

Optimal Financial Crises

FRANKLIN ALLEN and DOUGLAS GALE*

ABSTRACT

Empirical evidence suggests that banking panics are related to the business cycle and are not simply the result of "sunspots." Panics occur when depositors perceive that the returns on bank assets are going to be unusually low. We develop a simple model of this. In this setting, bank runs can be first-best efficient: they allow efficient risk sharing between early and late withdrawing depositors and they allow banks to hold efficient portfolios. However, if costly runs or markets for risky assets are introduced, central bank intervention of the right kind can lead to a Pareto improvement in welfare.

FROM THE EARLIEST TIMES, banks have been plagued by the problem of *bank runs* in which many or all of the bank's depositors attempt to withdraw their funds simultaneously. Because banks issue liquid liabilities in the form of deposit contracts, but invest in illiquid assets in the form of loans, they are vulnerable to runs that can lead to closure and liquidation. A financial crisis or *banking panic* occurs when depositors at many or all of the banks in a region or a country attempt to withdraw their funds simultaneously.

Prior to the twentieth century, banking panics occurred frequently in Europe and the United States. Panics were generally regarded as a bad thing and the development of central banks to eliminate panics and ensure financial stability has been an important feature of the history of financial systems. It has been a long and involved process. The first central bank, the Bank of Sweden, was established more than 300 years ago. The Bank of England played an especially important role in the development of effective stabilization policies in the eighteenth and nineteenth centuries. By the end of the nineteenth century, banking panics had been eliminated in Europe. The last true panic in England was the Overend, Gurney & Company Crisis of 1866.

*Allen is from the Wharton School of the University of Pennsylvania and Gale is from the Department of Economics at New York University. The authors thank Charles Calomiris, Rafael Repullo, Neil Wallace, and participants at workshops and seminars at the Board of Governors of the Federal Reserve, Boston College, Carnegie Mellon, Columbia, Duke-University of North Carolina, European Institute of Business Administration, Federal Reserve Bank of Philadelphia, Instituto Tecnológico Autónomo de México, University of Chicago, University of Maryland, University of Michigan, University of Minnesota, Nanzan University, New York University, the State University of New York, and the 1998 American Finance Association meetings. Financial support from the National Science Foundation, the C.V. Starr Center at New York University, and the Wharton Financial Institutions Center is gratefully acknowledged.

Table I
National Banking Era Panics

The incidence of panics and their relationship to the business cycle are shown. The first column is the NBER business cycle with the first date representing the peak and the second date the trough. The second column indicates whether or not there is a panic and if so the date it occurs. The third column is the percentage change of the ratio of currency to deposits at the panic date compared to the previous year's average. The larger this number the greater the extent of the panic. The fourth column is the percentage change in pig iron production measured from peak to trough. This is a proxy for the change in economic activity. The greater the decline the more severe the recession. The table is adapted from Gorton (1988, Table 1, p. 233).

NBER Cycle Peak-Trough	Panic Date	Percentage Δ (Currency/Deposit)	Percentage Δ Pig Iron
Oct. 1873-Mar. 1879	Sep. 1873	14.53	-51.0
Mar. 1882-May 1885	Jun. 1884	8.80	-14.0
Mar. 1887-Apr. 1888	No panic	3.00	-9.0
Jul. 1890-May 1891	Nov. 1890	9.00	-34.0
Jan. 1893-Jun. 1894	May 1893	16.00	-29.0
Dec. 1895-Jun. 1897	Oct. 1896	14.30	-4.0
Jun. 1899-Dec. 1900	No panic	2.78	-6.7
Sep. 1902-Aug. 1904	No panic	-4.13	-8.7
May 1907-Jun. 1908	Oct. 1907	11.45	-46.5
Jan. 1910-Jan. 1912	No panic	-2.64	-21.7

The United States took a different tack. Alexander Hamilton had been impressed by the example of the Bank of England and this led to the setting up of the First Bank of the United States and subsequently the Second Bank of the United States. However, after Andrew Jackson vetoed the renewal of the Second Bank's charter, the United States ceased to have any form of central bank in 1836. It also had many crises. Table I (from Gorton (1988)) shows the banking crises that occurred repeatedly in the United States during the nineteenth and early twentieth centuries. During the crisis of 1907 a French banker commented that the United States was a "great financial nuisance." The comment reflects the fact that crises had essentially been eliminated in Europe and it seemed as though the United States was suffering gratuitous crises that could have been prevented by the establishment of a central bank.

The Federal Reserve System was eventually established in 1914. In the beginning it had a decentralized structure, which meant that even this development was not very effective in eliminating crises. In fact, major banking panics continued to occur until the reforms enacted after the crisis of 1933. At that point, the Federal Reserve was given broader powers and this together with the introduction of deposit insurance finally led to the elimination of periodic banking crises.

Although banking panics appear to be a thing of the past in Europe and the United States, many emerging countries have had severe banking problems in recent years. Lindgren, Garcia, and Saal (1996) find that 73 percent

of the IMF's member countries suffered some form of banking crisis between 1980 and 1996. In many of these crises, panics in the traditional sense were avoided either by central bank intervention or by explicit or implicit government guarantees. This raises the issue of whether such intervention is desirable.

Given the historical importance of panics and their current relevance in emerging countries, it is important to understand why they occur and what policies central banks should implement to deal with them. Although there is a large literature on bank runs, there is relatively little on the optimal policy that should be followed to prevent or "manage" runs (but see Bhattacharya and Gale (1987), Rochet and Tirole (1996), and Bensaïd, Pages, and Rochet (1996)). The history of regulation of the United States' and other countries' financial systems seems to be based on the premise that banking crises are bad and should be eliminated. We argue below that there are costs and benefits to having bank runs. Eliminating runs completely is an extreme policy that imposes costly constraints on the banking system. Likewise, *laissez-faire* can be shown to be optimal, but only under equally extreme conditions. In this paper, we try to sort out the costs and benefits of runs and identify the elements of an optimal policy.

Before addressing the normative question of what is the optimal policy toward crises, we have to address the positive question of how to model crises. There are two traditional views of banking panics. One is that they are *random events*, unrelated to changes in the real economy. The classical form of this view suggests that panics are the result of "mob psychology" or "mass hysteria" (see, e.g., Kindleberger (1978)). The modern version, developed by Diamond and Dybvig (1983) and others, is that bank runs are self-fulfilling prophecies. Given the assumption of first-come, first-served, and costly liquidation of some assets, there are multiple equilibria. If everyone believes that a banking panic is about to occur, it is optimal for each individual to try to withdraw his funds. Since each bank has insufficient liquid assets to meet all of its commitments, it will have to liquidate some of its assets at a loss. Given first-come, first-served, those depositors who withdraw initially will receive more than those who wait. On one hand, anticipating this, all depositors have an incentive to withdraw immediately. On the other hand, if no one believes a banking panic is about to occur, only those with immediate needs for liquidity will withdraw their funds. Assuming that banks have sufficient liquid assets to meet these legitimate demands, there will be no panic. Which of these two equilibria occurs depends on extraneous variables or "sunspots." Although "sunspots" have no effect on the real data of the economy, they affect depositors' beliefs in a way that turns out to be self-fulfilling. (Postlewaite and Vives (1987) have shown how runs can be generated in a model with a unique equilibrium.)

An alternative to the "sunspot" view is that banking panics are a natural outgrowth of the *business cycle*. An economic downturn will reduce the value of bank assets, raising the possibility that banks are unable to meet their commitments. If depositors receive information about an impending down-

turn in the cycle, they will anticipate financial difficulties in the banking sector and try to withdraw their funds. This attempt will precipitate the crisis. According to this interpretation, panics are not random events but a response to unfolding economic circumstances. Mitchell (1941), for example, writes

when prosperity merges into crisis . . . heavy failures are likely to occur, and no one can tell what enterprises will be crippled by them. The one certainty is that the banks holding the paper of bankrupt firms will suffer delay and perhaps a serious loss on collection. [p.74]

In other words, panics are an integral part of the business cycle.

A number of authors have developed models of banking panics caused by aggregate risk. Wallace (1988, 1990), Chari (1989), and Champ, Smith, and Williamson (1996) extend Diamond and Dybvig (1983) by assuming the fraction of the population requiring liquidity is random. Chari and Jagannathan (1988), Jacklin and Bhattacharya (1988), Hellwig (1994), and Alonso (1996) introduce aggregate uncertainty, which can be interpreted as business cycle risk. Chari and Jagannathan focus on a signal extraction problem where part of the population observes a signal about future returns. Others must then try to deduce from observed withdrawals whether an unfavorable signal was received by this group or whether liquidity needs happen to be high. Chari and Jagannathan are able to show panics occur not only when the outlook is poor but also when liquidity needs turn out to be high. Jacklin and Bhattacharya also consider a model where some depositors receive an interim signal about risk. They show that the optimality of bank deposits compared to equities depends on the characteristics of the risky investment. Hellwig considers a model where the reinvestment rate is random and shows that the risk should be borne by both early and late withdrawers. Alonso demonstrates using numerical examples that contracts where runs occur may be better than contracts that ensure runs do not occur because the former improve risk sharing.

Gorton (1988) conducts an empirical study to differentiate between the "sunspot" view and the business-cycle view of banking panics. He finds evidence consistent with the view that banking panics are related to the business cycle and which is difficult to reconcile with the notion of panics as "random" events. Table I shows the recessions and panics that occurred in the United States during the National Banking Era. It also shows the corresponding percentage changes in the currency/deposit ratio and the change in aggregate consumption, as proxied by the change in pig iron production during these periods. The five worst recessions, as measured by the change in pig iron production, were accompanied by panics. In all, panics occurred in seven of the eleven cycles. Using the liabilities of failed businesses as a leading economic indicator, Gorton finds that panics were systematic events: whenever this leading economic indicator reached a certain threshold, a panic ensued. The stylized facts uncovered by Gorton thus suggest that banking

panics are intimately related to the state of the business cycle rather than some extraneous random variable. Calomiris and Gorton (1991) consider a broad range of evidence and conclude that the data do not support the “sun-spot” view that banking panics are random events.

In this paper, we develop a model that is consistent with the business cycle view of the origins of banking panics. Our main objective is to analyze the welfare properties of this model and understand the role of central banks in dealing with panics. In this model, bank runs are an inevitable consequence of the standard deposit contract in a world with aggregate uncertainty about asset returns. Furthermore, they play a useful role insofar as they allow the banking system to share these risks among depositors. In certain circumstances, a banking system under *laissez-faire* which is vulnerable to crises can actually achieve the first-best allocation of risk and investment. In other circumstances, where crises are costly, we show that appropriate central bank intervention can avoid the unnecessary costs of bank runs while continuing to allow runs to fulfill their risk-sharing function. Finally, we consider the role of markets for the illiquid asset in providing liquidity for the banking system. The introduction of asset markets leads to a Pareto reduction in welfare in the *laissez-faire* case. Once again, though, central bank intervention allows the financial system to share risks without incurring the costs of inefficient investment. This analysis is related to Diamond (1997) but he focuses on banks and financial markets as alternatives for providing liquidity to depositors and does not focus on the role of the central bank.

Our assumptions about technology and preferences are the ones that have become standard in the literature since the appearance of the Diamond and Dybvig (1983) model. Banks have a comparative advantage in investing in an illiquid, long-term, risky asset. At the first date, individuals deposit their funds in the bank to take advantage of this expertise. The time at which they wish to withdraw is determined by their consumption needs. Early consumers withdraw at the second date and late consumers withdraw at the third date. Banks and investors also have access to a liquid, risk-free, short-term asset represented by a storage technology. The banking sector is perfectly competitive, so banks offer risk-sharing contracts that maximize depositors' *ex ante* expected utility, subject to a zero-profit constraint.

There are two main differences with the Diamond–Dybvig model. The first is the assumption that the illiquid, long-term assets held by the banks are risky and perfectly correlated across banks. Uncertainty about asset returns is intended to capture the impact of the business cycle on the value of bank assets. Information about returns becomes available before the returns are realized, and when the information is bad it has the power to precipitate a crisis. The second is that we do not make the first-come, first-served assumption. This assumption has been the subject of some debate in the literature as it is not an optimal arrangement in the basic Diamond–Dybvig model (see Wallace (1988) and Calomiris and Kahn (1991)).

In a number of countries and historical time periods banks have had the right to delay payment for some time period on certain types of accounts. This is rather different from the first-come, first-served assumption. Sprague (1910) recounts how in the United States in the late nineteenth century people could obtain liquidity once a panic had started by using certified checks. These checks traded at a discount. We model this type of situation by assuming the available liquidity is split on an equal basis among those withdrawing early. In the context this arrangement is optimal. We also assume that those who do not withdraw early have to wait some time before they can obtain their funds and again what is available is split among them on an equal basis.

We begin our analysis with a simple case that serves as a benchmark for the rest of the paper. No costs of early withdrawal are assumed, apart from the potential distortions that bank runs may create for risk-sharing and portfolio choice. In this context, we identify the incentive-efficient allocation with an optimal mechanism design problem in which the optimal allocation can be made contingent on a leading economic indicator (i.e., the return on the risky asset), but not on the depositors' types. By contrast, a standard deposit contract cannot be made contingent on the leading indicator. However, depositors can observe the leading indicator and make their withdrawal decision conditional on it. When late-consuming depositors observe that returns will be high, they are content to leave their funds in the bank until the last date. When the returns are going to be low, they attempt to withdraw their funds, causing a bank run. The somewhat surprising result is that the optimal deposit contract produces the same portfolio and consumption allocation as the first-best allocation. The possibility of equilibrium bank runs allows banks to hold the first-best portfolio and produces just the right contingencies to provide first-best risk sharing.

Next we introduce a real cost of early withdrawal by assuming that the storage technology available to the banks is strictly more productive than the storage technology available to late consumers who withdraw their deposits in a bank run. A bank run, by forcing the early liquidation of too much of the safe asset, actually reduces the amount of consumption available to depositors. In this case, *laissez-faire* does not achieve the first-best allocation. This provides a rationale for central bank intervention. We show that the central bank can intervene with a monetary injection and this implements the first-best allocation. Suppose that a bank promises the depositor a fixed nominal amount and that, in the event of a run, the central bank makes an interest-free loan to the bank. The bank can meet its commitments by paying out cash, thus avoiding premature liquidation of the safe asset. Equilibrium adjustments of the price level at the two dates ensure that early and late consumers end up with the correct amount of consumption at each date and the bank ends up with the money it needs to repay its loan to the central bank. The first-best allocation is thus implemented by a combination of a standard deposit contract and bank runs.

One of the special features of the models described above is that the risky asset is completely illiquid. Since it is impossible to liquidate the risky asset, it is available to pay the late consumers who do not choose early withdrawal. We next analyze what happens if there is an asset market in which the risky asset can be traded. It is shown that this case is very different. Now the banks may be forced to liquidate their illiquid assets in order to meet their deposit liabilities. However, by selling assets during a run, they force down the price and make the crisis worse. Liquidation is self-defeating, in the sense that it transfers value to speculators in the market, and it involves a deadweight loss. By making transfers in the worst states, it provides depositors with negative insurance. In this case, there is an incentive for the central bank to intervene to prevent a collapse of asset prices, but again the problem is not runs per se but the unnecessary liquidations they promote.

This model illustrates the role of business cycles in generating banking crises and the costs and the benefits of such crises. However, since it assumes the existence of a representative bank, it cannot be used to study important phenomena such as *financial fragility* and *contagion* (Bernanke (1983), Bernanke and Gertler (1989)). This is a task for future research.

The rest of the paper is organized as follows. The model is described in Section I and a special case is presented that serves as a benchmark for the rest of the paper. In Section II we introduce liquidation costs and show how this provides a rationale for central bank intervention. In Section III we analyze what happens when the risky asset can be traded on an asset market. Concluding remarks are contained in Section IV.

I. Optimal Risk-Sharing and Bank Runs

In this section we describe a simple model to show how cyclical fluctuations in asset values can produce bank runs. The basic framework is the standard one from Diamond and Dybvig (1983), but in our model asset returns are random and information about future returns becomes available before the returns are realized. As a benchmark, we first consider the case in which bank runs cause no misallocation of assets because the assets are either totally illiquid or can be liquidated without cost. Under these assumptions, it can be shown that bank runs are optimal in the sense that the unique equilibrium of bank runs supports a first-best allocation of risk and investment.

Time is divided into three periods, $t = 0, 1, 2$. There are two types of assets, a safe asset and a risky asset, and a consumption good. The safe asset can be thought of as a storage technology, which transforms one unit of the consumption good at date t into one unit of the consumption good at date $t + 1$. The risky asset is represented by a stochastic production technology that transforms one unit of the consumption good at date $t = 0$ into R units of the consumption good at date $t = 2$, where R is a nonnegative random variable with a density function $f(R)$. At date 1 depositors observe a signal, which can be thought of as a leading economic indicator. This signal

predicts with perfect accuracy the value of R that will be realized at date 2. In subsection A it is assumed that consumption can be made contingent on the leading economic indicator, and hence on R . Subsequently, we consider what happens when banks are restricted to offering depositors a standard deposit contract—that is, a contract that is not explicitly contingent on the leading economic indicator.

There is a continuum of ex ante identical depositors (consumers) who have an endowment of the consumption good at the first date and none at the second and third dates. Consumers are uncertain about their time preferences. Some will be *early consumers*, who only want to consume at date 1, and some will be *late consumers*, who only want to consume at date 2. At date 0 consumers know the probability of being an early or late consumer, but they do not know which group they belong to. All uncertainty is resolved at date 1 when each consumer learns whether he is an early or late consumer and what the return on the risky asset is going to be. For simplicity, we assume that there are equal numbers of early and late consumers and that each consumer has an equal chance of belonging to each group. Then a typical consumer's utility function can be written as

$$U(c_1, c_2) = \begin{cases} u(c_1) & \text{with probability } 1/2, \\ u(c_2) & \text{with probability } 1/2, \end{cases} \quad (1)$$

where c_t denotes consumption at date $t = 1, 2$. The period utility functions $u(\cdot)$ are assumed to be twice continuously differentiable, increasing, and strictly concave. A consumer's type is not observable, so late consumers can always imitate early consumers. Therefore, contracts explicitly contingent on this characteristic are not feasible.

The role of banks is to make investments on behalf of consumers. We assume that only banks can distinguish the genuine risky assets from assets that have no value. Any consumer who tries to purchase the risky asset faces an extreme adverse selection problem, so in practice only banks will hold the risky asset. This gives the bank an advantage over consumers in two respects. First, the banks can hold a portfolio consisting of both types of assets, which will typically be preferred to a portfolio consisting of the safe asset alone. Secondly, by pooling the assets of a large number of consumers, the bank can offer insurance to consumers against their uncertain liquidity demands, giving the early consumers some of the benefits of the high-yielding risky asset without subjecting them to the volatility of the asset market.

Free entry into the banking industry forces banks to compete by offering deposit contracts that maximize the expected utility of the consumers. Thus, the behavior of the banking industry can be represented by an optimal risk-sharing problem. In the next three subsections we consider a variety of different risk-sharing problems, corresponding to different assumptions about the informational and regulatory environment.

A. The Optimal, Incentive-Compatible, Risk-Sharing Problem

Initially consider the case where banks can write contracts in which the amount that can be withdrawn at each date is contingent on R . This provides a benchmark for optimal risk sharing. Since the proportions of early and late consumers are always equal, the only aggregate uncertainty comes from the return to the risky asset R . Since the risky asset return is not known until the second date, the portfolio choice is independent of R , but the payments to early and late consumers, which occur after R is revealed, will depend on it. Let E denote the consumers' total endowment of the consumption good at date 0 and let X and L denote the representative bank's holding of the risky and safe assets, respectively. The deposit contract can be represented by a pair of functions, $c_1(R)$ and $c_2(R)$, which give the consumption of early and late consumers conditional on the return to the risky asset.

The optimal risk-sharing problem can be written as follows:

$$(P1) \begin{cases} \max & E[u(c_1(R)) + u(c_2(R))] \\ \text{s.t.} & \text{(i) } L + X \leq E; \\ & \text{(ii) } c_1(R) \leq L; \\ & \text{(iii) } c_1(R) + c_2(R) \leq L + RX; \\ & \text{(iv) } c_1(R) \leq c_2(R). \end{cases} \tag{2}$$

The first constraint says that the total amount invested must be less than or equal to the amount deposited. There is no loss of generality in assuming that consumers deposit their entire wealth with the bank, since anything they can do the bank can do for them. The second constraint says that the holding of the safe asset must be sufficient to provide for the consumption of the early consumers. The bank may want to hold strictly more than this amount and roll it over to the final period in order to reduce the uncertainty of the late consumers. The next constraint, together with the preceding one, says that the consumption of the late consumers cannot exceed the total value of the risky asset plus the amount of the safe asset left over after the early consumers are paid off; that is,

$$c_2(R) \leq (L - c_1(R)) + RX. \tag{3}$$

The final constraint is the incentive compatibility constraint. It says that for every value of R , the late consumers must be at least as well off as the early consumers. Since late consumers are paid off at date 2, an early consumer cannot imitate a late consumer. However, a late consumer can imitate an early consumer, obtain $c_1(R)$ at date 1, and use the storage technology to provide himself with $c_1(R)$ units of consumption at date 2. It will be optimal to do this unless $c_1(R) \leq c_2(R)$ for every value of R .

The following assumptions are maintained throughout the paper to ensure interior optima. The preferences and technology are assumed to satisfy the inequalities

$$E[R] > 1 \quad (4)$$

and

$$u'(0) > E[u'(RE)R]. \quad (5)$$

The first inequality simply states that the risky asset is more productive than the safe asset. This ensures that even a risk-averse investor will always hold a positive amount of the risky asset. The second inequality is a little harder to interpret. Suppose the bank invests the entire endowment E in the risky asset for the benefit of the late consumers. The consumption of the early consumers will be zero and the consumption of the late consumers will be RE . Under these conditions, the second inequality states that a slight reduction in X and an equal increase in L would increase the utility of the early consumers more than it reduces the expected utility of the late consumers. So the portfolio $(L, X) = (0, E)$ cannot be an optimum if we are interested in maximizing the expected utility of the average consumer.

An examination of the optimal risk-sharing problem shows us that incentive constraint (iv) can be dispensed with. To see this, suppose that we solve the problem subject to the first three constraints only. A necessary condition for an optimum is that the consumption of the two types be equal, unless the feasibility constraint $c_1(R) \leq L$ is binding, in which case it follows from the first-order conditions that $c_1(R) = L \leq c_2(R)$. Thus, the incentive constraint will always be satisfied if we optimize subject to the first three constraints only and the solution to (PI) is the first-best allocation.

The optimal contract is illustrated in Figure 1. When the signal at date 1 indicates that $R = 0$ at date 2, both the early and late consumers receive $L/2$ since L is all that is available and it is efficient to equate consumption given the form of the objective function. The early consumers consume their share at date 1 with the remaining $L/2$ carried over until date 2 for the late consumers. As R increases, both groups can consume more. Provided $R \leq L/X \equiv \bar{R}$ the optimal allocation involves carrying over some of the liquid asset to date 2 to supplement the low returns on the risky asset for late consumers. When the signal indicates that R will be high at date 2 (i.e., $R > L/X \equiv \bar{R}$), then early consumers should consume as much as possible at date 1, which is L , since consumption at date 2 will be high in any case. Ideally, the high date 2 output would be shared with the early consumers at date 1, but this is not technologically feasible. It is only possible to carry forward consumption, not bring it back from the future. Formally, we have the following result:

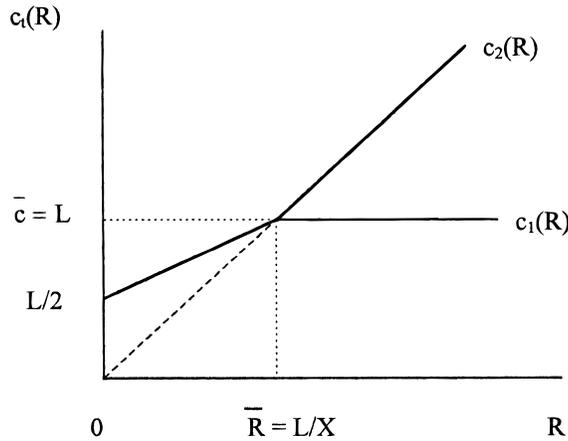


Figure 1. The optimal risk sharing allocation and the optimal deposit contract with runs. At date 0, the bank chooses the optimal investment in the safe asset, L , and the risky asset, X . The figure plots the optimal consumption for early consumers at date 1, $c_1(R)$, and for late consumers at date 2, $c_2(R)$, against R , the payoff of the risky asset at date 2. R can be observed at date 1 but not at date 0. When $R = 0$ the only consumption available is from the safe asset, L . To maximize the date 0 expected utility this is split equally between the two groups so $c_1(0) = c_2(0) = L/2$. The early consumers consume $L/2$ at date 1 and the remaining $L/2$ is carried over to date 2 for the late consumers. As R is increased both groups can consume more. At $\bar{R} = L/X$, L is consumed by the early consumers and $\bar{R}X$ is consumed by the late consumers. As R is increased above \bar{R} it is not possible for the early consumers to have more than L since this is the only consumption available at date 1. At date 2, the late consumers are able to consume $RX > L$. The optimal allocation can also be implemented by a deposit contract that promises \bar{c} to everybody withdrawing or, if that is infeasible, an equal share of L . For $R < \bar{R}$ the extent of the run on the bank in equilibrium ensures that early and late consumers receive equal amounts.

THEOREM 1: *The solution $(L, X, c_1(\cdot), c_2(\cdot))$ to the optimal risk-sharing problem P1 is uniquely characterized by the following conditions:*

$$c_1(R) = c_2(R) = \frac{1}{2} (RX + L) \quad \text{if } L \geq RX,$$

$$c_1(R) = L, c_2(R) = RX \quad \text{if } L \leq RX,$$

$$L + X = E,$$

and

$$E[u'(c_1(R))] = E[u'(c_2(R))R].$$

Under the maintained assumptions, the optimal portfolio must satisfy $L > 0$ and $X > 0$. The allocation is first-best efficient.

Proof: See the Appendix. ■

To illustrate the operation of the optimal contract, we adopt the following numerical example.

$$\begin{aligned}
 U(c_1, c_2) &= \ln(c_1) + \ln(c_2) \\
 E &= 2; \\
 f(R) &= \begin{cases} 1/3 & \text{for } 0 \leq R \leq 3; \\ 0 & \text{otherwise.} \end{cases}
 \end{aligned}
 \tag{6}$$

For these parameters, it can readily be shown that $(L, X) = (1.19, 0.81)$ and $\bar{R} = 1.47$. The level of expected utility achieved is $E[U(c_1, c_2)] = 0.25$.

B. Optimal Risk Sharing through Deposit Contracts with Bank Runs

The optimal risk-sharing problem (*P1*) discussed in the preceding subsection serves as a benchmark for the risk sharing that can be achieved through the kinds of deposit contracts that are observed in practice. The typical deposit contract is “noncontingent,” where the quotation marks are necessitated by the fact that the feasibility constraint may introduce some contingency where none is intended in the original contract. We take a standard deposit contract to be one that promises a fixed amount at each date and pays out all available liquid assets, divided equally among those withdrawing, in the event that the bank does not have enough liquid assets to make the promised payment. As discussed in the introduction, this rule of sharing on an equal basis is different from the Diamond and Dybvig (1983) assumption of first-come, first-served. Let \bar{c} denote the fixed payment promised to the early consumers. We can ignore the amount promised to the late consumers since they are always paid whatever is available at the last date. Then the standard deposit contract promises the early consumers either \bar{c} or, if that is infeasible, an equal share of the liquid assets L , where it has to be borne in mind that some of the late consumers may want to withdraw early as well. In that case, in equilibrium the early and late consumers will have the same consumption.

With these assumptions, the constrained optimal risk-sharing problem can be written as:

$$(P2) \left\{ \begin{array}{ll} \max & E[u(c_1(R)) + u(c_2(R))] \\ \text{s.t.} & \text{(i) } L + X \leq E; \\ & \text{(ii) } c_1(R) \leq L; \\ & \text{(iii) } c_1(R) + c_2(R) \leq L + RX; \\ & \text{(iv) } c_1(R) \leq c_2(R); \\ & \text{(v) } c_1(R) \leq \bar{c} \text{ and } c_1(R) = c_2(R) \text{ if } c_1(R) < \bar{c}. \end{array} \right.
 \tag{7}$$

All we have done here is to add to the unconstrained optimal risk-sharing problem (*PI*) the additional constraint that either the early consumers are paid the promised amount \bar{c} or else the early and late consumers must get the same payment (consumption).

Behind this formulation of the problem is an equivalent formulation that makes explicit the equilibrium conditions of the model and the possibility of runs. To clarify the relationship between these two formulations, it will be useful to have some additional notation. Let $c_{21}(R)$ and $c_{22}(R)$ denote the equilibrium consumption of late consumers who withdraw from the bank at dates 1 and 2, respectively, and let $\alpha(R)$ denote the fraction of late consumers who decide to withdraw early, conditional on the risky return R . Since early consumers must withdraw early, we continue to denote their equilibrium consumption by $c_1(R)$.

In the event that the demands of those withdrawing at date 1 cannot be fully met from liquid short term funds, these funds are distributed equally among those withdrawing. Those who leave their funds in the bank receive an equal share of the risky asset's return at date 2.

If a run does not occur, the feasibility conditions are

$$c_1(R) \leq L, c_1(R) + c_{22}(R) \leq L + RX, \tag{8}$$

as before. If there is a run, then the early consumers and the early-withdrawing late consumers share the liquid assets available at date 1,

$$c_1(R) + \alpha(R)c_{21}(R) = L, \tag{9}$$

and the late-withdrawing late consumers get the returns to the risky asset at date 2,

$$(1 - \alpha(R))c_{22}(R) = RX. \tag{10}$$

Since early consumers and early-withdrawing late consumers are treated the same in a run and all late consumers must have the same utility in equilibrium,

$$c_1(R) = c_{21}(R) = c_{22}(R). \tag{11}$$

If there is no run, then we can assume that $c_{21}(R) = c_{22}(R)$ without loss of generality. These conditions can be summarized by writing

$$\begin{aligned} c_1(R) + \alpha(R)c_2(R) &\leq L, \\ c_1(R) + c_2(R) &\leq L + RX, \end{aligned} \tag{12}$$

where $c_2(R)$ is understood to be the common value of $c_{21}(R)$ and $c_{22}(R)$.

Our final condition comes from the form of the standard deposit contract. Early withdrawers either get the promised amount \bar{c} or the demands of the early withdrawers (including the early-withdrawing late consumers) exhaust the liquid assets of the bank:

$$c_1(R) \leq \bar{c} \quad \text{and} \quad c_1(R) < \bar{c} \Rightarrow c_1(R) + \alpha(R)c_2(R) = L. \quad (13)$$

Now suppose that a feasible portfolio (L, X) has been chosen and that the consumption functions $c_1(\cdot)$ and $c_2(\cdot)$ satisfy the constraints of the risk-sharing problem (P2). Then define $\alpha(\cdot)$ as

$$\alpha(R) = \begin{cases} 0 & \text{if } c_1(R) \leq c_2(R); \\ \frac{L}{c_1(R)} - 1 & \text{otherwise.} \end{cases} \quad (14)$$

It is always possible to do so, since feasibility assures us that $c_1(R) \leq L$. Now it is easy to check that all of the equilibrium conditions given above are satisfied. Conversely, suppose the functions $c_1(\cdot)$, $c_{21}(\cdot)$, $c_{22}(\cdot)$, and $\alpha(\cdot)$ satisfy the equilibrium conditions above. There is no loss of generality in assuming that $\bar{c} \leq L$, so $c_1(R) < \bar{c}$ implies that $\alpha(R) > 0$ and it is easy to check that the constraints of the risk-sharing problem (P2) are satisfied. This proves that solving the risk-sharing problem (P2) is equivalent to choosing an optimal standard deposit contract subject to the equilibrium conditions imposed by the possibility of runs.

When we look carefully at the constrained risk-sharing problem (P2), we notice that it looks very similar to the unconstrained risk-sharing problem (P1) in the preceding section. In fact, the two are equivalent.

THEOREM 2: *Suppose that $\{L, X, c_1(\cdot), c_2(\cdot)\}$ solves the unconstrained optimal risk-sharing problem (P1). Then $\{L, X, c_1(\cdot), c_2(\cdot)\}$ is feasible for the constrained optimal risk-sharing problem (P2). Hence, the expected utility of the solution to (P2) is the same as the expected utility of the solution to (P1) and a banking system subject to runs can achieve first-best efficiency using the standard deposit contract.*

The easiest way to see this is to compare the form of the optimal consumption functions from the two problems. From (P1) we get

$$\begin{aligned} c_1(R) &= \min\left\{\frac{1}{2}(L + RX), L\right\}, \\ c_2(R) &= \max\left\{\frac{1}{2}(L + RX), RX\right\}, \end{aligned} \quad (15)$$

and from (P2) we get

$$\begin{aligned} c_1(R) &= \min\left\{\frac{1}{2}(L + RX), \bar{c}\right\}, \\ c_2(R) &= \max\left\{\frac{1}{2}(L + RX), L + RX - \bar{c}\right\}. \end{aligned} \quad (16)$$

The two are identical if we put $\bar{c} = L$. In other words, to achieve the optimum, we minimize the amount of the liquid asset, holding only what is necessary to meet the promised payment for the early consumers, and allow bank runs to achieve the optimal sharing of risk between the early and late consumers.

The optimal deposit contract is illustrated by Figure 1 with $\bar{c} = L$. For $R < \bar{R}$ the optimal degree of risk sharing is achieved by increasing $\alpha(R)$ to one as R falls to zero. The more late consumers who withdraw at date 1 the less each person withdrawing then receives. Early-withdrawing late consumers hold the safe asset outside the banking system. The return from doing this is exactly the same as the return on safe assets held within the banking system. The solution to the numerical example introduced above is unchanged with $\bar{c} = 1.19$. When $R = 1$, $\alpha(R) = 0.19$, and when $R = 0.5$, $\alpha(R) = 0.49$.

The total illiquidity of the risky asset plays an important equilibrating role in this version of the model. Because the risky asset cannot be liquidated at date 1, there is always something left to pay the late withdrawers at date 2. For this reason, bank runs are typically partial, that is, they involve only a fraction of the late consumers, unlike the Diamond–Dybvig (1983) model in which a bank run involves all the late consumers. As long as there is a positive value of the risky asset $RX > 0$, there must be a positive fraction $1 - \alpha(R) > 0$ of late consumers who wait until the last period to withdraw. Otherwise the consumption of the late withdrawers $c_{22}(R) = RX/(1 - \alpha(R))$ would be infinite. Assuming that consumption is positive in both periods, an increase in $\alpha(R)$ must raise consumption at date 2 and lower it at date 1. Thus, when a bank run occurs in equilibrium, there will be a unique value of $\alpha(R) < 1$ that equates the consumption of early-withdrawing and late-withdrawing consumers.

C. Standard Deposit Contracts without Runs

We have seen that the first-best outcome can be achieved by means of a “noncontingent” deposit contract together with bank runs that introduce the optimal degree of contingency. Thus, there is no justification for central bank intervention to eliminate runs. In fact, if runs occur in equilibrium, a policy that eliminates runs by forcing the banks to hold a safer portfolio must be strictly worse.

It is possible, of course, to conceive of an equilibrium in which banks voluntarily choose to hold such a large amount of the safe asset that runs never occur. Suppose that the incentive-efficient allocation involves no bank runs. Then we know from the characterization of the solution to (PI) that $c_1(R) = L$ and $c_2(R) = RX$ for all values of R . If we assume that the greatest lower bound of the support of R is zero, then the incentive-compatibility constraint requires that

$$L = c_1(0) \leq c_2(0) = 0. \quad (17)$$

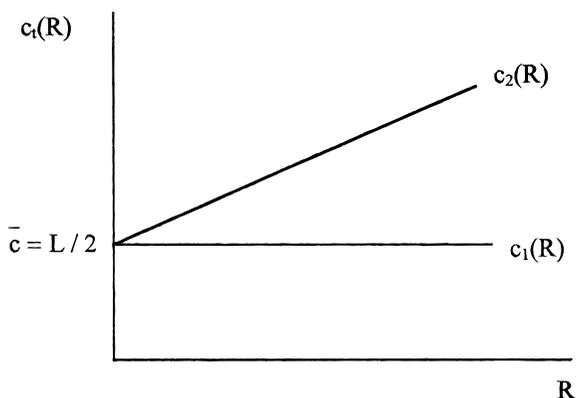


Figure 2. The optimal deposit contract without runs. At date 0, the bank chooses the optimal investment in the safe asset, L , and the risky asset, X , subject to the constraint that it can always provide the amount promised in the deposit contract \bar{c} to all depositors. The figure plots the optimal consumption for early consumers at date 1, $c_1(R)$, and for late consumers at date 2, $c_2(R)$, against R , the payoff of the risky asset at date 2. To ensure no runs the most that can be promised is $\bar{c} = L/2$.

So the entire endowment is invested in the risky asset, the early consumers receive nothing and the late consumers receive RE . But this means that $X = E$ must maximize

$$u(E - X) + E[u(RX)], \quad (18)$$

and the first-order condition for this is

$$u'(0) \leq E[u'(RE)R], \quad (19)$$

contradicting one of our maintained assumptions. Hence, runs cannot be avoided in the optimal risk-sharing scheme.

If the central bank were to prohibit holding portfolios that were vulnerable to runs, this would force the banks to guarantee a constant consumption level $c_1(R) = \bar{c}$ to early consumers, which they can only do by lowering the early consumers' consumption and/or by holding excess amounts of the safe asset. By the earlier argument, when $R = 0$ we have $2\bar{c} \leq c_1(0) + c_2(0) \leq L$ so either $\bar{c} = 0$ or $L > \bar{c}$, neither of which is consistent with the optimum.

THEOREM 3: *Assuming that the support of R contains zero, the deposit contract equilibrium implementing the first-best allocation involves runs. Hence, an equilibrium in which runs are prevented by central bank regulation is strictly worse than the first-best allocation.*

Theorem 3 shows that preventing financial crises by forcing banks to hold excessive reserves can be suboptimal. The optimal allocation requires early consumers to bear some of the risk. Figure 2 shows the constrained-optimal

contract when the bank is required to prevent runs by restricting its promised payout \bar{c} and increasing the level of reserves L . For the parameter values in our example, it can readily be shown that the constrained-optimal portfolio satisfies $(L, X) = (1.63, 0.37)$ and that $\bar{c} = 0.82$. The level of expected utility achieved is $E[U(c_1, c_2)] = 0.08$. In comparison with the case where the optimal allocation is implemented by runs, the consumption provided to early consumers is lower except when the return to the risky asset is very low ($R \leq 0.56$). As a result of this misallocation of consumption between early and late consumers, the ex ante welfare of all consumers is lower than in the first best.

The conclusion of Theorem 3 is consistent with the observation that, prior to central bank and government intervention, banks chose not to eliminate the possibility of runs, although it would have been feasible for them to do so. Under the conditions of Theorem 3, any intervention to curb bank runs must make depositors strictly worse off and, in any case, it cannot improve upon the situation, which is already first-best efficient according to Theorems 1 and 2.

D. Unequal Probabilities of Early and Late Consumption

The analysis so far has assumed that the probability of being an early consumer is 1/2. This is a matter of convenience only and it can be shown that with appropriate minor modifications the results above all remain valid when the probabilities of being an early or late consumer differ. To see this suppose depositors are early consumers with probability γ and late consumers with probability $1 - \gamma$. The probability of being an early (late) consumer is equal to the proportion of early (late) consumers, so the consumption of each type must be multiplied by γ ($1 - \gamma$) in the feasibility constraints. Then the optimal, incentive-compatible, risk-sharing allocation solves the following problem:

$$\begin{aligned}
 \max \quad & E[\gamma u(c_1(R)) + (1 - \gamma)u(c_2(R))] \\
 \text{s.t.} \quad & \text{(i) } L + X \leq E; \\
 & \text{(ii) } \gamma c_1(R) \leq L; \\
 & \text{(iii) } \gamma c_1(R) + (1 - \gamma)c_2(R) \leq L + RX; \\
 & \text{(iv) } c_1(R) \leq c_2(R).
 \end{aligned} \tag{20}$$

Since γ and $1 - \gamma$ appear symmetrically in the objective function and the constraints, they drop out of the Kuhn–Tucker first-order conditions. The characterization of the first-best allocation follows an exactly similar argument to the one given earlier. The total measure of consumers is now one rather than two, so the optimal consumption allocation is

$$c_1(R) = c_2(R) = L + RX \quad \text{if } L \geq RX \tag{21}$$

and

$$c_1(R) = \frac{L}{\gamma}, c_2(R) = \frac{RX}{1 - \gamma} \quad \text{if } L \leq RX. \quad (22)$$

With appropriate modifications, all the other arguments above remain valid. Similar extensions are available for the results in the following sections, but for convenience we continue to deal explicitly only with the case $\gamma = 1 - \gamma = 1/2$.

II. Costly Financial Crises

A crucial assumption for the analysis of the preceding section is that bank runs do not reduce the returns to the assets. The long-term asset cannot be liquidated, so its return is unaffected. By assumption, the safe asset liquidated at date 1 yields the same return whether it is being held by the early-withdrawing late consumers or by the bank. For this reason, bank runs make allocations contingent on R without diminishing asset returns. However, if liquidating the safe asset at date 1 involved a cost there would be a trade-off between optimal risk sharing and the return realized on the bank's portfolio.

To illustrate the consequences of liquidation costs, in this section we study a variant of the earlier model in which the return on storage by early-withdrawing late consumers is lower than the return obtained by the bank. Since there is now a cost attached to making the consumption allocation contingent on the return to the risky asset, incentive-efficient risk sharing is not attainable in an equilibrium with bank runs. Central bank intervention is needed to achieve the first-best.

A. Optimal Risk Sharing with Costly Liquidation

Let $r > 1$ denote the return on the safe asset between dates 1 and 2. We continue to assume that the return on the safe asset between dates 0 and 1 is one. This assumption is immaterial since all of the safe asset is held by the bank at date 0. As before, one unit of consumption stored by individuals at date 1 produces one unit of consumption at date 2. It will be assumed that the safe asset is less productive on average than the risky asset; that is,

$$E[R] > r. \quad (23)$$

The characterization of the incentive-efficient deposit contract follows the same lines as before. The bank chooses a portfolio of investments (L, X) and offers the early (late) consumers a consumption level $c_1(R)$ ($c_2(R)$), condi-

tional on the return on the risky asset. The deposit contract is chosen to maximize the ex ante expected utility of the typical consumer. Formally, the optimal risk-sharing problem can be written as:

$$(P3) \begin{cases} \max & E[u(c_1(R)) + u(c_2(R))] \\ \text{s.t.} & \text{(i) } L + X \leq E; \\ & \text{(ii) } c_1(R) \leq L; \\ & \text{(iii) } c_2(R) \leq r(L - c_1(R)) + RX; \\ & \text{(iv) } c_1(R) \leq c_2(R). \end{cases} \tag{24}$$

The only difference between this optimization problem and the original problem (P1) occurs in constraint (iii), which reduces to the earlier formulation if we put $r = 1$.

To solve problem (P3), we adopt the same device as before: remove the incentive-compatibility constraint (iv) and solve the relaxed problem. Then note that the first-order conditions for the relaxed problem require

$$u'(c_1(R)) \geq ru'(c_2(R)), \tag{25}$$

with equality holding if $c_1(R) < L$. Then $c_1(R) \leq c_2(R)$ for every R , so the incentive-compatibility condition is automatically satisfied.

The arguments used to analyze (P1) provide a similar characterization here. There exists a critical value of \bar{R} such that $c_1(R) < L$ if and only if $R < \bar{R}$. Then the consumption allocation is uniquely determined, given the portfolio (L, X) , by the relations

$$\begin{aligned} u'(c_1(R)) &= ru'(c_2(R)) \quad \text{if } R < \bar{R}, \\ c_1(R) &= L, c_2(R) = RX \quad \text{if } R \geq \bar{R}, \end{aligned} \tag{26}$$

where \bar{R} can be chosen to satisfy $u'(L) = ru'(RX)$. With this consumption allocation, we can show, using the maintained assumptions, that the portfolio will have to satisfy $L > 0$ and $X > 0$ and the first-order condition

$$E[u'(c_1(R))] = E[u'(c_2(R))R], \tag{27}$$

together with the budget constraint $L + X = E$, will determine the optimal portfolio.

In the case of the numerical example, it can be shown that if $r = 1.05$, $(L, X) = (1.36, 0.64)$ and $\bar{R} = 2.23$, the level of expected utility achieved is $E[U(c_1, c_2)] = 0.32$. Figure 3 illustrates the form of the optimal contract. Whereas in Figure 1 the two groups' consumption is equated for $R < \bar{R}$, now this is no longer the case because $r > 1$.

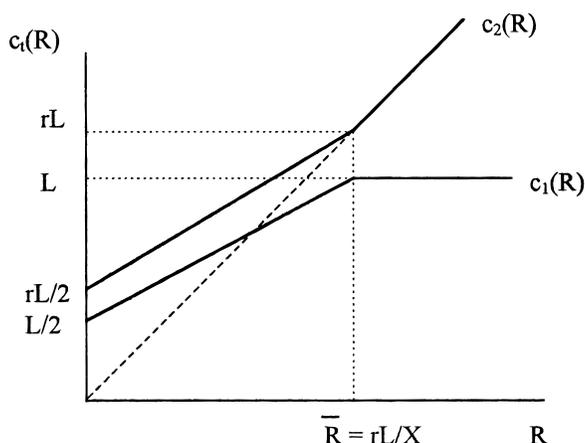


Figure 3. The optimal risk sharing allocation with costly liquidation. Between dates 1 and 2 the return on the safe asset within the banking system is $r > 1$. The figure plots the optimal consumption for early consumers at date 1, $c_1(R)$, and for the late consumers at date 2, $c_2(R)$, against R , the payoff of the risky asset at date 2. The case shown is when the utility function is $u(c_t(\cdot)) = \ln(c_t(\cdot))$. Maximizing date 0 expected utility now involves ensuring the ratio of the marginal utilities of consumption for early and late consumers is equated to the marginal rate of transformation, r . Consumption is higher for late consumers even for $R < \bar{R} = rL/X$.

B. Standard Deposit Contracts with Costly Liquidation

The next step is to characterize an equilibrium in which the bank is restricted to use a standard deposit contract and, as a result, bank runs become a possibility. The change in the assumption about the rate of return on the safe asset appears innocuous but it means that we must be much more careful about specifying the equilibrium. Let \bar{c} denote the payment promised by the bank to anyone withdrawing at date 1 and let $c_1(R)$ and $c_2(R)$ denote the equilibrium consumption levels of early and late consumers, respectively, conditional on the return to the risky asset. Finally, let $0 \leq \alpha(R) \leq 1$ denote the fraction of late consumers who choose to “run,” that is, to withdraw from the bank at date 1.

The bank chooses a portfolio (L, X) , the pair of consumption functions $c_1(R)$ and $c_2(R)$, the deposit parameter \bar{c} , and the withdrawal function $\alpha(R)$ to maximize the expected utility of the typical depositor, subject to the following equilibrium conditions. First, the bank’s choices must be feasible, and this means that

$$\begin{aligned}
 L + X &\leq E, \\
 c_1(R) + \alpha(R)c_2(R) &\leq L, \\
 (1 - \alpha)c_2(R) &\leq r(L - c_1(R) - \alpha(R)c_2(R)) + RX.
 \end{aligned} \tag{28}$$

The first two constraints are familiar. The final constraint says that withdrawals in the last period, which equal the consumption of the late-withdrawing fraction of the late consumers, cannot exceed the sum of the returns on the risky asset and the returns on the part of the safe asset that is carried over to the last period. The reason that we need to take explicit account here of the fraction $\alpha(R)$ of late consumers who withdraw early is that their decision affects the total amount of consumption available. A unit of consumption withdrawn at date 1 reduces consumption at date 2 by $r > 1$, so it is not a matter of indifference as it was under the previous assumption that $r = 1$.

The standard deposit contract requires the bank to pay the depositors who withdraw in the middle period either a fixed amount \bar{c} or as much as it can from liquid assets. Formally, this amounts to saying that

$$c_1(R) \leq \bar{c}, \tag{29}$$

$$c_1(R) + \alpha(R)c_2(R) = L \quad \text{if } c_1(R) < \bar{c}.$$

Finally, we have the incentive-compatibility condition:

$$c_1(R) \leq c_2(R), \tag{30}$$

and the equal-treatment condition:

$$c_1(R) = c_2(R) \quad \text{if } \alpha(R) > 0. \tag{31}$$

In other words, if some late consumers withdraw in the middle period, their consumption must be the same as the early consumers since they get the same payment from the bank and store it until the last period. In writing down these conditions, we have implicitly assumed that late consumers get the same consumption whether they withdraw early or late. This will be true in equilibrium, of course.

Having specified the constraints, the bank's problem is formally

$$(P4) \left\{ \begin{array}{ll} \max & E[u(c_1(R)) + u(c_2(R))] \\ \text{s.t.} & \text{(i) } L + X \leq E; \\ & \text{(ii) } c_1(R) + \alpha(R)c_2(R) \leq L; \\ & \text{(iii) } (1 - \alpha)c_2(R) \leq r(L - c_1(R) - \alpha(R)c_2(R)) + RX; \\ & \text{(iv) } c_1(R) \leq \bar{c}; \\ & \text{(v) } c_1(R) + \alpha(R)c_2(R) = L \quad \text{if } c_1(R) < \bar{c}; \\ & \text{(vi) } c_1(R) \leq c_2(R); \\ & \text{(vii) } c_1(R) = c_2(R) \quad \text{if } \alpha(R) > 0. \end{array} \right. \tag{32}$$

In principle, we could solve the bank's problem directly, but it will be convenient to simplify it first.

The simplification requires us to note that we are assuming the bank is implicitly allowed to choose the equilibrium that will result at dates 1 and 2, and this ensures that runs will not occur unnecessarily. More precisely,

$$\alpha(R) > 0 \quad \text{implies that } c_1(R) < \bar{c}. \quad (33)$$

To see this suppose, contrary to what is to be proved, that $\alpha(R) > 0$ and $c_1(R) = \bar{c}$. Now consider an alternative choice for the bank in which $c_2(R)$ is replaced by $\hat{c}_2(R)$ and $\alpha(R)$ is replaced by $\hat{\alpha}(R)$. Put $\hat{\alpha}(R) = 0$ and use constraint (iii) to define $\hat{c}_2(R)$:

$$\hat{c}_2(R) = r(L - \bar{c}) + RX > c_2(R). \quad (34)$$

Since all of the other conditions are satisfied, the fact that $\hat{c}_2(R) > c_2(R)$ contradicts the optimality of $c_2(R)$ and establishes the desired result.

Under the assumption that the bank can select the equilibrium in which no runs occur, if such an equilibrium exists, there are only two cases to be considered. Either there are no runs, $\alpha(R) = 0$, $c_1(R) = \bar{c}$, and $c_2(R) = r(L - \bar{c}) + RX \geq \bar{c}$; or else there are runs, $\alpha(R) > 0$, and $c_1(R) = c_2(R) < \bar{c}$. When there are runs, we have

$$(1 + \alpha(R))c_1(R) = L \quad (35)$$

from constraint (ii) and

$$(1 - \alpha(R))c_2(R) = RX \quad (36)$$

from constraint (iii), so using the equality of $c_1(R)$ and $c_2(R)$ gives us

$$\frac{L}{(1 + \alpha(R))} = \frac{RX}{(1 - \alpha(R))}, \quad (37)$$

or

$$\alpha(R) = \frac{L - RX}{L + RX}. \quad (38)$$

Substituting this value into the expression for $c_1(R)$ yields

$$c_1(R) = \frac{L}{(1 + \alpha(R))} = \frac{L + RX}{2}. \quad (39)$$

This is the same expression that we obtain in the costless case, which is not surprising once we recall that no safe asset is being held by the bank between dates 1 and 2 when there are bank runs.

Since we know that runs occur if and only if $c_1(R) < \bar{c}$, we know that runs occur if and only if $R < R^*$, where R^* is defined implicitly by the condition

$$\bar{c} = r(L - \bar{c}) + R^*X. \tag{40}$$

In other words, if there are no runs and the early consumers are paid the promised amount \bar{c} , there will be just enough to provide the late consumers with a level of consumption that satisfies the incentive-compatibility constraint. Clearly if $R < R^*$ there must be a run because it is not feasible to pay the late consumers \bar{c} and the early consumers cannot get less unless there is a run. Conversely, if $R \geq R^*$ then it is always feasible to avoid a run, and we have shown that in such cases the bank will find it optimal to do so. We focus on the interior case where $R^* > 0$.

Thus, the bank's decision problem can be simplified to the following:

$$\begin{aligned} \max \quad & \int_0^{R^*} 2u\left(\frac{L + RX}{2}\right) f(R) dR + \int_{R^*}^{\infty} \{u(\bar{c}) + u(r(L - \bar{c}) + RX)\} f(R) dR \\ \text{s.t.} \quad & \text{(i) } L + X \leq E; \\ & \text{(ii) } R^* = \frac{(1 + r)\bar{c} - rL}{X}. \end{aligned} \tag{41}$$

There are two types of solution for this problem. The first possibility is $\bar{c} = L$, in which case the optimal deposit contract is the same as the solution to (P2), which is illustrated in Figure 1. The second possibility is $\bar{c} < L$. This is illustrated in Figure 4. Note that since $\bar{c} > (L + R^*X)/2$, the functions $c_1(R)$ and $c_2(R)$ are discontinuous at $R = R^*$. Whether $\bar{c} = L$ or $\bar{c} < L$, it is clear that the first-order conditions for the solution of the incentive-efficient allocation are not satisfied; for example, for $R < R^*$ the first-order condition $u'(c_1(R)) = ru'(c_2(R))$ is violated. This can be seen directly by comparing Figures 1 and 4 with Figure 3.

The different types of equilibria can be illustrated in the context of the numerical example. As long as $r < 1.25$, the optimal deposit contract is the same as when $r = 1$ because $\bar{c} = L$ and so nothing is invested at rate r between dates 1 and 2. In other words it has $\bar{c} = L = 1.19$, $X = 0.81$, $\bar{R} = 1.47$, and $E[U(c_1, c_2)] = 0.25$. For $r \geq 1.25$ the optimal contract has $R^* = 0$. The representative bank finds it optimal to voluntarily prevent runs and the deposit contract is similar to the one shown in Figure 2. For example, in the case of $r = 1.25$, $\bar{c} = 1$, $L = 1.8$, $X = 0.2$, and $E[U(c_1, c_2)] = 0.25$. Hence small changes in r around $r = 1.25$ can cause large changes in the bank's optimal portfolio. The final possibility where $\bar{c} < L$ is illustrated by the case where the probability density function of R is uniform on $[0, 0.28]$ rather than $[0, 0.3]$ but everything else is as before. Here for $r = 1.04$, $\bar{c} = 1.30$, $L = 1.38$, $X = 0.62$, $R^* = 1.96$, and $E[U(c_1, c_2)] = 0.045$.

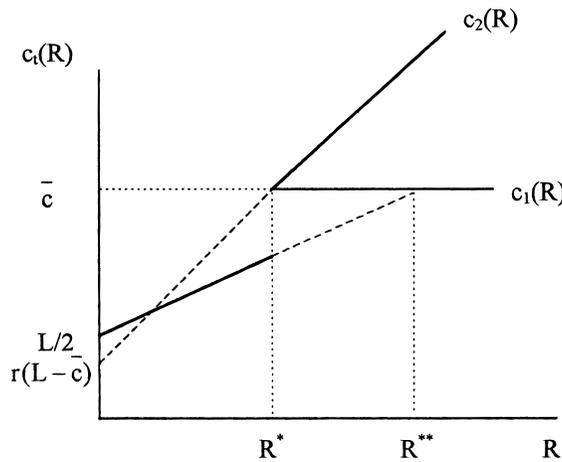


Figure 4. The optimal deposit contract with costly liquidation when $\bar{c} < L$. Between dates 1 and 2 the return on the safe asset within the banking system is $r > 1$ as in Figure 3. The figure plots the optimal consumption for early consumers at date 1, $c_1(R)$, and for late consumers at date 2, $c_2(R)$, against R , the payoff of the risky asset at date 2. The optimal deposit contract, which promises \bar{c} to everybody withdrawing or, if that is infeasible, an equal share of L , can no longer implement the optimal allocation. This is because for $R < R^*$ runs ensure that the consumption of the two groups is equated. At R^* there are two possibilities. The first is that there is a run and all the safe assets are withdrawn. The second is that only the early consumers withdraw and the remaining safe asset is kept in the bank and earns $r > 1$. This is higher than when the assets are withdrawn so the total amount both groups can consume is greater. For $R^* \leq R \leq R^{**}$ multiple equilibria exist because of these two possibilities.

C. Multiple Equilibria

As was noted earlier, the preceding analysis is based on the assumption that, when there are multiple equilibria at date 1, the bank is allowed to select the one that is preferred by depositors. In practice, this means that runs occur only if they are unavoidable—that is, only if there does not exist an equilibrium without runs. For any portfolio (L, X) and payout \bar{c} , it is clear that a run must occur if

$$\bar{c} > r(L - \bar{c}) + RX, \tag{42}$$

since it is impossible to pay the early consumers the promised amount \bar{c} and give at least as much to the late consumers. Conversely, if

$$\bar{c} < r(L - \bar{c}) + RX, \tag{43}$$

it is possible to give the late consumers more than \bar{c} , so there is an equilibrium without runs. However, if \bar{c} is close enough to $r(L - \bar{c}) + RX$, there is another possibility. If some late consumers decide to run, it will not be possible to pay out \bar{c} at date 1 to the early withdrawers, even if all the liquid

asset is paid out, and because the higher return on the safe asset held by the bank is lost through early liquidation, the late-withdrawing consumers will be worse off too. For an appropriate size of run the late consumers will be indifferent between running and waiting.

Let R^{**} denote the critical value of R below which this second type of equilibrium appears. Then R^{**} is determined by the condition that

$$\frac{R^{**}X + L}{2} = \bar{c}. \tag{44}$$

If a run occurred at this value of R then it would just be possible to give both types of consumer \bar{c} . The fraction of late consumers who run is determined by the condition that

$$(1 + \alpha(R^{**}))\bar{c} = L. \tag{45}$$

A simple calculation shows that

$$(1 - \alpha(R^{**}))\bar{c} = 2\bar{c} - L = R^{**}X, \tag{46}$$

so it is just feasible to give the late consumers \bar{c} at date 2. Figure 4 illustrates where R^{**} lies.

For values of R between R^* and R^{**} , we can choose $\alpha(R)$ so that

$$c_1(R) = \frac{L}{1 + \alpha(R)} = \frac{RX}{1 - \alpha(R)} = c_2(R), \tag{47}$$

which again satisfies the equilibrium conditions and allows a run. Both types of consumers are worse off in this situation than if a run had not occurred, since $c_1(R) = c_2(R) < \bar{c}$, but it is an equilibrium for the given values of L , X , and \bar{c} and so it cannot be ruled out. In the context of the numerical example where the probability density function of R is uniform on $[0, 2.28]$, $R^{**} = 1.97$, so there are multiple equilibria for $R \in [R^*, R^{**}] = [1.96, 1.97]$.

D. Optimal Monetary Policy

The inefficiency of equilibrium with bank runs arises from the fact that liquidating the safe asset at date 1 and storing the proceeds until date 2 is less productive than reinvesting them in safe assets held by the bank. A simple monetary intervention by the central bank can remedy this inefficiency. Essentially, it consists of giving to depositors the money provided by the central bank instead of goods. In the event of a run at date 1, the central bank gives the representative bank a loan of M units of money. The bank gives depositors a combination of money and consumption whose value equals the fixed amount promised in the deposit contract. Since early consumers want to consume their entire wealth at date 1, they exchange the money for

consumption with the early-withdrawing late consumers. The price level adjusts so that the early consumers end up with the first-best consumption level and the early-withdrawing late consumers end up holding all the money. At date 2, the representative bank has to repay its loan to the central bank. For simplicity we assume that the loan bears zero interest. The money now held by late consumers is just enough to allow the bank to repay its loan and the bank has just enough consumption from its remaining investment in the safe asset to give the early-withdrawing late consumers the second-best consumption level. The price level at date 2 adjusts so that the bank and the early withdrawers can exchange money for consumption in the correct ratio and the bank ends up with the amount of money it needs to repay the loan and the consumers end up with the first-best consumption level.

In order for this intervention to have the required effect on the choice of portfolio and the allocation of consumption, the deposit contract has to be specified in nominal terms. This means that a depositor is promised the equivalent of a fixed amount of money D if he withdraws in the middle period and whatever the representative bank can afford to pay if he withdraws in the final period. This intervention does not require the central bank to condition its policy on the return to the risky asset R . It is sufficient for the central bank to give the representative bank an interest-free line of credit that the representative bank can choose to draw on. Whatever part of the line of credit is used must be repaid in the last period. Without loss of generality, we can fix the size of the line of credit from the central bank and assume that the representative bank uses either none or all of it at date 1.

Let (L, X) be the portfolio and let $c_1(R)$ and $c_2(R)$ be the consumption functions derived from the optimal risk-sharing problem (P3). Let D be the nominal value of a deposit at date 1 and let M be the size of the loan available to the representative bank. (We assume that the bank will make use of the full line of credit or none of it.) In states in which the consumption of the early consumers is L there is nothing that the representative bank needs to do to prevent runs. As before, in states where $c_1(R) < L$, bank runs are valuable because they make the value of the deposits contingent on R , but here they operate through the price level, which is assumed to adjust so that

$$p_1(R)c_1(R) = D. \quad (48)$$

We do not want premature liquidation of the safe asset at date 1, so the late consumers must hold only money between dates 1 and 2. Since the nominal value of a withdrawal at date 1 is D , this implies that

$$\alpha(R)D = M. \quad (49)$$

Similarly, we want the early-withdrawing late consumers to be able to afford just $c_2(R)$ at date 2. To ensure this, we must have

$$\alpha(R)p_2(R)c_2(R) = M. \quad (50)$$

Clearly, there are many values of $\alpha(R)$, $p_1(R)$, and $p_2(R)$ that will satisfy these conditions. Furthermore, these conditions are sufficient for an equilibrium. At date 1, the bank hands out a mixture of goods and money to withdrawers. The early consumers do not want any money, so they exchange theirs with the late consumers. The late consumers do not want to hold any goods, since the return on money is greater than the return on goods:

$$\frac{p_1(R)}{p_2(R)} = \frac{c_2(R)}{c_1(R)} > 1. \tag{51}$$

Consequently, the late consumers end up holding only money between dates 1 and 2. At date 2, the early-withdrawing late consumers supply all their money inelastically to the representative bank in exchange for goods. The representative bank gets back just enough money to repay its loan from the central bank, and has enough goods left over to give each late-withdrawing consumer $c_2(R)$.

To see how the price level $p_1(R)$ is determined, consider the aggregate transactions at date 1. The bank gives each depositor $c_1(R)/[1 + \alpha(R)]$ units of consumption and $M/[1 + \alpha(R)]$ units of money. The early consumers supply a total of $M/[1 + \alpha(R)]$ units of money (in exchange for goods) and the late consumers supply a total of $\alpha(R)(c_1(R)/[1 + \alpha(R)])$ units of goods (in exchange for money). To clear the market the price adjusts to equate the value of goods supplied to the quantity of money:

$$\frac{M}{1 + \alpha(R)} = p_1(R)\alpha(R) \frac{c_1(R)}{1 + \alpha(R)}, \tag{52}$$

so

$$M = p_1(R)\alpha(R)c_1(R), \tag{53}$$

which is equivalent to the conditions above. The determination of $p_2(R)$ is similar.

THEOREM 4: *Suppose that the central bank makes available to the representative bank an interest-free line of credit of M units of money at date 1 which must be repaid at date 2. Then there exist equilibrium price levels $p_1(R)$ and $p_2(R)$ and an equilibrium fraction of early withdrawers $\alpha(R)$ for every value of R , which will implement the incentive-efficient allocation $\{(L, X), c_1(\cdot), c_2(\cdot)\}$.*

Although the central bank policy described in Theorem 4 removes the dead-weight costs of bank runs, it does not prevent the runs themselves. Injecting money into the banking system dilutes the claims of the early consumers so that they bear a share of the low returns to the risky asset. Without bank runs, first-best risk sharing would not be achieved.

To illustrate how the first-best allocation can be implemented in the context of the numerical example with $r = 1.05$, recall that the social optimum has $(L, X) = (1.36, 0.64)$, $\bar{R} = 2.23$, and $E[U(c_1, c_2)] = 0.32$. Suppose $D = 1.36$. For $R \geq \bar{R} = 2.23$ then $p_1(R) = p_2(R) = 1$. For $R < \bar{R} = 2.23$ the price levels at the two dates depend on the level of R . To illustrate, suppose $R = 2$. In that case $c_1(2) = 1.29$ so $p_1(2) = 1.36/1.29 = 1.05$ and $c_2(2) = 1.35$ so $p_2(2) = 1.36/1.35 = 1.01$. Similarly for other values of R . Note that it is optimal at these prices for the early withdrawers to hold money from date 1 to date 2 since the price of goods is falling. In other words, they do not use the storage technology available to them because they can do better holding money. The fraction of late consumers who withdraw from the bank and hold money will be determined by M . Suppose $M = 1$. Then $\alpha(R) = 1/1.36 = 0.74$.

III. Asset Trading and the Efficiency of Runs

As has been pointed out above, the assumption that the long-term risky asset is completely illiquid plays an important role in equilibrating bank runs, so that runs are typically partial, that is, they involve only a fraction of the late consumers. In this section, we introduce a competitive asset market in which the risky asset can be traded. The participants in the market are the banks, who use it to obtain liquidity, and a large number of wealthy, risk-neutral speculators who hope to make a profit in case some bank has to sell off assets cheaply to get liquidity. The speculators hold cash (the safe asset) in order to purchase the risky asset. The return on the cash is low, but it is offset by the prospect of speculative profits when the price of the risky asset falls below its "fair" value.

The impact of introducing the asset market can be illustrated using the consumption profiles in Figure 5. The graphs in this figure represent the equilibrium consumption levels of early and late consumers, respectively, as a function of the risky asset return R . For high values of R (i.e., $R \geq R^*$), there is no possibility of a bank run. The consumption of early consumers is fixed by the standard deposit contract at $c_1(R) = \bar{c}$ and the consumption of late consumers is given by the budget constraint $c_2(R) = r(L - \bar{c}) + RX$.

For lower values of R ($R < R^*$), it is impossible to pay the early consumers the fixed amount \bar{c} promised by the standard deposit contract without violating the late consumers' incentive constraint and a bank run inevitably ensues. However, there cannot be a partial run. The terms of the standard deposit contract require the bank to liquidate all of its assets at the second date if it cannot pay \bar{c} to every depositor who demands it. Since late withdrawers always receive as much as the early consumers by incentive compatibility, the bank has to liquidate all its assets unless it can give at least \bar{c} to all consumers. The value of R^* is determined by the condition that the bank can just afford to give everyone \bar{c} . Below R^* it is impossible for the

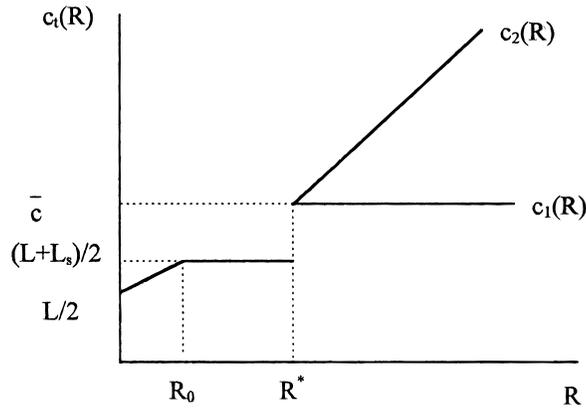


Figure 5. The optimal deposit contract when there is a market for the risky asset. The figure plots the optimal consumption for early consumers at date 1, $c_1(R)$, and for late consumers at date 2, $c_2(R)$, against R , the payoff of the risky asset at date 2, when the risky asset can be sold in a market. Liquidity is provided to this market by wealthy speculators who put L_s in the safe asset and X_s in the risky asset. The bank invests L in the safe asset and X in the risky asset. At R^* the bank has just enough of the safe asset to supply the amount promised in the deposit contract, \bar{c} , to the early consumers. For $R < R^*$ everybody withdraws and the bank has to liquidate all of its risky asset X . The cash available for purchasing the asset is L_s . For $R_0 < R < R^*$ there is “cash in the market pricing” and the price is $P(R) = L_s/X$. The safe asset and the proceeds from selling the risky asset are divided equally among the early and late consumers so everybody receives $(L + L_s)/2$. For $R < R_0$ the risky asset’s price is determined in the usual way.

bank to pay all the depositors \bar{c} , and the only alternative is to liquidate all its assets at the first date and pay all consumers less than \bar{c} . Since a late withdrawer will receive nothing, all consumers will choose to withdraw their deposits at the second date.

There is a discontinuity in the consumption profiles at the critical value of R^* that marks the upper bound of the interval in which runs occur. There are two reasons for this discontinuity. The first is the usual cost of liquidating the safe asset, which we study in the preceding section. The second is the effect of asset sales on the price of the risky asset. By selling the asset, the bank drives down the price, thus handing a windfall profit to the speculators and a windfall loss to the depositors. This windfall loss is experienced as a discontinuous drop in consumption.

To understand the pricing of the risky asset when there is a bank run, we have to distinguish two different regimes. For intermediate values of R ($R_0 < R < R^*$) the asset price is determined by the speculators’ holdings of cash. Since one unit of the safe asset is worth r in the last period, the “fair” value of the bank’s holding of the risky asset is RX/r . However, the amount of cash in the market is insufficient to pay the “fair” value of the risky asset, so the price is determined by the ratio of the speculators’ cash to the bank’s

holding of the risky asset. This price is independent of R , which explains why consumption is independent of R in this interval. The consumption available at date 1 consists of the bank's holding of the safe asset, L , and the speculators' holding L_s . This is split among the early and late consumers so each receives $(L + L_s)/2$.

For small values of R ($R < R_0$) the "fair" value of the risky asset is less than the amount of cash in the market, so the asset price is equal to the "fair" value.

To sum up, introducing a market for the risky asset has a number of important implications. It allows the bank to liquidate all of its assets to meet the demands of the early withdrawers, but this has the effect of making the situation worse. First, because a bank run exhausts the bank's assets at date 1, a late consumer who waits until date 2 to withdraw will be left with nothing, so whenever there is a bank run, it will involve all the late consumers and not just some of them. Secondly, if the market for the risky asset is illiquid, the sale of the representative bank's holding of the risky asset will drive down the price, thus making it harder to meet the depositors' demands.

The all-or-nothing character of bank runs is, of course, familiar from the work of Diamond and Dybvig (1983). The difference is that in the present model bank runs are not "sunspot" phenomena: they occur only when there is no other equilibrium outcome possible. Furthermore, the deadweight cost of a bank run in this case is endogenous. In addition to the explicit cost of liquidation ($r > 1$), there is a cost resulting from suboptimal risk sharing. To make this clear, in this section we assume that $r = 1$. When the representative bank is forced to liquidate the risky asset, it sells the asset at a low price. This is a transfer of value to the purchasers of the risky asset, not an economic cost. The deadweight loss arises because the transfer occurs in bad states when the consumers' consumption is already low. In other words, the market is providing negative insurance.

Once again, intervention by the central bank will be helpful, but the optimal policy will consist of eliminating the deadweight costs of runs that arise from premature liquidation, rather than eliminating the runs themselves.

A. The Bank's Decision

The bank chooses a portfolio (L, X) and a promised payout \bar{c} at date 1 subject to the usual feasibility and incentive constraints. The standard deposit contract requires the bank to pay an early withdrawer \bar{c} at date 1, if this is possible, and to liquidate all of its assets otherwise. If there is no run, the early consumers receive $c_1(R) = \bar{c}$ and late consumers receive whatever is left, that is, $c_2(R) = L - \bar{c} + RX$. We do not allow a run unless it is unavoidable: when there are multiple equilibria corresponding to a given value of R , we assume that the equilibrium without runs is chosen. There are two possible cases to consider. A run will occur if and only if it is impos-

sible to pay the early consumers \bar{c} and pay the late consumers an amount at least as great as \bar{c} . As we have seen, partial runs are no longer possible in equilibrium. If there is a run, the bank must liquidate all of its assets at date 1 and all late consumers will join the run. In that case, $c_1(R) = c_2(R) = \frac{1}{2}(L + P(R)X)$, where $P(R)$ is the market price of the risky asset.

Let R^* be implicitly defined by the condition

$$\bar{c} = \frac{1}{2}(L + R^*X). \tag{54}$$

Then a run occurs if and only if $R < R^*$. To see this, suppose that $R < R^*$. If there is no run, then $c_1(R) + c_2(R) \geq 2\bar{c} > L + RX$, contradicting the feasibility conditions. (Since $P(R) \leq R$ selling assets will not help either.) Conversely, if $R \geq R^*$, then it is clearly possible to choose $c_1(R) = \bar{c}$ and $c_2(R) = L - \bar{c} + RX \geq \bar{c}$.

The bank's decision problem can be written as follows:

$$(P5) \left\{ \begin{array}{ll} \max & E[u(c_1(R) + u(c_2(R)))] \\ \text{s.t.} & \text{(i) } L + X \leq E; \\ & \text{(ii) } c_1(R) = \begin{cases} \bar{c} & \text{if } R \geq R^* \\ \frac{1}{2}(L + P(R)X) & \text{if } R < R^* \end{cases}; \\ & \text{(iii) } c_2(R) = \begin{cases} L - \bar{c} + RX & \text{if } R \geq R^* \\ \frac{1}{2}(L + P(R)X) & \text{if } R < R^* \end{cases}; \end{array} \right. \tag{55}$$

where $R^* = (2\bar{c} - L)/X$.

B. The Asset Market

To create a market for the risky asset, we introduce a group of risk-neutral *speculators*, who make direct investments in the safe and risky assets. Speculators consume only in the last period and their objective is to maximize the expected value of their portfolio at date 2. The speculators are all identical, so they can be replaced by a representative individual who has an initial wealth W_s and chooses a portfolio $(L_s, X_s) \geq 0$ subject to the budget constraint $L_s + X_s = W_s$.

The assumption that holdings of the two assets must be nonnegative is important here. Risk neutrality is often interpreted as meaning that an individual can have unboundedly negative consumption and hence supply unboundedly large amounts of the safe asset. Such an interpretation would make no sense here, because we want to emphasize the consequences of restricted liquidity in the market. In particular, it will *not* be true that the price of the risky asset will be equal to its expected present value using the safe return $r = 1$ as the discount rate. Since the safe asset cannot be shorted,

the speculators may not be able to buy as much of the risky asset as they would like. In other words, the equilibrium discount rate is higher than $r = 1$ because speculators are liquidity constrained. In any case, the price of the asset is determined by the amount of cash the speculators supply in exchange for it. We call this “cash-in-the-market” pricing.

Since the risky asset has a higher expected return than the safe asset, the safe asset will be held only if the speculators can make a profit by buying the risky asset at a low price at date 1. If bank runs occur with positive probability in equilibrium, speculators must hold a positive amount of the safe asset. If speculators do not have a positive holding of the safe asset at date 1, then when the banks try to sell the risky asset the price will fall to zero in some states, which means that any speculator who had held the safe asset would make an infinite profit. (Note the importance for this argument of the assumption that speculators cannot short the safe asset.) Thus, in an equilibrium where runs occur with positive probability, $L_s > 0$.

If W_s is large enough (as we assume in the sequel) speculators must also hold the risky asset. If not, $L_s = W_s$, and if the price of the risky asset is less than its “fair” value R at date 1, this amount of the safe asset will be supplied in exchange for the amount of the risky asset offered by the banks. Since $X \leq E$, the price must be at least W_s/E . So the speculators only make a profit if $R > W_s/E$. However, as we shall see, the banks only sell the risky asset when the return R is sufficiently small, so by choosing W_s large enough we can ensure that the speculators profit only if $L_s < W_s$. To sum up, there is no loss of generality in assuming that $L_s > 0$ and $X_s > 0$ in any equilibrium in which bank runs occur with positive probability.

The necessary and sufficient condition for holding both assets to be an optimum for the speculator is that

$$E \left[\max \left\{ 1, \frac{R}{P(R)} \right\} \right] = E[R], \quad (56)$$

where $P(R)$ is the price of the risky asset at date 1. In other words, the expected return from holding the safe asset and buying the risky asset at date 1 when the price of the risky asset falls below R is equal to the expected return from a buy-and-hold strategy—that is, buying the risky asset at date 0 and holding it until date 2. Note that $P(R) \leq R$ for all values of R , because $P(R) > R$ implies that no one is willing to hold the risky asset and this cannot be an equilibrium. Therefore, we do not have to consider the possibility of switching from the risky to the safe asset at date 1 and the condition above reduces to

$$E \left[\frac{1}{P(R)} \right] = E[R]. \quad (57)$$

C. Equilibrium

An equilibrium for the model with an asset market consists of a portfolio (L_s, X_s) for the representative speculator, a price function $P(R)$ that satisfies the no-arbitrage condition (57), and a deposit contract $((L, X), \bar{c})$ that solves the decision problem (P5) given (L_s, X_s) and $P(R)$.

In the asset market, our earlier discussion shows that there are two cases to be considered: either $R \geq R^*$ and there is no run, or $R < R^*$ and there is a run. If there is no run, and hence no sale of assets in the market, the safe asset must have the same one-period return as the risky asset, so $P(R) = R$. Conversely, if there is a sale of assets, the representative bank supplies X inelastically. If $L_s > RX$, then the equilibrium price must be $P(R) = R$. If the price were lower, everyone would want to hold the risky asset and there would be an excess supply of the safe asset. If the price were higher, no one would want to hold the risky asset and there would be an excess supply of the risky asset. Similarly, if $L_s < RX$ then the price of the risky asset must be $P(R) = L_s/X$. At this price, the speculators supply the safe asset inelastically in exchange for the risky asset and the market clears because $L_s = P(R)X$. At any other price, this market-clearing condition is violated. (If $P(R) = R$, speculators may supply less than L_s , but this too violates market clearing.) Let R^0 be implicitly defined by the condition

$$L_s = R^0 X. \tag{58}$$

Then

$$P(R) = \begin{cases} R & \text{for } R \leq R^0 \text{ and } R \geq R^* \\ L_s/X & \text{for } R^0 < R < R^*. \end{cases} \tag{59}$$

In other words, the price collapses only if the return is low enough to provoke a run but not so low that the market is liquid enough to absorb the asset at its “fair” value. Figure 5 illustrates the equilibrium allocation for bank depositors.

In the numerical example it will be assumed that the wealth of the speculators $W_s = 1$ and that the other parameters are as in the standard case with $r = 1$. The optimal contract for depositors has $(L, X) = (1.06, 0.94)$, $R^0 = 0.25$, $R^* = 1.13$, with $P(R) = 0.25$ for $R^0 < R < R^*$ and $E[U(c_1, c_2)] = 0.09$. For the speculators $(L_s, X_s) = (0.24, 0.76)$ and their expected utility is $EU_s = 1.5$. Note that the depositors are significantly worse off in this equilibrium compared to the (P1) allocation where $E[U(c_1, c_2)] = 0.25$ and are only slightly better than in the case where the bank’s portfolio is such that no runs occur (as in Figure 2), in which case $E[U(c_1, c_2)] = 0.08$.

D. Optimal Monetary Policy

As a benchmark for judging the efficiency of the equilibrium with asset markets, we choose the allocation that solves (PI). This allocation can be implemented without relying on the asset market at all. It may not be the best the central bank can do, whatever one chooses to define as the “best,” but it provides a lower bound for the second-best, and for some parameter values we can show that it is significantly better than the equilibrium allocation. The essential idea behind the policy that implements the solution to (PI) is similar to the monetary intervention described in Section II, but here the central bank is interpreted as supporting the risky asset’s price, rather than making an unsecured loan to the bank. Specifically, the central bank enters into a repurchase agreement (or a collateralized loan) with the representative bank, whereby the bank sells some of its assets to the central bank at date 1 in exchange for money and buys them back for the same price at date 2. By providing liquidity in this way, the central bank ensures that the representative bank does not suffer a loss by liquidating its holdings of the risky asset prematurely.

As before, we assume that the standard deposit contract promises depositors a fixed amount of money D in the middle period and pays out the remaining value of the assets in the last period. The price level at date t in state R is denoted by $p_t(R)$ and the *nominal* price of the risky asset at date 1 in state R is denoted by $P(R)$. We want the risky asset to sell for its “fair” value, so we assume that $P(R) = p_1(R)R$. At this price, the safe and risky assets are perfect substitutes. Let (X, L) be the portfolio corresponding to the solution of (PI) and let $(c_1(R), c_2(R))$ be the corresponding consumption allocations. For large values of R , we may have $c_1(R) = L < c_2(R) = RX$; for smaller values we may have $c_1(R) = c_2(R) = \frac{1}{2}(L + RX)$. Implementing this allocation requires introducing contingencies through price variation: $p_1(R)c_1(R) = D < p_2(R)c_2(R)$ for $R > \bar{R}$ and $p_1(R)c_1(R) = D = p_2(R)c_2(R)$ for $R < \bar{R}$. These equations determine the values of $p_1(R)$ and $p_2(R)$ uniquely. It remains only to determine the value of sales of assets and the size of the bank run.

In the event of a bank run, only the late consumers who withdraw early will end up holding cash, since the early consumers want to consume their entire liquidated wealth immediately. If $\alpha(R)$ is the fraction of late consumers who withdraw early, then the amount of cash injected into the system must be $\alpha(R)D$. For simplicity, we assume that the amount of cash injected is a constant M and this determines the “size” of the run $\alpha(R)$. Since the safe asset and the risky asset are perfect substitutes at this point, it does not matter which assets the representative bank sells as long as the nominal value equals M . The representative bank enters into a repurchase agreement under which it sells assets at date 1 for an amount of cash equal to M and repurchases them at date 2 for the same cash value.

At the prescribed prices, speculators will not want to hold any of the safe asset, so $L_s = 0$ and $X_s = W_s$.

It is easy to check that all the equilibrium conditions are satisfied: depositors and speculators are behaving optimally at the given prices and the feasibility conditions are satisfied.

THEOREM 5: *The central bank can implement the solution to problem (P1) by entering into a repurchase agreement with the representative bank at date 1. Given the allocation $\{(L, X), c_1(R), c_2(R)\}$, corresponding to the solution of (P1), the equilibrium values of prices are given by the conditions $p_1(R)c_1(R) = D < p_2(R)c_2(R)$ for $R > \bar{R}$, and $p_1(R)c_1(R) = D = p_2(R)c_2(R)$ for $R < \bar{R}$. There is a fixed amount of money M injected into the economy in the event of a run and the fraction of late withdrawers who “run” satisfies $\alpha(R)D = M$. The price of the risky asset at date 1 satisfies $p_1(R)R = P(R)$ and the optimal portfolio of the speculators is $(L_s, X_s) = (0, W_s)$.*

Although Theorem 5 shows that central bank intervention can achieve the optimal solution to (P1), it does not show that this is strictly better than the market equilibrium, since the market equilibrium allows for possibilities, such as liquidating the risky asset at date 1, that are not available in (P1). However, it is easy to show that the solution to (P1) is (strictly) Pareto-preferred to the equilibrium of the model with asset markets.

COROLLARY 5.1: *The solution to (P1), implemented by the policy described in Theorem 5, is Pareto-preferred to the laissez-faire equilibrium outcome of the model with asset markets.*

Proof: See the Appendix.

Theorem 5 and its corollary can be illustrated with the standard numerical example. To illustrate how the incentive-efficient allocation (P1) can be implemented in the context of the numerical example with $r = 1$, recall that the optimum has $(L, X) = (1.19, 0.81)$, $\bar{R} = 1.47$, and $E[U(c_1, c_2)] = 0.25$. Suppose $D = 1.19$. For $R \geq \bar{R} = 1.47$ then $p_1(R) = p_2(R) = 1$. For $R < \bar{R} = 1.47$ the price levels at the two dates depend on the level of R . To illustrate, suppose $R = 1$. In that case $c_1(1) = c_2(1) = 1$ so $p_1(1) = p_2(1) = 1.19$. Similarly for other values of R . The lower the value of R , the higher $p_t(R)$, so that consumption is lowered by raising the price level. Also $P(R) = 1.19$. The fraction of late consumers who withdraw from the bank and hold money will be determined by M . Suppose $M = 1$, then $\alpha(R) = 1/1.19 = 0.84$. For the speculators $(L_s, X_s) = (0, 1)$ and their expected utility is $EU_s = 1.5$. The equilibrium with central bank intervention is clearly Pareto-preferred to the market equilibrium without intervention as indicated by the corollary.

IV. Summary

Empirical evidence provided by Gorton (1988) suggests that banking panics in the United States during the National Banking Era were not “sunspot” phenomena but rather the result of the business cycle. When depositors

observe leading economic indicators and perceive that a bank's receipts are going to be low, there is a run. This paper develops a simple model of this phenomenon and uses it to identify the optimal policy toward runs. It shows that financial crises can be optimal if the return to the safe asset is the same inside and outside the banking system. The reason is that the optimal allocation of resources involves imposing some risk on people who withdraw early. Allowing bank runs can be an efficient way of doing this. In this case, central-bank and other government policies that eliminate runs lower the welfare of depositors.

If the return to the safe asset is higher within the banking system than outside, so that bank runs are costly, runs alone cannot achieve the optimal allocation of risk. However, a monetary intervention by the central bank can allow the first-best to be achieved.

Finally, if the risky asset can be sold in an asset market, bank runs may be costly even when the return on the safe asset is the same inside and outside the banking system. Runs force the banks to liquidate their assets when prospects are bad. Simultaneous liquidation drives asset prices down and allows speculators in the asset market to profit. There is, in effect, negative insurance. Central bank intervention that prevents the collapse in prices in the asset market can allow a Pareto improvement.

Appendix

Proof of Theorem 1: If we ignore the incentive-compatibility constraint, the optimal risk-sharing problem becomes:

$$(PI)' \begin{cases} \max & E[u(c_1(R)) + u(c_2(R))] \\ \text{s.t.} & \text{(i) } L + X \leq E; \\ & \text{(ii) } c_1(R) \leq L; \\ & \text{(iii) } c_1(R) + c_2(R) \leq L + RX. \end{cases} \quad (A1)$$

A necessary condition for a solution to $(PI)'$ is that, for each value of R , the consumption levels $c_1(R)$ and $c_2(R)$ solve the problem

$$\begin{aligned} \max & \quad u(c_1(R)) + u(c_2(R)) \\ \text{s.t.} & \quad \text{(ii) } c_1(R) \leq L; \\ & \quad \text{(iii) } c_1(R) + c_2(R) \leq L + RX. \end{aligned} \quad (A2)$$

The necessary Kuhn–Tucker conditions imply

$$u'(c_1(R)) \geq u'(c_2(R)), \quad (A3)$$

with strict equality if $c_1(R) < L$. In any case, this implies that $c_1(R) \leq c_2(R)$, with strict equality if $c_1(R) < L$, so the incentive constraints (iv) will be satisfied automatically. Thus, a solution to $(P1)'$ is also a solution to the original problem $(P1)$.

Since we know that $c_1(R) = c_2(R)$ whenever $c_1(R) < L$, there are two regimes to be considered. Either $c_1(R) = L$ and, hence, $c_2(R) = RX$ or $c_1(R) = c_2(R) = \frac{1}{2}(RX + L)$. The first case can occur if and only if $L \leq RX$, so the optimal risk-sharing allocation must satisfy

$$c_1(R) = c_2(R) = \frac{1}{2} (RX + L) \quad \text{if } L \geq RX, \tag{A4}$$

and

$$c_1(R) = L, c_2(R) = RX \quad \text{if } L \leq RX. \tag{A5}$$

This allows us to write the optimal risk-sharing problem more compactly as follows:

$$\max \int_0^{\bar{R}} 2u\left(\frac{RX + L}{2}\right) f(R) dR + \int_{\bar{R}}^{\infty} (u(L) + u(RX)) f(R) dR \tag{A6}$$

$$\text{s.t. } L + X \leq E,$$

where $\bar{R} \equiv L/X$ is the value of the return on the risky asset at which the liquidity constraint begins to bind. Note that so far we have not established that the critical value of \bar{R} belongs to the support of R .

It remains to characterize the optimal portfolio. We first rule out two extreme cases. Suppose that $X = 0$. Then it is clear that $c_1(R) = c_2(R) = E/2$ and $\bar{R} = \infty$. This will be optimal only if $L = E$ maximizes

$$u(L/2) + E[u(R(E - L) + L/2)], \tag{A7}$$

and the first-order condition for this is

$$u'(E/2)/2 + u'(E/2)\left(\frac{1}{2} - E[R]\right) \geq 0, \tag{A8}$$

which implies $E[R] \leq 1$, contradicting one of our maintained assumptions.

Next suppose that $L = 0$. Then $c_1(R) = 0 \leq c_2(R) = RE$. For this to be an optimal choice, it must be the case that $X = E$ maximizes

$$u(E - X) + E[u(RX)], \tag{A9}$$

and the necessary first-order condition for this is

$$u'(0) \leq E[u'(RE)R], \quad (\text{A10})$$

which contradicts another of our maintained assumptions. Thus any optimal portfolio must satisfy $L > 0$ and $X > 0$.

Returning to the compact form of the risk-sharing problem above, we see that necessary conditions for an interior solution are:

$$\int u'(c_1(R))f(R)dR = \lambda \quad (\text{A11})$$

and

$$\int u'(c_2(R))Rf(R)dR = \lambda, \quad (\text{A12})$$

where λ is the Lagrange multiplier of the constraint $L + X = E$. Under the strict concavity of $u(\cdot)$, these first-order conditions uniquely determine the optimal values of L and X , which in turn determine \bar{R} , $c_1(R)$, and $c_2(R)$ through the relations described above. Q.E.D.

Proof of Corollary 5.1: We need to show that the solution to (P1), implemented by the policy described in Theorem 5, is strictly Pareto-preferred to the laissez-faire equilibrium outcome of the model with asset markets. Let (X_s, L_s) be the speculators' equilibrium portfolio, $P(R)$ the equilibrium asset-price function, and $\{(L, X), (c_1(R), c_2(R))\}$ the equilibrium deposit contract. The consumption functions solve

$$\begin{aligned} \max \quad & u(c_1(R)) + u(c_2(R)) \\ \text{s.t.} \quad & c_1(R) \leq L; \\ & c_1(R) + c_2(R) \leq L + RX, \end{aligned} \quad (\text{A13})$$

if $R \geq R^*$ and

$$\begin{aligned} \max \quad & u(c_1(R)) + u(c_2(R)) \\ \text{s.t.} \quad & c_1(R) + c_2(R) \leq L + P(R)X, \end{aligned} \quad (\text{A14})$$

if $R < R^*$. Note that $c_1(R) \leq \bar{c} \leq L$ for all values of R , so there is no loss of generality in combining these two problems and treating $(c_1(R), c_2(R))$ as the solution of

$$\begin{aligned}
& \max && u(c_1(R)) + u(c_2(R)) \\
& \text{s.t.} && c_1(R) \leq L; \\
& && c_1(R) + c_2(R) \leq L + \min\{P(R), R\}X,
\end{aligned} \tag{A15}$$

for all values of R .

Now suppose that the functions $c_1^*(R)$ and $c_2^*(R)$ solve the problem

$$\begin{aligned}
& \max && u(c_1(R)) + u(c_2(R)) \\
& \text{s.t.} && c_1(R) \leq L; \\
& && c_1(R) + c_2(R) \leq L + RX,
\end{aligned} \tag{A16}$$

for all values of R . Then, since $P(R) \leq R$, we must have $c_i(R) \leq c_i^*(R)$, for all R and $i = 1, 2$. The consumption functions $c_1^*(R)$ and $c_2^*(R)$ are feasible for (P1) if the portfolio (L, X) is chosen, so it follows that the solution to (P1) must be at least as good as the equilibrium outcome and strictly preferred by the depositors if the equilibrium involves selling the risky asset at a price $P(R) < R$ with positive probability.

The speculators get the same expected utility in either case, so we have proved the corollary. Q.E.D.

REFERENCES

- Alonso, Irasema, 1996, On avoiding bank runs, *Journal of Monetary Economics* 37, 73–87.
- Bensaid, Bernard, Henri Pages, and Jean-Charles Rochet, 1996, Efficient regulation of bank's solvency, Unpublished working paper, Institut d'Economie Industrielle, Toulouse.
- Bernanke, Ben, 1983, Nonmonetary effects of the financial crisis in the propagation of the Great Depression, *American Economic Review* 73, 257–263.
- Bernanke, Ben, and Mark Gertler, 1989, Agency costs, net worth, and business fluctuations, *American Economic Review* 79, 14–31.
- Bhattacharya, Suddipto, and Douglas Gale, 1987, Preference shocks, liquidity and central bank policy, in William Barnett and Kenneth Singleton, eds.: *New Approaches to Monetary Economics* (Cambridge University Press, New York, N.Y.).
- Calomiris, Charles, and Gary Gorton, 1991, The origins of banking panics, models, facts, and bank regulation, in Glenn Hubbard, ed.: *Financial Markets and Financial Crises* (University of Chicago Press, Chicago, Ill.).
- Calomiris, Charles, and Charles Kahn, 1991, The role of demandable debt in structuring optimal banking arrangements, *American Economic Review* 81, 497–513.
- Champ, Bruce, Bruce Smith, and Stephen Williamson, 1996, Currency elasticity and banking panics: Theory and evidence, *Canadian Journal of Economics* 29, 828–864.
- Chari, V., 1989, Banking without deposit insurance or bank panics: Lessons from a model of the U.S. National Banking System, *Federal Reserve Bank of Minneapolis Quarterly Review* 13(Summer), 3–19.
- Chari, V., and Ravi Jagannathan, 1988, Banking panics, information, and rational expectations equilibrium, *Journal of Finance* 43, 749–60.
- Diamond, Douglas, 1997, Liquidity, banks and markets, *Journal of Political Economy* 105, 928–956.
- Diamond, Douglas, and Philip Dybvig, 1983, Bank runs, deposit insurance, and liquidity, *Journal of Political Economy* 91, 401–419.
- Gorton, Gary, 1988, Banking panics and business cycles, *Oxford Economic Papers* 40, 751–781.

- Hellwig, Martin, 1994, Liquidity provision, banking, and the allocation of interest rate risk, *European Economic Review* 38, 1363–1389.
- Jacklin, Charles J., and Sudipto Bhattacharya, 1988, Distinguishing panics and information-based bank runs: Welfare and policy implications, *Journal of Political Economy* 96, 568–592.
- Kindleberger, Charles, 1978, *Manias, Panics, and Crashes: A History of Financial Crises* (Basic Books, New York, N.Y.).
- Lindgren, Carl-Johan, Gillian Garcia, and Matthew Saal, 1996, *Bank Soundness and Macroeconomic Policy* (International Monetary Fund, Washington, D.C.).
- Mitchell, Wesley, 1941, *Business Cycles and Their Causes* (University of California Press, Berkeley, Calif.).
- Postlewaite, Andrew, and Xavier 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.
- Sprague, Oliver, 1910, *A History of Crises Under the National Banking System* (U.S. Government Printing Office, Washington, D.C.).
- Wallace, Neil, 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.