

# Focus

## What is the rationale for regulating banks?

*Franklin Allen (University of Pennsylvania) and Elena Carletti (Center for Financial Studies at the University of Frankfurt)<sup>1</sup>*

In recent decades there has been significant deregulation in many industries. A sector that remains heavily regulated is banking. Why is this? In this article we argue that current banking regulation is the result of a sequence of reactions to historical events. Given that it is not designed to solve any particular problem, it is not clear that it is very effective. In what follows, we identify two important market failures that can justify intervention.

### Why is banking so heavily regulated?

In recent decades there has been significant deregulation in many industries. A sector that remains heavily regulated is banking. Why is this? One reason is consumer protection but this is relatively minor. The main reason for banking regulation is to prevent financial crises. However, banking regulation is unusual compared to other types of regulation in that there is not wide agreement on what the market failure is that justifies regulation.

With other types of regulation there typically is agreement. For example, antitrust regulation is necessary to prevent the pernicious effects of monopoly. The market failure is the lack of competition. With environmental regulation, there is a missing market. Polluters do not have to pay a price to compensate the people they harm. If there was a market where they did have to do this there would be an efficient allocation of resources and no need for intervention. But there isn't such a market and it is necessary to regulate instead. In contrast, with banking what is the market failure that justifies so much regulation?

In this article we argue that current banking regulation is the result of a sequence of reactions to historical events. Given that it is not designed to solve any particular problem, it is not clear that it is very effective. We suggest that there are two important market failures that can justify intervention. The first is a coordination problem that arises because of multiple equilibria. If people believe there is going to be a panic then that can be self-fulfilling. If they believe there will be no panic then that can also be self-fulfilling. The second market failure is that if there are incomplete markets the provision of liquidity is inefficient. In particular there must be significant price volatility in order for

the providers of liquidity to earn the opportunity cost of holding liquidity. Regulation and central bank intervention should be designed to solve these failures.

### The history of crisis prevention

During the latter part of the nineteenth century European central banks, particularly the Bank of England, developed techniques of liquidity provision both to financial markets and distressed financial institutions that prevented crises. Prior to the recent run at Northern Rock, the last true crisis in the U.K. was the Overend, Gurney and Company Crisis of 1866. It was during this period that Bagehot published his famous book *Lombard Street* outlining how central banks should intervene during times of crisis.

At this time the U.S. did not have a central bank. After the Revolution it had established the First Bank of the United States (1791-1811) and the Second Bank of the United States (1816-1836). In a report on the Second Bank, John Quincy Adams wrote "Power for good, is power for evil, even in the hands of Omnipotence". This mistrust of centralized financial power, led to a failure to renew the charter of the Second Bank. During the period without a central bank, the U.S. experienced several major financial crises and subsequent depressions. A particularly severe crisis in 1907 originating in the U.S. led a French banker to sum up European frustration with the inefficiencies of the U.S. banking system by declaring the United States was "a great financial nuisance". Finally, in 1913 the Federal Reserve System was created. However, the traditional distrust of centralized power led to a regional structure with decentralized decision making that was not very effective at preventing crises.

## Focus

The Federal Reserve was unable to prevent the banking crises that occurred in the early 1930's. There was a widespread perception that these crises were an important contributing factor to the severity of the Great Depression. The experience was so awful that it was widely agreed that this must never be allowed to happen again. The Federal Reserve System was reformed and the Board of Governors was given more power than had initially been the case. In addition, extensive banking regulation was introduced to prevent systemic crises. This regulation wasn't guided by theory but instead was a series of piecemeal reforms. In many European countries, such as France, the response was much stronger and involved government ownership of the banking sector. Either through regulation or public ownership the banking sector was highly controlled.

These reforms were very successful in terms of preventing banking crises. From 1945-1971 there was only one banking crisis in the world. That was in Brazil in 1962 when it occurred together with a currency crisis. Apart from that there was not a single banking crisis. The reason that crises were prevented is that risk taking and competition were controlled so much that the financial system ceased to perform its function of allocating resources efficiently. The financial repression that resulted from excessive regulation and public ownership eventually led to pressures for financial liberalization. Starting in the 1970's, regulations were lifted and in many countries with government ownership banks were privatized.

Financial liberalization not only allowed the financial system to fulfill its role in allocating resources. It also led to the return of banking crises and there have been numerous ones in the last three decades. Many have been in emerging countries but many have also been in developed countries such as those in Norway, Sweden, and Finland in the early 1990's. The frequency of crises in the recent period since 1971 is not that different from what it was before 1914, the last period when the global economy was as integrated as now.

### The costs of financial crises

Much of the debate on the costs of financial crises and their resolution has been concerned with how exactly to measure

costs. A large part of the early literature focused on the fiscal costs. This is the amount that it costs the government to recapitalize banks, reimburse insured depositors, and possibly other creditors. However, these are mostly transfers rather than true costs. The subsequent literature has focused more on the lost output relative to a benchmark such as trend growth rate.

There are two important aspects of the costs of crises when measured this way. The first is the high average cost and the second is the large variation in the amount of costs. One study found that the mean loss is at least 63 percent of real per capita GDP in the year before the crisis starts. The range of losses is very large. In Canada, France, Germany, and the U.S., which experienced mild non-systemic crises, there was not any significant slowdown in growth and costs were insignificant. However, at the other extreme the slowdown and discounted loss in output were extremely high. In Hong Kong the discounted PV of losses was 1,041 percent of real output the year before the crisis.

It is the large average costs and the very high tail costs of crises that make policymakers so averse to crises. This is why in most cases they go to such great lengths to avoid them. However, it is not clear that this is optimal. There are significant costs associated with regulations to avoid crises and in many cases the expected costs of crises are not very high. But what are these costs of regulation? Are crises always bad or can they sometimes be advantageous? Once again the key issue is what exactly is the market failure?

### What is the problem the Basel agreements are designed to solve?

The Basel agreements illustrate the lack of agreement on the basic underlying market failure. An enormous amount of effort has been put into designing these rules. Billions of dollars have been expended by the banks in setting up systems to implement them. They provide an example of regulation that is empirically rather than theoretically motivated. Practitioners have become experts at the details of a highly complex system for which there is no widely agreed rationale based in economic theory. What is the optimal capital structure? What market failure necessitates the imposition of capital adequacy requirements? Why can't

# Focus

the market be left to determine the appropriate level of capital? There are not good answers to these questions in the theoretical literature.

The key point is that just because there is asymmetric information of some kind does not necessarily mean there is a market failure and intervention is justified. It must be shown that the government can do better than the market. In the literature on capital adequacy, it is often argued that capital regulation is necessary to control the moral hazard problems generated by the existence of deposit insurance. Partial deposit insurance was introduced in the U.S. in the 1930s to prevent bank runs or, more generally, financial instability. Because banks issue insured debt-like obligations (e.g., bank deposits) they have an incentive to engage in risk-shifting behavior. In other words, the bank has an incentive to make excessively risky investments, because it knows that in the event of failure the loss is borne by the deposit insurance fund and in the event of success the bank's shareholders reap the rewards. The existence of bank capital reduces the incentive to take risks because, in the event of failure, the shareholders lose their capital. Thus, capital adequacy requirements are indirectly justified by the desire to prevent financial crises. However, it would seem that any capital adequacy measures should be coordinated with the extent of deposit insurance. The Basel agreements contain little on deposit insurance.

Any analysis of optimal policy must weigh the costs and benefits of regulation. This can only be done in a model that explicitly models the possibility of crises. In the absence of explicit modeling of the costs of financial crises, it is difficult to make a case for the optimality of intervention.

## The first market failure: panics

We suggest that the first important market failure that potentially justifies intervention is a coordination problem that arises because of multiple equilibria. The panics view suggests that crises are random events, unrelated to changes in the real economy. If everybody believes no panic will occur only those with genuine liquidity needs will withdraw their funds and these demands can be met without costly liquidation of assets. However, if everybody believes a crisis will happen then there is

a panic and it becomes a self-fulfilling prophecy as people rush to withdraw. Which of these two equilibria occurs depends on extraneous variables or "sunspots". Although sunspots have no direct effect on the real economy, they affect depositors' beliefs in a way that turns out to be self-fulfilling. In the bad equilibrium where bank runs occur, the banks may be forced into bankruptcy. If this happens the crisis may spill over to the real economy and a recession or depression may result.

The key issue in theories of panics is which equilibrium is selected and in particular what is the equilibrium selection mechanism. Sunspots are convenient pedagogically but this explanation does not have much content. It does not explain why the sunspot should be used as a coordination device. There is no real account of what triggers a crisis. This is particularly a problem if there is a desire to use the theory for policy analysis.

In recent years a new approach to this type of multiplicity has been developed. A small amount of asymmetric information can eliminate the multiplicity of equilibria in coordination games. These games with asymmetric information about fundamentals are called global games. Introducing random noise ensures that the fundamentals are no longer common knowledge and this prevents the coordination that is essential to multiplicity. Using a global games approach to ensure the uniqueness of equilibrium is theoretically appealing. It specifies precisely the parameter values for which a crisis occurs and allows a comparative static analysis of the factors that influence this set. This is the essential analytical tool for policy analysis. This kind of analysis is in its infancy. So far these tools have provided relatively little in terms of practical insights concerning regulation, and in particular the optimal mix of deposit insurance and capital adequacy requirements.

## The second market failure: inefficient liquidity provision

The second market failure is concerned with the inefficient provision of liquidity to financial markets. This kind of market failure appears to have been particularly important in the financial crisis of the summer and autumn of 2007. In order to understand how this market failure arises it is important to

## Focus

realize that the alternative to the panics view of crises is that crises are fundamental based. In this case depositors assess the situation of banks by observing leading economic indicators that provide information that is useful for assessing future bank asset returns. If the indicators suggest that returns will be high, then depositors are willing to keep their funds in the bank. However, if the indicators suggest returns will be sufficiently low they will withdraw their money in anticipation of low returns and there is a crisis.

In this fundamental based approach the question of whether there is a market failure turns on whether markets are complete or incomplete. If there are complete markets then there is no market failure. What is meant by complete markets is that it is possible for banks to pursue sophisticated risk management strategies using a wide set of derivatives. With incomplete markets where such sophisticated risk management strategies are not possible, there is a market failure.

The essential problem with incomplete markets is that liquidity provision is inefficient. The nature of risk management to ensure that the bank or intermediary has the correct amount of liquidity changes significantly from the case of complete markets. When markets are complete it is possible to use the full set of derivatives to ensure liquidity is received when it is needed. The banks pay for liquidity provision up front when they purchase the derivatives and the pricing of the derivatives ensures adequate liquidity is provided. In this case banks can buy liquidity in states where it is scarce and pay for it by selling liquidity in states where it is plentiful for them. The complete markets allow full risk sharing and insurance.

In contrast when markets are incomplete, liquidity provision is achieved by selling assets in the market when the liquidity is required. Asset prices are determined by the available liquidity or in other words by the "cash in the market". It is necessary that people hold liquidity and stand ready to buy assets when they are sold. These suppliers of liquidity are no longer compensated for the cost of providing liquidity ahead of time. Instead the cost must be made up on average across all states and this is where the problem lies. The providers of liquidity have the alternative of investing in high return long assets. There is thus an opportunity cost to

holding liquidity since this has a lower return than the high return long assets. In order for people to be willing to supply liquidity they must be able to make a profit in some states. If nobody held liquidity then when banks and intermediaries sold assets to acquire liquidity their price would collapse to zero. This would provide an incentive for people to hold liquidity since they can acquire assets cheaply. In equilibrium prices will be bid up to the level where the profit in these states where banks and intermediaries sell is sufficient to compensate the providers of liquidity for all the other states where they do not use the liquidity and simply bear the opportunity cost of holding it. In other words, prices must be volatile to provide incentives for liquidity provision. However, the low prices that are necessary for liquidity provision may cause banks to go bankrupt. There can thus be a crisis and this may again spill over to the real economy and lead to a recession or depression.

To summarize, when markets are incomplete asset prices must be volatile to provide incentives for liquidity provision. This asset price volatility can lead to costly and inefficient crises. There is a market failure that potentially provides the justification for regulation and other kinds of intervention to improve the allocation of resources. As the recent crisis that started in the subprime markets illustrates, current measures have been inadequate in solving the problem of liquidity provision to the interbank markets. Moreover, it is not yet clear what form optimal intervention should take.

## Regulation should be designed to correct market failures

There is no wide agreement on what the market failures are that justify government intervention. We have suggested two, one is a coordination problem and the other is inefficient liquidity provision. Regulations should be designed to correct these rather than being an accident of history.

### Endnotes

<sup>1</sup> This article is based on "Banks, Markets and Liquidity" to appear in the Reserve Bank of Australia's 2007 Conference Volume, *Financial System Structure and Resilience*. The paper is also Working Paper 07-24, Wharton Financial Institutions Center, University of Pennsylvania, and can be downloaded at <http://fic.wharton.upenn.edu/fic/papers/07/0724.pdf>. The paper contains references to the relevant literature that are omitted here.