Working Paper No. 86

Does Bank Regulation Retard or Contribute to Systemic Risk?

by

George G. Kaufman*

Kenneth E. Scott**

January 2001
DOES BANK REGULATION RETARD OR CONTRIBUTE TO SYSTEMIC RISK?

George G. Kaufman and Kenneth E. Scott*

Abstract

The prevention of "systemic risk" and a collapse of the banking system is often cited as an objective and rationale for bank regulation and government guarantees, but guarantees and bailouts change the incentives for bank managers and depositors and can lead to an increase in bank failures and losses. Is the net balance favorable, and what might be done to improve it?

The paper distinguishes among three different concepts of how a systemic collapse might occur: a large exogenous shock, a failure of one or a few major banks transmitted by a chain reaction among interconnected institutions, and the failure of one bank providing information causing a reassessment of other similar institutions. A key issue becomes whether market participants do a reasonably accurate job of distinguishing insolvent banks from temporarily illiquid ones; that is an empirical question, and some of the evidence is examined. The final section of the paper looks at ways, for each type of systemic risk, in which bank supervisors might improve the efficiency of their actions.

* Loyola University Chicago and Stanford Law School. Both authors are members of the U.S. Shadow Financial Regulatory Committee. Parts of the first part of the paper are based on Kaufman 2000 a and b.
One 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 the term "fire," the term "systemic risk" is less clearly defined, and unlike fire fighters, who are rarely accused of sparking or spreading rather than extinguishing fires, bank regulators have been accused of encouraging, albeit unintentionally, rather than retarding systemic risk. This paper discusses the alternative definitions of systemic risk, reviews the historical evidence of systemic risk in banking, describes how financial markets have traditionally protected themselves from systemic risk, evaluates the regulations adopted by bank regulators to reduce both the probability of systemic risk and the damage caused by it if and when it may occur, and makes recommendations for efficiently curtailing systemic risk in banking.

I Definitions of 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 country, a number of countries, or globally. Systemic risk may also occur in other parts of the financial sector - e.g., in securities markets as evidenced by simultaneous declines in the prices of a large number of securities in one or more markets in a country or across countries. Systemic risk may be either or both domestic and transnational.

A. Concepts of Systemic Risk

The precise meaning of systemic risk is ambiguous; it means different things to different people. A search of the literature reveals three frequent definitions. The first refers to a "big" or macro shock that produces near simultaneous adverse effects for most or all of the domestic economy or system. That is, 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, p. 4). Likewise, Mishkin (1995, p. 32) 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." How the transmission of effects from a macro shock to individual units occurs is unspecified.
Two other definitions focus on potential spillover from one unit to others. For example, Kaufman has defined systemic risk as 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, p. 47). This definition is consistent with that of the Bank for International Settlements (BIS) that systemic risk is "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, p. 177). This second definition emphasizes causation as well as correlation (correlation with causation) and requires strong direct interconnections or linkages among institutions or markets. When the first domino falls, it falls on others causing them to fall and in turn knock down others in a chain or "knock-on" reaction. Governor George of the Bank of England (1998) has recently 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." For banks, this may occur if Bank A, for whatever reason, defaults on a loan, deposit, or other payment to Bank B that produces a loss greater than B's capital and forces it to default on payment to Bank C with losses that are larger than C's capital, and so on down the chain (Crockett, 1997). Banks, particularly within a country, tend to be closely interconnected through interbank deposits and loans. Note that in direct causation only one bank need be exposed to the initial shock. All other banks along the transmission chain may be exposure free.

The smaller a bank's capital-asset ratio, the more leveraged it is, the more is it likely both to be driven into insolvency by insolvencies of banks located earlier on the transmission chain and to transmit losses to banks located later on the chain. What makes direct causation (chain reaction) systemic risk in financial sectors particularly frightening to many is both the lightning speed with which it is believed to occur and the belief that it can infect innocent as well as guilty parties, so that there is little or no way to protect against its damaging effects.

A third definition of systemic risk also focuses on spillover, from an initial shock, but does not involve direct causation and depends on weaker and more indirect interconnections. It emphasizes similarities in third party risk exposures among the firms involved. When one unit experiences an adverse shock from, say, the failure of a large financial or nonfinancial firm that
generates severe losses, uncertainty is created about the values of other units potentially subject to
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 with that of the initial unit economically, politically, or
otherwise, the greater is the probability of loss and the more likely are the participants to withdraw
funds as soon as possible. The response may induce liquidity and, even more fundamentally,
solvency problems. This may be referred to as a "common shock" or "reassessment shock" effect and
represents correlation without direct causation (indirect causation).

Because information on either the causes or magnitude of the initial shock or on the risk
exposures of each unit potentially at risk is not generally available immediately, accurately, or for
free, and analysis of the information is not immediate or free, participants require time and resources
to sort out the identities of the other units at risk and the magnitudes of any potential losses.
Moreover, as credit markets deteriorate, the quality of private and public information also
deteriorates and uncertainty increases further. Because many of the participants are risk averse and
would rather be safe than sorry, they will transfer funds, at least temporarily during the period of
confusion and sorting out, as quickly as possible to well recognized safe or at least safer units
without waiting for the final analysis. In the periods of great uncertainty and stress, market
participants increasingly tend to make their portfolio adjustments in quantities (runs) rather than in
prices (interest rates).¹ That is, at least temporarily, they will not lend at 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 analysis would identify them ex-post as having similar exposures that
actually put them at risk (guilty) or not (innocent). 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. Innocent parties cannot be differentiated from guilty parties. 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 financially institutions and markets. Any resulting liquidity

¹ An interesting theoretical explanation of such investor behavior is developed in Herring and Wachter,
1999.
problems are likely to temporarily spill over to banks not directly affected by the initial shock. That is, the initial domino does not fall directly on other dominos, but its fall causes players to examine nearby dominos until they can see whether they are subject to the same destabilizing forces as caused the initial domino to fall. Broad contagion occurs during such sorting-out periods.

At some later date, after the sorting out process is complete, some or all of these flows affecting innocent 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 and intensify the liquidity problems, particularly for more vulnerable units (Kaminsky and Schmukler, 1999). However, the more frequent are banking crises, the more are market participants likely to become both better prepared and better informed, the shorter the sorting out and liquidity problem periods are likely to be, and the briefer will be the duration of any overshooting.

B. Dangers of Systemic Risk

A distinction is often made between rational or information-based systemic risk and irrational, noninformation-based, random, or "pure" contagious systemic risk (Kaufman, 1994 and Kaminsky and Reinhart, 1998). Rational or informed contagion assumes that investors (depositors) can differentiate among parties on the basis of their fundamentals. Random contagion, based on actions by uniformed agents, is viewed as more frightening and dangerous as it does not differentiate among parties, impacting innocent as well as guilty parties, and is therefore likely to be both broader and more difficult to contain. Thus, Governor George (1998) 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 innocent as well as guilty banks on the transmission chain. Under common shock contagion systemic risk, it is likely that innocent parties may be impacted immediately during the sorting out period, but in time will be sorted out by investors and depositors from guilty parties. Thus, the empirical borderline between rational and irrational contagion is fuzzy and in part depends on the time horizon applied. Likewise, the definition of "innocent" and "guilty" is not always clear and precise. "Innocent" parties may be defined as units that are widely perceived to be economically well-behaved -- that is, banks

---

2 For banks, runs on the entire banking system into currency because no bank is perceived safe leads to a decline in aggregate bank reserves and, unless offset by the central bank, a multiple contraction in aggregate money and credit. (See also Davis, 1995 and Diamond and Dybvig, 1983.)
that are perceived to be economically solvent and not overly leveraged. "Guilty" parties then are economically insolvent, near-insolvent, or excessively leveraged banks.

Both the chain reaction and common shock definitions of systemic risk involve speedy contagion and require some actual or perceived direct or indirect interconnection among the parties at risk. Banks are interconnected directly through interbank deposits, loans, and payment system clearings and indirectly through serving the same or similar deposit or loan markets. Moreover, to the extent that banks operate across national borders, they link the countries in which they operate. Thus, an adverse shock that generates sufficiently large losses at one bank to drive it into insolvency may transmit the shock to other banks along the transmission chain. Moreover, as noted, 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 is greater for banking, the larger and more significant the bank experiencing the initial shock. The transmission and danger of systemic risk is likely to differ depending both 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 attempt to protect themselves from losses from shocks. For banks, this requires them to charge higher interest rates on riskier investments, monitor their counterparties more carefully, or accept more and better collateral, and to have sufficient capital to absorb any losses from their association with an infected bank or banks or from runs by their depositors. For the shock to be successfully transmitted and bring down other banks, losses must exceed capital at each bank along the chain. Banks with sufficient capital to absorb the transmitted losses will remain solvent, although they may be weakened, and stop the cascading. The amount of capital required to remain solvent depends both on the exposure of a particular bank to other units and on expectations about the magnitude of any shocks. These may be expected to vary both among banks and through time for any one bank. Nevertheless, ceteris paribus, the more banks or other institutions are leveraged, the smaller the adverse shock that is required to drive a bank or other institution into insolvency and the greater is the likelihood that any losses will be passed through 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 ahead of time. This differentiates the
financial sector 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 perceived innocent as well as guilty parties. Although it is relatively easy to distinguish the innocent from the guilty ex-post after the crisis, it is difficult in practice to distinguish between the two ex-ante before a crisis. Ex-ante information is frequently not sufficiently available, timely, or reliable to make the distinction with any great confidence. Banks, often with the active assistance and encouragement of their governments, frequently fail to disclose relevant information and tend to underreserve for loan losses and use other questionable and even fraudulent accounting procedures to inflate their reported capital ratios.

II Historic Evidence of Contagious Systemic Risk

Clusterings of bank failures are observed frequently. But do they reflect systemic risk? The empirical evidence depends on the definition used. Almost tautologically, systemic risk is observed most frequently when it is defined as a broad big shock. But, as noted earlier, this definition is silent on the transmission of contagion. Common shock systemic risk, particularly in the short-term, appears to be more frequent than chain reaction systemic risk. Except for the macro shock situation, systemic risk when it does occur appears to be both rational and confined primarily to "guilty" parties and not to randomly affect "innocent" parties fatally.

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 or a group of banks do spill over to other banks, but almost exclusively only to those 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 other, economically solvent banks, or that bank depositors run on economically solvent banks very often and that it drives these banks into insolvency when they do.
A. Potential exposure

A recent study has simulated the likelihood of direct causation or knock-on contagion in the U.S. through Fed funds transactions and other interbank exposures for the period February-March 1998 (Furfine, 1999). These funds are de-jure uninsured and, since the Depositor Preference Act of 1993, Fed funds are subordinated to all domestic deposits. The study finds that, if a high loss rate of 40% is assumed, the failure of the largest debtor bank in the U.S. Fed funds market would cause the economic insolvency of only 2 to 6 other banks holding less than 1% 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 10 other banks would fail. All other banks held sufficient capital to absorb the losses. If the assumed loss rate were reduced to 5%, approximately that experienced in the Continental Illinois Bank failure in 1984 (see below), no other banks would fail.

The results are not changed greatly when total interbank exposures are simulated. The simultaneous failure of the largest two debtor banks causes more than 15 other banks with more than 3% of total bank assets to fail only when the loss rate exceeds 65%. Such a loss rate would be exceedingly high for large resolved banks in the U.S. Even at the height of the banking crises in the 1980's, when regulators regularly forbore and delayed resolving insolvencies until after significant runs by uninsured depositors had effectively stripped the banks of their best assets and increased losses as a percent of the remaining assets, the losses at large commercial banks rarely exceeded 10% of assets (Kaufman, 1995b). At these loss rates, Furfine's simulations predict only minor knock-on effects. Moreover, these results overstate the damage to other banks as they assume failure when only tier 1 rather than total capital, including tier 2, 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 (Sheldon and Maurer, 1998 and Angelini, Maresca, and Russo, 1996).

B. Historical experience

Nor is the evidence any different for actual failures. When the Continental Illinois Bank, the seventh biggest bank in the U.S. at the time, 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 FDIC fully protected all creditors, no bank suffered any losses. But what would have happened if all creditors had not been fully protected? Not very much! Some 1,325 banks had exposure of less than $100,000 and were thus fully insured 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 60 cents on the dollar (a recovery rate on assets of only 40%), which was more than ten times either the estimated loss or the actual loss as of the time of its resolution, only 27 banks would have suffered losses in excess of their reported capital and become insolvent (U.S. Congress, 1984). These losses would have totaled only $137 million. Another 56 banks would have suffered losses equal to between 50 and 99% of their total capital in an amount totaling $237 million. If the Continental's loss had been as large as 10 cents on the dollar, more than twice the actual loss, no bank would have suffered a loss greater than its capital and only two banks would have suffered losses in excess of 50% of their capital. Banks had apparently acted to protect themselves by limiting their uninsured exposures relative to their capital and monitoring their positions carefully. It is also unlikely that any of the banks with insured deposits at the Continental would have failed had these deposits been uninsured.

Lastly, dangerous spillover losses to U.S. and some other foreign banks are often cited when the Herstatt bank in Germany was closed by the authorities in 1974. These banks had entered into foreign exchange transactions with Herstatt. Indeed, "Herstatt risk" has become a generic term to describe cross-border settlement risk for banks. But losses to these banks occurred not so much because of losses at Herstatt, but because the exchange in payments was not simultaneous due 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 then closed the bank at the close of business in Germany before Herstatt could 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 to the end of the business day in New York before closing the Herstatt bank, the counterparty losses would have been greatly reduced if not 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 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.
Except for fraud, clustered bank failures in the U.S. are almost always triggered by adverse conditions in the regional or national macroeconomies or the bursting of asset price bubbles, particularly in real estate (Benston and Kaufman, 1995, and Kaufman, 1999). That is, banks fail from exposure to the same common shock, such as a depression in agriculture, real estate or oil prices (Cottrell, Lawlor and Wood, 1995), not from direct spillover from other banks without themselves being exposed to the shock. Post mortems of failed banks in the U.S. 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 through 1936, shortly after the introduction of federal deposit insurance in 1933, reported that the most cited cause of failure was local financial distress followed by 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).

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 immediately affected adversely, the sorting out process appears to occur relatively quickly. To the extent deposit flows and particularly stock values of innocent banks (i.e., those with high capital or different risk exposures) are adversely affected 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 disclosure of supervisory actions by bank regulators finds 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, p. 298).

Even during the Great Depression of 1929-1933 and at the height of the banking crisis and bank runs in Chicago in June 1932, the evidence suggests that liquidity problems and depositor runs rarely, if at all, drove economically solvent independent banks into insolvency (Calomiris, 1999; Calomiris and Mason, 1997; and Wicker, 1996). Even in these difficult times, at the margin,

---

3 It is possible that the bad news event depressed all stock prices, so the innocent banks' stock prices were affected adversely but less so than those of the guilty banks.
depositors and other banks were able to differentiate economically solvent from insolvent banks rather quickly. Moreover, almost all the banks that failed during the depression were small unit banks. While the annual bank failure rate was 6, 11, 8, and 28% in 1930, 1931, 1932, and 1933, respectively, the percentage of deposits in these banks was only 2, 1, 2, and 12% of deposits in all banks. Benston et al. (1986, p. 62) concluded that "these failures occurred primarily because of adverse local business conditions rather than because of spillover from other failed banks outside their market areas." However, as in most previous severe U.S. banking crises, there were runs out of bank deposits and into currency, particularly by smaller depositors, so that the aggregate currency to deposit ratio increased and aggregate bank credit and deposits declined. Nevertheless, few, if any, innocent banks appear to be buried in the graveyard of failed U.S. banks. To the extent contagion exists in banking, at least in the U.S., it appears to be rational and information based, ignited by a common shock.4

Nor is there empirical evidence that bank failures 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; Mishkin, 1991; and Calomiris and Gorton, 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 are these sectors to losses and insolvencies from bank failures (Davis, 1995 and Kaufman, 2000a). Perhaps one of the reasons for the small negative effects of bank failures on other banks and 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 uninsured depositors at failed banks immediate access to the full amount of their insured funds and the estimated recovery value of their uninsured funds. In most other countries, both sets of depositors are generally not paid either type of claim until months, if not years, after the bank is resolved (Kaufman and Seelig, 2000).

Indeed, a European bank analyst has observed that:

---

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).
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, p. 680).

In many countries, particularly developing and transition economies, evidence of contagious systemic risk in banking is frequently confused with crises stemming from the freezing, confiscation, or devaluation of bank deposits (either in domestic or foreign currency) or the defaulting on bank-held government securities by governments. That is, the bank problems frequently arise not from the actions of the banks themselves in their banking activities, but from the use of the banks by the governments to pursue their nonbanking policies. And when the crises are bank made, they almost always reflect flagrant abuses that were permitted if not abetted by the government and the inability of the government to resolve insolvent banks in a timely and efficient manner. (Whitehouse, 1999, describes such a crisis recently in Russia.) These crises are more accurately defined as government created crises rather than bank created crises.

The above 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 greatly reduce the probability both of losses to themselves and of spillover to other banks. The externality of contagion appears to be priced by the market. This conclusion holds even when there appears to be some positive probability that some or all of the affected claimants may be ex-post partially or totally protected de-facto. It is also likely that the even stronger protective actions would have been taken by most bank stakeholders 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 either with more such regulation or other, nonregulated industries.

III  Dealing with Systemic Risk
In light of the foregoing discussion of theory and evidence, how should bank regulators and supervisors deal with systemic risk? We believe it to be clear 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). That is particularly true when it comes to the design and use of various "safety net" measures. But the issues are not easy ones, and it is useful to undertake a normative analysis in terms of the three, not mutually exclusive, concepts of systemic risk set forth in Part I.

A. Macro shock

If there are sharp drops in asset or currency values that affect the entire economy of a nation, banks will not be immune. Indeed, history has shown them to be particularly vulnerable, as 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 in the manner described below.

By far the most important contribution any government can make to preventing macro shocks and their effects is to avoid adopting monetary and fiscal policies that produce them. That is a subject far beyond the scope of this paper. However, it should be noted that the economies of some countries are small and undiversified, and hence quite vulnerable to disruptions from the world economy which they are in little or no position to control (Brock, 1992). This paper takes the occurrence of macro shocks, for whatever internal or external reasons, as given.

To protect themselves against such contingencies, banks can employ various risk management techniques, including maintaining higher capital ratios to absorb losses. But it is hard to predict the probability and magnitude of extreme events, and hence the level of capital that an individual bank, given its risk preferences, would wish to maintain, and the individual bank may not take into account the costs of contagion externalities. Indeed, in most countries, it does not even need to try, since the government has adopted de jure or de facto deposit insurance or other guarantee arrangements, which free the individual bank from pressure from depositors at risk and substitute regulatory capital requirements for market requirements. The evidence indicates that macro failures (as opposed to individual bank failures) are usually due much more to shortcomings in government monetary or fiscal policy than to shortcomings in bank management, and hence that those cost
incentives are more appropriately placed on the government than on the bank or its creditors (Scott and Mayer, 1971).

Federal deposit insurance has proven effective in stopping bank runs in the U.S. and that avenue of contagion spread—but at a price. The evidence indicates that it is accompanied by an increase in the costs of the initial insolvencies, in several ways (Gupta and Misra, 1999). First, those institutions were relieved of whatever market discipline could have been exerted by insured claimants. If the deposit insurance is underpriced, as is not uncommon, it contributes to a moral hazard problem—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, the likelihood is that the losses will grow at a rapid rate; incumbent management, if left in control, has every reason to take high risk (and even negative present value) investments, while government liquidators have limited expertise and weak incentives to profit-maximize.

The evidence in the US savings and loan debacle of the 1980s supports this bleak scenario. The negative net worth in 1983 of the S&L industry as a whole was estimated at around $25 billion (Ely, 1993), though not officially acknowledged at that time; by 1995, with the end of the long-deferred resolution process, the cost to taxpayers had climbed to around $160 billion (FDIC, 1998). There were few bank runs in the 1980s (mostly by uninsured depositors) under deposit insurance, but the aggregate losses of the institutions were of the same order of magnitude (about 3% of 1990 GDP) as in the Great Depression years of 1930-34 without deposit insurance and with numerous bank runs (Calomiris, 1998).

The undesirable side effects of deposit insurance have produced efforts to counteract them by regulation. The FDIC Improvement Act of 1991 ("FDICIA") changed a flat-rate deposit insurance assessment fee to a risk-related premium system, to address the moral hazard problem, and instituted a 'trip-wire' scheme of prompt, statutorily-mandated corrective actions and resolution of insolvencies that was intended also to counteract the bureaucratic tendencies toward forbearance and postponement (principal-agent conflict). The Basle Committee on Banking Supervision in July 1988 adopted a set of risk-based minimum capital standards for international banks, to offset the substitution of government guarantees for private capital in banks (Peltzman, 1970). Regulations,
however, in turn produce efforts by banks to take steps to avoid those that they find onerous and arbitrage against (game) those they find inadequate, and that leads to another layer of distortion costs (Jones, 2000). For example, the Basle Accord assigned only a 20% risk-weight to short-term interbank loans. Banks in the East Asian countries borrowed heavily in dollars in the early 1990s and relent at higher rates in their domestic currency, which helped precipitate a crisis when their exchange rates had to be devalued. But foreign exchange risk was not captured in the Basle standards and the lending bank creditors were generally protected in the ensuing IMF rescues, again to the impairment of market discipline. The Basle Committee in June 1999 proposed reforms in its standards, but many shortcomings remain (US Shadow Financial Regulatory Committee, 2000).

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 macro shock. In all other failures, claimants on the bank would not be protected by the government de-facto as well as de-jure and would in their own interest have to exert market discipline on its management at all times. As noted, it is more problematic to assign preventive responsibility in the case of macro policy failures to the bank or its depositors. But it would be difficult ex ante for regulators to ascertain the beginning of a macro crisis, 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 distinction difficult to sell.

But if it be not feasible to limit the government safety net to macro shocks, it is feasible to structure its operation to reduce the negative side effects. That was the goal of the Congress 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 sharply narrowed, 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 particular, when an institution became "critically undercapitalized" (with a ratio of tangible equity to total assets of less than 2%), the supervisor was to set in motion a process of relatively speedy sale or closure (Benston and Kaufman, 1994 and Scott, 1993). And in resolving a failed institution, FDIC was enjoined to employ the least costly method of meeting its insurance obligation and not to protect creditors or uninsured depositors if that would increase its losses. There is an exception for cases of systemic risk, but it is
viewed skeptically; to invoke it, FDIC must have the concurrence in writing of two-thirds of the Federal Reserve Board and the Secretary of the Treasury (after consultation with the President), and then recover its loss by a special assessment on the banking industry. It is unlikely that "too big to fail" in which uninsured depositors are fully protected against loss will be much relied on in the future; the Federal Reserve Bank of Minneapolis has proposed that the law be further amended in order to make it impossible (Feldman and Rolnick, 1998).

There are other aspects of the current deposit insurance system in the United States that deserve comment when discussing the handling of macro shocks. The impact of bank failures (on deposit holders, the money supply and the economy) is reduced by two features. First, as noted earlier, depositors are not cut off from their funds for any appreciable period of time at the time their institutions are resolved; insured deposits are paid within a business day or two, and advance dividends on uninsured claims are paid at about the same time, based on the estimated recovery value of the failed bank's assets (FDIC, 1998; Kaufman and Seeley, 2000). Simply shutting down a failed bank, as supervisors have often done in some countries, feeds incentives to run at all possibly affected banks at the first suggestion of trouble. Second, the policy of prompt resolution of insolvent or near-insolvent banks, if properly implemented by the supervisory agencies, should result in relatively small losses. If bank failure produces no or only moderate losses (except to shareholders and management), they can be absorbed by the capital buffer at other banks and there should be little contagion. This 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.

B. Failure chains

With respect to chain reaction 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, but banks have many ways to protect themselves against defaults by their counterparties, and it is important that regulation not undermine their incentives to do so. There should be no deposit-insurance coverage of interbank transactions; there could be no weaker claim for customer protection than that of another institution in the same business engaging in informed and voluntary dealings. A fortiori, there should be no safety-net policy of too big to fail (meaning too big not to pay off in full all depositors and even other creditors), which 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 is now often viewed 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 about $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, then there might be fear of a cascade of defaults throughout the system, producing a 'gridlock.' The Fed's response was to guarantee payment of transfers made by a bank, thereby assuming the credit risk that they might not be fully collectible at the end of the day; up until 1994, the guarantee of such 'daylight overdrafts' was free. That meant, of course, that banks had little reason to pay close attention to the financial condition of their interbank payments counterparties, and the Fed's exposure on daylight overdrafts grew accordingly (Hancock, Wilcox and Humphrey, 1996).

Since 1994, the Fed has tried to rein in the problem by making a charge (at a current annualized rate of 0.36%) for daylight overdrafts and by setting caps on net debit positions, but it still funds about 40% 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 importantly, 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 vs. payment and delivery vs. payment) without government credit guarantees.

Governments can impair, sometimes severely, an institution's efforts to manage its risks and portfolio prudently by imposing policies of credit allocation toward "favored" borrowers, be they cronies or politically potent voter groups. This has infected the banking systems of most countries, to varying degrees. To stay close to home, the United States for half a century legally restricted thrift institutions for the most part to investment in local residential construction and ownership. This left
them woefully undiversified in both a geographical and a portfolio sense. The consequences were no small factor in the S&L catastrophe of the 1980s (Scott, 1990).

C. Reassessment failures

The other mechanism of contagion spread identified in Part I was 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. The debate over this category has been whether the reassessment of risk, in light of the new information revealed by the initial failures, is rational and discriminating or panic-driven and undifferentiated.

The evidence reviewed in Part II indicates that depositors have done much better than they are usually given credit for in distinguishing insolvent from solvent banks, and shutting down the former through runs faster than supervisors might have been inclined to do. But it is not necessary to definitively resolve that debate in order to draw lessons from it for the banking agencies.

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

One step by supervisors to enhance bank transparency would be to permit, rather than forbid, banks disclosing contents from their examination reports and supervisory rating (Jones and King, 1995). The current practice of mandatory secrecy, a skeptic might argue, is apparently 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. Neither position is reassuring.

Another step would be to encourage banks to disclose market values of 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 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, there will be greater market demand for disclosure
of such information. FDICIA enjoined the 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).) Nothing came of the Congressional mandate.

IV Conclusion

The point of this paper has been that many bank regulatory actions have been double-edged, if not counterproductive. With regard to systemic risk, there may well be circumstances in which there cannot be complete reliance on private ordering, but excessive reliance on deposit insurance and other government safety-net measures has also been very costly worldwide.

Our purpose has been to emphasize some of those costs, and urge bank regulators to be more sensitive than they have been to how their actions can impair private market incentives. Indeed, we suggest a strategy of seeking to minimize the scope of the government's backup role and to maximize the effectiveness of private actors. That approach was not much in evidence through the latter half of the 20th century; we hope it will play a larger role in the 21st.
REFERENCES


