What Is Systemic Risk, and Do Bank Regulators Retard or Contribute to It?

GEORGE G. KAUFMAN AND KENNETH E. SCOTT

O

e of the most feared events in banking is the cry of systemic risk. It matches the fear of a cry of “fire!” in a crowded theater or other gatherings. But unlike fire, the term systemic risk is not clearly defined. Moreover, unlike firefighters, who rarely are accused of sparking or spreading rather than extinguishing fires, bank regulators at times have been accused of contributing to, albeit unintentionally, rather than retarding systemic risk. In this article, we discuss the alternative definitions and sources of systemic risk, review briefly the historical evidence of systemic risk in banking, describe how participants in financial markets traditionally have protected themselves from systemic risk, evaluate the regulations that bank regulators have adopted to reduce both the probability of systemic risk and the damage it causes when it does occur, and make recommendations for efficiently curtailing systemic risk in banking.

Systemic Risk

Systemic risk refers to the risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by comovements (correlation) among most or all the parts. Thus, systemic risk in banking is evidenced by high correlation and clustering of bank failures in a single coun-
try, in a number of countries, or throughout the world. Systemic risk also may occur in other parts of the financial sector—for example, in securities markets as evidenced by simultaneous declines in the prices of a large number of securities in one or more markets in a single country or across countries. Systemic risk may be domestic or transnational.

Definitions of Systemic Risk in Banking

The precise meaning of systemic risk is ambiguous; it means different things to different people. A search of the literature reveals three frequently used concepts. The first refers to a “big” shock or macroshock that produces nearly simultaneous, large, adverse effects on most or all of the domestic economy or system. Here, systemic “refers to an event having effects on the entire banking, financial, or economic system, rather than just one or a few institutions” (Bartholomew and Whalen 1995, 4). Likewise, Frederic Mishkin defines systemic risk as “the likelihood of a sudden, usually unexpected, event that disrupts information in financial markets, making them unable to effectively channel funds to those parties with the most productive investment opportunities” (1995, 32). How the transmission of effects from a macroshock to individual units, or contagion, occurs and which units are affected are generally unspecified. Franklin Allen and Douglas Gale (1998) model one process through which macroshocks can ignite bank runs.

The other two definitions focus more on the microlevel and on the transmission of the shock and potential spillover from one unit to others. For example, according to the second definition, systemic risk is the “probability that cumulative losses will accrue from an event that sets in motion a series of successive losses along a chain of institutions or markets comprising a system. . . . That is, systemic risk is the risk of a chain reaction of falling interconnected dominos” (Kaufman 1995a, 47). This definition is consistent with that of the Federal Reserve (the Fed). In the payments system, systemic risk may occur if an institution participating on a private large-dollar payments network were unable or unwilling to settle its net debt position. If such a settlement failure occurred, the institution’s creditors on the network might also be unable to settle their commitments. Serious repercussions could, as a result, spread to other participants in the private network, to other depository institutions not participating in the network, and to the nonfinancial economy generally. (Board of Governors of the Federal Reserve System 2001, 2)

Likewise, the Bank for International Settlements (BIS) defines systemic risk as “the risk that the failure of a participant to meet its contractual obligations may in turn cause other participants to default with a chain reaction leading to broader financial difficulties” (BIS 1994, 177). These definitions emphasize correlation with cau-
What Is Systematic Risk?

sation, and they require close and direct connections among institutions or markets. When the first domino falls, it falls on others, causing them to fall and in turn to knock down others in a chain or “knock-on” reaction. Governor E. A. J. George of the Bank of England has described this effect as occurring “through the direct financial exposures which tie firms together like mountaineers, so that if one falls off the rock face others are pulled off too” (1998, 6). For banks, this effect may occur if Bank A, for whatever reason, defaults on a loan, deposit, or other payment to Bank B, thereby producing a loss greater than B’s capital and forcing it to default on payment to Bank C, thereby producing a loss greater than C’s capital, and so on down the chain (Crockett 1997). Banks, especially within a country, tend to be connected closely through interbank deposits and loans. Note that in this second definition, unlike in the first macroshock definition, only one bank need be exposed in direct causation to the initial shock. All other banks along the transmission chain may be unexposed to this shock. The initial bank failure sets off the chain or knock-on reaction.

The smaller a bank’s capital-asset ratio—the more leveraged it is—the more likely it is that it will be driven into insolvency by insolvencies of banks located earlier on the transmission chain and will transmit losses to banks located later on the chain. What makes direct-causation systemic risk in financial sectors particularly frightening to many is both the lightning speed with which it occurs and the belief that it can affect economically solvent (innocent) as well as economically insolvent (guilty) parties, so there is scarcely any way to protect against its damaging effects.

A third definition of systemic risk also focuses on spillover from an initial exogenous external shock, but it does not involve direct causation and depends on weaker and more indirect connections. It emphasizes similarities in third-party risk exposures among the units involved. When one unit experiences adverse effects from a shock—say, the failure of a large financial or nonfinancial firm—that generates severe losses, uncertainty is created about the values of other units potentially also subject to adverse effects from the same shock. To minimize additional losses, market participants will examine other units, such as banks, in which they have economic interests to see whether and to what extent they are at risk. The more similar the risk-exposure profile to that of the initial unit economically, politically, or otherwise, the greater is the probability of loss, and the more likely it is that participants will withdraw funds as soon as possible. This response may induce liquidity problems and even more fundamental solvency problems. This pattern may be referred to as a “common shock” or “reassessment shock” effect and represents correlation without direct causation (indirect causation).

Because information either on the causes or the magnitude of the initial shock or on the risk exposures of each unit potentially at risk is not generally available immediately or accurately and is not without cost, and because analysis of the information is not immediate or free, participants generally require time and resources to sort out the identities of the other units at risk and the magnitudes of any potential losses. Moreover, in banking, as credit markets deteriorate, the quality of private and public
information available also deteriorates as the cost of accurate information increases and as uncertainty increases further. Because many of the participants are risk averse and would rather be safe than sorry, they quickly will transfer funds, at least temporarily during the period of confusion and sorting out, to well-recognized safe or at least safer units without waiting for the final analysis. In addition, in periods of great uncertainty and stress, market participants tend increasingly to make their portfolio adjustments in quantities (runs) rather than in prices (interest rates). That is, at least temporarily, they will not lend at almost any rate. Thus, there is likely to be an immediate flight or run to quality away from all units that appear potentially at risk, regardless of whether further and more complete analysis might identify them ex post as having similar exposures that actually put them at risk of insolvency. At this stage, common-shock contagion appears indiscriminate, potentially affecting more or less the entire universe and reflecting a general loss of confidence in all units. Solvent parties are not differentiated from insolvent. Because these runs are concurrent and widespread, such behavior by investors is often referred to as “herding” behavior.

The runs are likely to exert strong downward pressure on the prices (upward pressures on interest rates) of the securities of affected financial institutions and markets. Any resulting liquidity problems are likely to spill over temporarily to banks not directly affected by the initial shock. Thus, the initial domino does not fall directly on other dominos, but its fall causes players to examine nearby dominos to see whether they are subject to the same destabilizing forces that caused the initial domino to fall. Broad contagion is likely to occur during such sorting-out or reassessment periods.

At a later date, after the sorting-out process is complete, some or all of these flows affecting solvent banks may be corrected or reversed. Nevertheless, during the sorting-out period, the fire sale–driven changes in both financial quantities (flows) and prices (interest rates) are likely to overshoot their ultimate equilibrium levels because of an uncertainty discount and thus to intensify the liquidity problems, particularly for more vulnerable units (Kaminsky and Schmukler 1999). However, the more frequent banking crises are, the more likely are market participants to become both better prepared and better informed, the sorting-out and liquidity-problem periods to be shorter, and the duration of any overshooting to be briefer.

A distinction is often made between rational or information-based, directly or indirectly caused systemic risk and irrational, noninformation-based, random, or “pure” contagious systemic risk (Aharony and Swary 1996; Kaminsky and Reinhart 1998; Kaufman 1994). Rational or informed contagion assumes that investors (depositors) can differentiate among parties on the basis of their fundamentals. Random contagion, based on actions by uninformed agents, is viewed as more frightening and dangerous because it does not differentiate among parties, affecting solvent as well as insolvent parties, and therefore is likely to be both broader and more diffi-

---

1. An interesting theoretical explanation of such investor behavior is developed in Herring and Wachter 1999.
cult to contain. Thus, Governor George (1998, 6) of the Bank of England considers systemic risk as exceptionally costly because “the danger that a failure of one financial business may infect other, otherwise healthy, businesses.” Direct, knock-on contagion is perceived as knocking over solvent as well as insolvent banks on the transmission chain. Common-shock contagion systemic risk is likely to affect solvent banks immediately during the sorting-out period, although in time investors and depositors will sort these banks out from the insolvent banks. Thus, the empirical borderline between rational and irrational contagion is fuzzy and depends in part on the time horizon applied. Likewise, the definition of solvent and insolvent is not always clear and precise. Solvent parties may be defined as units that are perceived widely to be economically well behaved—that is, banks that are perceived to be economically sound and not overly leveraged. In contrast, insolvent banks are those perceived as insolvent or solvent but near insolvency or excessively leveraged.

**Dangers of Systemic Risk**

Both the chain-reaction and the common-shock concepts of systemic risk involve speedy contagion and require some actual or perceived direct or indirect connection among the parties at risk (Kaufman 1994). Banks are connected directly through interbank deposits, loans, and payment-system clearings and indirectly through serving the same or similar deposit or loan markets. In addition, to the extent that banks operate across national borders, they link the countries in which they operate. Thus, an adverse shock that generates losses at one bank large enough to drive it into insolvency may transmit the shock to other banks along the transmission chain. Moreover, adverse shocks in the financial sector appear to be transmitted more rapidly than similar shocks in other sectors. Both theory and evidence suggest that the probability, strength, and breadth of any contagious systemic risk are greater for banking, the larger and more significant is the bank experiencing the initial shock. It follows that the transmission and danger of systemic risk are likely to differ depending on the strength of the initial shock and on the characteristics of the bank initially affected.

In the absence of guarantees, units on the transmission chain reasonably may be expected to attempt to protect themselves from losses caused by shocks. For banks, this attempt requires them to charge higher interest rates on riskier investments, to monitor their counterparties carefully, to require more and better collateral, and to have sufficient capital to absorb any losses from their association with an infected bank or from runs by their depositors. Jean-Charles Rochet and Jean Tirole (1996) model such a structure. In general, for the initial shock to be transmitted successfully and to bring down other banks, losses must exceed capital at each bank along the chain.

---

2. Because no bank is perceived to be safe, runs on the entire banking system into currency lead to a decline in aggregate bank reserves and, unless offset by the central bank, to a multiple contraction in aggregate money and credit. See Davis 1995 and Diamond and Dybvig 1983.
Banks with sufficient capital to absorb the transmitted losses will remain solvent, although they may be weakened, and thus will stop the cascading. The amount of capital required to remain solvent depends on the exposure of a particular bank to other units and on the expectations regarding the magnitude of any shocks. Both the exposure and the expectations vary among banks and through time for any one bank. Nevertheless, ceteris paribus, the more leveraged are the banks or other institutions, the smaller is the adverse shock required to drive a bank or other institution into insolvency, and the greater is the likelihood that any losses will be passed along the transmission chain. In addition, the faster the transmission occurs, the more difficult it is for units to develop their protection after the shock has occurred, and the more important it is for them to have sufficient protection in place beforehand. In these regards, the financial sector differs from most other sectors, where the transmission of adverse shocks is slower and units generally have time to act to protect themselves after the initial shock has occurred.

Random contagious systemic risk is considered particularly dangerous and undesirable because it spills over to and damages both banks that are perceived to be economically solvent and those that are considered insolvent. Although it is relatively easy to distinguish the solvent from the insolvent after the crisis, it can be difficult in practice to do so before a crisis. Ex ante information is frequently not sufficiently available, timely, or reliable to make the distinction with much confidence. Banks, often with the active assistance and encouragement of their governments, frequently fail to disclose relevant information and, especially as they approach insolvency, tend to provide insufficient reserves for loan losses and to use questionable and sometimes even fraudulent accounting procedures to inflate their reported capital ratios.

**Historical Evidence of Contagious Systemic Risk**

Clusterings of bank failures occur frequently, but do they reflect systemic risk? The empirical evidence depends on the definition of systemic risk used. Almost tautologically, systemic risk is observed most frequently when it is defined as a big, broad shock. As noted earlier, however, this definition is silent on the existence or transmission of contagion. Common-shock systemic risk, particularly in the short term, appears to be more frequent than chain-reaction systemic risk. Systemic risk, when it does occur, appears both to be rational and to be confined primarily to “insolvent” institutions and not randomly to affect solvent banks fatally (Kaufman 2000a).

With respect to banks, at least in the United States, there is little if any evidence of contagious systemic risk that causes economically solvent banks to become economically or legally insolvent, either before or after the introduction of federal government guarantees and insurance (Kaufman 1994). U.S. banks have been studied most thoroughly because of their large number, good historical data, and minimum government ownership or control. The evidence indicates that problems at one bank or at a group of banks do spill over to other banks in general, but almost exclusively
only to banks with the same or similar portfolio-risk exposures and subject to the same shock. There is little if any empirical evidence that the insolvency of an individual bank directly causes the insolvency of economically solvent banks or that bank depositors run on economically solvent banks very often or that, when they do, they drive these banks into insolvency.

**Potential Exposure**

A recent study simulated the likelihood of direct causation or knock-on contagion in the United States through Fed funds transactions and other interbank exposures for the period February–March 1998 (Furfine 2003). These funds are de jure uninsured and, since the Depositor Preference Act of 1993, are subordinated to all domestic deposits. The study found that if a high loss rate of 40 percent is assumed, well above average bank loss rates experienced even in the crises of the 1930s and 1980s, the failure of the largest debtor bank in the U.S. Fed funds market would cause the economic insolvency of only two to six other banks holding less than 1 percent of total bank assets. The failure of smaller debtor banks would have lesser effects. If the two largest debtor banks failed at the same time, fewer than ten other banks would fail. All other banks held sufficient capital to absorb the losses. If the assumed loss rate were reduced to 5 percent, approximately that experienced in the Continental Illinois Bank failure in 1984, no other banks would fail.

The results did not change much when total interbank exposures were simulated. The simultaneous failure of the largest two debtor banks causes more than fifteen other banks with more than 3 percent of total bank assets to fail only when the loss rate exceeds 65 percent. Such a loss rate would be exceedingly high for large resolved banks in the United States. Even at the height of the banking crises in the 1980s, when regulators regularly forbore and delayed resolving insolvencies until after significant runs by uninsured depositors effectively had stripped the banks of their best assets and had increased losses as a percent of the remaining assets, the losses at large commercial banks rarely exceeded 10 percent of assets (Kaufman 1995b). At these loss rates, Furfine’s (2003) simulations predict only minor knock-on effects. Moreover, these results overstate the damage to other banks because they assume failure when only tier 1 (basically equity capital), rather than total capital, including tier 2 (basically subordinated debt and limited loan-loss reserves), is depleted. Similarly, simulation studies of the Swiss and Italian domestic interbank markets also report a relatively small “threat to financial market stability” from default by one bank (Angelini, Maresca, and Russo 1996; Sheldon and Maurer 1998).

**Historical Experience**

*Chain Reactions.* The evidence does not differ for actual failures. When the Continental Illinois Bank, the seventh biggest bank in the United States at the time, with
assets of more than $32 billion, failed in mid-1984, it was the largest correspondent bank in the country. Nearly 2,300 other banks held deposits at or loaned funds to the Continental. Because the Federal Deposit Insurance Corporation (FDIC) fully protected all creditors when it failed, no bank suffered any losses. But what would have happened if all creditors had not been protected fully? Not very much! Some 1,325 banks had exposure of less than $100,000 and thus were insured fully by the FDIC. Although the remainder had some risk exposure, a study by the staff of the House Banking Committee found that had Continental’s loss been as large as sixty cents on the dollar (a recovery rate on assets of only 40 percent), which was more than ten times either the estimated loss or the actual loss as of the time of its resolution, only twenty-seven banks would have suffered losses in excess of their reported capital and thus would have become insolvent (U.S. Congress 1984). These losses would have totaled only $137 million. Another fifty-six banks would have suffered losses equal to between 50 and 99 percent of their total capital, in an amount totaling $237 million. If the Continental loss had been smaller, say, ten cents on the dollar—still more than twice the actual loss—no other bank would have suffered a loss greater than its capital, and only two banks would have suffered losses in excess of 50 percent of their capital. Banks apparently had acted to protect themselves by limiting their uninsured exposures relative to their capital and by monitoring their positions carefully. Given the relatively small size of the loss, it is also unlikely that any of the banks with $100,000 or less in deposits at the Continental, which were fully insured, would have failed had those deposits been uninsured because they maintained capital well in excess of that amount.

Spillover losses to U.S. and some foreign banks when the Herstatt Bank in Germany failed and was closed by the authorities in 1974 are cited often as evidence of systemic risk. Indeed, Herstatt risk has become a generic term to describe cross-border settlement risk for banks. Losses were suffered primarily by banks that had entered into foreign-exchange transactions with Herstatt, and they occurred not so much because of losses at Herstatt as because the exchange in payments between these banks and Herstatt was not simultaneous, owing to differences in time zones. The counterparty banks paid the mark side of the transactions to Herstatt during its working day, but the German authorities closed the bank at the close of business in Germany before Herstatt was scheduled to make the corresponding dollar payments to the counterparty banks during their business day, primarily in New York, many hours later (Eisenbeis 1995). If the German authorities had waited until the end of the business day in New York before closing the Herstatt bank, the counterparty losses would have been reduced greatly or perhaps avoided altogether. Instead, they would have accrued to Herstatt depositors and the German bank deposit insurance fund. Thus, much of the spillover from the Herstatt Bank to other, primarily foreign, banks from these transactions represents more of a government risk than a market risk. Even so, no other bank failed as a result of this debacle.
Common-Shock Reassessment. Except for fraud, clustered bank failures in the United States almost always are triggered by adverse conditions in the regional or national macroeconomies or by the bursting of asset-price bubbles, especially in real estate, and not by exogenous “sunspot” effects (Allen and Gale 1998; Benston and Kaufman 1995; Kaufman 1999). Banks fail because of exposure to a common shock, such as a depression in agriculture, real estate, or oil prices (Cottrell, Lawlor, and Wood 1995), not because of direct spillover from other banks without themselves being exposed to the shock. Post mortems of failed U.S. banks indicate that in almost every instance since the introduction of deposit insurance, the bank was already economically insolvent for many months and, on occasion in the 1980s, even for years before it was resolved by the regulators (Kaufman 1995b).

A study of national bank failures from 1865 to 1936, shortly after the introduction of federal deposit insurance in 1933, reported that the most cited cause of failure was local financial distress, and the next most cited was incompetent management. Runs or loss of public confidence were cited in less than 5 percent of all 4,449 causes listed for the 2,955 failures surveyed (O’Connor 1938, 90).

Sudden unexpected bad news about a particular bank or group of banks appears to ignite a round of reexamination of other banks by market participants to determine their risk exposures. Although deposit flows and stock values of a large group of banks may be affected adversely immediately, the sorting-out process appears to occur relatively quickly. To the extent that deposit flows and (especially) stock values of innocent banks (those with high capital or different risk exposures) are affected adversely by a bank failure or other adverse event, they rebound within a day or two so that no lasting significant announcement effects on stock values are observed (Kaufman 1994). Similarly, a recent study of stock-market reaction to the disclosure of supervisory actions by bank regulators found that the announcements can cause spillover effects to other banks. However, “only banks in the same region . . . [or] with similar exposures are affected” (Jordan, Peek, and Rosengren 2000, 298).

The evidence suggests that even during the Great Contraction of 1929–33 and at the height of the banking crisis and bank runs in Chicago in June 1932, liquidity problems and depositor runs rarely, if ever, drove economically solvent independent banks into insolvency (Calomiris 1999; Calomiris and Mason 1997, 2000; Wicker 1996). In those difficult times, at the margin, depositors and other banks were still able to differentiate economically solvent from insolvent banks rather quickly. Moreover, almost all the banks that failed during the Depression were small unit banks. Although in 1930, 1931, 1932, and 1933 the annual bank failure rate was 6, 11, 8, and 28 percent, respectively, the percentage of deposits in the failed banks was only 2, 1, 2, and 12 percent of deposits in all banks. An analysis of this period concluded that “these failures occurred primarily because of adverse local business conditions rather than...
than because of spillover from other failed banks outside their market areas” (Benston et al. 1986, 62). However, as in most previous severe U.S. banking crises, there were runs out of bank deposits and into currency, especially by smaller depositors, so that the aggregate currency-deposit ratio increased, and aggregate bank credit and deposits declined. Nevertheless, few if any initially solvent banks appear to be buried in the graveyard of failed U.S. banks. To the extent that contagion exists in banking, at least in the United States, it appears to be rational and information based, ignited by a common shock.4

Nor is there empirical evidence that bank failures ever ignited downturns in the macroeconomy. Rather, again at least for the United States, the direction of causation appears to be primarily from downturns in the macroeconomy and the stock market (asset price bubbles) to increases in bank failures (Benston et al. 1986; Benston and Kaufman 1995; Calomiris and Gorton 1991; Mishkin 1991). Bank failures, however, are likely to exacerbate the magnitude of the downturns that caused them. The extent of adverse spillover from the banking sector to other sectors depends on the degree of leverage elsewhere. The higher the leverage of business firms and households, the more vulnerable they are to losses and insolvencies from bank failures (Davis 1995; Kaufman 2000a). Perhaps one of the reasons for the small negative effects of bank failures on other banks and on the macroeconomy, at least in recent years in the United States, relative to those that might have been feared, is the unique policy of effectively giving both insured and often uninsured depositors at failed banks immediate access to the full amount of their insured funds and the estimated recovery value of their uninsured funds, respectively. Thus, there is no (or at worst only a brief) loss of liquidity to depositors or to the economy (Kaufman and Seelig 2002).

In most other countries, both insured and particularly uninsured depositors generally are not paid either their claims until months, if not years, after the bank is resolved as the funds are collected by the receiver. Indeed, a European bank analyst has observed:

The issue is not so much the fear of a domino effect where the failure of a large bank would create the failure of many smaller ones; strict analysis of counterparty exposures has reduced substantially the risk of a domino effect. The fear is rather that the need to close a bank for several months to value its illiquid assets would freeze a large part of deposits and savings, causing a significant negative effect on national consumption. (Dermine 1996, 680)

Depositors fear the loss of liquidity in bank failures as much as, if not more than, the loss of credit value, especially when the credit losses are nonexistent if the deposits are fully insured and relatively small for uninsured depositors.

4. The same conclusion holds in cases where a lesser adverse shock did not lead to bank failures but only to reduced profits as reflected in reductions in dividends (Bessler and Nohel 2000).
In many countries, especially those with developing and transition economies, evidence of contagious systemic risk in banking frequently is confused with crises stemming from the freezing, confiscation, or devaluation of bank deposits (either in domestic or foreign currency) or from the defaulting on bank-held government securities by governments. The bank problems frequently arise not from the actions of the banks themselves in their banking activities, but from the governments’ use of the banks to pursue their nonbanking policies. The recent bank closures in Argentina are a good example of such government behavior. When the crises are bank made, they almost always reflect flagrant abuses that the government permitted, if not abetted, and the government’s inability to resolve the insolvency in a timely and efficient manner. (Whitehouse [1999] describes such a crisis recently in Russia.) These crises are defined more accurately as government created rather than bank created.

The preceding evidence strongly suggests that in the absence of de jure deposit insurance, depositors and other bank creditors take sufficient protective action on their own to reduce greatly the probability of losses to themselves and of spillover to other banks. Much if not all of any externality of contagion appears to be priced by the market and internalized. This conclusion holds even when there appears to be some positive probability that some or all of the affected claimants may be protected partially or totally ex post de facto. It is also likely that most bank stakeholders would have taken even stronger protective actions in the absence of regulations or other regulatory actions that project a perception of safety. In practice, private banking appears to be no less stable in an atmosphere of little government prudential regulation than with more such regulation; nor does it appear any less stable than other nonregulated industries.

Dealing with Systemic Risk

In light of the foregoing discussion of theory and evidence, how should bank regulators and supervisors deal with systemic risk? The preceding analysis clearly indicates that private-market incentives can and do play a major role in limiting systemic risk and that the government should always be highly sensitive to whether its actions are undermining or reinforcing the private mechanisms (Kaufman 1996). The latter is especially important in relation to the design and use of various safety-net measures. The issues, however, are not easy ones, and it is useful to undertake a normative analysis in terms of the three not mutually exclusive definitions of systemic risk set forth earlier.

Macroshock

If asset or currency values drop sharply and affect a nation’s entire economy, banks will not be immune. Indeed, history has shown them to be particularly vulnerable because debtors default and collateral depreciates. The most recent example is the banking and currency crises that hit Indonesia, Korea, Malaysia, and Thailand in 1997.
and Russia and Brazil in 1998. All banks will incur losses in severe depressions or when asset bubbles burst; weaker banks will become insolvent, and failures may spread beyond them.

By far the most important contribution any government can make to preventing macroshocks and their effects is to avoid adopting monetary and fiscal policies that produce them or to introduce policies that moderate them. Such policies lie beyond the scope of this article. It should be noted, however, that many countries have small, undiversified economies highly vulnerable to external disruptions that they have little or no ability to control or offset (Brock 1992). In this article, we take as given the occurrence of macroshocks for whatever internal or external reasons.

To protect themselves against such contingencies, banks employ various risk-management techniques, including the maintenance of higher capital ratios to absorb unexpected losses. It is difficult, however, to anticipate the probability and magnitude of extreme events and hence the amount of capital that an individual bank, given its risk preferences, ought to maintain. Indeed, in most countries, banks do not even need to try to protect themselves against “one in a hundred years” events because their governments have adopted de jure or de facto deposit insurance or other guarantee arrangements that in large part free the individual bank from pressure by depositors at risk and that substitute regulatory capital requirements for market requirements. The evidence indicates that macrofailures (as opposed to individual bank failures) usually arise more from shortcomings in government monetary, fiscal, or regulatory policy than from shortcomings in bank management. Hence, the cost of those shortcomings is placed more appropriately on the government than on the bank or on its depositors (Scott and Mayer 1971). Nevertheless, the bank’s and depositors’ responses to damaging government policies are likely to exacerbate risk taking, the fragility of the financial sector, and the magnitude and damage of the macroshock (Crockett 2000).

For example, federal deposit insurance has proved effective in stopping bank runs in the United States and in blocking that avenue of contagion spread—but at a price. The evidence indicates that deposit insurance is associated with an increase in the costs of the initial insolvencies in two ways (Gupta and Misra 1999). First, institutions were relieved of whatever market discipline might have been exerted by insured claimants. If the deposit insurance is underpriced, as is not uncommon, it contributes to a moral-hazard problem in which bank management is induced to take on greater risk. Second, bank supervisors have strong incentives to delay recognition of insolvencies and payment for their losses. In any political regime, it is advantageous to defer costs beyond one’s term in office, if possible. As recognition and resolution are delayed, losses are likely to grow rapidly. Incumbent management, if left in control, has every reason to take high-risk (and even negative present-value) investments, and government liquidators have limited expertise and weak incentives to maximize profits.

The evidence on the U.S. savings-and-loan (S&L) debacle of the 1980s supports this bleak scenario. The 1983 negative net worth of the S&L industry as a whole was
estimated at approximately $25 billion after the sharp decline in interest rates had reduced much of the earlier losses attributable to interest-rate risk (Ely 1993; Kane 1980). Nevertheless, by 1995, at the end of the long-deferred resolution process, the cost to taxpayers had climbed to approximately $160 billion, almost all of it attributable to losses from credit risk (FDIC 1998). A few bank runs (by uninsured depositors for the most part) occurred in the 1980s under deposit insurance, but the aggregate losses of the institutions were of the same order of magnitude (approximately 3 percent of 1990 GDP) as in the Great Depression years 1930–33 without deposit insurance and with numerous bank runs (Calomiris 1999).

The undesirable side effects of deposit insurance have produced efforts to counteract them by regulation. The FDIC Improvement Act (FDICIA) of 1991 changed a flat-rate deposit-insurance assessment fee to a risk-related premium system to deal with the moral-hazard problem, and it instituted a “trip-wire” scheme of prompt, statutorily mandated corrective actions and resolution of insolencies that was intended also to counteract the bureaucratic tendencies toward forbearance and postponement (principal-agent conflict). In July 1988, the Basel Committee on Banking Supervision adopted a set of risk-based minimum-capital standards for international banks, in part to offset the substitution of government guarantees (public capital) for private capital in banks (Peltzman 1970). However, banks often take steps to avoid those regulations that they find onerous and to arbitrage against (to “game”) those they find inadequate, and such reactions give rise to another layer of distortion costs (Jones 2000). For example, the initial Basel Accord assigned only a 20 percent risk weight to short-term interbank loans. Banks in the East Asian countries borrowed heavily in dollars in the early 1990s and lent at higher rates in their domestic currency, which helped to precipitate a crisis when their exchange rates had to be devalued. But foreign-exchange risk was not captured in the Basel standards, and the lending bank creditors generally were protected in the ensuing International Monetary Fund (IMF) rescues, again to the impairment of market discipline. In 1999 and 2001, the Basel Committee proposed reforms in its standards to meet these objections. It refined the risk categories and weights; added capital requirements for operational risk; permitted the use of bond ratings assigned to borrowers by recognized rating agencies to categorize risk classes; permitted more sophisticated banks to use their own internal models to evaluate credit risk; and expanded the sole emphasis on minimum-capital requirements (pillar one) to include provisions for improving supervisory review (pillar two) and market discipline (pillar three). Nevertheless, many shortcomings remain. The U.S. Shadow Financial Regulatory Committee (2000 and 2001), among others, has made recommendations for correcting these shortcomings, but many problems remain.

The moral-hazard and principal-agent problems that poorly priced deposit insurance creates, or at least exacerbates, suggest that the cost-benefit balance would be
improved if insurance coverage were provided beyond small accounts at most only in
the event of a macroshock. In all other failures, claimants on the bank would not be
protected by the government de facto as well as de jure and in their own interest
would have to exert market discipline on bank management at all times. As noted, it
is more problematic to assign preventive responsibility to the bank or to its depositors
in the case of macro–policy failures, but it would be difficult ex ante for regulators to
ascertain the beginning of a macrocrisis or to draw the line as to when a number of
individual failures fall into that category. And politically it would no doubt be a diffi-
cult distinction to sell.

If it is not feasible to limit the government safety net to macroshocks, however,
it is feasible to restructure its operation to reduce the adverse side effects. Such
restructuring was Congress’s goal in enacting FDICIA in 1991. The FDIC was
instructed to establish a risk-based assessment system for deposit insurance, replacing
the half-century-old uniform flat rate and its contributions to moral hazard (Shiers
1994). Supervisory discretion to forbear was intended to be narrowed sharply,
though hardly abolished, by specification of a structure of mandatory, presumptive,
and optional corrective actions, geared to a set of five declining capital levels. In par-
ticular, when an institution became “critically undercapitalized” (with a ratio of tan-
gible equity to total assets of less than 2 percent), the supervisor was to set in motion
a process of relatively speedy sale or closure (Benston and Kaufman 1994; Scott
1993). In resolving a failed institution, the FDIC was enjoined to employ the least
costly method of meeting its insurance obligation and not to protect creditors or
uninsured depositors if doing so would increase its losses. An exception is made for
cases of systemic risk, but it is viewed skeptically; to invoke it, the FDIC must have the
concurrence in writing of two-thirds of the Federal Reserve Board and of the secre-
tary of the Treasury (after consultation with the president), and then it must recover
its loss by a special assessment on the banking industry. It is unlikely that a “too big
to fail” policy, in which uninsured depositors are protected fully against loss, will be as
much relied on in the future as in the past.

Other aspects of the current U.S. deposit-insurance system also deserve com-
ment in relation to the handling of macroshocks. Two features reduce the impact of
bank failures on deposit holders, on the money supply, and on the economy. First,
as noted earlier, depositors are not cut off from their funds for long when their
institutions are resolved; insured deposits are paid within a business day or two, and
advance dividends on uninsured claims often are paid at approximately the same
time, based on the estimated recovery value of the failed bank’s assets (FDIC 1998;
Kaufman and Seelig 2002). Simply shutting down a failed bank for an indetermi-
nate period and freezing deposits, as supervisors often have done in some countries,
feeds incentives to run on all possibly affected banks at the first suggestion of trou-
ble. Second, the policy of prompt resolution of insolvent or nearly insolvent banks,
if properly implemented by the supervisory agencies, should result in relatively
small if any losses to depositors. If bank failure produces no or only moderate losses
(except to shareholders), those losses can be absorbed by the capital buffer at other banks, and there should be little contagion or systemic risk. This consideration underlines the importance of banking agencies’ having and enforcing credible and predictable closure (resolution) rules prior to the development of massive losses, as in the 1980s.

Failure Chains

With respect to chain-reaction or direct-causation failures flowing through interconnected institutions, there are two lines of attack. Supervisors, as just noted, can reduce the amount of loss in the initial failure by prompt closure rules. Private banks also have many ways, such as careful monitoring and exposure ceilings, to protect themselves against defaults by their counterparties, and it is important that regulation not undermine their incentives to do so (Rochet and Tirole 1996). Deposit insurance should not cover interbank transactions; no weaker claim for customer protection can exist than that of another institution in the same business engaging in informed and voluntary dealings. A fortiori, there should be no safety-net “too big to fail” policy (meaning too big to pay off in full all depositors and even other creditors at failed institutions)—a policy that eliminates entirely the need for counterparties to the largest banks to take even elementary measures to reduce their risk exposure.

In the current technological environment, the greatest volume of interbank transactions takes place through the large-value-payments system, and it now is viewed often as a focal point of systemic risk (Corrigan 1987). In 1999, the average daily value of funds transfers through Fedwire was almost $1.4 trillion and of government securities approximately $700 billion (Federal Reserve Board 2000). If the failure and resolution of a major bank caused it to be unable to meet its payment obligations in these transactions, fear of a cascade of defaults throughout the system might arise, producing a “gridlock.” The Fed’s response was to guarantee payment of transfers made by a bank on Fedwire, thereby assuming the credit risk that the transfers might not be fully collectible at the end of the day. Until 1994, the Fed provided this guarantee of such daylight overdrafts without charge. Therefore, of course, banks had little reason to pay close attention to the financial condition of their interbank payments to counterparties, and the Fed’s exposure on daylight overdrafts grew accordingly (Hancock, Wilcox, and Humphrey 1996).

Since 1994, the Fed has tried to limit the problem by making a charge (at a relatively low current annualized rate of 0.36 percent) for daylight overdrafts and by setting caps on net-debit positions. Still, it funds approximately 40 percent of funds transfers by extending daylight overdraft credit (McAndrews and Rajan 2000), which in 1999 ran at an average magnitude of $50 billion per minute (Zhou 2000). Once again, regulation has served to weaken banks’ incentives to protect themselves. Without payment finality, banks would themselves limit their exposure by monitoring and
rating their counterparties, charging accordingly for credit extended, limiting the size of their credit positions, and requiring collateral. Most important, U.S. banks would have strong reasons to push for the full implementation of a real-time gross-settlement system that transfers only good funds (payment versus payment and delivery versus payment) without government credit guarantees.

By imposing policies of credit allocation toward “favored” borrowers—be they cronies, perceived socially desirable sectors, or politically potent voter groups—governments can impair, sometimes severely, an institution’s efforts to manage its risks and portfolio prudently. Such pressure has affected the banking systems of most countries to varying degrees, particularly in countries that permit state-owned banks. To stay close to home, the United States for half a century legally restricted thrift institutions for the most part to investing in local residential construction and ownership and to financing long-term, fixed-rate, residential mortgage loans with short-term deposits. This requirement left them woefully undiversified in both a geographical and a product sense. The consequences were no small factor in the S&L catastrophe of the 1980s (Scott 1990).

In addition to losses to uninsured depositors at an affected bank, another chain of transmission of adverse shocks to banks is sometimes said to be complex transactions, particularly on derivatives markets, between a very large bank and other banks and nonbank parties. The banks need to unwind these positions quickly before maturity may generate large fire-sale losses and disorderly markets. The Prompt Corrective Action (PCA) provisions of the FDICIA reduce, even if they do not eliminate, this possibility by requiring bank supervisors to become progressively more familiar with financially troubled banks as their capital ratios decline through the undercapitalized zones. This process should provide the regulators with sufficient time to plan and prepare for the sale of an institution before it reaches the 2 percent equity-to-capital ratio closure rule or shortly thereafter within the permissible 90-day (extendable to 270-day) period to minimize any disorderly ramifications of the resolution. If successful, the regulators can achieve the dual public-policy goal of having the uninsured depositors at risk and maintaining orderly markets without invoking the systemic risk or “too big to fail” exemption. Indeed, if the regulators need some additional time to unwind very large and complex banks in an orderly way, provisions exist for the chartering and temporary operation of a bridge bank for this purpose.

Common-Shock or Reassessment Failures

The other mechanism of contagion identified earlier is the failure or near failure of one or several institutions from losses originating elsewhere and the reassessment by depositors, creditors, and shareholders of other institutions (common-shock contagion). Debate over this category has concerned whether the reassessment of risk, in light of new information revealed by the initial failures, is rational and discriminating or panic driven and undifferentiated.
The evidence reviewed earlier indicates that depositors have done much better than they usually are given credit for in distinguishing insolvent from solvent banks and in shutting down the former through runs more quickly than supervisors might have been inclined to do. It is not necessary, however, to resolve that debate definitively in order to draw lessons from it for the banking agencies.

The obvious lesson is that banking supervisors should not impede but instead should enhance the disclosure of information about the financial condition of banking institutions. Bank depositors, like bank counterparties, in many instances can protect themselves if all reason to do so is not destroyed. At the same time, supervisors should facilitate their ability to differentiate among banks in a time of crisis or uncertainty.

One step supervisors can take to enhance bank transparency would be to permit, rather than forbid, banks to disclose the contents of their examination reports and supervisory ratings (Jones and King 1995). The banking agencies, viewing examination reports as their private property, usually refuse to allow outside auditors access to them. In 1989, Congress required such access by statute but eliminated that provision two years later in the FDICIA. The current practice of mandatory secrecy, a skeptic might argue, apparently is founded either on the notion that depositor confidence must be based on ignorance or on the proposition that management is willing to reveal negative information to examiners because they believe nothing much will result from it, compared to the consequences of telling the world at large, or perhaps on the reluctance of regulators to face a market test. None of these positions is reassuring.

Another step would be to encourage banks to disclose market values of all assets and liabilities in financial statements, at least in footnotes. Not all items can be so valued with precision, but many more can be estimated reasonably accurately and already are in banks’ internal risk-management models and calculations. If proposals for larger banks to issue uninsured subordinated debt (U.S. Shadow Financial Regulatory Committee 2000) bear fruit, the market will demand more disclosure of such information. The FDICIA enjoined banking agencies to develop within a year a method “to provide supplemental disclosure of the estimated fair market value of assets and liabilities, to the extent feasible and practicable, in any balance sheet, financial statement, report of condition, or other report” (12 U.S.C.A. §1831n[a][][3][D]). Unfortunately, nothing came of this congressional mandate.

Conclusion

Many bank regulatory actions have been double-edged, if not counterproductive. With regard to systemic risk, circumstances may exist in which complete reliance cannot be placed on private ordering; however, excessive reliance on deposit insurance and other government safety-net measures, even if well intentioned, has been very costly.
Our purpose in this article has been to emphasize some of those costs and to urge bank regulators to be more sensitive than they often have been to how their actions can impair private-market incentives and thus reduce the benefits of their actions. Indeed, we suggest a deliberate strategy of seeking to minimize the scope of the government’s backup role and to maximize the effectiveness of private actors as the first line of defense against systemic risk. That approach was not much in evidence through the latter two-thirds of the twentieth century. It is not possible either theoretically or empirically to draw up a comprehensive balance sheet of all the benefits and costs produced by bank regulation and intervention over that period, but, in our own view, it is arguable that the costs outweighed the benefits, and the regulators may well have contributed to systemic risk as much as they retarded it. We hope that a new strategy that reduces potentially counterproductive government policies will play a larger role in the twenty-first century.

References


Brock, Philip, ed. 1992. *If Texas Were Chile—A Primer on Banking Reform*. San Francisco: ICS.


Hancock, Diana, James Wilcox, and David Humphrey. 1996. Intraday Management of Bank Reserves. *Journal of Money, Credit, and Banking* 28, no. 4: 870.


Acknowledgments: Both authors are members of the U.S. Shadow Financial Regulatory Committee. Parts of the first section of this article are derived from Kaufman 2000a and Kaufman 2000b. A previous version of the article was presented at a conference following the Third Annual Joint Meeting of the European, Japanese, Latin American, and U.S. Shadow Financial Regulatory Committees in Amsterdam, June 2001. The authors are indebted to the conference participants and to Bill Bergman (Federal Reserve Bank of Chicago) and Edward Kane (Boston College) for helpful comments on earlier drafts.