the same in both states, as is the supply. Excess supply implies that $p_s = 1$, for $s = H, L$, which is inconsistent with equilibrium (the short asset is dominated). Thus, demand must equal supply and there is a single price $p_H = p_L = p$ that clears the asset market at date 2. Investors will only hold the short asset if $p\rho = 1$, but in that case $p\bar{R} < 1$, so no one will be willing to hold the long asset. This cannot be an equilibrium. So there is no provision of liquidity by the investors. Since capital is costly, we want to minimize the amount holding constant the consumption of the depositors. Thus, $\pi_L = 0$. Further, since there is no liquidity provision by investors, we can assume without loss of generality that the bank holds enough of the short asset to pay the consumers at date 1 and enough of the long asset to provide consumption for the late consumers at date 2. The bank's DP reduces to

$$
\begin{array}{ll}
\max & \lambda U(d_1) + (1-\lambda)U(d_2) \\
\text{s.t.} & x + y \leq 1 + e \\
& \lambda d_1 \leq y \\
& (1-\lambda) d_2 \leq R_L x \\
& d_1 \leq d_2 \\
& (1-\varepsilon)(R_H - R_L)x \geq \rho e.
\end{array}
$$

This leads to first-best risk sharing, but the depositors' expected utility is reduced relative to the equilibrium with Arrow securities because of the cost of capital. To see this, consider the first order conditions for this problem (as usual ignoring the incentive constraint):

$$\begin{aligned}
&U'(d_1) = \mu_2 \\
&U'(d_2) = \mu_3 \\
&\mu_1 = \mu_2 \\
&\mu_1 = \mu_3 R_L + \mu_4(1-\varepsilon)(R_H - R_L) \\
&\mu_1 = \rho\mu_4
\end{aligned}$$

or

$$U'(d_1) = U'(d_2)R_L + \frac{U'(d_1)}{\rho}(1-\varepsilon)(R_H - R_L),$$

which implies

$$\rho = \rho\frac{U'(d_2)}{U'(d_1)}R_L + (1-\varepsilon)(R_H - R_L).$$

Now if $\rho = \bar{R}$ then $\rho U'(d_2) = U(d_1)$. If $\rho > \bar{R}$ then $\rho U'(d_2) > U(d_1)$. As $\rho$ increases, $d_1$ and $d_2$ draw closer together until at last the incentive constraint is binding. A further increase in $\rho$ will make default in state $L$ an optimal response.

Now consider the case where there is default in state $L$ at date 2. The representative bank will sell the long asset in exchange for liquidity in the bad state, so investors must be willing to hold the short asset. The demand for liquidity in the good state is lower than the demand in the bad state, so the prices of future consumption at date 1 are $p_H = 1$ and $p_L = p$. In order to induce investors to hold the short asset, we must have

$$\rho = (1-\varepsilon) + \varepsilon\frac{1}{p}$$

The decision problem of the representative bank can be written as follows:

$$\begin{aligned}
\max \quad & \lambda U(d_1) + (1-\lambda)\left\{(1-\varepsilon)U(d_2) + \varepsilon U(c_L)\right\} \\
\text{s.t.} \quad & x + y \leq 1 + e \\
& \lambda d_1 + (1-\lambda)d_2 + \pi_H \leq y + R_H x \\
& \lambda d_1 + (1-\lambda)pc_L \leq y + pR_L x \\
& d_1 \leq c_L \\
& (1-\varepsilon)\pi_H \geq \rho e.
\end{aligned}$$

If $\rho$ gets even higher, it may be optimal to consider default at date 1 in state $L$, but I won't attempt to analyze that case here.

### 6.3 Constrained efficiency

Rather than conduct a full analysis of the constrained efficiency of the model, in this section I consider the welfare impact of imposing a capital requirement $\bar{e}$ above the equilibrium level. Each case considered in Section 6.2 requires a different argument.

If there is no default at date 2, the banks do not use the asset market at date 1. At the market-clearing price, the investors do not want to hold the short asset, so there will be no trade at date 1 in the new equilibrium. Forcing the banks to hold more capital by imposing a constraint $e \geq \bar{e}$ reduces expected utility.

If there is default in the bad state at date 2, the investors provide liquidity by holding speculative balances of the short asset. It is optimal for the investors to hold the short asset only if $p_H = 1$, $p_L = p$ and

$$\rho = (1 - \varepsilon) + \varepsilon \frac{1}{p}.$$

This condition uniquely determines the price $p$. Thus, forcing the banks to hold more capital by imposing a constraint $e \geq \bar{e}$ does not change the prices at which banks can sell the long asset. At these prices, the banks can sell as much as they wish, just as they could in the original equilibrium. Thus, the maximum expected utility they can achieve is the solution to the DP given in Section 6.2 with the added constraint $e \geq \bar{e}$. Obviously, adding a constraint to the problem will not increase expected utility.

A similar argument applies if there is default at date 1.

## References


Allen, F. and D. Gale (1998). "Optimal Financial Crises," *Journal of Finance* **53,** 1245-1284.

Allen, F. and D. Gale (2000a). "Financial Contagion," *Journal of Political Economy* **108,** 1-33.

Allen, F. and D. Gale (2000b). "Optimal Currency Crises" *Journal of Monetary Economics*, Carnegie-Rochester Series on Public Policy, to appear.

Allen, F. and D. Gale (2000c). "Bubbles and Crises," *The Economic Journal* **110,** 236-256

Allen, F. and D. Gale (2000d). "Banking and Markets," New York University, unpublished.

Allen, F. and D. Gale (2000e). *Comparing Financial Systems.* Cambridge, MA: MIT Press.

Calomiris, C. and C. Kahn (1991). "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *The American Economic Review* **81**, 497-513.

Chari, V. and R. Jagannathan (1988). "Banking Panics, Information, and Rational Expectations Equilibrium," *Journal of Finance* **43**, 749-60.

Diamond, D. and P. Dybvig (1983). "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* **91**, 401-419.Hellwig, M. (1994). "Liquidity Provision, Banking, and the Allocation of Interest Rate Risk," *European Economic Review* **38**, 1363-1389.

Geanakoplos, J. and H. Polemarchakis (1986). "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete," in Walter, Heller, Ross Starr, and David Starrett eds. *Essays in honor of Kenneth J. Arrow: Volume 3, Uncertainty, information, and communication.* Cambridge, New York and Sydney: Cambridge University Press, 65-95.

Jacklin, C., and S. Bhattacharya (1988). "Distinguishing panics and information-based bank runs: Welfare and policy implications," *Journal of Political Economy* **96**, 568-592.

Macey, J. and M. O'Hara (2000). "Solving the Corporate Governance Problems of Banks: A Proposal," Cornell University, unpublished.

Postlewaite, A. and X. Vives (1987). "Bank Runs as an Equilibrium Phenomenon," *Journal of Political Economy* **95**, 485-491.

Rochet, Jean-Charles and Jean Tirole, (1996). Interbank Lending and Systemic Risk, *Journal of Money, Credit, and Banking* 28, 733-762.

Wallace, N. (1988). "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously," *Federal Reserve Bank of Minneapolis Quarterly Review* **12** (Fall), 3-16.

Wallace, Neil, 1990, A Banking Model in which Partial Suspension is Best, *Federal Reserve Bank of Minneapolis Quarterly Review* 14(Fall), 11-23.