The Political Foundations of Scarce and Unstable Credit

Charles W. Calomiris

Henry Kaufman Professor of Financial Institutions, Columbia University

Presented at the Federal Reserve Bank of Atlanta

2013 Financial Markets Conference

Maintaining Financial Stability: Holding a Tiger by the Tail

Atlanta, Georgia

April 9, 2013

The Political Foundations of Scarce and Unstable Credit

Charles W. Calomiris and Stephen Haber *

March 2013

Preliminary Conference Draft

Please Do Not Cite without the Authors' Permission

* Calomiris is Henry Kaufman Professor of Financial Institutions, Columbia Business School, and a Research Associate of the National Bureau of Economic Research. Haber is A.A. and Jeanne Welch Milligan Professor in the School of Humanities and Science, Stanford University, and Peter and Helen Bing Senior Fellow at the Hoover Institution. This draft was prepared for the Federal Reserve Bank of Atlanta's April 2013 Financial Markets Conference, "Maintaining Financial Stability: Holding a Tiger By the Tail." The authors thank Larry Wall for helpful suggestions. This article is largely based on the authors' book, *Fragile By Design: Banking Crises, Scarce Credit and Political Bargains*, Princeton University Press, forthcoming.

I. Introduction

Since at least the days of Adam Smith, economists have noticed that some countries persistently create more prosperity than others. Banking systems also differs across countries. Banking system performance generally is indicated by two criteria: credit abundance and the propensity for crises. These measures display dramatic and persistent cross-country differences. Some countries create stable banking systems (that is, banking systems in which systemic crises are absent), while others do not. Some countries create banking systems capable of providing abundant arms-length credit, while others do not. A few countries manage to achieve both stable and abundant credit, while many others achieve neither.

Despite the importance of these persistent differences, the dominant theoretical models of banking do not include “country subscripts” that could explain them. With respect to persistent differences in the ability to supply credit, economists do understand that monopolization of banking will tend to result in the reduced supply of credit. And, they also understand that if the holder of the monopoly bank charter is able to lend exclusively to himself and his business associates, that setup will dramatically magnify the scarcity of credit experienced by arms-length borrowers. But economic theory does not explain why such a system of monopolized, insider banking would arise in some countries, given that its consequences for growth are so pernicious. In a world where banking institutions are designed optimally to achieve prosperity, monopolized insider banking systems should not exist.

With respect to banking crises, economists have constructed models capable of explaining such events – despite their undesirable consequences – but those models cannot explain why the influences they identify express themselves in some countries much more than in others. Theories of banking crises posit three aspects of banking, some combination of which are presumed to explain the origins of crises: bank structure, interbank connections, and human nature. Bank structure refers to the maturity and liquidity mismatch between banks’ relatively illiquid and long-lived loans and their relatively liquid and short-lived liabilities. Structural theories regard banking crises as the result of the inherent exposure of banks to “liquidity risk” arising from this mismatch. Interbank connections can exacerbate crisis risk through the problem of risk “externalities.” Individual banks will choose their holdings of liquid assets

and their leverage (debt relative to equity capital) based on their individual interests, and will not take into account the social costs that arise from the spillover effects (externalities) created by interbank connections. According to this class of crisis theories, the primary role of regulation is to "internalize" externalities by forcing banks to hold more cash assets and to maintain lower leverage ratios than they would choose voluntarily. The absence of sufficient regulation to internalize externalities, therefore, is the cause of banking crises. Other theories see banking crises as symptoms of human nature's myopia, which drives destructive cycles of greed and fear.

The basic facts of banking history reveal an obvious problem with all three theories of banking crises: they are *general* theories that conceive of banking crises as arising from aspects of banking common to all times and places. The frequency of banking crises is not the same across time and space, and therefore, all of the dominant economic theories of crises are incomplete.

Consider the example of Canada. Canadian banks historically had balance sheets like other banks, and participated in complex global interbank networks since the early 19th century. Yet Canadian banks, throughout their history, avoided systemic banking crises – with the exception of two short-lived suspensions of convertibility in 1837 and 1839 in response to crises originating in the United States. Moreover, prudential regulation was absent for much of Canada's history, as was a central bank (until 1935). According to the structural theory of crises, the exposure of Canadian banks to liquidity risk should have been higher than in many other countries, given that the Bank of Canada did not come into existence until 1935. According to the externalities theory and the myopia theory, the absence of activist prudential regulation in Canada during most of its history should have been associated with a higher frequency of banking crises, but it was not.

It is not that the core ideas of the economic theories of banking crises are "wrong" in identifying sources of risk in banking systems – on the contrary, there is substantial evidence that the influences that they identify are important. The history of banking crises, including the subprime crisis in the United States, show that the exposure to liquidity risk associated with banks' reliance on short-term debts to fund their operations can put financial institutions at risk if debtholders refuse to rollover those debts.

Similarly, the subprime crisis showed that counterparty risks in complex financial systems can lead to the transmission of adverse shocks among financial institutions.

The variation across countries and over time in banking instability does not show that theories of banking crises are irrelevant; rather, it shows that the economic understanding of systemic risk is *insufficient* for understanding when and where banking crises will actually occur. The structure of banks entails liquidity risk, but banks can overcome that risk by acting prudently – specifically, by maintaining sufficient amounts of cash assets, by managing their portfolio risk, limiting their leverage, and by entering into arrangements of mutual liquidity risk insurance with other banks. Interbank linkages can create externalities within the banking system, but again, proper risk management and the creation of effective institutions to manage those risks can be created to deal with those risks.

In our forthcoming book (Calomiris and Haber 2013), we show that the decisive influences for determining whether credit is abundant or scarce, and whether banking crises will be likely to occur, are *political*. Each country's banking policies are the result of what we call their particular Game of Bank Bargains. For example, countries often favor various sorts of limitations on the supply of credit because such limitations serve the interests of the dominant political coalition that controls the rules governing the banking system. Governments tolerate instability in banking as a means to channel subsidies to powerful interests. For example, if the government decides to establish generous safety nets protecting banks, and does not accompany those safety nets with effective prudential regulation, then banks will be encouraged, or sometimes even forced, to become less cautious in their management of risk. We emphasize that the problem is political, since politics will determine the extent of the safety net, the extent to which prudential regulation will be effective, and other influences on banks' risk preferences.

Furthermore, outcomes in the Games of Bank Bargains are constrained by the ability of the government to commit to abide by its agreements. The parties in control of the government are always part of the dominant coalition of interests that determine banking system outcomes. In the modern world of the past four centuries, bankers, bank stockholders and depositors rely on the State to grant bank charters, enforce loan contracts, and decide on the policies that regulate – or sometime expropriate –

them. All modern governments rely on chartered banks as key elements of their financing arrangements. This can be a major problem for banks because governments cannot commit not to expropriate, especially under particularly trying fiscal circumstances. The risk of expropriation is especially high in autocracies, where the check of democratic approval is absent.

In particular, the mutual dependency between banks and sovereigns contributes to the risk of banking crises driven by problems in state finances. Taking account of political influences that operate through the Game of Bank Bargains also gives rise to shocks that are generally not envisioned in economic approaches to bank risk. Wars, coups, and other stresses on sovereign finances can produce expropriations of banks by governments, or regulatory regimes intended to tax banks heavily – often through the inflation tax.

Just as important, the risk of expropriation that results from the difficulty of committing not to expropriate can motivate entry barriers in banking, which in turn, to limit credit availability. In autocracies, where expropriation risk is especially high, limiting entry into banking allows the government to raise expected returns for bank equity investors to compensate them for the expropriation risk inherent in the political environment.

Leaving politics out of the theory of banking crises truly omits the “prince” from the “play.” Political outcomes of the Game of Bank Bargains shape the rules under which banks operate, and define the shocks to which they will be subject. The dramatic political differences between the times and places where banking crises are either absent or frequent show that a political economy approach to modeling banking crises is essential to understanding the most basic facts about persistent differences across countries in the supply of credit and the probability of banking crises.

Section II explores some basic cross-country patterns with regard to current differences in the abundance of credit and the stability of banking systems, and lays out a bit of the logic of the political foundations of under-performing banking systems developed in Calomiris and Haber (2013). In Sections III and IV we discuss two pairs of examples drawn from more extensive discussions in Calomiris and Haber (2013), which show how exogenous historical circumstances that shaped politics produced

different outcomes with respect to the relative abundance of credit and the probability of banking crises. England vs. Scotland prior in the 18th and 19th centuries (Section III), and the United States vs. Canada from the mid-18th century to the present (Section IV). The countries within each of these pairs have much in common, and yet their banking system outcomes were dramatically different as the result of different political conditions that shaped banking. Section V concludes.

II. The Unavoidable Partnership between Banks and Governments

The political foundations of banking system underperformance are illustrated by striking differences across countries that are visible in the simplest statistics about bank performance. First, banking crises are not uniformly distributed across the globe. Of the 124 non-communist countries Calomiris and Haber (2013) analyze in their cross-country comparison of banking fragility, 41 were crisis-free from 1970 to 2010. Sixty two countries had one crisis. Nineteen countries experienced two crises. One country underwent three crises, and another weathered no less than four crises. That is to say, countries that underwent systemic banking crises out-numbered countries with stable banking systems by two-to-one; and 17 percent of the countries in the world appear to have been preternaturally crisis prone.

The non-random nature of this pattern is underlined by the fact that the country that experienced the most crises was Argentina, a nation so badly governed for so long that its political history is practically a synonym for mismanagement. Its close competitor for most crisis-prone country (with three crises since 1970) is the Democratic Republic of the Congo, the nation whose brutal colonial experience served as the inspiration for Joseph Conrad's *Heart of Darkness*, which was governed after independence by one of the third world's most long-lived and avaricious despots (Mobuto Sese Seko, who ruled from 1965 to 1997), and whose subsequent history is a template for tragedy.

The 19 countries that had two banking crises are also far from a random draw. The list speaks for itself: Chad, Nigeria, the Central African Republic, Cameroon, Kenya, the Philippines, Thailand, Turkey, Bolivia, Ecuador, Brazil, Mexico, Colombia, Costa Rica, Chile, Uruguay, Spain, Sweden, and…the United States. One of the striking features of this list is the paucity of high income, well-governed

countries on it. Of the 124 countries in our dataset, one-third are categorized by the World Bank as being in the high income group. But only three of the 21 crisis prone countries, 14 percent, are in this group. This suggests that, for the most part, being crisis-prone is connected to other relatively undesirable traits and outcomes. But that raises another troubling question. Why is the United States on this list?

Equally problematic are banking systems that provide too little credit relative to the size of the economy—a phenomenon known as under-banking. This unfortunate outcome, too, appears not to be randomly distributed. Consider the striking contrast between Canada and Mexico, the two NAFTA partners of the United States. From 1990 to 2009, private bank lending to firms and households averaged 97 percent of GDP in Canada, but in Mexico the ratio was only 19 percent. The stark difference between Canada and Mexico captures a recurring pattern: poor countries have much smaller banking systems than wealthy countries. In fact, the most extremely under-banked countries, where the ratio of bank credit to GDP only averages roughly 12 percent, are among the World's poorest, including Chad, Haiti, and Nepal. The most credit-abundant group of countries, where the ratio of bank credit to GDP averages 83 percent, are among the World's richest, including Denmark, the Netherlands, and New Zealand.

Crucially, there is also substantial variance across countries within groups defined by per capita income, which suggests that the amount of credit extended within countries is not solely a function of demand for credit, but also reflects constraints on the supply of credit. In other words, the fact that some countries with the largest banking systems within each income group extend much more credit than others in the same income group (or even the next highest income group) suggests that many countries are under-banked after controlling for income differences. For example, Mexico appears to be under--banked even by the standards of Upper Middle Income Countries. Similarly, the United States appears to be under-banked relative to other High Income Countries, such as Canada, which has nearly twice the U.S. ratio of private bank credit to GDP.

If there are few countries that have been free of banking crises since the 1970s, and if much of the world is under-banked, how many countries enjoy plentiful credit and banking stability? Answering this question requires us to draw a line between those economies where bank credit is abundant and those

economies that are under-banked. If we set the dividing line at those countries whose average ratio of bank credit to GDP is one standard deviation above the mean (81 percent), which corresponds to the level of Australia, and define a stable banking system as being free of systemic crises since 1970, we arrive at a shocking answer: only six out of 124 countries—five percent —meet those criteria.

There is a huge literature in financial economics and economic history showing that stable and abundant bank credit are conducive to growth and equality of opportunities within a society. If creating a stable banking system capable of providing abundant credit to talented entrepreneurs and responsible households is such a good idea, then why are such banking systems so rare? Our answer to this question is that differences in the fragility of banks and the scarcity of bank credit across countries persist because they reflect durable differences in the ability of political institutions to limit arbitrary actions by the parties in control of the government and rent seeking by interest groups. The crux of the problem is that all governments face inherent conflicts of interest when it comes to the operation of the banking system, but some types of government—most particularly democracies whose political institutions limit the influence of populist coalitions—are better able to mitigate those conflicts of interest than others.

These inherent conflicts of interest are of three basic types: 1) Governments simultaneously supervise and regulate banks, and look to them as a source of government finance; 2) Governments enforce the 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; and 3) Governments allocate losses among creditors in the event of bank failures, but they may simultaneously look to the most numerically significant group of those creditors—bank depositors—for political support. The implication, we hope, is inescapable: the property rights system that structures banking is not a passive response to some efficiency criterion, but rather it is the product of political deals that determine which laws are passed, which groups of people have licenses to contract with whom, for what, and on what terms. These deals are guided by the logic of politics, not the logic of the market.

The fact that the property rights system underpinning banking systems is an outcome of political deal-making means that there are no fully "private" banking systems; rather, all modern banking is best

thought of as a partnership between the government and a group of bankers, and that partnership is shaped by the institutions that govern the distribution of power in the political system. Government regulatory policies toward banks reflect the deals that gave rise to those partnerships, as well as the power of the interest groups whose consent is politically crucial to the ability of the parties in control of the government to sustain those deals. 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 in the Game of Bank Bargains are the actors with a stake in the performance of the banking system: the parties in control of the government, bankers, minority shareholders, debtors, and depositors. The rules governing play are set by the society's political institutions: those rules determine which other groups have to be included in the government-banker partnership; or, alternatively, who can be left out in the cold because the rules of the political system make them powerless. Coalitions among the players 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, the permissible activities of banks, and the allocation of losses when banks fail. What is at stake in the Game of Bank Bargains is, therefore, the distribution of the benefits that come from a system of chartered banks. The parties in control of the government always receive a share of those benefits, and the coalition that forges a partnership with the government splits the remainder.

The struggle among political coalitions determines who gets to play what roles in the financial system; that is, who is granted what kind of banking charter, and which groups of borrowers get government-favored access to credit. A central aspect of the Game of Bank Bargains, therefore, is deciding the rules for entry into banking. It is rarely the case that government chooses a fully "open-access" chartering regime. Being selective about who gets to be a banker, how much they are allowed to lend, and to whom, are crucial elements of the game.

In Calomiris and Haber (2013), we seek to understand the process by which different sets of rules of play emerge in different countries, as well as to understand how the players operate within those differing sets of rules. We show how differences in fundamental political institutions across times and places produce differences in the rules of the game, and how those politically-based rules sometimes result predictably in stable and plentiful bank credit, and sometimes in unstable and scarce bank credit, or (as in the case of the United States), in unstable and plentiful 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 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.

We summarize some of those narrative histories in Sections III and IV below, but before turning to them, it is also worth noting that political differences across countries are powerful predictors of the cross-sectional differences in the abundance of bank credit and the fragility of banks noted above. The "Happy Six" countries referenced above, which provide abundant credit without crises, have much in common politically. They are Australia, Canada, Hong Kong, Malta, New Zealand, and Singapore. Note that three of these countries – Hong Kong, Singapore, and Malta – are city-states or small islands, rather than large countries. We would hypothesize that it should be much easier to avoid banking bargains that are harmful to the general society (that is, scarce credit or banking instability) in a society where lack of

economic diversity makes for much simpler politics. What coalition might be formed among agrarian populists, small bankers, and politicians representing rural districts in a country that has virtually no agricultural sector, such as Hong Kong or Singapore? Indeed, in an entrepot city-state, the interests of manufacturers, bankers, and politicians are naturally aligned: they must maintain a stable, efficient, open economy, or become so poor and backward that they get swallowed up by a neighboring country. There is simply too much at stake in terms of the survival of the economic and political elites, and too few other interest groups lobbying for alternative uses of the banking system, to push the society away from a path that leads to stable and abundant credit.

The three large countries on the list of the "Happy Seven" – Australia, Canada, and New Zealand—share two features. First, they were all part of the British Empire. Second, they are among the world's most stable, and long-lived democracies. Political scientists often use the "Polity Score" (a metric of political competition and political institutions) to measure the extent of democracy and autocracy.[1] Although single, quantitative, measures of democracy are imperfect, they provide a useful means of capturing important differences among political systems, especially when one takes long-run averages of yearly scores. The polity scores for the period 1925-2010 for Australia, Canada, and New Zealand all are 100 (the maximum). The average polity score for that period worldwide is 52.6. Indeed, even if we include the city-states and small islands that are part of the "Happy Six" we are struck that there is only one autocracy in the group, Singapore (with a Polity Score of 40 on average since independence from Britain).

The three large countries on the list of the "Happy Six" have something else in common: the structure of their political systems tended to mitigate the ability of populists and bankers to form coalitions that disadvantage everyone else. In Section IV below, we will discuss the role of centralized economic policy-making and an unelected Senate in structuring the Game of Bank Bargains in Canada, and the role of a large French population in encouraging that political structure. Similarly, New Zealand's political institutions developed under the threat of a majority that might have been hostile to the

[1] See Calomiris and Haber (2013), Chapter 14.

commercial aspirations of its British settlers (the Maori). Thus, in order to constrain the influence of a numerical majority whose interests might not be aligned with an English minority, New Zealand established veto gates in the decision structure of the government, including an unelected upper house (which remained in existence until 1950) and a Governor with veto power over legislation. Australia did not have an unelected upper house, but we note that it became independent from the United Kingdom in stages; it was not until 1986 that it achieved full independence from the British courts and Parliament. Moreover, like Canada and New Zealand—and unlike the United States—Australia's constitution granted the national government centralized control over economic and banking policy; populists could not form coalitions with bankers and then enact policies to their liking by winning statehouse by statehouse, as happened in the United States throughout much of its history.

What happens if we set a standard for success that only requires countries to have been crisis-free since 1970 and have an average level of bank credit to GDP equal to the mean for all countries? The list of successes now expands to 13 countries: the "Happy Six" are joined by the Bahamas, Barbados, Belize, Brunei Darussalam, Macao, Mauritius, and South Africa. Six of the additions are small islands or city-states. With the exception of Macao, all of the countries on the list of the "Happy 13" were British colonies or protectorates, and all of them achieved their independence peacefully.

South Africa is the additional large country on the list of the "Happy 13." Like Canada and New Zealand, South Africa is a democracy that for most of its history gave English settlers power beyond their numbers. Indeed, the 1910 constitution of the Union of South Africa, as well as its subsequent revisions and enabling laws, went to great lengths to disenfranchise the country's numerically-dominant African and Indian populations. The British in South Africa, ironically in cooperation with the Boer's who they had compelled to be part of the Union by force, did not just create an unelected and mal-apportioned Senate (the way they did in Canada and New Zealand), they basically whittled away at the laws governing the franchise so that by 1956 "blacks" and "coloureds" had virtually no political voice at all. South Africa's 1997 constitution finally established full and equal suffrage for all South Africans, but it also contains a number of institutions designed to limit populist currents. First and foremost among these is a

weak president who is not directly elected by the population, but is indirectly elected by the lower house of parliament. His right to veto legislation is then limited by a constitutional court. Second, the upper house of parliament, which must approve all legislation, is also not elected, but is appointed by the provincial legislatures, with each province having the same number of representatives regardless of its population. Moreover, depending on the type of bill under consideration, the passage of legislation by the upper house may require a super-majority. In fact, under some circumstances, bills in the lower house must also be passed by super-majorities. In sum, although South Africa's political institutions were not crafted with bank regulation necessarily in mind (much as was also the case in Canada and New Zealand), those institutions make it very hard for a coalition of populists and bankers to enact bank regulatory policies that benefit them, at the expense of everyone else.

If the "Happy Six" and "Happy 13" tend to be either city-states, small islands, or democracies with institutions that limit populist currents, then what do their mirror images look like? That is, what are the characteristics of the countries that are preternaturally crisis-prone and that provide very low levels of private credit relative to their GDP? These countries are also not a random draw from the world. If we invert the "Happy Six" threshold, and ask which countries have had at least two systemic banking crises since 1970, and also have had an average ratio of private credit to GDP at least one standard deviation below the mean since 1991, we find that there are two "very unhappy" countries: Chad and the Democratic Republic of the Congo. It would not take lengthy argumentation to show that these are among the least democratic countries in the world (their average polity scores since 1970 have been 30 and 26, respectively). If we invert the rules for the "Happy 13," and ask which countries have had at least two banking crises and also have had a ratio of private credit to GDP equal to or less than the mean since 1991, we obtain a list of 16 "unhappy" countries—most of which have either been persistently autocratic, had short experiments with democracy that ended in coups, or have only recently begun to experiment with democracy after long periods of autocratic rule. In addition to Chad and the Democratic Republic of the Congo, the "Unhappy 16" includes Argentina, Bolivia Brazil, Cameroon, the Central African Republic, Colombia, Costa Rica, Ecuador, Kenya, Mexico, Nigeria, the Philippines, Turkey, and

Uruguay. Since 1970, the average polity score for this group is only 58; it would be lower still if we took the average over a longer time span. Of the 16 countries in this group, only three maintained average polity scores above 80 for the period 1970-2011: Costa Rica (100), Colombia (88) and Turkey (82). In sum, we are struck by how few exceptions there are to what appears to be a general pattern: countries with stable banking systems that provide abundant credit tend to be stable democracies with institutions that limit the opportunities for bankers and populists to form rent-seeking coalitions; countries with unstable banking systems that provide low levels of credit tend to either be autocracies, democracies of very recent vintage, or democracies in which the institutions that limit rent seeking are weak.

A central question emerges from these comparisons: Is politics truly the cause of persistent differences in banking performance, or are politics and banking simply indicators of some other common causal factor? In our view, narrative histories provide the best means for investigating the causal links between politics and banking. The following summarizes some of the narratives in Calomiris and Haber (2013) that we think prove the importance of politics as a causal factor in banking system performance.

III. England and Scotland

Beginning in the 1690s, as a consequence of the Glorious Revolution in which William of Orange became king of England and Scotland, both countries chartered banks. Not only were these banks were chartered under the same British crown, after 1707 Scotland and England were governed by a single Parliament (as the result of Scotland's decision to dissolve its separate Parliament). Yet political decisions governing the structure of the two countries' banking systems resulted in two completely different banking systems until the middle of the 19th century. The structural differences in banking policies had important implications for the relative abundance of credit and the propensity for bank failures.

In England, from 1694 to 1825, the Bank of England was the only bank that was allowed to take the joint-stock corporate form; all other banks had to be organized as partnerships, and were legally constrained to have no more than six partners, which meant that they could not grow to a sizable scale. Banks were also subject to usury laws, which discouraged them from expanding their circle of borrowers

(if a bank cannot charge a new client a higher interest rate to compensate for the fact it does not know much about him, it will not lend to that client at all). The English government exempted itself from these usury laws, thereby channeling credit to itself, rather than the private sector.

England's repressed banking system had adverse consequences for the private sector, but it served the State well by providing a captive source of savings to fund the sovereign's war needs. England industrialized in spite of the fact that the government had quite consciously constrained the growth of the banking system in order to favor itself through its exclusive partnership with the Bank of England. This constrained capital accumulation by the private sector during the early years of the industrial revolution, as investment was financed out the pockets of tinkerers and manufacturers, not through bank lending.

Not only did repressive policies limit bank credit, they also made the English banking system unusually unstable. Prior to 1870 England had one of the most crisis-prone banking systems in the world. The instability of English banking reflected the limits on the size, branching and diversification of banks. The limited private access to credit and unstable structure of English banking produced public complaints, which prompted the government to pursue risk-subsidizing policies in support of the brokered bills market (administered through government pressures on the Bank of England) to make private credit cheaper. The moral-hazard consequences of these policies encouraged excessive risk taking by small banks and bill brokers, which further contributed to banking instability.

In sharp contrast to England, by the middle of the 18th century Scotland had developed a highly competitive and innovative banking system, and enjoyed abundant credit and stable banks, which provided credit to all sectors of the economy through wide-reaching networks of branching banks. From 1694 through 1825, while England's banking system consisted of the Bank of England and the small country banks and goldsmith banks that operated on a small scale and under restrictive regulation, a completely different sort of banking system operated nearby in Scotland. The Scottish system, in sharp contrast to the English one, was the very model of competition, innovation, accessibility of credit for the private sector, and stability. In all of those respects, Scotland's system was a constant reminder to its neighbors in England of what a banking system could be, and what the English banking system was not.

The comparison was all the more galling to the English considering how relatively backward the Scots had been at the turn of the 18th century.

Thanks in large part to its banking system, Scottish backwardness did not persist. Not only was Scotland at the heart of the British Enlightenment, it became central to the Atlantic trade between Britain and its colonies. Scottish manufacturing, husbandry, fisheries, and agriculture also thrived. Scotland's competitive system of banks, many of which operated networks of branches throughout the country, contributed to this success story by providing broad access to credit. A branch of an existing bank could be opened in a remote location with much lower overhead cost than establishing an entirely new bank, because the branch can rely on headquarters to provide managerial, accounting, and operational services; all the branch needed was a clerk and till cash.

Unlike the Bank of England, which initially operated only from its Threadneedle Street headquarters, or the small English country or goldsmith banks, Scotland's branching banks linked urban headquarters with branches that operated in areas that could not otherwise have supported a banking presence. Scottish banking innovation went beyond branch banking; its bankers also invented interest-bearing deposits, the interbank clearing of bank notes, and lines of credit (the so-called "cash credit account" invented by the Royal Bank of Scotland in 1728) that permitted borrowers to arrange for credit in advance and draw upon their accounts as needed. Scotland's system was also among the first to rely upon small-denomination bank notes as a source of financing.

At the heart of the innovativeness of Scottish banking and the abundance of credit in Scotland was the free chartering of banks. There were three specially chartered banks: the Bank of Scotland (1695), the Royal Bank of Scotland (1727), and the British Linen Company (1746). In addition, freely chartered Provincial Banks – that is, chartered under common licensing rules, rather than by government action – also operated as of 1747, and freely chartered Joint Stock Banks entered the system beginning in 1810.

Competition among Scottish banks was fierce. A few data points make the extent of that competition clear. In 1825, the three specially chartered banks operated an average of 15 branches each, while the freely chartered banks in the system averaged almost 3 branches per bank. In 1802, the amount

of bank assets per capita in Scotland was 7.5 pounds sterling, in comparison to 6.0 in England. Scottish banks paid remarkably narrow spreads (roughly one percentage point) between rates of interest paid on loans and rates paid on deposits. The fact that they were able to do so, and simultaneously earn respectable rates of return for their shareholders implies a high level of efficiency.

Just as remarkable, Scottish banks were less likely to fail or impose losses on their debtholders than were English banks. Bank failure rates in England were almost five times those of Scotland from 1809 to 1830, and while losses to English debtholders were sometimes significant there were virtually no losses on Scottish deposits or notes. The lower failure risk of Scottish banks reflected their greater size, competitiveness, and portfolio diversification. Unlike English banks that were limited to the six-partner rule (which kept them small and local in scope) Scottish banks were typically large and diverse.

Why did the political process governing bank chartering result in Scotland's having such a different banking structure than England? After all, the King of England and the King of Scotland were one-and-the-same person—and had been so since 1603. Why was the outcome in the Scottish Game of Bank Bargains so dramatically different from the bargain that had been struck in England?

At the dawn of the era of bank chartering, when the Bank of England and the Bank of Scotland were created in 1694 and 1695, respectively, England and Scotland had separate Parliaments (they were separate kingdoms sharing the same sovereign). At the time that King William was hunting for a way to finance his war against France, Scotland was poor and remote. Little would be gained from creating a monopoly Scottish bank that would help to finance the Crown. At the same time, the creation of such a bank would have required negotiating with the Scottish Parliament. While they generally favored King William over the deposed James II, the Scottish Parliament was not as committed to the idea of financing the king's imperial ambitions as the Parliament of England. In point of fact, the charter of the Bank of Scotland prohibited it from lending to the Crown without an Act of Parliament—a fact that suggests that the Scottish Parliament was quite conscious of the problems that could arise if the Bank of Scotland were turned into a vehicle of public finance. From the point of view of the British Crown, it was simply easier

to adopt a policy of laissez faire with respect to the more distant Scots, and use the Bank of England (as well as other English joint stock companies) to finance the Crown's war efforts.

The Bank of England proved sufficient to finance the Crown's wars with France. Thus, even after the Parliaments of England and Scotland were fused into a single Parliament of Great Britain in 1707, there was no attempt by the Crown to re-craft the Scottish banking system to serve its war aims. For its part, Scotland committed itself not to become a separate base of political power that might have withdrawn from or otherwise acted contrary to the interests of the British Crown. Scotland responded to its favored status in banking by what might be described as an act of unilateral political disarmament. In 1707, in the Act of Union, Scotland disbanded its national Parliament in a demonstration of allegiance to Britain. Political union meant that the Crown had little cause to worry about the positive economic consequences of its laissez faire policy toward Scottish banks, trade and industry. Despite political union, Scotland's banking system remained separate from England and Wales, and quite distinct from them, until liberalization in those countries ultimately led them to imitate Scottish success, beginning in 1825.

What political changes underlay the decision to liberalize English banking? Just as war finance needs had produced financial repression in England prior to 1825, the end of the Napoleonic Wars reduced the financing pressures that had produced the partnership between the Crown and the Bank of England. The victory of the Duke of Wellington at the Battle of Waterloo in June 1815 brought an end to nearly 130 years of war with France. The British government had succeeded in accomplishing its single most important strategic goal: making sure that Continental Europe remained politically divided and weak, and thus unable to contest England's ability to build and maintain a world empire.

The Pax Britannica came with two big challenges to the British State. The first was how to address the demands for political reform that had been percolating since the 18th century. The ideals of Republicanism that fueled the American Revolution were not entirely homegrown; many of them were English imports (though with plenty of colonial value-added). Calls for reform could be set aside while Britain waged its "Second Hundred Year's War" against France. Once Napoleon was defeated, however, it was difficult to ignore them. Indeed, the mobilization of a nation to fight a two-decade long war against

a revolutionary government that conceived of its population as citizens instead of "subjects," encouraged a rethinking of what, exactly, it meant to be British. Even anti-populist writings such as Edmund Burke's *Reflections,* had to make an argument for a liberal conception of government and had to advocate for gradual constitutional reform as a substitute for revolution. The fact that this rethinking of citizenship came on the heels of a long fight against another revolution—this one culturally closer to home, the American Revolution of 1775-1783—could only have made it even more profound.

Reforms of English banking coincided with deep electoral and procedural reforms of government. From the 18th through the early 19th centuries, numerous attempts had been made by the Whigs to revise the voting rules for electing members of the House of Commons. The proposed reforms were intended to expand the franchise and shift voting power to the newly expanded urban population centers, which had grown as a consequence of the Industrial Revolution and increased commerce. These attempts had often met defeat, but by 1832 they could be resisted no longer: there were violent demonstrations; governments were formed, failed, and re-reformed; even the king interceded by threatening to create 80 new peerages (appointments to the House of Lords) that would be awarded to supporters of electoral reform. The Representation of the People Act of 1832 (commonly known as the Reform Act or the Great Reform Act) increased the Parliamentary representation of the burgeoning industrial cities at the expense of locations with miniscule populations (the so-called "rotten boroughs"), and expanded the voting franchise.

In the absence of war, reforms also were undertaken to end mutual back scratching between the government and the Bank of England. Some measures improved accounting; others regulated the mode by which the Bank lent to the Treasury. These measures also precluded the Bank of England from making advances to the government without the express approval of Parliament. As one contemporary noted: "These regulations … put it out of the power of the Government to obtain pecuniary accommodation from the Bank of an irregular character….The relations between the Government and the Bank in regard to the management of the funds at the disposal of the latter body, must now be regulated by mutual

consideration of the general interests of the community, and are so amenable to public control as to render irregular, or even questionable, proceedings practically impossible."[2]

The most important step in the reconfiguration of the relationship between the Bank of England and the government was an end to the bank's monopoly charter. Despite the Bank of England's desire to maintain its monopoly privileges, the rising tides of democracy and industrialization proved too much for it. The advocates of competition in chartering ultimately won the day because their power was growing, and because they could point to three powerful facts to support their position: (1) England lacked sufficient sources of credit for the private sector; (2) the English financial system produced an unstable credit market that was unusually prone to commercial and bank failures; and (3) there was ample evidence from outside England (especially in Scotland) that competitive chartering would produce greater access to credit with greater banking stability.

The timing of the reform of English bank chartering illustrates the importance of specific historical events for precipitating long-contemplated changes. Political trends and valid arguments in favor of reform may favor change, but as is often the case, major changes in the structure of financial systems often coincide with crises that galvanize the movement for reform. In the case of English banking reform, that galvanizing event was the Banking Crisis of 1825, arguably the first truly global banking crisis of the modern era and one with severe consequences for English merchants, manufacturers, artisans, and bankers. Critics blamed the Bank of England (and the structure of the banking system, which its privileges had produced) for the severity of the crisis. The lack of available liquidity outside London during the crisis produced commercial insolvencies and bank failures on a spectacular scale. Seventy-three out of 770 country banks failed during the crisis. Bankruptcy filings in 1826 reached an unprecedented level.[3] The inherent vulnerability of the country's small banks was clearly demonstrated, as were the costs to the private sector of the absence of liquidity providers (large banks) that could have supported the smaller banks operating outside of London.

---

[2] Schuster (1923) pp. 10-11, quoting G. Arbuthnot's memorandum to the Report of the Select Committee on the Bank Act, 1858.

[3] Neal (1998), pp. 63-69.

The lessons that the British public took away from the Crisis of 1825 were threefold. First, the Bank of England was an unreliable repository of trust for managing the supply of money and credit in the economy. Second, it was insufficiently committed to the public good (after all, it was a for-profit bank, and could not have been expected to put aside its own interests). Third, the structure of the English banking system that the bank's monopoly privileges had engendered was too centralized in London and potentially very destabilizing for other English towns and cities.

The policy response to the crisis was the Country Bankers' Act of 1826. This Act allowed (and strongly encouraged) the Bank of England to open branches outside of London. Over the ensuing years, it opened branches in numerous cities, and two of those new locations (Manchester and Birmingham) were particularly active. Additionally, the Bank's monopoly over bank chartering was ended. Competing joint-stock banks could be chartered, although they would not be permitted to operate within a 65-mile radius of London, and their shareholders were not granted limited liability. These banks were not prohibited by law from issuing bank notes, but the Bank of England made it known that it would not deal with banks that entered that business, and this threat was sufficient to prevent the new joint-stock banks from issuing notes. The Bank of England made similar agreements with the pre-existing country banks: in exchange for the country bank consenting not to issue banknotes, the Bank of England agreed to discount bills for those banks at its new branches. When all was said and done, the 1826 Act had basically forced the Bank of England to increase the amount of credit available by working with both the new joint-stock banks and the pre-existing country banks. The Bank of England was, however, able to craft arrangements that allowed it to maintain its effective monopoly over note issues.

In 1833 the Bank of England's charter was up for renewal, and the government took this opportunity to erode its monopoly still further, and to demand further concessions from the Bank. The government renewed the Bank's charter through 1855, but the charter would be revocable at the pleasure of Parliament after an initial grace period of 12 years. The 1833 Charter Act also removed the prohibition against joint-stock banks operating in London. The Bank of England now faced competition even within its home market, although that competition was somewhat attenuated by the fact that the Act prohibited

the new joint-stock banks of London from issuing banknotes. In addition, the Act effectively eliminated the five percent usury limit, thereby broadening the pool of borrowers to whom the Bank of England and the new joint-stock banks could lend—in effect making the credit market more competitive still. In 1844 Parliament further eroded the rents enjoyed by the Bank of England. The Peel Act of that year gave the Bank an airtight monopoly on new note issues, but required the Bank to maintain 100 percent specie reserves against those note issues. The Peel Act meant that the Bank could not benefit financially from issuing notes paid zero interest, because the gold reserves held against its notes also paid zero interest.

The English banking system began as a crony enterprise primarily serving the fiscal interests of the State and the personal interests of a small group of well-connected private citizens. By the second half of the 19th century, the political pressures that democratization and industrialization had put on the banking system, coupled with an end to the government's need for war finance, undid that monopoly arrangement. As the result of fundamental changes in the industrial organization of the banking system (the creation of a nationwide system of branching banks), reforms in the nature of the government safety net policies for banks and borrowers, and a reduced need for the English government to rely on the banking system to finance its wars, by the late 1860s England's and Scotland's banking systems had converged. The newly integrated British banking system would enjoy many decades in which banking crises were absent and banks were an important source of credit for the private sector.

IV. The United States and Canada

The United States and Canada were both North American colonies, and were under British control by 1760. But their political histories before and after that date were quite different. Before 1759, Canada had been a French colony, while the United States – despite some Dutch, French, and Spanish settlements – had long been dominated by the British. Ultimately, the colonies that became the United States revolted against British rule and established a new republic in the 1770s and 1780s. Canada's constitution evolved within the British Empire, and the British Crown remains an important influence in Canada today. Those political differences between the United States and Canada, and the fundamental

historical and geographic differences that underlie them, are also able to account for the dramatic differences in their banking histories.

The Crippling Influence of Populism in U.S. Banking History

In the United States, from the time of the Revolutionary War through the early nineteenth century, the dominant coalition in control of bank chartering consisted of an alliance between political elites in both federal and state governments, many of whom were members of the Federalist Party, and a privileged group of financiers. This crony era of U.S. financial history was successful at financing a newly emerging State, as well as providing capital for the first stages of American commercial, agricultural, and industrial development.

A coalition of elites could not be sustained, however, in a political system characterized by federalism, populism, and an expanding franchise. Those realities, in turn, reflected the deeper historical roots of the United States. The 13 Colonies that became the United States were a backwater surrounded by hostile neighbors. British colonists faced the constant threat posed by the Spanish in Florida, the French in Quebec and the Ohio Valley, and the Indians, who often allied with the French, nearly everywhere else. In order to be viable, the 13 Colonies had to be an armed camp. There was no obvious source of wealth, however, that could pay for a large standing army to keep the enemy at bay. The 13 Colonies neither contained a Potosi that produced piles of silver coins, nor a Pernambuco that yielded prodigious quantities of highly valuable sugar. The one thing that the 13 Colonies did have, however, was seemingly endless expanses of farmland suitable for tobacco, maize, and wheat—provided that those lands could be cleared of Indians. Crucially, those crops share characteristics that allow them to be grown efficiently on family farms: they are highly storable and exhibit modest scale economies in production.

This combination of—hostile neighbors, abundant land, and storable crops that could be grown on small production units—made British North America a society of small, free-holding farmers who were armed to the teeth. America's common people proved able to achieve something that no group of small farmers had achieved since the Roman Republic: they demanded, and obtained, the right to vote. America's colonies might have had governors appointed by the king, but they also had colonial

assemblies—Houses of Burgesses--that were popularly elected. Though the exact rules varied by colony and locality, male suffrage was widespread, with typically 40 to 50 percent of early 18th century colonists eligible to vote for colonial assemblies in the mid-Atlantic states, and 70 to 80 percent in New England and the South.[4] Large landowners initially dominated the elected assemblies, but by the mid-18th century America's small farmers had begun to break away from the tradition of voting for their "social betters."[5]

When America's colonial elites, the large landowners and merchants, decided to move for independence they had to mobilize this class of armed, independent farmers against the British. That was no easy trick: they were essentially asking them to go toe-to-toe against the most disciplined, and best equipped, army in the world. Motivating them to do so required appeals based on liberty, freedom, and equality. Once they had done so, there was no putting the genie back in the bottle. The process of revolution was a profoundly radicalizing experience. Thus, to the shock of America's elites, after independence the small farmers who Washington had led across the Delaware would no longer doff their hats to their "social betters" nor defer to them in matters of politics.[6]

Although Federalist elites were able to control the first decades of political organization, including bank chartering, this crony era soon gave way to a second era of banking policy, which ran from the 1810s to roughly 1980. During this period, banking policies were determined by a durable alliance between small "unit bankers" (banks with no branches) and agrarian populists (farmers who distrusted corporations of nearly every type, as well as the elites that controlled them). The government's share of the benefits from chartering banks shifted from direct taxation or ownership interests in banks, to regulatory requirements forcing banks to lend to both state and national governments.

The geographically segmented, unit-bank organization of the banking system that emerged from this bargain reflected the interests of the coalition partners: bankers who wanted to create local monopolies, and populists who disliked corporations of any type, but most especially those associated with big city "aristocrats." The populist support for unit banking reflected, in part, the advantages that

[4] Keyssar (2000), p. 7.
[5] Wood (1991), p. 173.
[6] Wood (1991), pp. 175-182.

some classes of local borrowers received from limiting bank entry through unit banking. Because unit banking tied local banks to the local economy it made bankers more willing to continue to provide credit to their existing borrowers during lean times. Unlike a branch bank that might move funds to other locations in pursuit of greener pastures, unit bankers' lending opportunities were limited to the local economy. If a banker was based, for example, in central Illinois, a fall in the price of corn reduced the income and property values of everyone in the local economy. Bankers could reduce the supply of loans to locals in response to that shock, but if they did the only viable alternative investment was cash. Limited alternatives favored greater continuation of credit during lean times. This "credit insurance" advantage can explain why states with relatively high agricultural wealth tended to favor unit banking; relatively prosperous farmers who were already on the client list of the local unit banker had a stake in preserving branching restrictions.

The economic organization of U.S. banking during this "agrarian populist-unit bank era," however, entailed significant costs: it was inherently unstable, non-competitive, and inefficient in its allocation of credit. Its most salient characteristic was that the U.S. banking system, with some exceptions, was composed of tens of thousands of small banks operating local monopolies. The absence of branches meant that banks could neither spread risk across regions, nor could they easily move funds in order to head off bank runs, or coordinate a collective response to problems that arose. The United States became and remained the most banking crisis-prone economy in the world. Major banking panics, involving either widespread bank insolvency or systemic bank runs, occurred repeatedly throughout the 19th and 20th centuries. High costs of credit were also prevalent. Unit banking was not a cost-effective way to supply credit, as it entailed high overhead costs and little ability to spread risk. Furthermore, the fact that these unit banks operated local monopolies meant that they were able to charge more for loans and pay less for deposits than would have been the case had they had to compete with one another.

Contributing to the resilience of the unit banker-agrarian populist coalition was the decentralization of authority over bank chartering. Had America's political institutions granted the federal government the sole right to charter banks, bank chartering policy would have been a matter of national

policy, and it may have been harder for agrarian interests to prevent the chartering of nationwide branching banks. The federal organization of the U.S. government prevented that from happening, however. From the beginning of the nation, America's states saw great value in maintaining control over bank chartering within their states. States needed ways to finance themselves and the Tenth Amendment to the Constitution provided one: any power not explicitly delegated to the federal government, or explicitly denied to the states, resided with the states. The Constitution prohibited states from issuing legal tender paper money—but it said nothing about the right of states to charter banks, whose banknotes could circulate as currency (although they could not be made a legal tender). State-chartered banks soon became a source of significant financing for state governments. Governments received stock in their chartered banks, taxed their banks, and also passed laws requiring their banks to hold state government debts as backing for their note issues.

States, therefore, had strong fiscal incentives to create bank charters—and strong incentives to do whatever was necessary to maximize the value of those charters. States often were major owners of bank shares, especially prior to the 1830s, and they typically paid for those shares with a loan from the bank, which they then repaid out of the dividend stream. States received a larger stream of dividends when the banks earned monopoly rents: they therefore constrained the number of banks within their borders. States might extract additional income from banks by threatening them with new entrants to the banking market: they therefore accepted "bonuses" from incumbent banks to deny the charter applications of potential competitors. To further insure that the interest of bankers remained aligned with the governments that chartered them, banks typically were not granted perpetual charters. When their charters expired, they had to pay a fee or "bonus" for a renewal. Not surprisingly, circa 1810-30, bank dividends and bank taxes often accounted for one-third of total state revenues.[7]

States received no charter fees or other economic benefits from banks incorporated in other states: they therefore had an interest in prohibiting interstate branching. Challenges to those prohibitions were rejected by the Supreme Court in the landmark Bank of Augusta v. Earle case (1839), in which the Court

[7] Sylla, Legler, and Wallis (1987); Wallis, Sylla, and Legler (1994).

affirmed that states could limit banking activities by out-of-state banks so long as those prohibited activities were specifically enumerated by statute.

Banking in the Early Republican United States was therefore characterized by segmented monopolies. The four largest cities in the United States in 1800—Boston, Philadelphia, New York, and Baltimore—had only two banks apiece. Smaller markets typically had only one bank, if they had a bank at all. In 1800 there were only 28 banks (with a total capital of only $17.4 million) in the entire country. These banks, it should be emphasized, discriminated on the basis of both economic and non-economic criteria, such as political party affiliation.[8]

The national government also saw value in making use of its "implied" constitutional power to charter banks. Beginning with Alexander Hamilton's First Bank of the United States (BUS), a single national bank and state-chartered institutions coexisted. The system of a nationwide BUS and segmented state-granted banking monopolies was not sustainable, however, under America's political institutions. There were two political conflicts related to chartering – the conflict between the national and state chartering authorities, and the conflict over who could obtain a state charter.

Bankers with state charters had opposed the BUS from the time of its initial chartering in 1791. The reason for their opposition was straightforward: branches of the BUS—which were located in New York, Baltimore, Boston, Charleston, Norfolk, Savannah, New Orleans, Washington, D.C., and Philadelphia--undermined local banking monopolies. State bankers therefore had incentives to form a coalition with the Jeffersonians, who were ideologically opposed to chartered corporations and "aristocratic" bankers, to eliminate the BUS. They succeeded in blocking its rechartering in 1811.

Parallel to the political fight over national chartering, was the competition among the states resulting from the interaction of federalism, an expanding frontier, and a broad suffrage, all of which helped to propel a growing demand for more banks and a revulsion to the original practice of permitting only a few banking charters to the favored few. Political competition within and among states undermined the incentives of state legislatures to constrain the numbers of charters they granted.

[8] Wallis, Sylla, and Legler (1994), pp. 135- 39; Bodenhorn (2003), p. 142; Majewski (2004).

Massachusetts began to increase the number of charters it granted as early as 1812, abandoning its strategy of holding bank stock as a source of state finance and instead levying taxes on bank capital. Pennsylvania followed Massachusetts's lead with the Omnibus Banking Act of 1814. The act, passed over the objections of the state's governor, ended the cozy Philadelphia-based oligopoly that, until then, had dominated the state's banking industry. Rhode Island also followed Massachusetts' lead: in 1826 it sold its bank shares, increased the numbers of charters it granted and began to tax bank capital as a replacement for the income it had earned from dividends. It soon became, on a per capita basis, America's most heavily banked state.

The final blow to the crony era of bank chartering in the United States came in 1832, when Andrew Jackson successfully vetoed the rechartering of the Second Bank of the United States. That bank had been chartered in 1816, ironically by the very Republicans who had opposed the rechartering of the BUS in 1811. The War of 1812 demonstrated the importance to the government of a bank that could serve as its fiscal agent. The Second Bank of the United States was founded on the same principles as the first bank, and it faced opposition from the same local banker-agrarian populist coalition. The populist presidency of Andrew Jackson, and the risky political gambling of the Second BUS's President, Nicholas Biddle, during the struggle over the re-chartering of the bank, resulted in its demise: Jackson vetoed the renewal of the bank's charter in 1832, forcing it to close in 1836.

By the 1830s, the coalition of small banks and agrarian populists had won. The successful veto of the Second Bank of the United States marked a definitive end to the crony era of American banking history. From this point forward—until roughly 1980—the coalition of small bankers and agrarian populists would dominate the politics of bank chartering and bank regulation in the United States. While there was considerable variation across states, chartering of banks by special legislative acts tended to be replaced by so-called "free banking" acts from the late 1830s onwards. Under free banking, bank charters no longer had to be approved by state legislatures. Rather, individuals could open banks provided that they registered with the state comptroller and deposited state or federal bonds with the comptroller as a guarantee of their note issues.

Free banking was not, however, a complete rethinking of the earlier system of segmented monopolies, and it did not result in a fully "open-access" system of banking within or across states. Most importantly, free banks were not permitted to branch. Given the high overhead costs of establishing a bank, this effectively limited the entry of banks in sparsely populated areas. Given the regulatory costs and limited entry under free banking, the free banking era was not associated with a dramatic increase in competition. Indeed, some research even suggests that the passage of free banking statutes, per se, were not critical to the process of banking growth.[9] Other research helps to explain why: despite the democratization of chartering under free banking, the unit structure of New York's free banking system significantly limited the extent to which free banking was able to produce an expansion of bank credit in that state. Because branching restrictions persisted, free banking simply expanded the number, and reduced the size, of local bank monopolies. It also removed the government from the position of awarding bank charters to its political allies. The results were twofold: some of the rents that had been earned by bankers were dissipated; and borrowers who had earlier been closed out of credit markets now had access to finance, though it came from a bank that still had a great deal of local market power.

During the Civil War, under the pressure of war financing needs, the national government once again began to charter banks, but in doing so it did not challenge the political equilibrium that had been constructed by the coalition of unit bankers and agrarian populists. Rather, it conformed to the unit banking structure, and initially prohibited branching by national banks. National banks, like state-chartered free banks, were required to hold a large amount of government bonds and new government legal tender currency, which provided a captive market for U.S. government debts.

The decentralized U.S. banking system structure of thousands of unit banks persisted for many decades. The banking crises of the late 19th and early 20th centuries gave rise to much soul searching, recorded in the many volumes of material produced by the National Monetary Commission in 1910. But rather than challenge the political equilibrium of unit banking, the government created the Federal

[9] Ng (1988).

Reserve System to provide an elastic supply of reserves – which was widely understood as a second-best means of achieving stability, but the only politically feasible means of doing so.

Even the banking crises of the Great Depression were unable to undermine the coalition that supported this inefficient and unstable system. Rather, the agrarian populist-unit banker coalition used the depression to create a new set of institutions designed to prop up the preexisting fragile system. These institutions included deposit insurance and a set of laws, including the notorious Regulation Q, that made it illegal for banks to pay interest on checking accounts and that limited the interest rates they could pay on other types of accounts.

Deposit insurance and Regulation Q changes managed to stabilize the system for a time, but they did not cure the structural inefficiencies and high credit costs of unit banking. Furthermore, the ability of the post-Depression banking structure to survive depended on the government's commitment to conservative monetary and fiscal policies. Once the U.S. government departed from fiscal and monetary conservatism in the 1960s the system began to break down. Indeed, the public began to pull their savings out of banks in favor of various bank competitors, such as mutual fund money market accounts. Technological innovations, such as the invention of the ATM, further undermined the system by allowing banks to skirt laws against branching. Thus, by the 1980s, the coalition that had supported unit banking began to crumble. For the first time in U.S. history—centuries after they appeared in other countries—large banks with nationwide branching networks, came into existence.

Restrictions on intrastate and interstate branch banking finally began to be undermined in the 1970s, but not until the 1990s was the banking market to be completely opened up to competition. Five forces worked to undo the unit banker-agrarian populist coalition. The first was demographic: during the 20th century, the United States was transformed from a predominantly rural to a predominantly urban country, which meant that voting power shifted toward America's cities. The second force was technological progress that eroded banks' ability to extract rents from borrowers and depositors. The computer revolution drove down the cost of information storage and retrieval, allowing prospective lenders anywhere in the country to assess a borrower's credit-worthiness reasonably well without having

to rely as much on locally obtained information. With respect to depositors, technology also spurred much greater competition, especially via networked automated teller machines (ATMs that are linked via computer). Third, accelerating price inflation in the 1960s and 1970s spurred disintermediation from the regulated banking system, and created the first of the post-1960 "shadow banking systems" of relatively unregulated finance companies and money market mutual funds. Regulation Q – established in 1933 – limited the interest rate that could be paid on bank deposits. As inflation and nominal market rates of interest rose, the real interest rate payable on regulated deposits became increasingly negative, making it hard for banks to attract deposits. Fourth, U.S. banks – which were relatively small and constrained in their geographic reach and permissible product lines, compared with the banks of other developed countries – were losing global market share in banking. The Fed and many U.S. politicians became advocates of the deregulation of interest rate ceilings, the removal of branching restrictions, and the elimination of limits on bank powers (especially the limits on corporate securities underwriting by banks) as a means of allowing U.S. banks to compete with their foreign counterparts. Finally, a fifth force driving reform was a wave of banking distress in the 1980s, which set into motion a political movement in favor of bank consolidation. The extreme banking distress of the 1980s even encouraged many unit bankers, as well as bank borrowers, and government officials, to favor the relaxation of branching restrictions. A unit banker facing the failure of his bank saw acquisition by a branching bank as a way to exit with some stock wealth and perhaps even a job in the new bank, a desirable alternative to losing everything. The borrowers at failing unit banks saw the branching banks that were willing to buy weak banks as a crucial source of funding. For the FDIC and federal government officials, the big banks willing to acquire small failing banks reduced the costs of paying off failed banks depositors. For state governments, the new bank entrants were a welcome means of restoring local economic growth.

A major blow to the state laws that prohibited interstate branching came in 1982, when, in response to the Savings and Loan crisis, Congress amended the Bank Holding Company Act of 1956 to allow failed banks to be acquired by any bank holding company, regardless of state laws. This induced many states to enter into regional or national reciprocal arrangements whereby their banks could be

merged (not just purchased by a holding company) with banks from another state. Between 1984 and 1988, 38 states joined one of these reciprocal arrangements.[10] Banks operating national branching networks accounted for only ten percent of the U.S. banking system in the early 1980s. By the mid-1990s, they accounted for more than 70 percent.[11] The final blow to the unit banks came in 1994, when Congress codified the process that had been taking place at the state level by passing the Riegle-Neal Interstate Banking and Branching Efficiency Act. Banks could now branch both within states and across state lines. This was, in short, the death knell of the unit banker-agrarian populist coalition that had shaped American banking institutions since the 1830s.

The Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 marked the demise of the unlikely political coalition between unit bankers and agrarian populists that had dominated banking policy for over a century and a half. It permitted a wave of mergers and acquisitions that created the megabanks that now have a branch in nearly every city or town in the United States.

The creation of the new megabanks generated tremendous profits for merging banks—from economies of scale, economies of scope, the potential for market power, and too-big-to-fail government protection. Political actions that create profit also create new opportunities for deciding how to divide them. In the new U.S. Game of Bank Bargains defined by branching deregulation, populist politics continued to play a role in determining the allocation of profits, although the center of populist power had shifted from rural to urban areas.

Each merger and acquisition required approval from regulators, most particularly from the Federal Reserve Board. The process of approval required that banks show that they had been good citizens of the communities in which they operated, and this fact provided a source of leverage for activist groups, such as the Association of Community Organizations for Reform Now (ACORN), who could block or delay a merger by claiming that the banks were not in compliance with the Community Reinvestment Act of 1977. What had been a largely moribund piece of legislation now became a very

[10] Kroszner and Strahan (1999).
[11] Calomiris (2010).

valuable chip in the Game of Bank Bargains, which perhaps explains why the Act was revised eight times once the merger wave got underway, each revision usually increasing its stringency. Bankers seeking to become nationwide enterprises had to ally with activist groups to obtain their political blessing. In exchange, the activist groups obtained contractual guarantees from the would-be merging banks to direct mortgage and other credit, as well as cash contributions, to themselves and their constituents.

The incentives to become a megabank were multiple. Potential advantages included diversification, the ability to spread overhead costs over a larger operation, economies of scope (a large bank could afford to provide a broader range of products and services). Additional potential advantages of becoming a mega-bank was were the potential for obtaining market power and the potential implicit subsidy of too-big-to-fail protection.

The Federal Reserve Board had the key decision making authority over mergers, as the regulator of bank holding companies, but other bank regulators and the Justice Department also could weigh in to oppose mergers, if they chose to do so. There were several criteria that could be used to block approval of a merger. First, an acquiror had to be financially strong. Second, the merged bank could not have excessive market power. This was not much of a constraint because the Fed assessed market power by looking at a merged bank's deposit market share, rather than its ability to set prices in credit markets.

The Fleet Financial-BankBoston merger of 1999 is a telling example: by combining the only two New England banks of significant size it created a megabank that could set prices for business borrowers. Mid-sized businesses that were too big to borrow from the remaining small, local banks, and too small to be able to borrow in global markets, were particularly affected. They objected to the merger on these grounds, as did the Mayor of Boston and the Attorney General of Massachusetts. All to no avail: the Fed approved the merger, and interest spreads for business borrowers rose by a full percentage point.

The third criterion by which a merger could be blocked was "good citizenship" (as regulated under the CRA) and, unlike market power, this was indeed a binding constraint. The language of the CRA focuses on making sure that banks serve their local communities, but this largely translated into ensuring that low-income urban communities with minority populations were not subjected to discrimination in

lending. The early years of the CRA do not appear to have produced much in the way of results. From 1977 to 1992, only $43 billion in CRA commitments by banks had been announced, and almost all of that occurred after 1989. As of 1995, however, revisions to the CRA meant that banks faced adverse consequences for running afoul of federal government bank supervisors who monitored and rated their CRA compliance. As President Clinton boasted in a July 1999 speech, "[CRA] was pretty well moribund until we took office. Over 95 percent of the community investment…made in the 22 years of that law have been made in the six and a half years that I've been in office."[12]

Why did banks care about their CRA ratings? Banks could receive a range of CRA "grades"—Outstanding, Satisfactory, Needs to Improve, and Substantial Noncompliance—and these depended on the degree to which a bank was serving the needs of the low-income and minority groups in the communities where it operated. The main penalty for getting a weak rating was that it could potentially scuttle a bank merger on the basis of "bad citizenship." A bank that was not pursuing an aggressive strategy of mergers and acquisitions did not, therefore, need to pay much attention to its CRA rating. A bank with big ambitions to grow, however, needed to a good rating from CRA.

It was, therefore, often in a bank's interest to enter into an explicit partnership with an activist group in advance of a Fed merger hearing. Given the existence of the CRA, both sides had incentives to strike a deal, because failure to do so meant that the bank merger might be blocked, thereby forcing the bank to forego the opportunity to increase its profits and forcing the activist group to forego the opportunity to serve its members and increase the resources at its disposal. The politicians whose policies made these deals possible saw no reason to get in the way of them. As President Clinton proudly proclaimed in a 1999 speech, the banking reform legislation of that year "establishes the principles that, as we expand the powers of banks, we will expand the reach of the [Community Reinvestment] Act."[13]

There was nothing subtle about the manner in which the deals between merging banks and activist groups were arranged. In fact, an umbrella organization for activist groups, the National

[12] Calomiris and Haber (2013), Chapter 7.
[13] Calomiris and Haber (2013), Chapter 7.

Community Reinvestment Coalition (NCRC), actually put together a 101-page guide on how to negotiate with banks that were in the process of merging. The NCRC guide did not shy away from encouraging activist organizations to take advantage of their leverage over a prospective bank merger: "When a lender desires to merge with another institution or open a branch, the lender must apply to the Federal Reserve Board and/or to its primary regulator for permission. If the lender has received low [sic] CRA rating, the federal agency reviewing the lender's application has the authority to delay, deny, or condition the lender's application."[14] The guide goes on to say: "Merger and acquisition activity presents significant opportunities for community groups to intervene in the approval process and raise CRA concerns and issues. Some banks are very desirous of Outstanding ratings so that they can present a clean reinvestment record to regulators when they ask for permission to merge….Activists should keep in mind that changes from Outstanding to Satisfactory ratings (and back again) is effective in leveraging reinvestment as well as changes from passing to failing ratings (and back again to passing). This is true regardless of whether the movement in ratings is the overall rating for the bank or a rating for particular geographical areas."[15] The guide explains how to affect a bank's grade: "…community organizations can offer written comments on a bank's CRA and fair lending performance when a bank has submitted an application to merge or acquire another bank or thrift."[16]

Activist groups were successful in negotiating many long-term contracts with banks, in which they received specific monetary and other commitments for their organizations. Between 1977 and 2007 there were no fewer than 376 such agreements, involving scores of groups. The commitments that activist organizations obtained from banks came in two forms. First, banks committed to supply mortgage and small business credit to borrowers identified by the activist organizations. Over the period 1977-2007, these directed credit commitments totaled $867 billion, with almost all of that growth coming in the years after 1992. Banks also provided a second source of support to activist groups, by paying them fees for administering the directed credit programs into which they had entered or by making direct contributions

[14] Calomiris and Haber (2013), Chapter 7.
[15] Calomiris and Haber (2013), Chapter 7.
[16] Calomiris and Haber (2013), Chapter 7.

to those groups. As of 2000, the U.S. Senate Banking Committee estimated that the total of such fees and contributions to all activist groups came to $9.5 billion.[17]

Had it not been for the CRA, banks would have made fewer and less risky loans in the 1990s and 2000s. A recent study, by Agarwal, Benmelech, Bergman and Seru (2013) compares the portfolios of banks in the six quarters prior to a CRA evaluation relative to the portfolios of other banks not slated for an evaluation, and finds that an impending a CRA examination caused banks to increase their lending by 5%, and increased the default risk of those banks mortgage loans by more than 15 percentage points. This approach provides lower-bound estimates of both increased lending and increased levels of default risk resulting from the CRA. Another approach to measuring the impact of CRA compliance is to focus on the increase in the level of CRA commitments over time. This is the approach taken by Pinto (2011), who assumes, conservatively, that the CRA had no binding effects on bank lending until the Clinton Administration's CRA policy push. Under that assumption, Pinto concludes that, by 2007, there were $2.2 trillion dollars in CRA commitments that would not have been undertaken by banks voluntarily.

The partnership between megabanks and activist groups became even more ambitious as it drew in a third set of partners—Fannie Mae and Freddie Mac. Banks would not make limitless commitments to their activist partners: CRA loans implied higher levels of risk for the bank than more traditional mortgage loans. Thus, the activist groups used their political power in Washington to generate regulatory mandates on housing GSEs, which included the Federal National Mortgage Association (FNMA, commonly known as Fannie Mae), and the Federal Home Loan Mortgage Corporation (FHLMC, commonly known as Freddie Mac), and the Federal Home Loan Banks (FHLBs). Fannie Mae and Freddie Mac, in particular, were required to repurchase mortgage loans that had been made to low income, urban, and minority constituencies. This change was a win-win for activist groups and mega-banks; more credit could be directed to targeted constituencies at less cost to the banks because the banks were now able to resell some of their CRA-related mortgages to a GSE on favorable terms.

[17] Calomiris and Haber (2013), Chapter 7.

These government mandates on Fannie and Freddie were not vague statements of intent, they were specific targets; and in order to meet those targets Fannie and Freddie had little choice but to weaken their underwriting standards. By the mid-1990s, Fannie and Freddie were agreeing to purchase mortgages with downpayments of only three percent (instead of the 20 percent that had been the industry standard). Soon after they were buying mortgages with weak credit scores. By 2003, they were agreeing to purchase massive quantities of loans with no documentation of income (so called liar, or no-doc, loans). In exchange, they obtained valuable concessions from Congress, most particularly capital standards (minimum ratios of equity capital to assets) that were only 60 percent that of commercial banks holding similar loan portfolios. That is, the managers and shareholders of the GSEs joined the megabank-urban activist coalition. They became a crucial ingredient to the growth of the coalition's resources, a crucial part of the institutional glue that held the coalition together.

Weak underwriting standards – a key ingredient in the subprime crisis of 2007-2009 – were not an excludable good (or bad); they were available to *everyone*. Fannie and Freddie, by virtue of their size and their capacity to repurchase and securitize loans made by banks, set the standards for the entire industry. Thus, large swathes of the American middle class—whether they realized it or not—were soon pulled into this large bank-urban activist-GSE coalition by jumping on the easy credit bandwagon. This fact cannot be emphasized strongly enough: when Fannie and Freddie agreed to purchase loans that only had a three percent downpayment, no documentation of income or employment, and a far from perfect credit score they changed the risk calculus of large numbers of American families, not just the urban poor.

The Canadian Mirror

The extraordinary stability of the Canadian banking system has been one of its most visible and oft-noted characteristics for nearly two centuries. Since 1840 the United States has had 12 major banking crises, while Canada has had none--not even during the Great Depression. This Canadian achievement is especially remarkable in light of the fact that Canada is a staples-based economy, heavily reliant on exports, and thus largely at the mercy of international variation in its "terms of trade." More remarkable

still, the stability of Canada's banks was accomplished with little government intervention, either in the form of prudential regulation or assistance to distressed banks.

How did Canada do it? The short answer is that the Canadian banking system has a very different structure than that of the United States; it has always been composed of very large banks with nationwide branches. This has not only allowed Canadian banks to diversify their loan portfolios across regions, it has also allowed them to transfer funds in order to shore up banks in regions affected by an adverse economic shock. Nationwide branch banking has also allowed Canada's banks to capture scale economies in administration, while competing among themselves for business in local markets. Canadian banks, historically, charged lower interest rate spreads (the difference between loan interest rates and deposit interest rates) than U.S. banks, especially in remote areas.

One potential shortcoming of a concentrated system such as Canada's is that it could undermine competition among banks, resulting in less credit at higher prices for households and business enterprises. Canada's democratic political institutions—particularly, a popularly elected lower house—however, put a limit on the extent to which the banks could earn monopoly profits. The threat of political backlash against the banks is institutionalized in a rather peculiar practice: for over a century and a half, the Canadian Parliament has carried out periodic legislative reviews and re-chartering of its banks. Until 1992, this occurred every ten years; since 1992, it occurs every five years. The practice of revising the Bank Act and rechartering the banks is not solely a stick with which to threaten bankers; it is also a carrot that rewards sound business practices by giving the bankers themselves a voice in the crafting of the new legislation. In short, widespread suffrage in Canada has given Canadian bankers strong reasons to follow the dictum: Pigs get fat, hogs get slaughtered.

This short answer, however, begs a more fundamental question: What factors allowed Canada to achieve these positive outcomes relative to the experience of the United States? Canada's success is even more striking in light of the fact that when Canadian policy makers looked for models on which to base their system they looked across the border, to the United States. The earliest banks in Canada (beginning

with the Bank of Montreal in 1817) were modeled closely and explicitly on the nationally chartered, multi-branch Bank of the United States (founded in 1791).

Why did Canadians retain that structure for the next two centuries, while Americans rejected it? America's rejection of branch banks during the 19th century, was not the result of ignorance—rather it was the result of politics. Specifically, it was the outcome of a particular set of political coalitions that existed in the United States and the strongly decentralized federal political structure within which they bargained over banking. Understanding why Canada retained a system based on a small number of banks with the power to branch across provincial lines therefore requires us to understand the factors that shaped the political coalitions that emerged in Canada and the political structures in which they bargained.

A single, overarching factor shaped Canada's political economy: following the American Revolution, British policy makers were determined to hold onto Canada, but doing so was difficult because the vast majority of the Canadian population in the late 18th century was of French origin. The British Crown accomplished its goal. Canada had no American-style revolution, in which a commercial elite made common cause with farmers against British rule. Rather, Canada remained part of the British Empire, gradually obtaining greater autonomy over time.

Holding Canada in the empire required British policy makers to engage in a series of institutional experiments designed to simultaneously concede increased self-government to their Canadian subjects while limiting the political power of the numerically-large French population. They did not do so simply out of chauvinism; these steps were essential if Canada was to be an economically viable colony. The solution that they eventually hit upon was a federal system that gave the central government control over economic policy making. Canada's 1867 Constitution also allowed the national government to nullify laws that had been passed by individual provinces (e.g., French-speaking Quebec). Systematic malapportionment in the upper house of the Canadian Parliament ensured that the French would always hold a minority of the seats. In short, the need to solve a difficult problem of empire—crafting political institutions so that the French population in Quebec could not hold up the economic development of the Canadian interior—gave rise to a political system in which the central government had the exclusive right

to charter banks. Provincial governments could not create local, territorially-demarcated banks that could serve as the nucleus of an anti-national bank coalition, as happened in the United States.

Like the U.S. founding fathers, Canada's early merchants were entrepreneurial men whose wealth was of recent vintage and who were a progressive force lobbying for policies to promote economic growth and a more open political and economic system (that is, one not controlled by the British overlords). Unlike the founding fathers, however, Canada's merchants were not clamoring for a revolution to overthrow British rule. The reason for their reluctance to call for independence is not hard to divine: there was a third group that was a threat to both the merchants and the overlords that was populist, agrarian—and overwhelmingly French.

Finding a way to craft a viable colony was made even more difficult because of several features of Canada's geography. These were: its East-West orientation; the fact that much of its natural resource wealth, including its best farmland, was located deep in the interior; and that interior was geographically contiguous with the United States. These geographic factors mattered multiplicatively, not just additively; that is, each factor magnified the importance of the others, and made finding a solution to the problem of a large French-origin population even more crucial. The locational centrality of the French province of Quebec meant that without centralized government control of economic policy, British efforts to improve cross-regional transportation might be blocked by a hostile French population. The fact that Canada's western provinces were geographically contiguous with the United States made the problem facing British policy makers even more serious. What would prevent agricultural and commercial groups from those interior provinces, frustrated by decisions made by French-dominated Quebec, from choosing to become part of the United States, or from supporting a U.S. invasion of Canada? Thus, control over the Province of Quebec was not just economically important to Great Britain in its own right; Quebec held the key to the future of Canada as a viable British colony.

An 1837 rebellion prompted the key political changes that would define Canadian political structure, and banking policy. The 1840 Union Act unified Upper and Lower Canada into a single province with a single legislature, hereafter called the United Province of Canada. The Union Act went

beyond simple unification, however. It gave each of the former provinces the same number of seats in the new legislature—regardless of the fact that predominantly French-speaking Lower Canada had a much larger population than English-speaking Upper Canada—thereby ensuring that the French could not muster a majority in the lower assembly in order to block English proposals. In addition, the British Parliament remained a direct source of centralized British control through its own authority, which it had a direct interest in preserving. The British Parliament retained rights of review and approval over a range of matters, appointed governors and also members of the Legislative Council. The Union Act did, however, grant some local voice in the selection of the members of that upper house.

These arrangements were expanded under the British North American Act of 1867, also known as the Constitution Act, which became the basis for the modern nation of Canada. The act conceded greater political autonomy to Canadians—so much so that Canada effectively became a sovereign nation—while simultaneously limiting the power of the French. What had been Upper Canada, was renamed Ontario, while what had been Lower Canada was renamed Quebec. At the same time, two maritime provinces, New Brunswick and Nova Scotia, which had been separately governed by Britain since the 18th century, were folded in, creating a confederation of four provinces named the Dominion of Canada. The French were now outnumbered, both as a proportion of the population and as a proportion of the provinces. Within a few years, the British completely drove the nails into the French coffin by adding three more (virtually unpopulated) provinces; Manitoba (1870), British Columbia (1871), and Prince Edward Island (1873). Quebec was now just one of seven provinces.

Furthermore, the membership of Canada's Senate under its Constitution of 1867 ensured that merchant and banker interests, and more generally, the interests of property owners, would be protected against populist attacks. All legislation had to be approved by the Senate, which was designed to be a body of "second sober thought" to check the excesses of the House of Commons. In order to ensure sufficient sobriety of thought, membership in the Senate was subject to a wealth test: Senators had to possess $4,000 in net worth in the form of land—a threshold that was quite high in the nineteenth century, though it has ceased to be a binding constraint. The selection mechanism for Senators, however,

continues to be binding: Senators are not elected, but rather, are appointed by provincial governors, subject to the approval of the Monarch of Great Britain (although in practice, the Canadian Prime Minister has a great deal of say in who is appointed to the Senate).

Finally, to further limit the French in Quebec, the Act of 1867 centralized power in the national government. Any power not specifically delegated to the provinces was left to the central government, making it the polar opposite of the U.S. system, in which non-enumerated rights are reserved for the states. The Act specifically granted authority over commercial, economic, and banking policy to the national government. The Act also effectively gave the national government the right to nullify any action by a provincial government provided that such nullification is "declared by the Parliament of Canada to be for the general advantage of Canada or for the advantage of two or more of the provinces."[18]

The implications of these aspects of the Canadian political system for banking legislation were far-reaching. In the United States, bank chartering was dominated by fragmented state-level decisions until the 1980s, allowing local coalitions in favor of unit banking to shape policies. In contrast, in Canada, authority over banking was centralized from the very beginning, and that authority was protected from populist pressures so that it was able to serve the broader social objective of ensuring abundant and stable credit. Any political bargaining among coalitions had to take place at the national level, meaning that the initial system of large banks with multiple branches was never seriously threatened.

Canada's particular set of institutions and coalitions shaped the Canadian banking system, and resulted in a remarkably stable banking history over two centuries. One of the most striking things about the comparative banking history of the United States and Canada is that both began as an alliance of merchant-financiers and political elites. In the United States, that narrow coalition was quickly undermined by the rise of populism. In Canada, it was not.

Canadian bankers and politicians were conscious of the differences in performance and stability between the Canadian and U.S. banking systems. They attributed those differences to the fundamental structural difference between the Canadian branching system and the U.S. unit bank system. As Canadian

---

[18] McInnis (1959), p. 294.

banker E.L. Stewart Patterson wrote in 1917: "Practically every country in the world except the United States has recognized the utility, if not the absolute necessity, of the branch system of banking in handling commodities as liquid as money or credit. A bank system without branches is on par with a city without waterworks or a country without a railroad so far as an equable distribution of credit is concerned."[19]

Circa 1911, Western Canada (Alberta, British Columbia, Manitoba, and Saskatchewan) had 21 banks operating roughly 900 branches spread across the region's cities, towns, and villages—most with miniscule populations. The average incorporated town had only 900 inhabitants, but amazingly had two branches of competing banks. The average incorporated village had only 230 inhabitants and yet typically had a bank branch.[20] While not lending against real estate, these branches did provide credit to farmers—so much so that by the 1920s the hundreds of small, private (non-incorporated) banking houses that had dotted the Canadian countryside during the late 19th century had disappeared.

There were populist challenges to Canadian banking laws, but the political institutions of Canada made it difficult to enact legislation to change the basic rules of the game of Canadian banking. Reforms to banking law had to be worked out first in the House of Commons, where it was much harder to create and sustain a winning agrarian coalition than would have been the case had bank laws been enacted at the state level, as was the case in the United States. Any reform that got through the House of Commons then had to be approved by the Senate, whose members were appointed by the Governor General for life. Groups favoring the status quo in banking law were therefore able to delay reforms, propose watered down compromises, or block change entirely.

A particularly telling example occurred in 1911. A coalition of populists and conservatives was loaded for bear: four banks had failed from 1906-08, and then in 1910 the Farmer's Bank went under, losing every last cent of its depositor's savings.[21] Depositors in the bank made common cause with its shareholders, who were subject to additional assessments of capital because Canadian banks were chartered with double liability. They demanded that the government compensate them, alleging that the

---

[19] Patterson (1917), p. 60.
[20] Eckardt (1913), pp. 159-163.
[21] Carr, Mathewson, and Quigley (1995), p. 1140.

Minister of Finance had been negligent in granting the bank the Treasury Board Certificate it needed to activate its charter in 1906. The issue emerged as a hot button topic in the 1911 elections: Conservative Party candidates even promised compensation for depositors if they were elected. The coalition of conservatives and populists also demanded sweeping changes to the Bank Act, including requiring parliamentary approval for any further bank mergers or acquisitions, instituting government supervision of the banks, and empowering banks to lend on real estate.

Opponents of these reforms saw them for what they were: the outcome of a coalition of special interests. No one could reasonably argue that allowing banks to lend on real estate would increase bank stability, given the association in other countries between real estate lending and the greater risk of bank failure. Opponents of reform used their influence in Parliament to continually delay the revision of the Banking Act. Their delaying tactics apparently worked; when the revision to the Bank Act was finally passed in 1913 all of the demands for reform had been eviscerated. Instead of government supervision of the banks, there would be compulsory annual audits of the banks' financial statements by their own shareholders. Instead of being able to lend on real estate, banks would be allowed to lend on farm stock and grain in storage. And, instead of requiring that bank mergers and acquisitions had to be approved by parliament, they only needed the consent of the Minister of Finance. As for the idea that depositors in the Farmer's Bank deserved compensation for their losses; it, too, died a slow death. A royal commission in 1913 determined that the loss to depositors was the fault of the bank's managers, not the Treasury Board. A subsequent attempt by a Conservative Minister of Finance to pass a bill compensating depositors failed when the Senate tabled the motion for six months—and then never got around to taking it up again.

A similar example of the power of Canada's institutions to blunt populist demands can be seen in the case of the Home Bank failure in 1923. Here again, depositors claimed that their losses (equal to roughly 75 percent of their deposits) should be fully compensated by the government. The House of Commons voted in 1925 to reimburse some depositor losses, but the Senate was unwilling to go along.

The political institutions that underpinned the Canadian banking system—and hence the basic organization of the banking system itself—also proved to be remarkably resilient to the Great Depression.

Canada was a primary product producer, and as the demand for its most important products—timber, minerals, and cereals—collapsed, the economy contracted sharply. No Canadian banks actually failed during the depression. Nevertheless, Canadian populists blamed the country's orthodox monetary and banking policies for the state of the economy. Nowhere were these sentiments stronger than in the West, where populists had long resented the Eastern "money trust" that they claimed ran the country. In the Province of Alberta, a fringe political party espousing a crank economic theory, the Social Credit Party, took over the legislature in a landslide victory in 1935. The federal government used its constitutional powers to nullify much of the subsequent legislation passed by the Alberta legislature—which included the payment of "social dividends" to all adult citizens and the establishment of price controls based on the notion of "just prices"—but it was clear that mainstream national parties needed to do more than just block confused, if well-intentioned, economic reforms.

Canada's distinctive banking rules also proved to be robust to demands for increased access to credit, and the rise of new non-bank intermediaries, in the post World War II period. After the war, the return of hundreds of thousands of Canadian soldiers, sailors, and airmen drove the process of urbanization—and the demand for urban housing. The government took a number of steps designed to provide housing credit to this obviously crucial urban constituency. Importantly, it did not do so through the banks, which remained unable to make loans against real estate—and which did not lobby for the right to do so. Rather, the government attempted to placate demands for real estate lending with its rewriting of the National Housing Act in 1944, and with the establishment of the Central Mortgage and Housing Corporation the following year. When all was said and done, Canadian taxpayers subsidized insurance companies to get into the mortgage business, and when the market looked unattractive to them (even with government guarantees), taxpayers directly subsidized homebuilders.[22]

Canada's unusual political institutions, which were themselves an outcome of its geography and colonial history, forged a durable partnership between bankers, merchants, and the parties in control of the government, and throughout its history, prevented populist capture of the banking system in pursuit of

[22] Calomiris and Haber (2013), Chapter 9.

special interest subsidies. Many observers of Canadian banking credit its superior performance to its regulatory structure. As we have emphasized, however, regulation is the outcome of political bargains. In the United States, instability was permitted by regulators because it served powerful political interests (rural populists and unit bankers in the pre-1980 period, and urban populists and mega-banks afterward). In Canada, the banking system was not used as a means of channeling subsidized credit to a favored political constituency, so there was no need to tolerate instability.

V. Conclusion

Our summaries of the role of politics in shaping banking system outcomes in Britain, the United States and Canada illustrate many broader patterns that we explore further for those and other countries in Calomiris and Haber (2013). In concluding we would emphasize three political realities that are fundamental to the Game of Bank Bargains, which are particularly well illustrated by the country histories summarized above.

First, successful banking systems cannot develop without the active encouragement of government. If a government lacks the desire or ability – for example, because of its fiscal needs – to allow a competitive, stable system of bank credit supply for the private sector to develop (as in England until the 1820s), the banking system will be repressed.

Second, scarce credit and a high propensity for crises can arise even in the absence of such fiscal pressures, and often do arise in autocracies and populist democracies. In populist democracies, such as the United States, the regulation of banking is used as a political tool to favor some parties over others. It is not that the dominant political coalition in charge of banking policy desires instability, per se, but rather, that it is willing to tolerate instability as the price for obtaining the benefits that it extracts from controlling banking regulation. A broader and related principle, which Calomiris and Haber (2013) show applies to countries around the world, is that in countries where it is difficult to form coalitions to give special access to bank credit to a subset of the population, favorable banking systems outcomes – stable and abundant bank credit – are much more likely.

Third, regulatory failures – e.g., the decision not to permit nationwide branch banking in the United States prior to the 1990s, or the Fed's permissive approach to bank mergers in the 1990s – like regulatory successes are themselves outcomes of political bargains. This also implies that the prospects for regulatory reform are often bleak because meaningful reform requires commensurate supporting shifts in the political coalitions that make the enactment and enforcement of those reforms possible.

References

Bodenhorn, Howard (2003). *State Banking in Early America: A New Economic History.* New York: Oxford University Press.

Calomiris, Charles W. (2010). "The Political Lessons of Depression-Era Banking Reform," *Oxford Review of Economic Policy*, 26, 2010, 540-560.

Calomiris, Charles W., and Stephen Haber (2013). *Fragile By Design: Banking Panics, Scarce Credit, and Political Bargains.* Princeton: Princeton University Press, forthcoming.

Carr, Jack, Frank Mathewson, and Neil Quigley (1995). "Stability in the Absence of Deposit Insurance: The Canadian Banking System, 1890-1966." *Journal of Money, Credit and Banking* 27: 1137- 1158.

Eckardt, H. M. P. (1913). "Canadian Banking." *Annals of the American Academy of Political and Social Science* 45: 158-170.

Keyssar, Alexander (2000). *The Right to Vote: The Contested History of Democracy in the United States*. New York: Basic Books.

Kroszner, Randall S., and Philip E. Strahan (1999). "What Drives Deregulation? Economics and Politics of the Relaxation of Bank Breaching Restrictions." *Quarterly Journal of Economics* 114: 1437-67.

Majewski, John (2004). "Jeffersonian Political Economy and Pennsylvania's Financial Revolution from Below, 1800-1820." Mimeo, University of California, Santa Barbara.

McInnis, Edgar (1959). *Canada: A Political and Social History*. New York: Rinehard & Co.

Neal, Larry (1998). "The Financial Crisis of 1825 and the Restructuring of the British Financial System." *Review, Federal Reserve Bank of St. Louis*. 80: 53-76.

Ng, Kenneth (1988). "Free Banking Laws and Barriers to Entry in Banking, 1838-1860." *The Journal of Economic History* 48: 877-889.

Patterson, E. Stewart (1917) [2009]. *Banking Principles and Practice*. London: Dodo Press.

Schuster, Sir Felix (1923). *The Bank of England and the State.* London: University of Manchester Press.

Sylla, Richard, John B. Legler, and John J. Wallis (1987). "Banks and State Public Finance in the New Republic, 1790-1860." *Journal of Economic History* 47: 391-403.

Wallis, John J., Richard Sylla, and John B. Legler (1994). "The Interaction of Taxation and Regulation in Nineteenth-Century U.S. Banking." In *The Regulated Economy: An Historical Approach to Political Economy*, ed. Claudia Goldin and Gary Libecap, 121-44. Chicago: University of Chicago Press.

Wood, Gordon S. (1991). *The Radicalism of the American Revolution*. New York: Alfred A. Knopf.