

**Fragile Banks, Durable Bargains:
Why Banking is All About Politics and Always Has Been**

Charles W. Calomiris and Stephen Haber

Date of First Draft: October 30, 2010

Date of this Draft: March 23, 2011

Do not cite without permission

Chapter 1

If Finance is Such a Good Idea, Why do We See So Little of It?

Why is it so hard to create stable banking systems that allocate credit efficiently?

Why do we observe a world in which banking systems often allocate too much credit to firms and households that are poor risks, resulting in systemic banking crises (widespread loan defaults that threaten the existence of the banking system itself)? Why do we also observe banking systems that persistently constrain credit so tightly that economic growth and social mobility are choked off? Why is it so hard for countries to find a comfortable middle ground in which credit flows to talented entrepreneurs and responsible households, and in which taxpayers are not forced to periodically rescue bankers from the consequences of their own mistakes?

More puzzling still, why is it the case that the same banking systems that are prone to systemic crises also tend to tightly constrain credit? A simple exercise, using the World Bank's Financial Structure Database and the database on banking crises compiled by Laeven and Valencia (2008), makes the pattern clear: countries that experienced banking crises during the 36-year period between 1970 and 2006 not only had much lower levels of bank lending (as a percent of GDP) compared to non-crisis countries, but they also had much slower rates of growth of bank lending.¹ Moreover, their slower rates of credit growth were not caused by

¹ We divided the world into three groups: those that did not experience any systemic banking crises, those that experienced one banking crisis, and those that experienced more than one banking crisis between 1970 and 2006. Countries that were crisis free had an average ratio of bank credit to GDP of 78 percent in 2006, while those that had a single crisis had a ratio of 41 percent, and those that multiple crises had a ratio of 29 percent. We then measured the rate of

a simple mechanical linkage between banking crises and cyclical contractions of credit. Rather, being prone to crises appears to have been an outcome of under-banking (very low levels of bank credit, relative to GDP)—a result that is very curious indeed, even paradoxical. Generally speaking, widespread defaults occur after banks have allocated too much credit to borrowers who are bad risks. Being under-banked implies, however, that most firms and households cannot get credit. The troubling apparent resolution of this paradox is that many “under-banked” economies repeatedly supply credit imprudently: after a crisis is resolved, banks appear to once again misallocate scarce credit to firms and households that are likely to default.

Even more puzzling, endemic crises and persistent under-banking occur even though banking systems are subject to close regulation and supervision by governments. In most countries, banks are regulated much like public utilities, such as electricity generation: entry is controlled by government agencies in order to assure that the firms providing the service remain profitable; and the government inspects their operations to make sure that they are providing efficient service to their customers while not taking imprudent risks. Why then do governments often allow lending to spiral out of control, permitting banks to make imprudent loans to firms and households that have a high probability of default? In the same vein, why do governments often allow those same imprudent banks to deny service to customers—to the point that in many places banks only lend to the enterprises controlled by their own board

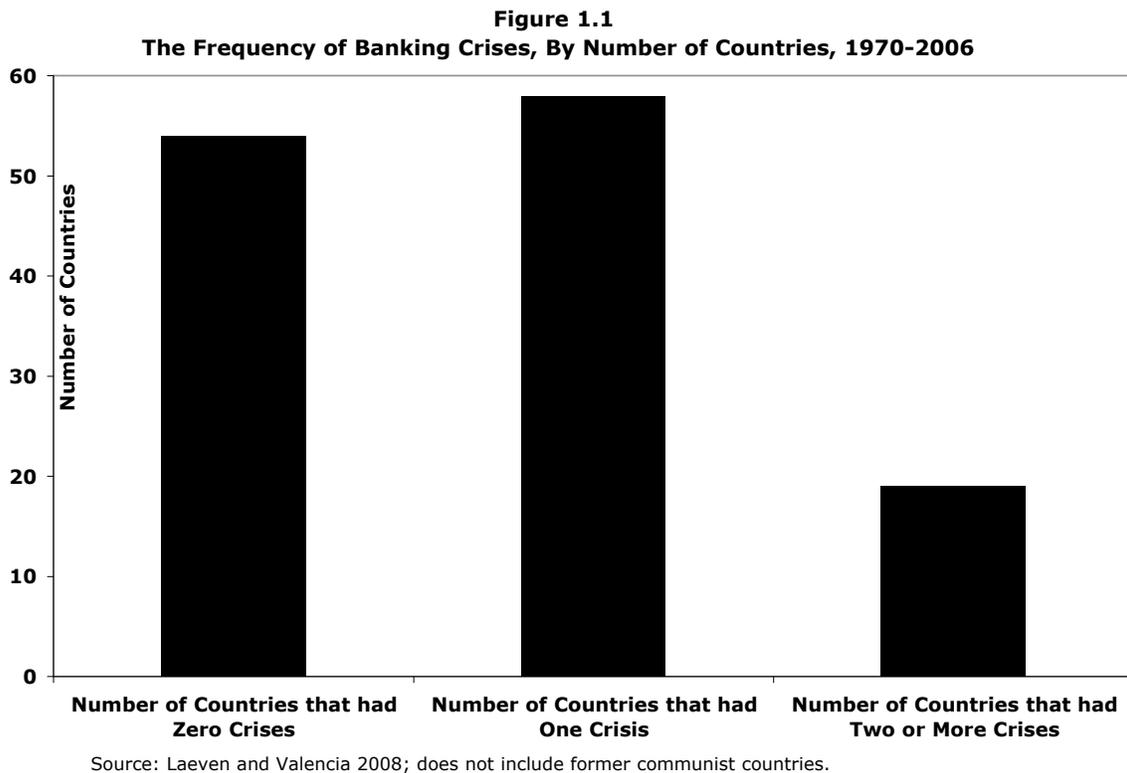
growth of credit in each group across the years 1993 to 2006 (we could not go back before 1993 because there are many missing observations in the World Bank’s Financial Structure Database before that year), and found that the non-crisis countries grew by 50 percent, the single crisis countries grew by 17 percent, and the multiple crisis countries grew by four percent. These patterns are not sensitive to the exclusion of countries that had a crisis either at the end or beginning of our sample period (1991-93 or 2004-06). If anything, truncating the sample makes the differences between the crisis and non-crisis countries even larger.

members? How can it be that a sector of the economy that is highly regulated, closely supervised, and crucial for millions of households and firms work so badly in so many countries? Why is it the case that societies do not seem to learn from their past mistakes, or from each other, about the best way to organize and operate their banking systems?

Potential answers to these questions fall into two broad categories. The first is that bankers and regulators do the best they can, but there are so many unforeseen circumstances that it is extremely hard to get it right. In this view, banking systems are like Tolstoy's unhappy families: they are all imperfect in their own way. The implication is that every systemic crisis or case of under-banking should be studied in isolation, because the reasons for their occurrence are particular and unpredictable. The second is that banking crises and under-banking are the result of systematic and predictable factors. In this view, fragile banking systems that extend credit narrowly exist not by accident, but by design. The implication is that evidence and reason should allow us to identify those factors and understand how they interact to produce socially unfortunate outcomes.

While the first class of explanation is very much favored by Finance Ministers and Central Bank Directors, there is abundant evidence that it cannot possibly be right. Banking crises have not been rare, one-off events that defy prediction; they have been endemic—and not just in the period 2007-10, when the banking systems of the United States, the United Kingdom, and several other developed countries had to be rescued at taxpayer expense. The database of banking crises constructed by Laeven and Valencia (2008) reveals that there were far more countries that experienced a banking crisis than there were countries with stable

banking systems during the period 1970-2006.² We present some of their data in Figure 1.1: it shows that 77 countries had at least one crisis, while 54 countries were crisis-free. Even more shocking, 19 of those 77 had multiple crises, and two of those 19 were especially crisis-prone: Argentina had three banking crises; while the Democratic Republic of the Congo had four.



Those banking crises were costly. The Laeven and Valencia (2008) data show that the negative net worth of failed banks averaged roughly 16 percent of GDP per banking crisis—a cost that was typically borne by taxpayers who funded the clean-up of the failed banks. Their

² We exclude former and current communist countries from this analysis, because their state-run banking systems do not allocate credit, but rather act as an accounting system for the state-controlled allocation of investment. The concept of a banking crisis has no real analytic meaning in such a system.

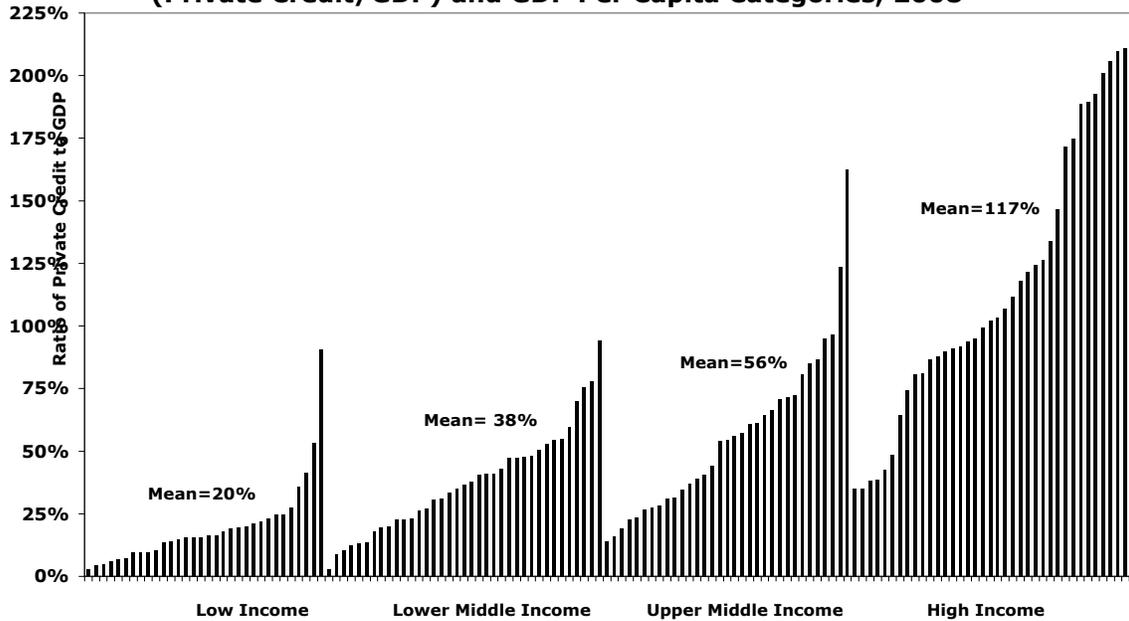
data also show that the cost of banking crises in terms of lost GDP (the effects of credit contractions, heightened sovereign debt risk, and currency collapse on economic activity) was larger still: on average, lost output amounted to roughly 20 percent of GDP per crisis.

Not only have crises been endemic, but “under-banking” has been ubiquitous as well. Indeed, throughout most of the world, firms and households have been perennially starved for credit. Consider the striking contrast between Canada and Mexico, the two NAFTA partners of the United States. In 2008, bank lending to firms and households totaled 130 percent of GDP in Canada, but in Mexico the ratio was only 19 percent—a ratio that had not changed much in 100 years (Haber 2008). The dramatic difference in those ratios means that Mexican families have a much more difficult time financing home, automobile, and consumer goods purchases, and Mexican business enterprises have a much more difficult time obtaining working capital, than their Canadian counterparts. These facts explain why Mexico “exports” roughly 500,000 workers per year to the United States: because credit is scarce, the economy grows too slowly to generate sufficient employment; Mexico cannot “import” the U.S. banking system, but Mexican workers can export themselves to the United States.

As Figure 1.2 shows, the stark difference between Canada and Mexico captures a recurring pattern: poor countries have much smaller banking systems than wealthy countries. In fact, the countries on the far left hand side of Figure 1.2, where the ratio of bank credit to GDP is less than ten percent, are among the World’s poorest, including Chad, the Democratic Republic of the Congo, Laos, Yemen, and Haiti. The countries on the far right hand side of Figure 1.2, where the ratio of bank credit to GDP is close to 200 percent, are among the

World's richest, including Denmark, the Netherlands, Great Britain, the United States, and Luxembourg.

**Figure 1.2:
The Relationship Between Financial Depth
(Private Credit/GDP) and GDP Per Capita Categories, 2008**



Source: World Bank Financial Structure Database

There is, in fact, a large body of research, spanning a diverse range of approaches to evidence, which demonstrates that Figure 1.2 depicts a *causal* relationship: economic growth is an outcome of the availability of finance. Economic historians have shown that Holland, Great Britain, and the United States experienced revolutions in financial intermediation and financial institutions *before* their rise to global economic hegemony in the 18th, 19th, and 20th centuries, respectively. They also found that Russia, Germany, and Japan underwent similar revolutions in financial intermediation before they narrowed the gap with the world's economic leaders in the late 19th and early 20th centuries (Gerschenkron 1962; Cameron 1967; Sylla 1975, 2006, 2008; North and Weingast 1989; Neal 1990; de Vries and van der Woude 1997; Rousseau and Wachtel 1998; Rousseau 2003; Rousseau and Sylla, 2003, 2004).

Financial economists, most notably King and Levine (1993), Levine and Zervos (1998), Taylor (1998), and Beck, Levine, and Loayza (2000) have shown that higher levels of financial development across countries are causally related to faster rates of physical capital accumulation and economic growth. A somewhat later generation of scholars—most notably Rajan and Zingales (1998), Wurgler (2000), Cetorelli and Gamberra (2001), Fisman and Love (2004), and Beck, Demirgüç-Kunt, Laeven, and Levine (2007)—focused on the development of industries as well as countries, and they reached the same conclusion: finance leads growth. Research focusing on the growth of regions within countries by Jayaratne and Strahan (1996), Black and Strahan (2002), Guiso, Sapienza, and Zingales (2004), Cetorelli and Strahan (2006), Dehejia and Lleras-Muney (2007), and Correa (2008) produced broadly similar results. The evidence, in short, is pretty clear: when countries are under-banked they forego more rapid technological progress, faster job creation, and increased opportunities for social mobility. Given the relationship between economic growth and the ability to project power internationally, under-banked countries also put themselves at a disadvantage when it comes to defending their sovereignty and influencing events abroad.

What systematic factor might be driving endemic crises and persistent under-banking? Our answer is that government actors have inherent conflicts of interest when it comes to the operation of the banking system, and these can produce banking instability and under-banking. First, governments regulate banks – ostensibly to limit banks’ exposures to risks— but governments also look to banks as a source of risky public finance. That is, they simultaneously supervise and regulate the banks—and borrow from, or tax, them. Second, governments enforce credit contracts that discipline debtors on behalf of banks (and in the

process assist in the seizing of debtor collateral), but they rely on those same debtors for political support. Third, governments allocate losses among creditors in the event of bank failures, but they simultaneously look to the most numerically significant group of those creditors—bank depositors—for their own political survival. These conflicts of interest are unavoidable: the business of banking is the creation and trading of contracts—between banks and governments, between banks and their creditors, between banks and their debtors, and between bank insiders (founders and majority shareholders) and bank minority shareholders—and in the final instance the parties in control of the government must arbitrate those contracts and allocate the losses when contracts cannot be fulfilled.

These inherent conflicts of interest among government actors imply that regulatory policies toward banks often reflect the influence of the political coalitions that support the government. Banks are regulated and supervised according to technical criteria, and banking contracts are enforced according to abstruse laws, but those criteria and laws are not created and enforced by robots programmed to maximize social welfare; they are the outcomes of a political process—a game as it were—with the stakes being wealth and power. The players are the actors with a stake in the performance of the banking system: bank insiders, minority shareholders, debtors, depositors, and the parties in control of the government. The rules governing play are set by the society's political institutions: those rules determine who has to be included in a winning political coalition; or, alternatively, who can safely be left out in the cold because the rules of the political system make them powerless. Coalitions form as the game is played, and those coalitions determine the rules governing bank entry (and hence the competitive structure and size of the banking sector), the flow of credit and its terms, and the allocation of losses when banks fail.

Our goal in this book is to explain this game. We show how political differences across times and places produce differences in the rules of the game, and how those politically-based rules result predictably either in stable and plentiful bank credit, or in unstable and scarce bank credit. Which players favor vertiginous increases in credit, and which players favor placing tight constraints on the amount of credit available? Under what circumstances can they forge durable political coalitions with other players that have an interest in the organization of the banking system? What are the terms of exchange among the members of these coalitions? Are there differences in the way the game is played depending on whether the political system is democratic or authoritarian? Do those differences have implications for the durability of coalitions, the size and structure of banking systems, and the fragility of the banks?

In order to understand this inherently complex game we do something that, to our knowledge, has not been done before: we trace the co-evolution of politics and banking in detail for several countries, one country at a time, over long periods of time. We spell out how coalitions were formed, why some endured while others were undermined, how they brought about specific and important changes in the policies governing banking, how those policies determined which groups were able to access credit and which could not, and how some of those policies produced disastrous banking crises.

Facts, even when organized chronologically, do not speak for themselves. They require a logical framework that connects them. We therefore begin, in Chapter Two, by presenting a theoretical framework. Theory comes first in our presentation because we rely on it to organize the facts presented in the subsequent chapters. We do not pretend that the

country histories are tests, in the positivist sense, of this theoretical framework; rather the theoretical framework was itself distilled from—indeed, one might even say discovered through—our research into the country histories.

We then develop seven historical narratives, each one covering a country case study, in order to show how differences in political coalitions gave rise to differences in the regulations governing bank entry and bank operations. We then show how those differences in regulation produced dramatic differences in the industrial organization of banking, the availability of bank credit, and the stability of the banking system. Our historical approach allows us both to draw comparisons across countries and to draw comparisons within countries over time. The fact that we can pinpoint the sequence of events within any country means that we can be certain about causal inference: we can know whether changes in political coalitions caused changes in regulation, and whether changes in regulation caused changes in the structure and performance of banking, because we know the order in which these events occurred.

We begin with Scotland, in Chapter Three, because it was the place where many key aspects of modern banking—such as laws that permit shareholders to limit their liability—were first developed. Scotland's early banking history was characterized by an unparalleled degree of competition among banks. Three banks were chartered by special acts of the government, and those coexisted with a large number of other banks, many of which themselves operated large branching networks throughout the country. Scotland's early decision to permit many banks to operate throughout the country reflected its unique political and economic history. Scottish banking law and chartering remained independent of England's dominance until 1845.

We then focus on England, in Chapter Four, which eventually adopted many of Scotland's innovations and used them to become the world's financial leader in the 19th century. But England began its banking history quite differently from Scotland. England's early bank chartering took place in the context of changes in the country's fundamental political institutions in the late 17th century. Those changes allowed a group of wealthy financiers to form a durable coalition with the parties in control of parliament in order to establish the Bank of England, which enjoyed unique privileges from its charter. That coalition helped solve a problem that had perennially plagued Britain's wealthy (the government could easily expropriate them), but it produced a new problem: it made it very difficult to obtain a bank charter to operate any large-scale bank other than the Bank of England until well into the 19th century. That is, the political coalition underpinning the Bank of England succeeded in creating a banking system, but it was a system based on the maintenance of a privileged monopoly that allocated credit narrowly. It was also a highly unstable banking system, which we show reflected the political deals inherent in a monopoly bank charter. Over time, in reaction to scarce and volatile credit, political coalitions shifted to favor greater competition in chartering, which produced a system of stable and abundant bank credit.

We focus on the United States in Chapter Five, covering the period from the revolutionary war in the 1770s until the repeal of restrictions on interstate branching in the 1990s. The United States is a fascinating case because the initial organization of its banking system was strikingly different from that which we observe today, in which banks are free to open branches more or less as they please and grow to enormous proportions. Indeed, the

initial organization of the U.S. banking system was modeled largely on that of England, but the system of monopolies was quickly undermined by two distinctively American political institutions: federalism and a broad suffrage. The United States did not, however, just jump from one system to another. Rather, the United States is a striking example of how bank insiders can form a coalition with a group of very unlikely partners—in this case, populists who distrusted banks and corporations of any type—in order to slow the transition between systems. In fact, that transition moved at an almost glacial pace: until well into the 20th century, the vast majority of U.S. banks were legally enjoined from opening branches; rather there were tens of thousands of single office “unit banks.” The whole point of this system was to create a system of local, segmented monopolies. Banks that we think of as national entities, such as the eponymously named Bank of America, was, for most of its history, actually nothing of the kind: it was a *California* bank enjoined from operating outside the state until the removal of entry barriers during the 1980s and 1990s ultimately allowed it and other U.S. banks to operate on a nationwide basis. Because of the high cost of information, these limits on branching limited competition, driving up the price of credit.

Chapter Six highlights the crucial role played by politics in the organization of the U.S. banking industry by contrasting it with the history of its northern neighbor, Canada. Indeed, Canada presents something of a counterfactual experiment: it had a similar colonial and cultural heritage, but it did not have the particular set of political institutions and circumstances that drove the U.S.’s bizarre, “unit banking” outcome. The structure of the Canadian banking system was therefore strikingly different: from its beginnings, the Canadian system was characterized by a small number of very large banks with tremendous branch networks. The result was not just lower costs of credit, but a much more stable banking

system. In point of fact, the United States has suffered three systemic banking crises over the past 80 years—the widespread bank failures of the Great Depression, the Savings and Loan Crisis of the 1980s, and the subprime crisis of 2007-09—while Canada has suffered none.

We focus on Mexico in Chapter Seven, and do so because it allows us to understand the differences between banking systems in authoritarian and democratic political systems. Unlike Scotland, England, the United States, and Canada, where elections have been part of the political system for centuries, and in which the right to vote was gradually expanded over the course of the 19th and early 20th centuries, Mexicans were denied the right to effective suffrage until the late 1990s. During most of its history, Mexico was governed by one type of authoritarian system or another. Moreover, those authoritarian governments engaged in either partial or total expropriations of the banking system on multiple occasions. Thus, Mexico not only allows us to understand how dictators form coalitions with bank insiders and minority shareholders to create a banking system, but it also allows us to understand the conditions under which autocrats break those coalitions by seizing the wealth of those same insiders and minority shareholders. Precisely because the Mexican government did this on multiple occasions, the Mexican case also allows us to understand how it managed to coax bankers into forming a new coalition to create a banking system all over again. The key, as we shall show, is that throughout Mexican history the government tightly regulated bank entry in order to drive rates of return up high enough to compensate bank insiders and shareholders for the risk that, at some point, they would be expropriated. In short, in Mexico, banking was indeed all about politics, and the politics dictated a highly concentrated system that allocated credit to bank insiders themselves.

In order to more deeply understand how the political coalitions that underpin banking systems vary by regime types, we focus on the case of Brazil in Chapter 8. Over the course of the past two centuries Brazil has experimented with constitutional monarchy, a federal republic with a tightly limited suffrage, various military dictatorships, and, since 1988, a multi-party democracy with universal suffrage. One of the striking aspects of Brazil is that until it democratized in 1988, its banking system was similar to that of Mexico: it was small, unstable, and allocated credit narrowly.

Chapter Nine draws on what we have learned from the historical analysis of these six cases to provide an analysis of the U.S. subprime banking crisis of 2007-09. Chapter ten concludes, and provides a discussion of the implications of our results—and of our approach to evidence—for the study of finance broadly defined.

Some readers may wonder why we confine ourselves to a small number of case studies. Why six countries, instead of eight, ten, or twenty? As a practical matter, exploring how a country's institutions changed over time is not an enterprise characterized by increasing returns to scale; there is an obvious tradeoff between the number of cases covered and the depth with which they can be discussed. Moreover, our purpose is not to provide an exhaustive history of the political economy of banking in every country on the globe, rather it is to develop and illustrate a general framework that we believe has wide explanatory power. We invite other researchers to test that framework against additional country cases, as well as against large-N datasets.

Other readers may wonder why we base our analysis on historical narratives, rather than basing them on the statistical analysis of ready-made datasets that cover every country in the world over the very recent past—an approach that has become quite common in the social

sciences. In the concluding chapter of this book we show that some of the patterns unearthed by our historical analyses can be detected, though somewhat imperfectly, using large-N statistical techniques and recent data.

We do not base our analysis on those techniques, however, because any such large-N analysis must be based on a coherent narrative. Identifying causation *requires telling a story*—a sequence of events and factors that logically connects a cause and an outcome. Indeed, any econometric strategy designed to isolate causal influences ultimately must be based on the plausibility of a model, and that model is based on a story: it is the story that justifies the model’s “identification restrictions.” There are also practical limits to employing large-N analyses to the study of the politics of bank regulation. Banking history stretches far back in time, but the available cross-national datasets are truncated with respect to time. Indeed, most of those off the shelf datasets extend no farther back than 1960. Furthermore, the history of bank chartering and other crucial political outcomes that matter for banking history (the nature of entry barriers, the legal rules governing banks, and the operations that banks undertook) have been highly idiosyncratic. In order to capture all of these influences, one would need to code an impractically large number of indicator variables in a complex causal system of equations for a relatively small number of countries, making statistical identification of cause and effect virtually impossible.

Several themes emerge from the historical country narratives developed in this book. The first is that a key dimension that can explain variation in both banking development and banking stability—depicted in Figures 1.1 and 1.2—is the prevalence of politically-created

entry barriers into banking. One of the most important points of intersection between banking and politics is the process that governs the licensing (chartering) of banks. In particular, a crucial difference across countries is the degree to which the political system allowed the creation of *durable* barriers to entry: in some countries, the government allowed open entry from the very beginning; in other countries, entry was initially blocked, but the government could not sustain those barriers indefinitely; and in still other countries, the initial barriers persisted.

A second major theme is that the nature of the coalitions that generate entry barriers varies across political regime types. Briefly stated, in an autocracy it is far easier to create a stable coalition in support of the tight regulation of bank entry. The reason is not hard to divine: in an autocracy, potential debtors cannot vote, and so they have more difficulty finding their way into a winning political coalition. As a consequence, autocracies are capable of creating highly stable banking systems, but those systems allocate credit narrowly. They are designed to channel credit to the government and to the enterprises owned by an elite class of government-selected bankers.

This is not to say, however, that banking systems in countries with authoritarian governments make up in stability what they lack in inefficient allocation of credit. As we shall see, both as a matter of logic and as a matter of history, the obverse is true: banking systems in authoritarian regimes tend to be less stable than in democracies. When times get tough in an autocratic political system, bank insiders can expropriate minority shareholders and depositors with impunity, so long as they share some of the rents with the autocrat. When times get tougher still, autocrats can expropriate the bank insiders with impunity.

A third major theme relates to the limitations of mass suffrage in creating stable and abundant bank credit. Although mass suffrage tends to make it more difficult to sustain a banking system that allocates credit narrowly to an elite group—mass suffrage is far from a panacea when it comes to creating a stable banking system. What is to keep debtors from voting for representatives who promise to dramatically increase the supply of credit, improve the terms on which it is offered, and then forgive those debts when they prove difficult to repay? As we shall see in our discussion of the United States subprime crisis, such a scenario is far from a distant theoretical abstraction.

The implication, we hope, is clear: banking systems are fragile plants. There are some societies that have managed to create stable banking systems that allocate credit broadly—but they are very few in number, for a good reason: they require political institutions that simultaneously allow for mass suffrage, but that also limit the authority and discretion of the parties in control of the government. Demonstrating why and how this is so, we think, is important not only for understanding the history of banking crises, but for understanding the ones that will come in the future.

Chapter 2

Why Banks are Always About Politics

Every reader of this book doubtless has at least one bank account. Most of us give little, if any, thought to the security of those accounts or the solvency of the banks where they are located. The fact that we do not is an outcome of a rather complicated set of institutions that have been hammered out over the past three centuries. Without those institutions there would be very good reasons to lay awake at night worrying about our money because a bank is, by design, a potentially unstable enterprise.

Our purpose in this chapter is to explain why it has proved so challenging to create banking systems that provide abundant and stable credit. We describe the solutions—the durable bargains—that different types of societies have crafted in order to mitigate the fragility of banks, some more successfully than others. In order to do that we begin by stripping away all of the institutions that have been crafted by banks and governments (e.g., the legal system’s enforcement of rights, limited liability for shareholders, government supervisory and regulatory agencies) in order to lay bare the property rights problems that are inherent in the business of banking. We then explore how those problems are mitigated—but not solved—by the creation of those institutions. Our goal is not to encourage you to move your savings under your mattress, but to provide an understanding of two crucial issues: why there are no banking systems without the active participation of governments; and why *no government* can be a completely neutral arbiter when it comes to the banking system. In short, our goal is to explore why banking is all about politics—and always has been.

Let us then start with the basics: what outputs does a bank produce, and what inputs does it consume? The major outputs of a bank—its product—are contracts, which primarily take the form of IOU's payable to the bank. The major inputs to the bank are contracts, which primarily take the form of IOU's payable to depositors. That is the basic business model: IOU's in, and IOU's out.

Any enterprise whose inputs and outputs consist primarily of IOU's is, by design, potentially unstable and risky. In the first place, it is extraordinarily difficult, if not impossible, to exactly match the term structures of the contracts on either side of the balance sheet. Bankers therefore engage in a difficult balancing act: they borrow money on a short-term basis from depositors and lend it on a long-term basis to debtors; depositors might show up to withdraw their money, but there might not be enough in the till.¹ In the second place, it is extraordinarily difficult, if not impossible, for the banker to accurately predict the probability with which the contracts on the assets side of the balance sheet will be fulfilled. Bankers therefore engage in a difficult process of forecasting: they write loan contracts funded primarily by deposit contracts, but they do not know with absolute certainty that the loan contract will be fulfilled; even the most honest borrower might default because of unforeseen circumstances, potentially leading to a loss of the depositors' money.²

¹ They may also fund their operations by borrowing from non-depositor creditors, which includes other banks. In developing countries, these sources of loanable funds tend to be of less quantitative importance than deposits. Moreover, the general problem—that banks are highly leveraged—remains regardless of the source of their indebtedness.

² The banker could attempt to solve both of these problems by offering the depositors very high interest rates to compensate them for risk, but that would only make the situation for the depositors riskier still: in order to honor the high interest contracts with depositors the banker

The obvious solution to both of these problems (which are referred to in the banking literature as “liquidity risk” and “credit risk”) is to create cushions in the balance sheet to absorb those risks. The first is called equity capital, a block of money that is invested in the bank by stockholders, rather than lent to it. This “capital cushion” mitigates credit risk by giving the bank the ability to honor its IOUs to depositors even if it experiences some losses on its IOUs from borrowers. A second cushion is the cash assets that banks maintain in addition to their loans. The “cash cushion” allows banks to reduce liquidity risk by using cash to finance depositor withdrawals. The cash cushion also limits credit risk to depositors because the higher the proportion of cash assets relative to loans the lower the riskiness of the overall portfolio of the bank (since the value of cash assets is riskless). The larger the bank—the greater the stock of deposit contracts and loan contracts—the larger the required capital and cash cushions it will need.

By raising equity capital, the bank has solved one set of problems, but in doing so, it has created another. Why would anyone want to invest in this capital cushion? In the first place, the banker—or more commonly a group of bank insiders (the founders, who are often its managers, as well as its directors)—is asking equity investors to trust them. But, what is to keep the bank insiders from losing the investor’s money, by taking wild bets or embezzling it? In technical terms, the bank insiders are offering minority shareholders (the outside equity investors) “cash flow rights” (the right to be paid, as a function of income earned), but not

would have to write contracts bearing even higher interest rates with debtors; but the only debtors who would be attracted to a such a contract would be those engaged in a venture promising a return even higher than the interest rate, which is to say one that is speculative, and extremely risky.

giving them “control rights” (voice in running the firm). The bank insiders may come up with various ways for the minority shareholders to monitor their activities, but inevitably there will be an asymmetry: the bank insiders will have control rights far in excess of their cash flow rights.³ In the second place, the bank insiders are asking investors to put their *entire stock* of wealth at risk: if the enterprise fails, each and every partner is singly liable for all of the debts of the firm.⁴ Making matters worse, the bank is, by design, a very highly leveraged enterprise: it borrows money from one group of people and then lends it to another group of people. Minority shareholders are therefore putting themselves at enormous risk. In short, the bank insiders need minority shareholders to invest their wealth in the bank in order to make it stable enough to attract depositors, but by raising equity from outsiders they have simply transferred the risk from the depositors to the minority shareholders!

³ In the United States, bank insiders are separated into two groups: the managers and the board of directors. In theory, the managers report to the board, who are the agents of the minority shareholders. In practice, boards are often captured by managers. There is a substantial corporate governance literature examining conflicts of interest between stockholders and “entrenched” managers (who may engage in self-serving conduct in conflict with the interests of stockholders, including insufficient use of debt finance, insufficient risk taking, and a low supply of effort). In the United States, and similar economies, where minority shareholders are protected from fraudulent behavior in the management of corporate assets (e.g., sales of corporate assets to firms controlled by management at below market prices – typically referred to as “tunneling”) this is the main agency problem of corporate governance. In developing economies, where protections for minority shareholders are weak, and where laws against fraudulent transfers are less prevalent or less enforced, managers are typically the controlling shareholders of the firm: both the board and the management tend to be drawn from the families that initially founded the bank and continue to have a controlling interest. Here, tunneling and other fraudulent behavior tend to be the focus of corporate governance failures, whereby majority stockholders/managers steal value from the firm, thereby transferring wealth from minority shareholders to controlling shareholders.

⁴ This is why business partnerships, with unlimited extended personal liability, tend to limit the number of partners, why they create complicated covenants and conditions for partners, and why partners engage in costly monitoring of one another.

Why on earth would any equity investor put himself in such a risky situation? There are several ways to solve this problem. Bank insiders can reassure outsiders about the riskiness of the loan portfolio, at least in part, by pledging a large amount of their own wealth as equity capital. The legal system can also provide protections against embezzlement and self-dealing by insiders at the expense of outsiders. Finally, limited liability of stockholders can reduce the extent of the risk borne by outside investors. The Scots invented the legal concept of limited liability banking in the 18th century. Limited liability means that the equity investor cannot lose more than the amount of capital he has actually invested in the bank.

Without some combination of these protections, it is very difficult to attract outside equity investors, and thus very difficult to grow the scale of a bank much beyond that of a medieval money-lender—someone who could make loans and charge interest, but not mobilize savings en masse. As it eventually developed, the limited liability bank broke the size barrier by bounding a shareholder's risk in proportion to the capital that he invests. Not surprisingly, the idea spread beyond its origins in Scotland and is now the basic organizational form of large-scale business enterprise around the planet. In fact, in the vast majority of countries the first enterprises to seek charters granting their shareholders a limit on liability were banks: the special limited liability acts for banks typically antedated general incorporation laws in most countries by decades.⁵

⁵ This is not to say that limited liability is an absolute requirement for the existence of banks. If the risk faced by minority shareholders is sufficiently low, they may deploy their capital even without it. There were, for example, large banks chartered without limited liability in Scotland in the 18th and 19th centuries—a subject to which we shall return in Chapter 3. It is to say, however, that as a general rule, outside investors are reluctant to deploy their capital unless their liability is limited. It is also not to say that limited liability necessarily requires

A bank cannot simply declare that its shareholders have limited liability or other legal protections. Only the government can offer minority shareholders legal protections, including limited liability. The reason is straightforward: governments create and enforce laws; governments create and enforce privileges, such as charters that grant limited liability. A charter, we hasten to point out, is not just a license; it is a contract between the bank and the government. On the one hand, that contract creates a set of obligations—such as taxes on bank profits or capital, requirements for banks to maintain sufficient holdings of government fiat currency or government bonds, and requirements that banks submit to government supervision of their operations. On the other hand, that contract confers a set of valuable privileges, such as the right to create money that serves as a legal tender for the payment of public and/or private debts, the right to hold government deposits, and, of course, limited liability for its shareholders. Not surprisingly, potential bankers will pay handsomely for a charter—especially if the government limits the number of bank charters granted. Scarce charters not only make banks more profitable, they also help fund banks at the time they are chartered. If charters are scarce, then the equity capital of the bank will command a premium over its “paid in” book value. That premium is a form of free financing for bank insiders, an

that the shareholders’ risk is bounded to the value of their invested capital. Limits on liability varied across countries and time: there were systems of double and triple liability, for example, where stockholders were liable for capital calls equal to one to two times their paid-in capital; there were systems in which the limits on a stockholder’s liability increased the longer he or she owned the shares; and there were systems that resemble the one that currently exists in the United States, in which liability is fully limited to the invested capital. It is to say, however, that the equity structure of banks generally involves the creation of some limit on the liability of stockholders.

extra buffer of value that reduces the amount of paid in capital needed to attract depositors to the bank.

The Three Property Rights Challenges of Banking

The granting of scarce bank charters helps to solve the problem of funding banks, but it exacerbates three other fundamental property rights problems that have been looming in the background, but which now come to the fore precisely because the government has taken an active role in the creation of the bank as a legal entity and in its subsequent regulation and supervision.

First, the banks and the government have to create mechanisms that either prevent the government from expropriating the banks once they are created, or that compensate the banks' investors and depositors for accepting the risk that they may be expropriated. The bank cannot now hide its wealth from the government, as it could before it had a charter; it has given the government the right to inspect or supervise its operations, as a condition of the charter.

Second, minority shareholders and depositors have to create mechanisms that either prevent the bank insiders from expropriating their capital through fraud or “tunneling”—the expropriation of the value of the bank by channeling bank resources to firms owned by bank insiders—or that compensate them for accepting the risk that they may be expropriated. The

charter from the government did not address this problem; it only limited the liability of shareholders toward depositors and other creditors.⁶

Third, the bank has to create mechanisms that prevent bank borrowers from expropriating bank insiders, minority shareholders, and depositors by reneging on loan contracts, or it has to create mechanisms that compensate those groups for accepting the risk that they may be expropriated by borrowers.

These problems are not independent of each other. Enforcing laws against fraud or tunneling by the bank insiders, and enforcing loan contracts with debtors, necessarily involves the police power of the government. But any government strong enough to carry out those functions is also strong enough to expropriate the bank. For a banking system to operate effectively, minority shareholders and depositors need the government to pass and enforce laws against tunneling, ensure that banks fulfill their contractual obligations to them, and create accounting standards and regulatory and supervisory agencies that improve the chances that the bank can be evaluated accurately by outsiders and disciplined by the competition among banks for outsiders' funds. Similarly, the bank needs the courts and police to enforce loan and other counterparty contracts; without them, for example, collateral cannot be repossessed. Finally, depositors and shareholders may look to the government to reduce their risks by creating protections against loss through deposit insurance or bank bailouts. Can a

⁶ Depositors and minority shareholders do not have to receive the same compensation for risk. Depositors are trapped: they need the banks as a means of payment, as thus may be forced to accept negative real returns. In this case, they limit their exposure by limiting their deposits to the bare minimum necessary to meet current payments. We return to this issue later in this chapter.

government with the physical enforcement capacity to do all these things be trusted not to disenfranchise all the various contracting parties in its own selfish interest?

Conflicted Government

Further complicating the government's problem of credibly solving these three key property rights challenges, the government is not a disinterested, independent party. In fact, the parties in control of the government typically have multiple conflicts of interest when it comes to the banking system. First, they may have preexisting economic stakes, the value of which depend on the ways financial rights are structured and enforced. Second, the individuals in control of the government seek to remain in power, and in order to do so they need to finance the state. That is, they simultaneously supervise and regulate the banks—and borrow from, or tax, them. Third, at the same time that the government is charged with enforcing loan contracts against debtors, the parties in control of the government may need the political support of those same debtors. That need for political support may erode the government's commitment to enforcing repayment. Fourth, at the same time that the government is charged with liquidating failed banks and allocating their losses according to the laws governing the suspension of bank operations and receivership, the parties in control of the government may need the political support of depositors, whose savings are at risk. In short, government officials may have tremendous incentives to behave opportunistically toward the property rights system that underpins the banking system.

As if this were not problematic enough, the government is not a unitary actor. The actual work of government is done by bureaucrats whose interests rarely coincide with

maximizing the wealth of the state, much less maximizing social welfare. In fact, the selfish objectives of bureaucrats may lead them to promulgate costly and ineffectual regulations simply in order to extract bribes from market participants to relax those same regulations.

Making matters worse still, the government has multiple margins through which it can behave opportunistically. It can raise funds by expropriating the banks outright; it can seize resources by printing fiat money that it uses to purchase goods and services; it can borrow from the banks and then renege on its debts; it can impose low- or zero-interest reserve requirements, forcing banks to hold government bonds, currency or deposits at the central bank that yield negative real interest rates; or force banks to direct low-interest loans to government enterprises or other government-favored borrowers (who, in turn, find ways to reward the government for its favoritism). The government also has multiple channels through which it can be a source of largesse. It can grant favors to particular groups of bank insiders, depositors, or debtors that are deemed politically crucial to maintaining power. Examples of these actions by government include regulatory forbearance towards bank insiders, insuring depositors beyond the statutory limits, or forcing banks to forgive debts.

Even worse, in response to adverse economic shocks, the parties in control of the government may decide that the path of least resistance politically is to favor all of these groups—insiders, minority shareholders, depositors, and debtors—through rescues and bailouts that come at the expense of taxpayers. The government may signal these groups that it will do so by creating a government safety net for credit market participants, most often in the form of a deposit insurance and bank resolution system, which is typically designed to minimize the visibility to the public of the costs of bailouts and the allocations of losses.

The existence of a system of deposit insurance, ironically, generates incentives for the parties in control of the government to use the banks to reward politically-favored constituencies. In the absence of deposit insurance and government intervention into bank loss-sharing, the combination of the first-come first-served rule for depositors and the laws governing suspension of bank operations and receivership for failed banks, as enforced by the courts, determines the allocation of losses associated with delinquent loans and failed banks. In other words, the savings of depositors are at risk. Depositors therefore discipline bank insiders by withdrawing their funds when the insiders take imprudent risks—and it is this ability to do so that gives depositors control rights over the bank. The seniority of deposits, depositors’ rights to withdraw their deposits, and the transfer of control rights over banks in liquidation all have an essential social function, namely to encourage banks to be prudent and honest in managing their risks (Calomiris and Kahn 1991). The problem—if it may be called that—is that prudent lending practices may conflict with the desire of the parties in control of the government to channel credit to politically crucial groups. Deposit insurance paid for by taxpayer absorption of losses ex post “fixes” the problem: depositors no longer have an incentive to discipline the bank insiders, because their funds are no longer at risk.⁷ The bank insiders are now free to take riskier bets. This loss allocation arrangement may have adverse

⁷ We are not making the claim that deposit insurance necessarily increases bank risk or imposes on losses on taxpayers. It is possible, as a matter of theory, to design a deposit insurance systems in such a way that it does not increase bank risk taking or become a burden on taxpayers. For example, three antebellum U.S. states—Indiana, Ohio, and Iowa—operated deposit insurance systems that were self-funded and that actually increased bank stability. (See Calomiris 1990).

consequences in encouraging excessive risk taking by banks, but it can also entail large gains to politicians and bankers.

The reason that taxpayers are often left holding the bag in bailing out insolvent banks is that the people who are arranging the rules that govern risk-taking and loss-sharing do not have to obtain taxpayers' active agreement. Minority shareholders have to take the active step of deploying their capital. Depositors have to take the active step of placing their funds in a bank. Both are interested parties to the deal. Taxpayers, however, are just along for the ride. This does not mean that their preferences can be completely ignored: there is a limit to the size of the bag that they can be left holding. Imposing too great a loss on taxpayers can induce them to rise up against the parties in control of the government, as we discuss in Chapter Seven with respect to the decision of Mexico's citizens to throw out the PRI in 1997. It does mean, however, that the parties in control of the government, bank insiders, minority shareholders, and depositors have plenty of leeway to fleece taxpayers.

Three factors generate that leeway. First, as Stigler (1971) famously argued, it is difficult to coordinate opposition to any program that benefits few at the expense of many because of the transactions costs of political activity. Second, taxpayers are not able to identify easily the allocation of costs and benefits from bailouts. The allocation of loss in a bailout is not determined by the courts, but by a deposit insurance/resolution authority operating within the government under opaque circumstances and ad hoc arrangements—and that resolution authority can allocate the losses in ways that are not transparent to taxpayers. Third, because taxpayers are sometimes depositors, it will not necessarily be clear to

taxpayers whether they are better or worse off as a result of the government's opaque intervention.

This may all seem tremendously complicated—but that is exactly the point! Precisely because the parties in control of the government are operating on multiple margins across a number of different constituencies—bank insiders, minority shareholders, debtors, taxpayer-depositors, and taxpayer-non-depositors—it is difficult, if not impossible, for many of the individuals affected to calculate whether the actions of the government have made them better or worse off. People cannot monitor the government on each and every policy dimension that affects the value of their property rights. Both the intent of government financial policies and their actual economic consequences can be difficult to determine *ex ante*. This is especially the case if the government is simultaneously reforming multiple regulatory institutions, some of which potentially enhance the value of property rights and some of which reduce them.

The Politics of Financial Property Rights

The implication, we hope, is clear: the property rights system that structures banking is not a passive response to some efficiency criterion, demanded in an anonymous “market” for institutions; rather it is the product of political deals hammered out by coalitions of market participants, which are intended to improve the welfare of the members of those coalitions, not the society at large. Those deals are possible because the economic incentives of the bank insiders, minority shareholders, depositors, debtors, and taxpayers are not inherently aligned—and because the parties in control of the government have ample incentives to exploit those differences in incentives for their own political or economic ends.

The allocation of political power determines the composition of coalitions and the deals that they structure. The deals determine which laws are passed, which judges are appointed, which groups of people have which licenses to contract with whom, for what, and on what terms. They also determine the distribution of the burden of taxation, the allocation of public spending, the regulation of entry, the licensing of banks, the supervision of publicly traded companies, and the flow of credit and its terms. These bargains are exceptionally complex—they involve sets of explicit trades and implicit alliances—but at root they are about the creation and distribution of economic rents and the maintenance of political power.

Given the number of agents with cash flow or control rights in banks, the heterogeneity of bank debtors across countries and time, and the diversity of political systems, the set of potential coalitions and deals is quite large. Several characteristics are, however, clear as a matter of logic—and as we shall see in the chapters that follow, as a matter of evidence, as well.

First, while the number of possible combinations of bank insiders, minority shareholders, taxpayer-depositors, taxpayer-non-depositors, debtors, and parties in control of the government is very large, not all *combinations* can constitute a *coalition*. A coalition is a group of people who agree on an arrangement to allocate the totality of the rent created by the rules that govern an economic system, and who are in a unique position to observe the actual allocation of the rent being earned. That means that many possible combinations of parties are not viable: forming a coalition with one group may require the parties in control of the government simultaneously to behave opportunistically against other groups.

Second, authoritarian political systems are likely to generate different sets of coalitions than democracies. In an autocracy, the government cannot make a credible (“time-consistent”) commitment to the bank insiders, minority shareholders, and depositors that it will not expropriate them. In addition, in an autocracy, debtors (and individuals who wish to be debtors) cannot influence the parties that control the government through the ballot box. This means that it is difficult for potential debtors—firms and households that seek credit, but cannot obtain it—to force the government to liberalize the rules governing the allocation of charters.

Third, precisely because creating a stable banking system that allocates credit broadly requires three distinct property rights problems to be solved, transitions from autocracy to democracy do not imply an automatic and instantaneous reorganization of the banking system simply because debtors now have the right to vote. This is a point that cannot be emphasized strongly enough: the property rights problems that are endemic to all banking systems are mitigated by creating institutions, and, by definition, institutions (which act as repositories of rules and discretionary authority, created by previous rounds of political bargains) change slowly. This is particularly the case with regard to the institutions that permit banks to sanction debtors who attempt to expropriate the bank: efficient courts, honest police, and accurate property and commercial registers cannot be created with the stroke of a pen; they take considerable time (and require considerable expense) to develop—and as we will discuss below, there are good reasons why they tend not to develop under authoritarian political systems. The implication is that regime change will bring about a reorganization of the banking system, but that reorganization will occur with a considerable lag.

Fourth, there are a range of possible coalitions within both democracies and autocracies—and the results of these coalitions are not neutral with respect to the industrial organization of banking. That is, democracies can generate coalitions that will undermine the banking system (imagine, for example, a coalition between the parties in control of the government and bank debtors, in which political support is exchanged for a moratorium on debt repayment). Autocracies can generate banking systems that are stable (imagine a coalition between an autocrat, bank insiders, and minority shareholders in which loans to the government are exchanged for binding limits on entry that raise rates of return high enough to compensate the insiders and minority shareholders for the risk of expropriation). Thus, there is not a bimodal distribution in the industrial structure of banking between democracies and autocracies. Rather, each generates different sets of *possible coalitions* such that, on average, democracies are more likely to give rise to stable banking systems characterized by relatively open entry and fewer restrictions on banks, while autocracies are more likely to give rise to unstable banking systems characterized by relatively limited entry and more restrictions on banks.

In order to understand the central tendencies of each type of political system (democracy or autocracy), we sketch out below the logic by which each system structures the incentives of both political elites and the various participants in the banking system. These models are meant to be understood as heuristic devices that allow readers to understand the country case studies that follow. They are simplified representations of reality—not stand-ins for reality itself.

One Equilibrium: Banking as Crony Partnership

We start with a very simple organizational structure of government, one in which the parties that wield political power have tremendous authority and discretion. As shorthand, we will refer to this government as an autocracy, and the parties that wield power as the autocrat. But this shorthand is not to be taken too literally. There are all kinds of political arrangements, ranging from hereditary monarchies, to military juntas, to single-party “democracies,” in which the parties that hold power operate with few constraints.

The autocrat faces a dilemma: on the one hand, he needs a banking system to improve his own chances of political survival; and, on the other, the fact that he is an autocrat discourages the population from deploying their capital in a bank over which he holds sway. Permit us to explain. An economy without banks grows slowly, which means that there is less for the autocrat to tax. An economy without banks also finds it more difficult to finance unforeseen fluctuations in public expenditures—such as quelling an armed uprising—through fiscal deficits. At times of high spending, the only options available may be to confiscate property or to print fiat money, which is effectively the same as a property confiscation, except the incidence of this confiscation falls on the holders of cash. In short, without banks, state finance is far more difficult—so difficult that it may jeopardize the autocrat’s survival, in the most literal sense of the term.

The autocrat cannot, however, simply will banks into existence. He needs private individuals to deploy their capital to finance the banks chartered by the government. What are their incentives for doing so, when the autocrat can expropriate their banks at will? In fact, the greater the autocrat’s need for a bank—which is to say, the more desperate his

financial situation—the more reluctant individuals will be to deploy their capital, because they know that their bank will be the first target of confiscation.

In sovereign financial emergencies, banks are particularly attractive targets of expropriation because of their rapidly redeployable liquid assets, and because of the multitude of government policies that can be used to expropriate, tax, or extort banks (outright nationalization, rules requiring directed loans, reserve requirements, transaction taxes, and the granting or revoking of valuable privileges). In the extreme case of nationalization, it is not hard to see why governments find banks a more attractive target than other firms. If the autocrat confiscates a farm, factory, or mine it is only valuable because it can produce a stream of income, and that happens if, and only if, he can figure out how to run it effectively. The autocrat has the option, of course, of selling the confiscated asset for immediate cash, but he has to do so at a steep discount—because the buyer of the asset knows that there is little that prevents the autocrat from expropriating it all over again. The autocrat is therefore in the position of a thief selling his goods to a fence: he will receive only a small portion of the value of what he has taken, because the buyer is bearing a serious expropriation risk. This problem does not loom large, however, when the assets being confiscated (or stolen) take the form of cash or other forms of liquid wealth: they are valued, without delay, at one-hundred cents on the dollar.

The autocrat therefore faces a commitment problem: given the desirability of expropriating banks (ex post), how does he get bank insiders—the individuals who provide the initial equity capital—to invest in banks in the first place? Promising not to expropriate the banks will not work. His unlimited authority means that no promise he makes is credible.

The implication is that he has to *compensate* bankers for the probability that they will be expropriated, which is to say that he has to raise the rate of return on their capital above what they would obtain in a competitive market without expropriation risk by providing them with what economists call rents—a stream of income from a scarce privilege. Those privileges must benefit the bankers in enough of the potential future states of the world that they are willing to supply funding willingly, even though they expect to be expropriated in other future states of the world. One obvious privilege is the bank charter itself—especially if there is only one charter being granted. Other privileges might include the right to collect taxes, run the mint (thereby earning seignorage), issue currency, or monopolize the export or import of particular goods (Maurer and Gomberg, 2004). In short, the autocrat and the bank insiders form a *coalition* in which the autocrat gets access to bank finance in exchange for, among other things, restricting entry into banking.

To take advantage of the potential profitability of their banking license, bank insiders must mobilize sufficient equity capital to serve as the “footings” on which asset and deposit growth are based. Bank insiders, of course, like to diversify their investments (rather than invest all their wealth in a risky bank), and in any case, often do not possess enough personal wealth to provide adequate bank footings on their own. Thus, bank insiders look to outside equity investors to supply some of their capital needs.

How can bank insiders get minority shareholders to buy stock in the bank? The insiders, of course, have every incentive to sell shares—because doing so reduces their exposure in the event that the autocrat decides to expropriate them or reduce their rents by renegeing on the privileges he has granted. The problem is that the prospective minority

shareholders know this as well! Minority shareholders also know that, to the degree that bank insiders have interests in non-financial enterprises, their incentive is to lend to those same companies (typically, on concessionary terms) rather than lend to the non-financial companies who are their competitors. In addition, minority shareholders know that in an economic crisis insider borrowers will not face harsh enforcement of loan contracts from their parent banks, and thus will have strong incentives to default on their loans—a subject to which we shall return shortly. In short, the minority shareholders face a problem of *double* expropriation: from the autocrat; and from the insiders, via concessionary insider lending and weak enforcement of insider delinquencies. Worse, they know that the insiders are in a coalition with the autocrat, which means that he may allow the insiders to expropriate the minority shareholders with impunity!

The minority shareholders are not sheep to be fleeced. These problems are predictable, given the rules of the game in which banks are operating, and minority shareholders will demand some form of *expected compensation* for the risk of double-expropriation. That compensation must take the form of a sufficiently high expected rate of return on their investment to compensate them for the risk of double-expropriation. For the bank to afford to pay that return to its stockholders, the special privileges awarded to the bank must be very lucrative indeed.

Is a competitively structured banking system possible under these circumstances? Yes, but it is not optimal from the standpoint of the autocrat or the bankers. In the presence of high expropriation risk, if free entry were allowed, the high returns demanded by stockholders would naturally constrain entry and the growth of the system. But a system that grants special

licenses and privileges to favored bank insiders, and restricts entry, as part of a larger, grand deal between bank insiders and the autocrat will dominate the free-entry equilibrium. The reason is that the grand deal between the autocrat and bank insiders can provide a share of rents to the government on a constant basis, thereby reducing somewhat the risk of expropriation by giving the autocrat a vested interest in the favored banks. That reduced expropriation risk will allow the banking system to be larger and less risky than under free entry, and the gains of that improvement would accrue as rents to the insiders and the autocrat.

The result of the combination of constraints that characterizes the autocratic government's banking system, therefore, is a coalition of interests composed of three parties: the autocrat, the minority shareholders, and the bank insiders. The autocrat receives a steady source of public finance (directed credit, reserve requirements, bribes, and taxes) and an occasional emergency source of funding (various forms of expropriation in times of fiscal stress). The bank insiders earn rents, first, as high-paid bank managers in a non-competitive market, and second, through the super-normal returns they earn in the non-financial enterprises that they own, which benefit from low interest charges on bank loans and special access to credit (which serves as a barrier to competition from other firms). The minority shareholders earn compensation in the form of super-normal stock returns for the double-expropriation risk they face.

Within this arrangement, what is the purpose of the state's role as a regulator and supervisor of banking system activities? Why, one may ask, is there a need at all for government regulation in this system? Why, if there are just the three parties to the coalition,

and a few favored banks, does the autocrat bother with banking regulations? The answer is that the existence of a framework of banking laws and regulations means that the parties to the coalition do not have to bargain over the distribution of rents at each and every decision point. The laws and regulations also serve as a signal to potential bank competitors: don't bother; you cannot compete with the favored insiders, because the laws are written in such a way as to make it nearly impossible to get a bank charter, and if you do get one, the regulators have ways of making you regret it!

Another party has to be induced to participate in the small and risky banking system that arises under autocracy: depositors. Bank depositors are in search of a low-risk, liquid means of saving. In a well-developed banking system, depositors tend to be highly selective, demanding a positive real return on their deposits, and intolerant of any significant risk. As we will see in our discussions of Scotland, England, the United States, and Canada, even from an early date, prior to any government protection of depositors, depositors earned small, positive real returns and suffered very low risk of loss. But, in an autocratic banking system, where returns to equity holders must be high, where loans to insider firms are subsidized, where governments and bank insiders extract significant rents, and where periodic fiscal crises result in expropriation, providing depositors with a positive real return on a low-risk deposit is much more challenging. How do the insiders and minority shareholders induce depositors to put their cash in the bank under these conditions?

One possibility is that government will try to protect depositors with taxpayer-financed deposit insurance. But the scale of that protection is inherently limited by the fiscal resources of the autocrat: governments that rely on banks as a primary means of fiscal support

cannot simultaneously tax depositors and subsidize them! Thus, deposit insurance will not be able to create a deep financial system within an autocratic banking equilibrium.

For the most part, therefore, depositors simply will not be attracted to banks in an autocratic banking system. Deposits in such systems often earn negative real returns and are subject to substantial risk of loss. In these systems, many depositors stay out of the formal financial system, holding foreign currency or commodities or other forms of wealth in lieu of bank deposits. But some deposit funds will still be supplied to banks, even in these adverse circumstances. After all, firms have to maintain bank balances in order to cover payrolls and accounts receivables; and some households must maintain minimal bank balances in order to execute certain payments. The transactions costs and legal constraints of avoiding the deposit market are prohibitively high for some purposes. A manufacturing firm in Zimbabwe, for example, cannot, as a practical matter, pay its workers with checks drawn on a bank in Ontario, Canada, and those workers cannot, as a practical matter, pay their phone bills with a check drawn on a bank in New York. In short, some depositors are trapped (what economists describe as a situation of highly “inelastic demand”).

The implications of trapped depositors are multiple—and they ultimately reinforce the tendency of the banks to allocate credit to the autocrat and the insiders. Precisely because they earn less than they would through alternative investments, depositors will not put any more of their liquid wealth into the banking system than they have to. Relative to the size of the economy, the deposit base will be small. The small deposit base will, in turn, mean that the banks have little need to search for new lending opportunities beyond the credit already

granted to the autocrat and the insiders. A small group of insiders will control all the bank credit in this society.

Now for the really bad news: this system of underdeveloped banks and scarce credit is self-reinforcing in the long run. Despite the lack of credit available to firms with highly profitable investment opportunities, there will be little tendency for the supply of credit to improve. Even informal, non-bank money-lenders will have a hard time supplementing the supply of credit in this repressed financial system. The reason is that autocratic banking systems also tend not to develop the legal foundations that would permit successful arms-length lending to occur.

Whether institutions that ensure the enforcement of arms-length loans will emerge depends on the preferences of the political coalition that runs the banking system. Banks and money-lenders develop costly systems of internal credit analysis only when they have to get into the business of making loans to non-related parties. Credit reporting services only develop when banks are willing to pay for their services, which is to say when they need to be able to sanction debtors that they can destroy their reputational capital in the case of default. Efficient property and commercial registers, which are necessary in order to credibly offer collateral that creditors would be able to repossess upon borrowers' default, only develop when there is a demand for them. The same is true for honest police and judges. The bottom line is this: the institutions that enforce contract rights are expensive to create and maintain. If banks do not need them, because they do not need to make arms-length loans, they are not likely to emerge.

In autocratic banking systems, not only are banks unwilling to incur costs that would facilitate their own ability to engage in arms-length lending, they may actively oppose the creation of improved legal and informational institutions that would facilitate arms-length financing by others. The reason is not hard to divine: any improvements in the institutions that permit arms-length financing would erode the rents enjoyed by the banker's own non-financial firms by virtue of their privileged access to credit. In Mexico in 1995, for example, one of us was involved in an initiative to improve the registration of collateral interests in inventories and accounts receivable (which is routine under the United States' Uniform Commercial Code registration procedures). That initiative, championed by the World Bank and other advisors, did not receive the support of Mexico's insider-lending banks. It was only once those banks collapsed, and were purchased by foreign banks, that significant reforms of collateral registration were adopted.

Not only will the "autocratic-crony" banking system allocate credit narrowly, it will also tend to be unstable. One source of instability is the fact that the autocrat faces few constraints on his authority and discretion. During normal times he receives rents from the banking system from a variety of explicit or implicit taxes on banks, including forced lending to the government and non-remunerative reserves at the central bank. If, however, the autocrat requires finance in excess of what he can generate from taxation, and in excess of what he can borrow from the banks, he has an incentive to expropriate the banks. This is a desperate move, because the autocrat is foregoing loans tomorrow in exchange for a one-time infusion of cash today. The advantage of this move, however, is that it may allow the autocrat to live to see tomorrow.

A second source of instability is the propensity of the bank insiders to lend to their own enterprises. During an economic crisis the insiders may have incentives to loot their own banks. The reason is simple, and the phenomenon has been extensively studied: the insiders have all of the control rights in the bank, but they have only a portion of its cash flow rights. To the extent that their cash flow rights in their other enterprises are of greater magnitude than their cash flow rights in the bank—which they almost certainly will be, because the bank, by design, minimizes the insider’s investment exposure—the insiders will use the bank to rescue their other enterprises. That is, they will make loans to their other enterprises, and then default on those loans (Akerlof and Romer, 1993; La Porta, López-de-Silanes, Shleifer, and Vishny, 1997, 1998; Rajan and Zingales, 1998; Johnson, Boone, Breach, and Friedman, 2000; Johnson, La Porta, López-de-Silanes, and Shleifer, 2000; Laeven, 2001; Bae, Kang, and Kim, 2002; Mitton, 2002; Habyarimana, 2003; La Porta, López-de-Silanes, and Zamarripa, 2003). Not surprisingly, there is a strong empirical relationship connecting weak rule of law (which is to say countries governed by autocrats), insider lending, and slow credit growth (Cull, Haber, and Imai, 2011).

When stripped of all of the institutional complexities, the political deal that shapes bank risk-taking decisions and loss allocations under autocracy can be seen as a rent distribution system, and the way rents are distributed varies according to the state of the world: Good Times, Bad Times, and Disaster. Good Times and Bad Times refer to business cycle fluctuations within the normal range of economic outcomes for that society. Disaster refers to bad times that are so extreme that the autocrat’s hold on power is in question. This is not just an exogenous economic state of the world, it is a political state of the world. The

same severe economic shock that would be considered a serious setback for a democratic government can be a matter of survival for an autocratic government: for the democratic leader the stakes are removal from office in the next election; for an autocrat the stakes can be a bullet in the back of the head.

The Autocrat shares in the cash flows earned by the banking system through the various means of extracting rent at his disposal (e.g., taxation, subsidized credit, reserve requirements, bribes). He earns more rent in Good Times than in Bad Times. During Disaster, however, his fiscal needs are so severe that his best move is to expropriate the banking system. If it exists, deposit insurance increases the variance of bank cash flows across the different states of the world. Therefore during Bad Times, the autocrat's rents are smaller in the presence of deposit insurance.

The Bank Insiders, like the government, earn greater rents during Good Times than during Bad Times, and deposit insurance increases the variance of their rents across the different states of the world. During disasters, expropriation makes even insiders a source of government rent extraction.

Minority Shareholders must be compensated for the risks they bear. They earn positive returns in Good Times, and negative returns in Bad Times. During Disaster their stakes are expropriated. The effects of deposit insurance on this group are ambiguous: its exact effects depend upon the degree to which the deposit insurance system is used to rescue insolvent banks during Bad Times .

Excluded Debtors are a constant source of rent extraction across all regimes and states of the world. This reflects the lost investment opportunities and competitive disadvantage they suffer as a result of being excluded from the credit system.

Depositors are also a source of rent extraction for the other players in the banking game. In all states of the world, and in both regimes with and without deposit insurance, depositors are trapped and earn low or negative real returns. Absent deposit insurance, depositors do not necessarily lose more in Bad Times than in Good Times because depositor discipline constrains bank risk taking. In equilibrium, Bad Times without deposit insurance do not necessarily translate into runs in which the deposits are expropriated. In the presence of deposit insurance, depositors also do not necessarily lose more in Bad Times than in Good Times, as government protection of depositors shields them from losses even though their banks may be insolvent. Under both the deposit insurance and non-deposit insurance regimes, depositors experience severe losses under Disaster. In Disaster states of the world even depositors get expropriated by fiscally-strained governments, despite de jure deposit insurance. This can take various forms, including inflating away the value of deposits by expanding the money supply, or “redenominating” them as we discuss in the Mexican case in Chapter Seven.

Taxpayers are only a source of rent extraction from the banking system when there is deposit insurance and/or government bailouts of banks. In that regime, in both Bad Times and Disaster states of the world, taxpayers suffer substantial losses. In Bad Times they are transferring resources to depositors, insiders, and the government. In Disaster, they are

transferring resources to the government and depositors, with the exact proportions determined by the extent to which the government expropriates the depositors.

Summary

The political institutions of autocracy generate a banking system with a credit market that strongly departs from any notion of allocative efficiency. Competition in the credit market is limited, with affiliates of bank insiders receiving favored and subsidized access to finance. The market will not allocate credit broadly because the suppliers of credit do not benefit from doing so. Banks will not search out talented entrepreneurs or hard working families; those groups will be starved for credit, while a few cronies connected to the bank insiders are awash in funds. The system will be small and inherently prone to crises: periodically, the insiders will expropriate the minority shareholders and the depositors; and periodically, the autocrat will expropriate everybody.

Observers confronting the realities of autocratic, crony banking systems for the first time may be outraged by their durable political bargains and their fragile, thin, and inefficient banks. These arrangements are likely to contradict readers' notions of fairness and incite them to decry their inherent inefficiencies and the disastrous implications they have for economic growth and social mobility. But these are not "failures" of the system. Rather, they are inherent in its design. Repressed, autocratic banking systems are far from optimal, but they may be the best that the societies that create them can do under the predetermined political circumstances.

A Different Political Equilibrium: Banking in an Open Political System

The defining feature of democracies are representative institutions. This means, by definition, that the parties in control of the government can be removed by election. All other things being equal, they have less authority and discretion than their counterparts in autocracies. It also means, by definition, that prospective borrowers are politically consequential because they have the right to vote. Prospective borrowers that are denied credit in the crony partnership equilibrium did not figure in our model of autocracy. In democracies, however, they can be members of the political coalition that creates and divides rents in banking.

This last feature of representative government—prospective debtors who can vote—creates a thorny problem for the insiders and their minority shareholders. On the one hand, debtors and potential debtors can demand that the parties in control of the government broaden access to credit and lower its cost by removing entry restrictions in banking or by removing restrictions on bank operations. That is, they have the potential power to undermine a coalition consisting only of favored bank insiders, their minority shareholders, and the parties in control of the government. On the other hand, however, the very fact that debtors can form a coalition with the parties in control of the government means that they are potentially in a position to expropriate the insiders, minority shareholders, and depositors. That is, debtors (e.g., farmers with high leverage facing the prospect of foreclosure in the United States in the 1920s, or sub-prime mortgage borrowers in the 2010's) can trade votes for the cancellation or rescheduling of debt contracts.

In the face of government interventions that facilitate transfers to debtors, bank insiders, minority shareholders, and depositors will seek to insulate themselves from the costs of debt forgiveness. Among other things, they may seek government insurance of bank deposits against loss, and subsidized loans or capital assistance to the bank from the government. In other words, a coalition of bank debtors, depositors, bank insiders, minority shareholders, and the parties in control of the government may choose to expropriate taxpayers—a subject to which we shall return at length in Chapter 9.

Thus, the existence of suffrage, in and of itself, is a necessary, but not sufficient condition for the creation of a stable banking system that allocates credit efficiently. Importantly, therefore, a political system in which suffrage is the *only* constraint on the parties in control of the government will be unlikely to produce a banking system characterized by high levels of entry and competition: only a sufficiently concentrated banking system may be able to defend itself politically in that environment. A banking system with very limited competition – where powerful bank insiders buy protection in the market for political influence – may be the endogenous economic and political response to the prospect of populist favoritism of bank debtors at the expense of banks.

This problem is not a distant theoretical abstraction. Indeed, “the tyranny of the majority” is a central concern in the design of democratic systems, because it puts the holders of all wealth at risk. That risk explains why long-lasting democratic political systems, paradoxically, include sets of institutions that are designed to protect the holders of wealth by granting them power beyond their numbers. Such protective institutions typically include restrictions on suffrage on the basis of property or income, electoral rules that favor centrist

political parties (e.g., first-past-the-post electoral systems for representatives in the United States), multi-cameral legislatures, indirect election of legislators and federal executives, judicial review of legislation, and bills of rights that protect property. Often safeguarding these institutions are rules that require legislative super-majorities in order to change any of the institutions (e.g., the rules governing Constitutional amendments).

Protective institutions work by creating multiple veto-gates in the creation of laws. But, it is not just the number of gates that a law has to get through that matters. It is the fact that societies tend to be riven by multiple issue areas, each of which has non-coincident constituencies—what political scientists refer to as “cross-cutting cleavages.” Getting past each veto gate requires that proponents exploit a cross-cutting cleavage to get sufficient support (e.g., A supports B’s law to teach creation science, while B supports A’s law that makes it harder for debtors to declare bankruptcy). Designing such grand bargains is the primary function of political parties. The more veto gates there are, and the more cross-cutting cleavages there are, the lower the probability that a law can be passed that puts a minority at risk—including a minority composed of bank insiders and minority shareholders.

The implication, we hope, is clear: suffrage is not enough to ensure a deep and competitive financial system; the holders of wealth need to be able to create sufficient veto gates so that their wealth will not be confiscated; but if wealthy bankers are too successful in protecting their privileges, then debtors will not be able to use their voting rights to demand that the parties in control of the government loosen the restrictions on bank entry or bank operations sufficiently to permit competition among banks. A successful coalition to ease restrictions on bank entry may require, ironically, the assent of bank insiders and minority

shareholders. As we shall soon see, when we look at the development of the U.S. banking system, privileged insiders are only likely to go along with relaxing entry barriers if they are compensated on some other margin. For example, circa 1800, an insider in one of the few Philadelphia banks might have been willing to go along with the relaxation of bank entry barriers, if he or his associates expected to get one of the new charters being issued for banks in Pittsburgh, where there, as yet, were no banks. Striking such a deal is the essential political foundation of a fully developed banking system.

The problem, however, is that while it is possible to strike such a deal at any one point in time, it is very difficult to make those deals time consistent. The number of constituent groups is large, and each group has incentives to incessantly lobby the parties in control of the government to renegotiate the deal at the margin. Moreover, demands for changes in the deal can be cast, not unreasonably, as necessary because of changes in the technology of banking. The fact that rule changes tend to be incremental, can be cast as in the “public interest,” and are necessarily written in abstruse, technical language, makes creating a time consistent deal more problematic still. As a result, although democracies with multiple veto gates may be more likely than autocracies to produce stable banking systems that allocate abundant credit efficiently, there is no necessary guarantee that they *must* do so. Even in a highly developed democracy banking systems are fragile plants.