KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

10 OCTOBER 2022

Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2022

FINANCIAL INTERMEDIATION AND THE ECONOMY

The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel

BOX 50005 (LILLA FRESCATIVÄGEN 4 A), SE-104 05 STOCKHOLM, SWEDEN
TEL +46 8 673 95 00 ⋆ WWW.KVA.SE

Financial Intermediation and the Economy

To Ben S. Bernanke, Douglas W. Diamond, and Philip H. Dybvig
"for their research on banks and financial crises"

1. Introduction

The 2022 Sveriges Riksbank Prize in Economic Sciences in Honor of Alfred Nobel rewards foundational research on the role of banks in the economy, particularly during financial crises. Financial intermediaries such as traditional banks and other bank-like institutions facilitate loans between lenders and borrowers, and thereby play a key role for the allocation of capital. They enable households to get a mortgage to buy a home, farms to get a loan to buy a harvesting machine, and firms to get a loan to build a new factory.

However, financial intermediaries also play a key role during times of significant economic distress. For example, during the Great Depression in the 1930s, a large number of banks failed and the credit supply contracted significantly, further deepening and prolonging the recession. Another example is the 2007-2009 Global Financial Crisis, which arguably started in the financial sector, and financial intermediaries were at the core as the crisis unfolded. This crisis led to a long recession, sometimes called the Great Recession. The fact that banks and other financial intermediaries perform important functions but at the same time can be associated with devastating crises poses a critical challenge to policymakers.

Two parallel research projects that originated in the early 1980s, both motivated by the experiences of the banking sector during the Great Depression, have significantly advanced our understanding of the role banks play in the economy. Douglas Diamond and Philip Dybvig developed theoretical models to explore the role banks play in the economy and why they are vulnerable to bank runs. Specifically, Diamond and Dybvig (1983) presented a theory of *maturity transformation* and showed that an institution using demand deposits to finance long-term projects is the most efficient arrangement, but that, at the same time, this arrangement has an inherent vulnerability: bank runs may arise.

Diamond (1984) developed a theory of a bank's provision of *delegated monitoring* services and showed that banks can ensure that projects with high (but risky) long-run returns obtain funding by monitoring borrowers on behalf of lenders. This was an entirely new approach to understanding banks; earlier

researchers had taken the functions of banks as given and had not attempted to explain their fundamental role in society.[1]

The empirical research by Ben Bernanke addressed the same questions. His study object was the Great Depression: the deepest and longest economic downturn in recorded history, one that began in the U.S. but also became global. Motivated by a combination of theoretical arguments, he provided historical documentary evidence and empirical data to uncover the importance of the credit channel for the propagation of the depression.

Bernanke (1983) showed, in particular, that the downturn became so deep and so protracted in large part because bank failures destroyed valuable banking relationships, and the resulting credit supply contraction left significant scars in the real economy. These were new insights; earlier economic historians had viewed bank failures merely as a consequence of the downturn, or mattering to the rest of the economy only by contracting the money supply, rather than directly damaging investments through severed credit arrangements.[2] Thus, Bernanke's work was not only relevant for understanding the Great Depression, but also more generally for providing evidence on the critical role of banks in the economy.

The theoretical and empirical findings of Bernanke, Diamond, and Dybvig thus reinforce each other. Together they offer important insights into the beneficial role that banks play in the economy, but also into how their vulnerabilities can lead to devastating financial crises. The findings have proven extremely valuable for policymakers: the actions taken by central banks and financial regulators around the world in confronting two recent major crises – the Great Recession and the economic downturn that was generated by the COVID-19 pandemic – were in large part motivated by the laureates' research.

1.1 The role of financial intermediaries

Institutions such as banks and similar financial intermediaries exist arguably because financial markets fundamentally channel savings toward real investment. In the aggregate economy, savings must equal investments, but investment opportunities and the willingness and ability to save usually do not coincide at the individual level.

At a given point in time, some firms and households want to invest in, say, a house or a factory building that costs more than they can pay with their current income. At the same time, others are in the opposite situation and prefer not to spend all their income as it accrues.

[1] See below and the discussion in Section 3.1.
[2] See, e.g., Keynes (1936) and Friedman and Schwartz (1963).

Moreover, different savers have different liquidity needs and differ in their willingness to take on risk, while investment projects differ greatly in these same dimensions: some are long- and some are short-term, and some are very risky, while others are comparatively safe. The role of financial markets is to solve the problem of coincidence of saving and investment, while taking into account the needs of different savers and investors.

Financial contracting sometimes involves ultimate borrowers and lenders directly, such as when a supplier provides trade credit to a buyer, or when a corporation issues new stocks and bonds to the general public. But it would be highly costly and impractical if all financial contracts were of this form. For example, it would likely be prohibitively costly for a home buyer to write a separate financial contract with every individual lender that ultimately finances her mortgage. Furthermore, if every lender required the contract to stipulate that she had the right to get her money back on demand, costs would escalate quickly, as the borrower may repeatedly have to seek refinancing.

To solve this problem, financial intermediaries such as banks and mutual funds exist. These institutions channel funds from savers to investors, receiving funds from some customers and using the funds to finance others. They also make it possible for the borrower to have a long-term financing agreement at the same time as lenders can withdraw the money they lent on demand. The latter function is, precisely, the maturity transformation referred to above. On the household side, short-maturity "demand deposits" (or "checking accounts"), are transformed into long-maturity mortgages.

Thus, financial intermediaries perform essential functions for society, but history has demonstrated that they also can be fragile, admitting phenomena such as bank runs. These occur when depositors "panic" and rush to withdraw their funds, leaving the bank without sufficient assets. Bank runs were frequent during the Great Depression – and occurred before that, as well as after; for a Hollywood illustration, watch the 1946 classic movie *It's a Wonderful Life*.[3] During the Great Recession, the runs involved other financial intermediaries than traditional banks – so-called shadow banks – but the phenomenon was fundamentally the same.

Bank runs can be contagious, driving large parts of financial intermediation to a halt. Such systemic financial crises are typically followed by deep economic downturns, as was the case during the Great Depression, the Great Recession, and many other financial crises around the world (e.g., the banking crises in Scandinavia in the early 1990s).

[3] Capra, Frank, and James Stewart, *It's a Wonderful Life*. Los Angeles, CA: Liberty Films, 1946.

1.2 Three complementary insights

The laureates' research can be said to have generated three complementary insights, each with long-lasting impact both on research and policy.

Valuable maturity transformation is inherently vulnerable (Diamond and Dybvig, 1983). Bank lending is short-term because a regular household typically wants its savings to be available "on demand," while investors typically need to borrow money over long horizons. By pooling funds from many savers, who may each demand liquidity but who are unlikely to all do so at the same time, the intermediary can fund long-term projects that offer higher returns.

The bank deposit contract enables savers to share the risk they face because the timing of their consumption needs is uncertain, and it makes all savers better off than they would be without the bank. Effectively, the bank transforms long-term borrowing into short-term lending: they create liquidity. However, the very nature of maturity transformation makes banks vulnerable to self-fulfilling runs; concerns that other people will demand their deposits back, leading the bank to have insufficient funds, may lead all savers to run to the bank to withdraw their money.

Even fundamentally healthy banks may get into trouble if such bank runs become widespread. Government policies, such as insuring deposits or a central bank acting as a lender of last resort, can help prevent bank runs.

Delegated monitoring allows savers to get access to safe, high returns (Diamond, 1984). By using their expertise in evaluating and monitoring borrowers, and by pooling funds from many savers and diversifying across borrowers, banks reduce the aggregate monitoring costs that would otherwise have been borne by borrowers. This enables households' savings to be channeled to productive investments at a lower cost.

Banks, by lending to many different risky investments and financing this by essentially safe debt from many different savers, obtain the right incentive to monitor the borrowers. The key is that depositors hold debt, and can force the bank into bankruptcy if it does not repay their deposits. Without this feature, savers would have to monitor the banks themselves and little gain would result from having banks as intermediaries. Pooling is thus key in solving the problem of "who should monitor the monitor."

Jointly, these two theoretical contributions explain how financial intermediaries create liquidity in the economy. On the liability side, banks pool many savers together, which enables savers who end up needing liquidity to have their demand met by long-term savers who do not need liquidity at the same

moment in time. On the asset side, banks pool many loans together and monitor them on behalf of savers, which makes it possible to finance risky and illiquid loans with much less risky and liquid deposits. The analysis on the liability side takes as given that there are highly productive, but long-term, investment projects to fund; the analysis on the asset side shows how banks can make such projects feasible, despite the fact that these projects often involve important elements of risk, even to the borrower.

> **Financial intermediation is key for real activity (Bernanke, 1983).** Historical documentary evidence and empirical data from the Great Depression show that bank runs had major real and long-run economic consequences. Banks provide important screening and credit-monitoring services, and they develop crucial long-term relationships with borrowers.
>
> When banks failed in large numbers, existing banking relationships were destroyed, and it took time to build new relationships. The result was a severe credit crunch affecting primarily households, farms, and small businesses, and this deepened and prolonged the Great Depression. In other words, the effects of the recession were exacerbated through the credit channel.

While Bernanke (1983) took issue with core economic theories that built on an absence of credit frictions, he did not provide a specific new theory to replace the frictionless benchmark. Such a theory was instead provided by Diamond and Dybvig. In fact, Bernanke's empirical evidence can be viewed as supporting the narrative told in Diamond and Dybvig's (1983) paper on bank runs and as providing support for the screening and monitoring roles emphasized in Diamond (1984).

1.3 Fundamental impact on economic research

When the work of Diamond and Dybvig appeared in the early 1980s, the role of financial intermediaries had been discussed for a long time. While economists had pointed to the roles financial intermediaries play, most of this discussion had been informal.[4] More formal theories had emerged during the 1970s, but they assumed rather than explained the existence and structure of banks.

By contrast, Diamond and Dybvig's research provided logically consistent mathematical models, where the existence and structure of banks were derived rather than assumed. By providing formal models based on microeconomic foundations, the key assumptions and economic mechanisms were laid bare. This in turn enabled others to later modify and extend the ideas. In some cases, follow-on work would

[4] One notable exception is Edgeworth (1888).

criticize certain assumptions and implications of the original models, and in doing so, developed important complementary perspectives.

Together, the original contributions and the subsequent literature gave rise to the modern theory of banking, which today forms one of the pillars of economics and finance. Its emphasis on microeconomic foundations has made this theory applicable to many different types of financial intermediaries and helps researchers analyze the consequences of the evolving financial landscape. As changes in technology and regulation lead to new types of institutions entering financial markets, competing with or even replacing traditional banks, these new actors still perform the same fundamental functions and are subject to the same underlying fragilities.

Similarly, before Bernanke's (1983) work, the role of the financial sector had (with only a few exceptions) been neglected in macroeconomic analysis and policy, and the importance of bank failures and credit crunches in the Great Depression, emphasized by contemporary researchers such as Fisher (1933), had been largely forgotten. Following Bernanke's work, macroeconomists started analyzing the role of financial markets and intermediaries in propagating economic fluctuations and recessions. The basic mechanisms that Bernanke (1983) shows were at work during the Great Depression were shown to be of first-order importance in subsequent financial crises around the world.

Research is still ongoing, and we are gradually achieving a deeper understanding of the role that financial intermediation plays in the macroeconomy. Researchers have used recent crises to test the mechanisms proposed by Bernanke (1983) using better data and more sophisticated econometric methods, and the evidence supports his credit-channel hypothesis.

1.4 Importance for financial regulation and policy

The research by Bernanke, Diamond, and Dybvig, along with all the work that followed, delivers a framework that guides regulation of the financial system and can be used to evaluate various policy tools aimed at combating economic crises. The financial system is continuously changing, however – new types of players enter and new financial instruments are developed. Thus, financial regulation must evolve, and this requires continuous analysis of how policies should be designed to be most effective.

Policy interventions such as deposit insurance come not only with benefits, but also with potentially significant costs. Many observers have argued, for example, that excessive protection of banks can lead to moral hazard and may contribute to inequality. In this debate, however, the theoretical frameworks rewarded in this year's Prize help policymakers by highlighting the relevant economic mechanisms and trade-offs that need to be considered when designing financial policy.

1.5 Outline of document

The remainder of the document is laid out as follows. Section 2 discusses what banks do and briefly discusses the historical role of banks. Section 3 considers how the laureates' research relates to earlier and contemporaneous work in banking and discusses the laureates' contributions in detail. Section 4 turns to subsequent research and evidence provided both by the laureates themselves and by others. Section 5 discusses the Global Financial Crisis, specifically how policy responses were influenced by the laureates' research and how these policy interventions have led to a new approach to "macroprudential" financial regulation. Section 6 concludes this scientific background.

2. What banks do

In this section, we briefly describe the well-known core functions of banks. We also provide a historical perspective on banks and banking panics and some comments on modern banks and bank-like institutions. The latter will be more thoroughly discussed in Section 5 below.

2.1 Central functions

Most of the money that households own is in the form of bank deposits: as we work and earn wages, our employers transfer funds to our bank accounts, typically monthly or bi-weekly. We then pay bills by making electronic transfers from our accounts, and we use debit and credit cards, or checks, connected to our accounts to make purchases.

In this way, bank deposits provide "liquidity" for households: the money deposited in the bank can be withdrawn whenever we need to use it for purchases and consumption, without prior notice. Since we are typically unsure in advance of when and exactly how much money we will need to spend, bank deposits – which are a safe way of storing wealth and also typically pay interest – strike a balance between convenience and generating a return on our savings.

Thus, on a bank's balance sheet, a central component of liabilities consists of deposits from households. On the asset side are bank loans granted to households and firms. Unlike deposits, most of these loans are long-term and illiquid.

When a household borrows from the bank to buy a house, the loan typically takes the form of a multi-year mortgage; maturities of 20 years or more are quite common. Similarly, when firms borrow from banks to finance investment, it is often in the form of a term loan that the firm is required to repay over several years.

When banks lend money, they also *monitor* the borrower, i.e., examine the risk that the borrower is unable to repay the loan. Why it is efficient to delegate this monitoring function to banks, and what implications this has for their financial structure, is a central question addressed in Diamond's (1984) fundamental contribution.

Banks make money from the fact that the interest rate they charge on the loans they grant is higher than the rates they offer on bank deposits and other debt: their assets (bank loans) pay higher returns than they pay for the liabilities (bank deposits). Banks also have other assets and liabilities; on the asset side, they have "reserves" (traditionally cash in a vault but now held as low-interest deposits at the central bank), and on the liability side is equity as well as other debt instruments. This additional debt is called "wholesale funding" and includes, for example, certificates of deposits/commercial paper traded in financial markets, typically of short maturity, and repurchase agreements or "repos", which are also short-term.

The general description of a banks' balance sheets is thus that (i) their assets have longer maturity than their liabilities; and (ii) their assets pay higher return than they pay on their liabilities, thus generating returns to the bank's equity holders. The first property is often referred to as *maturity transformation*: short-term liquid assets (deposits) are transformed into long-term illiquid assets (bank loans).

As we shall see momentarily, maturity transformation is coupled with an inherent, and much-discussed, vulnerability. On the surface, the second property makes banks look like arbitrageurs: they borrow at a lower rate than they lend and hence are "money-printing machines": the higher the volume of activity, the more money they make.

Banks' money-printing abilities and their vulnerability begs the question what the societal value of banks is, if any. This question is at the heart of the research of this year's laureates: while emphasizing the fragility of banks, but also their importance in the economy particularly during financial crises, their research helps us understand the fundamental role that banks play for households and firms.

Why does the maturity transformation of the kind banks provide gives rise to fragility? Because the bank's deposits are "on demand," it is possible that all the depositors will need liquidity, and hence withdraw their funds at the same time. And if they do, banks will need to sell their assets, most of which typically return much less if liquidated early than if they were held to maturity.

It is even possible that the bank will be unable to honor its obligations to depositors and other investors, in which case the bank would fail. Fortunately, this is unlikely to happen under normal circumstances, but as we shall see, there are many examples of bank failures throughout history. A particularly troubling situation is a bank run or (bank panic), where depositors rush to the bank to withdraw their

funds because they expect others to do the same, i.e., they hope to be first in line while there are still some funds left in the bank.

How bank runs can occur, as an outcome of rational decision making, is a central question to which Diamond and Dybvig (1983) provides an answer. Bank runs have happened regularly all over the world, even before the Great Depression.[5] We now offer a brief history of banks.

2.2 Historical background

Modern-day banks perform a variety of functions, but their central features – allowing savings to turn into investments and accepting deposits and lending out money – can be traced very far back in time, including to various parts of Asia thousands of years ago, ancient Egypt, and the Greek and Roman empires.[6] Regulation of the borrowing and lending activities is also an old phenomenon and has a variety of origins; for example, several religions have adopted restrictions on the payment of interest.

The origin of public banks performing the basic intermediation functions can be traced back to Medieval Italy. During the 17th century, banks also started issuing assets that were used as a means of payment: "inside money," i.e., a means of payment whose supply is generated through private markets (and thus not by the government). Toward the end of the 17th century, central banks emerged in several countries, i.e., government-controlled banks that issue money ("outside" money); in the U.S., President Abraham Lincoln had his government produce paper money in the form of "greenbacks," which were legal tender, to finance the American Civil War.

In many instances, banks successfully created widely circulating, inside money: they issued IOUs stating that the bank owed the holder a nominal amount ("I owe you"). These IOUs enabled banks to effect maturity transformation, but also made banks vulnerable to bank runs and ultimately bank failures.

Bank failures are costly not just because liquidating the bank's long-term claims prematurely would rarely bring about nearly as much as they would had these assets been held to maturity, but also because of the costs involved in carrying out bankruptcy proceedings. In the U.S., during the second half of the 19th century, bank runs were actually fairly common, as can be seen from Figure 1; in 1873 and 1893 there were more general banking panics, involving runs against many banks at the same time (over 100 banks failed in the former year and over 500 in the latter). In other words, a bank run on one bank appears to be able to spread and "contaminate" other banks.

[5] See for example Bernanke and James (1991) Table 2.7.
[6] For the origins of banking, see, e.g., Kindleberger (1993).

Policymakers understood that a temporary suspension of bank operations – suspension of convertibility of the bank's deposits, during which depositors would keep their claims but not be able to withdraw funds – could remedy the situation, but temporary suspensions were not implemented systematically and efficiently until the 20th century. The need to find effective policies that could address bank runs and the threat of systemic banking panics was never more pressing than during the Great Depression.[7]

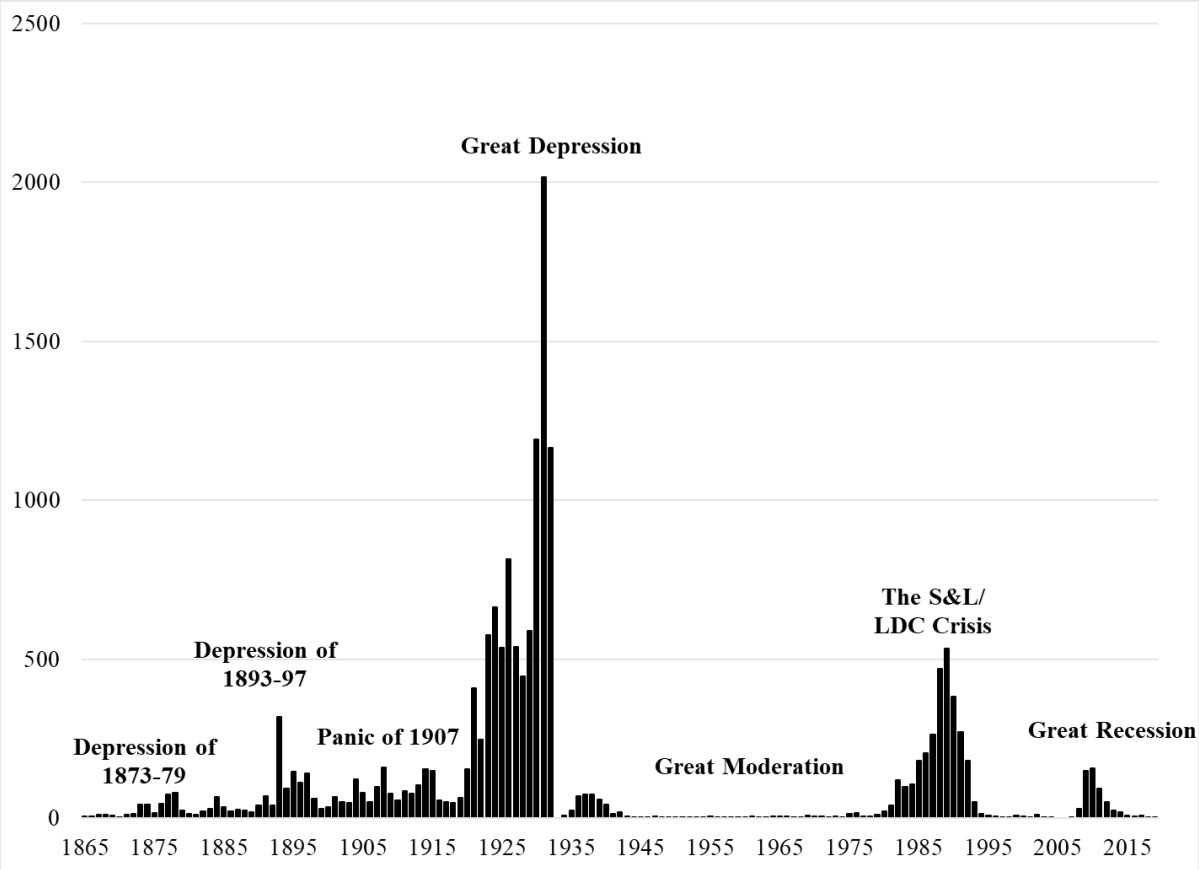

Figure 1: Bank failures 1865–2018, as reported by the Federal Deposit Insurance Corporation. Data from 1865–1932 are drawn from the *Annual Report for the Federal Deposit Insurance Corporation for the Year Ending December 31, 1934*, Tables 37–40, pp. 92–95. Data for 1934–2018 are drawn from FDICs BankFind Suite: Bank Failures & Assistance Data.

The Great Depression is by far the deepest and longest recession since systematic data on economic variables such as Gross Domestic Product (GDP) became available. It began with a stock market crash in the U.S. in 1929 – within a few years the market lost about 90% of its value – and lasted for a long time (10 years in the U.S.). The Great Depression spread rapidly around the world, in the form of stock market crashes, large decreases in international trade, and a dramatic contraction of economic activity.

[7] While this historical background focuses on the U.S., systemic banking crises were far from just a U.S. phenomenon. Reinhart and Rogoff (2009) document 54 banking crises in other countries than the U.S. over the period 1800-1930; in total, they identify 268 such crises across the world between 1800 and 2008.

Figure 2 illustrates the importance of the Great Depression in a historical context for the U.S. The figure shows output measured as logarithm of GDP per capita in the U.S.; output grows at a remarkably constant rate of a couple of percent per year from the first data points in the mid-19th century. The only large blip that is visible is the Great Depression (and the rebound from it). Other recessions are visible but are, in a historical context, minor from the perspective of the aggregate output produced. The figure ends in 2018; we will discuss the Global Financial Crisis of 2007–2009 and the recession that followed it in Section 5 below.

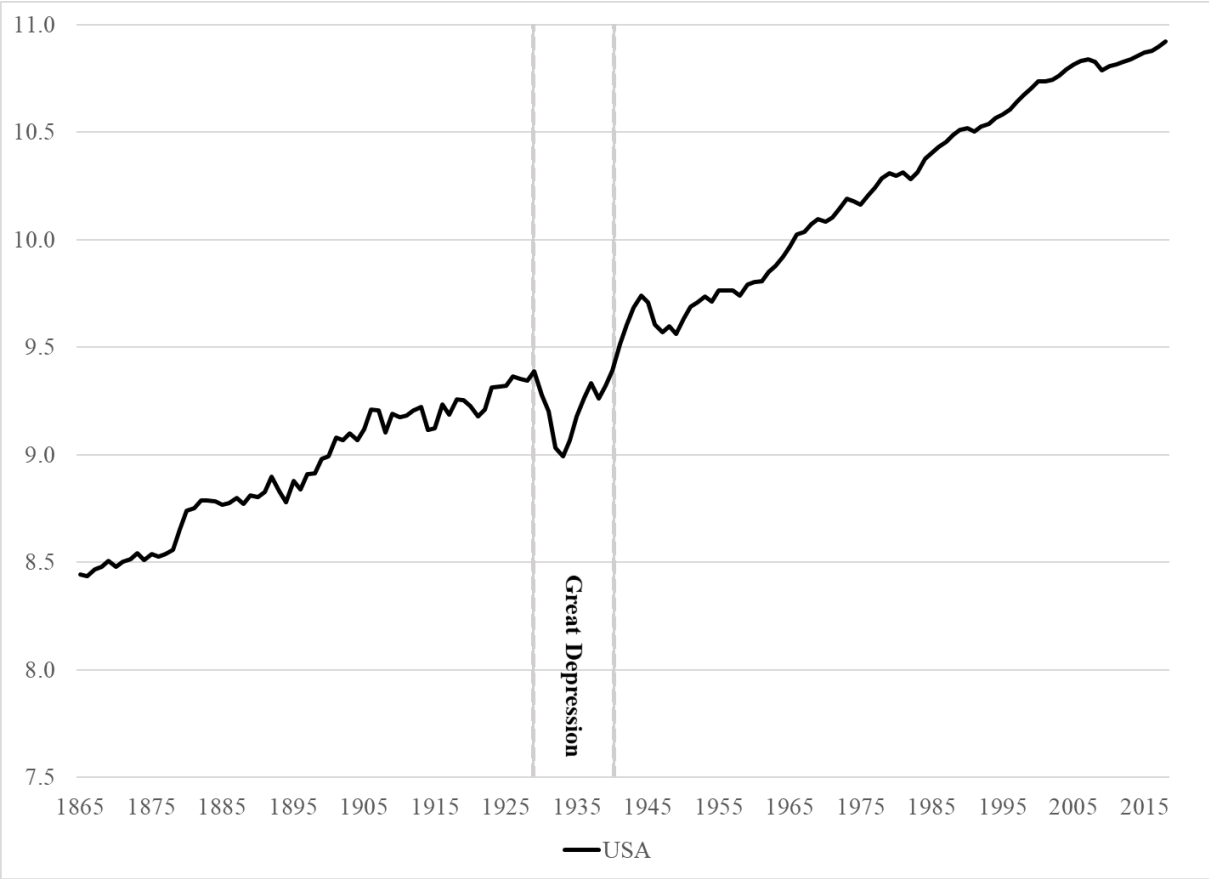

Figure 2: Logarithm of U.S. per capita GDP, 1865–2018. GDP data are adjusted for differences in the cost of living between countries and for inflation. GDP is measured in constant 2011 international dollars. Data were drawn from the Maddison Project Database, version 2020.

The U.S. was not the only country affected by the Great Depression. Figure 3 covers 1930–1935 and shows the timing and magnitude of the maximum decline in GDP per capita from last peak for select countries, expressed in percent of the last peak. Countries such as the U.K. and Japan hit their lowest point in 1931, while output in Mexico declined by 31.4% and in Singapore by 41.2% in 1932. In 1933, the U.S. hit its lowest point representing a 32.7% decline relative to the most recent peak, but the drop in output was even larger for Canada (34.8%) and Cuba (39.41%).

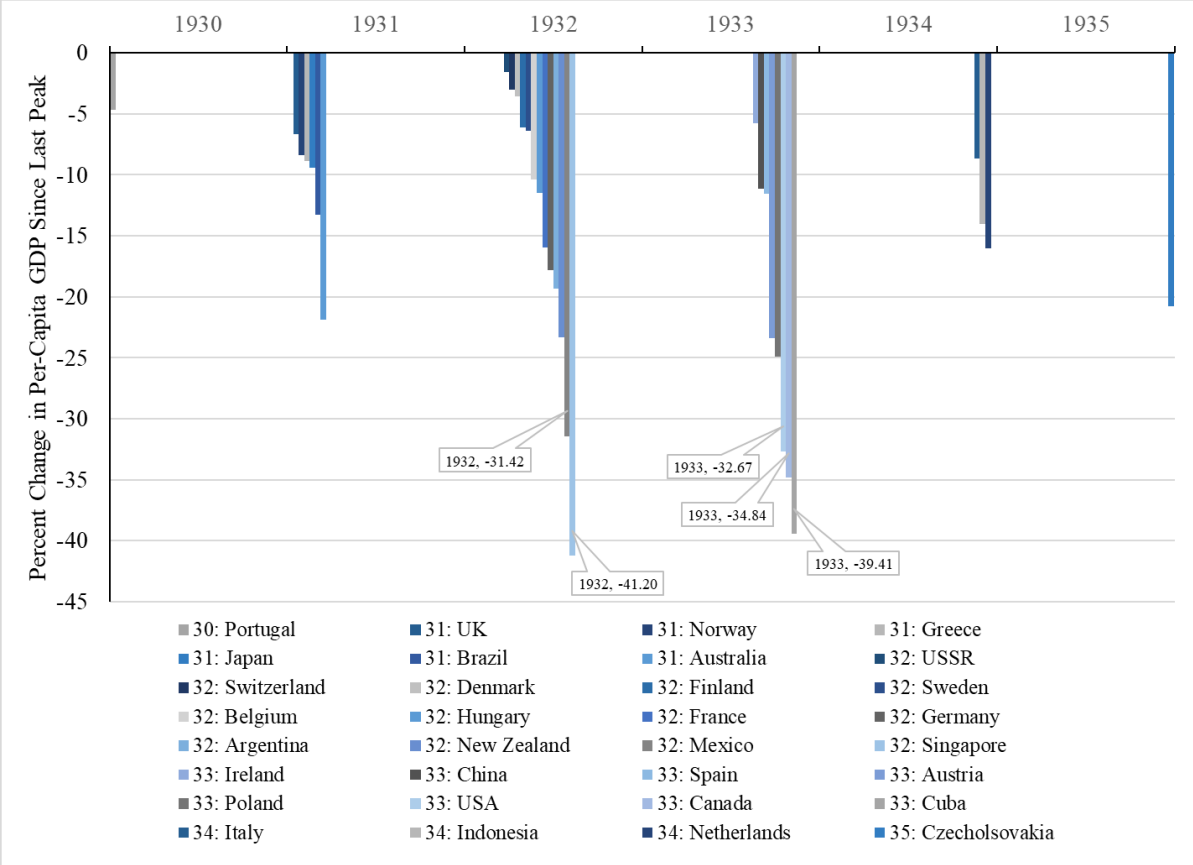

Figure 3: Percent change in per capita GDP since last peak for select countries, 1930–1935. Countries are sorted by percent change in per capita GDP within each year. Callouts for countries each year with more than a 30% change in per capita GDP are from left to right: 1932, Mexico, Singapore; 1933: USA, Canada, and Cuba. GDP data are adjusted for differences in the cost of living between countries and for inflation. They are measured in constant 2011 international dollars. Data were drawn from the Maddison Project Database, version 2020.

The fact that the Great Depression was so significant, through its major impact on so many economies around the world and on the welfare of so many people, has of course been reflected in economic research in many ways. The quest to understand the Great Depression – its roots, its depth, and its duration – has generated major new insights into how economies work. In particular, we have learned about a number of weaknesses in market economies and how to handle them with regulation and economic policy; Keynesian macroeconomics grew out of these insights.[8]

Economic historians paid less attention to the role of banks and imperfectly functioning credit markets during the Great Depression. Banking panics erupted in 1930 and kept occurring at a historically high

[8] Keynes (1936) argued that recessions were primarily due to drops in aggregate demand, moving economic output below the production capacity of the economy. According to this view, governments should counter recessions through an expansionary fiscal policy that boosts aggregate demand.

rate until 1933, with an accumulated number of bank failures in the thousands (see Figure 1 above); nearly half of all the banks failed. In 1933, a Banking Holiday was pronounced, i.e., banks were temporarily closed and this became a turning point.

Outside of the U.S., Austria, Belgium, Estonia, France, Germany, Hungary, Italy, Latvia, Poland, and Romania suffered serious banking panics during 1930–1933. Several countries in Latin America (e.g. Argentina, Chile, and Mexico) and Asia (e.g. British India, Indonesia, and Singapore) also suffered dramatic economic downturns during the Great Depression.[9]

In the subsequent research on the Great Depression, bank failures were typically viewed as a consequence of the depression rather than as an important element in explaining its evolution. One notable exception was Friedman and Schwartz (1963), who argued that the bank failures where of first-order importance through their effect on money supply. Since the money stock consists both of outstanding currency and bank deposits, the dramatic drop in deposits caused by the bank panics – and the U.S. Federal Reserve (Fed) not countering this drop with expansionary monetary policy – caused a drop in money supply that in turn led to deflation that exacerbated the economic downturn. They also argue that the actual credit losses of banks going into the panics were relatively small, but that the subsequent run on bank deposits, which also affected solvent banks, was the main reason for the dramatic contraction of the money stock.

A new perspective, however, came about with Bernanke's fundamental paper, published in *American Economic Review* in 1983. Similar to Friedman and Schwartz, he argued that banks and banking panics stood in the center of why the recession became so deep and so long-lived, but not only because of its effect on money supply. According to his narrative and empirical analysis, supported with documentary evidence from the period, he argues that the contraction of banking activity was detrimental because it disrupted the intermediation of credit between lenders and borrowers. Since bank relationships where not easy to replace, this lead to severe credit constraints for bank-dependent borrowers – particularly farms, small firms, and households – which depressed economic activity and contributed to a deeper and longer downturn.

Around the same time as Bernanke wrote his paper, and also motivated in part by the role of banks during the Great Depression, Diamond and Dybvig set out to understand the fundamental role banks play in the economy and, at the same time, their apparent vulnerability. Those theories led to path-breaking papers that, together with Bernanke's work, provided a new framework for understanding the crucial role banks play in the economy. The combination of Diamond and Dybvig's conceptual

[9] See Figure 3 and Rothermund (1996).

theoretical work and Bernanke's careful study of the role of banks in an economic downturn have provided us with a fundamental understanding of how banks operate and their role in the economy. We will describe their contributions in Section 3 and discuss their influence on subsequent research in Section 4.

2.3 Contemporary financial markets

Today's financial markets include several institutions aside from banks that, broadly speaking, channel savers' money to investors, and some of them also engage in maturity transformation. Stock markets provide direct links whereby savers fund publicly traded companies; publicly traded stocks are liquid in that they can be sold quickly, but they do not offer a safe return. Another source of investor funding is bond issuance; bonds of larger companies can often also be traded in centralized markets and thus offer a degree of liquidity. A more recent phenomenon is the growth of venture capital and private equity firms that offer funding for companies that are not publicly traded, but these firms do not engage in maturity transformation and are typically not accessible to most individual savers.

Today there are also a number of intermediaries – such as securitization vehicles and money market mutual funds – that provide debt-financing like banks do but operate largely outside of the regulated banking system. By financing long-term illiquid investment with shorter-term and more liquid instruments, these non-bank intermediaries also engage in maturity transformation. Because of this similarity, they are often referred to as *shadow banks.* These shadow banks now account for a significant share of intermediation activity in the economy and the failure of shadow banks were at the heart of the Great Recession, sometimes also called the Global Financial Crisis.[10] Thus, also in terms of their vulnerability to runs and panics to these institutions resemble banks.

Finally, commercial banks have developed in a number of ways, not least of which is their reliance on "wholesale funding," a significant new liability item on their balance sheet: short-term borrowing in money markets from financial institutions. We return to the more recent banking history, shadow banks included, in Section 5.

3. Banks and the economy

In this section, we first discuss banking research predating the laureates' contributions. We then in turn discuss Diamond and Dybvig's (1983) model of maturity transformation, Diamond's (1984) model of

[10] Poszar et al., 2010; Acharya et al., 2013; Buchak et al., 2018.

delegated monitoring, and Bernanke's (1983) study of financial intermediation during the Great Depression.

3.1 Banking research before the laureates' contributions

While banks have been discussed by economists since at least the 18th century, in particular by David Hume and Adam Smith, this discourse has taken place mostly in the context of monetary economics. In particular, banks created "inside money" in the form of bank deposits, which added to the "outside money" supplied by the government. Since not all depositors would need to withdraw their deposits at the same time, banks only needed to hold relatively little currency (or "reserves") in order to create a much larger quantity of deposits.

Based on statistical theory, Edgeworth (1888) provided the first formal model of this mechanism, which inspired a separate literature on cash inventory management. Bagehot (1873) argued for the role of the central bank as a "lender of last resort" in cases when there was not enough reserves to cover withdrawals.[11]

In contrast, relatively few of the early economists emphasized the role of banks for the allocation of capital in the economy. Important exceptions were Böhm-Bawerk (1911) and Schumpeter (1911), who argued that the services provided by financial intermediaries – mobilizing savings, evaluating projects, managing risk, monitoring managers, and facilitating transactions – are essential for technological innovation and economic development.

In the decades before Bernanke's (1983) contribution, the macroeconomic discussion on financial intermediation was dominated by the question of whether the inside money creation of banks was desirable. On one side of the debate, some economists were in favor of the real bills doctrine (going back to Adam Smith, 1776). They argued that banks should be allowed to freely create inside money and market forces would prevent excessive "credit creation" by private banks. Proponents of this view (see, e.g., Gurley and Shaw, 1960) argued that limiting the ability of banks to create inside money, e.g., by requiring minimum reserve ratios, "forces the private banking sector frequently to become a 'disequilibrium system' – for it is forcefully prevented in this way from achieving its optimum portfolio" (Patinkin, 1961, p. 99).

On the other side, proponents of the quantity theory of money (building on, e.g., Wicksell, 1898) argued that a real bills regime permits excessive fluctuations in the supply of money and, hence, in the price

[11] In particular, Bagehot asserts that the lender of last resort should lend freely during a banking crisis; however, it should only lend to solvent banks and at a penalty interest rate, to promote good *ex ante* behavior among bankers (or as modern economists would put it, avoid "*ex ante* moral hazard"). This policy was coined the "Bagehot doctrine" and is commonly referred to in modern policy discussions (see, e.g., Goodhart, 2010).

level, which amplified business cycle fluctuations. One of the most important arguments of this view was provided by Friedman and Schwartz (1963), which was discussed above, who argued that the adherence of the Fed to the real bills doctrine was what caused the drop in money supply following the bank failures in 1930–1933. As a result, many economists advocated legal restrictions on banks in order to separate "money creation" from the process of private intermediation (e.g., Despres et al., 1950).

Milton Friedman, one of the main proponents of the quantity theory, went as far as to argue in favor of 100% reserve banks, i.e., that banks should only be allowed to invest deposits in cash or risk-free government debt (Friedman, 1948). A few economists, such as Brainard and Tobin (1963a, 1963b), criticize this view for ignoring the role of financial intermediaries for transferring savings to investment:

> "[I]ntermediation permits borrowers who wish to expand their investments in real assets to be accommodated at lower rates and easier terms than if they had to borrow directly from the lenders. If the creditors of financial intermediaries had to hold instead the kinds of obligations that private borrowers are capable of providing, they would certainly insist on higher rates and stricter term" (Brainard and Tobin, 1963a).

The macroeconomic discussion on banks was largely informal, however, and when formalized, macroeconomic models that were founded on microeconomics typically abstracted from the role of financial intermediaries. The few exceptions, such as Sargent and Wallace (1982), focus on banks' role for money creation and abstract from their role in financing real investment.

The view of banks in the microeconomics and finance literature during this time was very much influenced by the seminal insights of Miller and Modigliani (Modigliani and Miller, 1958; Miller and Modigliani, 1961). They show that in a world without frictions, the way firms were financed was irrelevant.

In such a world, firms make investment decisions in order to maximize their value, which in turn is simply equal to its discounted expected cash flow. The discount rate is given by the firm's cost of capital, which is equal to the required return of investors in competitive capital markets. Expected cash flow is determined by rational managers who make decisions in order to maximize firm value. A given capital structure is simply one way of splitting this value between different capital providers – providers of debt and equity – but has no impact on cash flows or the firm's average cost of capital, and is thus irrelevant for firm value.

Their theorem turned out to have far-reaching consequences: in the absence of frictions, financial intermediation would not affect firm value either, and would thus be irrelevant for capital provision. As

long as an investment project is valuable, firms would be able to obtain capital for it in a competitive capital market.[12]

In order to generate a role for financial intermediation, then, there had to exist some capital market friction that intermediaries were able to overcome. The banking literature that emerged in the 1970s explained banks as having a cost advantage relative to direct lending by savers, and as being able to efficiently intermediate the maturity mismatch between borrowers and lenders. One branch of the literature focused on banks having access to technology that enables them to reduce transaction costs when matching borrowers and savers (see, e.g., Kaufman, 1973, Ch. 4). Specifically, they can issue higher yielding securities (e.g., deposits) than alternatives because they have an advantage in finding profitable investments.

Banks were also thought to have special credit evaluation skills, enabling them to make loans at lower costs than savers if they were to make loans directly to borrowers. Together, these features implied that banks had a special appeal not just to savers but also to borrowers.

Another strand of the early banking literature emphasized the "preferred habitats" of borrowers and lenders. Goodhart (1975, Ch. 6) describes the preference of borrowers to match the length of their borrowing to the length of their investment, while savers are hypothesized to prefer safer, shorter maturity. This creates a potential mismatch between the horizons of the borrowers and lenders, resulting in borrowers having to pay a premium to borrow long-term, and savers having to accept a discounted return to have access to short-term liquidity. According to this theory, banks can reduce the mismatch in preferences by issuing short maturity securities and investing in longer maturity loans, thereby profiting from the return differences.

While these early arguments foreshadowed some of the insights in Diamond and Dybvig's research and what became the modern theory of banking, their accounts were incomplete in a variety of ways and lacked a coherent foundation in microeconomics. The more formalized models of banks assumed some exogenous transaction cost or exogenously given difference between borrowing and lending rates, and these models took the institutional setup of banks as given.

In his review of the banking literature written just before the contributions of Diamond and Dybvig, Baltensperger (1980) wrote: "There exist a number of rival models and approaches which have not yet been forged together to form a coherent, unified and generally accepted theory of bank behaviour." Why would savers necessarily prefer safer, shorter-dated investments? Why would firms desire to

[12] In the words of Fama (1980, p.39): "[W]hen banking is competitive, [the bank's] portfolio management activities in principle fall under the Modigliani-Miller theorem on the irrelevance of pure financing decisions. It follows that there is no need to control the deposit creation or security purchasing activities of banks to obtain a stable general equilibrium with respect to prices and real activity."

match maturity of financing with maturity of their investments? The logic of Modigliani and Miller (1958) suggests that financial structure should be irrelevant to firms, in the absence of frictions and transaction costs. What could account for the large transaction costs that would justify the existence (and dominance) of banks in financial markets?

Diamond and Dybvig were the first to provide a coherent theory, founded on microeconomics, showing that an outcome where banks offer short-term demand deposits to fund long-term investments arises endogenously as an optimal contract given reasonable economic assumptions. In the process of doing so, their work provides the foundation for what is now recognized as the modern theory of financial intermediation.

They offered two critical insights. One is that there are fundamental reasons why bank loans are a dominant source of financing in the economy and why banks are funded by short-term, demandable debt. The other is that banks are inherently fragile and thus subject to runs. We first describe each theoretical contribution below. We then turn to Bernanke's contribution, where he argues that failure to understand banks' role in the economy can have devastating consequences, as exemplified by the Great Depression.

3.2 Maturity transformation

Diamond and Dybvig (1983) modelled the maturity transformation role of banks. Many investment opportunities are long-term, while investors value short-term liquidity, i.e., the ability to withdraw their savings for immediate consumption if needed. The role of banks is to aggregate the savings of investors and invest in long-term projects. Only a fraction of investors will actually need to exercise their option to withdraw their savings early, since only a fraction of investors will be subject to short-term liquidity needs. This makes it possible for the bank to meet the liquidity needs of short-term investors, while investing their savings in productive long-term projects.

Diamond and Dybvig argue, however, that this maturity transformation makes banks inherently fragile and subject to self-fulfilling bank runs. The problem arises from the fact that if the bank had to liquidate all long-term investments early (at a loss), there would not be enough funds to cover all deposits. If a depositor believes that the other depositors will withdraw their funds from the bank, thereby forcing the bank to liquidate its long-term investments prematurely, she will also run to withdraw her deposits before the bank runs out of funds.

Diamond and Dybvig show how government regulation, such as deposit insurance or lender of last resort policies, can help avoid such coordination failures. Their model provided a unified and logically consistent framework for many of the informal arguments in the previous literature, and it stimulated a

large subsequent literature that has yielded new fundamental insights on issues such as financial contagion, inside money creation, financial propagation, and financial regulation.

Model setup

Diamond and Dybvig formulated a problem of optimal risk sharing across risk-averse agents who want to invest their savings for future consumption, and they showed that banks can enhance welfare by creating "liquidity." As their basic model is rather stripped down, it will be described in some detail here.

The model has three periods, $T = 0, 1, 2$, and a single consumption good. There are *ex ante* identical agents with unit endowment at $T = 0$, who value consumption c_T over the two periods 1 and 2. Their utility function is defined over these two consumption levels, each at the two states of nature relevant to them.

In period 1 in particular, an agent may become "impatient" and, at that moment, only value consumption in period 1; this occurs with probability λ. With probability $1 - \lambda$, the agent instead becomes patient and cares about the equally weighted sum of consumption in the two periods.[13] As of $T = 0$, the agent's utility is the probability-weighted average of a function $u(c)$, where c is a weighted sum of consumption in the two periods; this function is assumed to satisfy $u'(0) = \infty$ and $-cu''(c)/u'(c) > 1$, i.e., it has relative risk aversion larger than 1.

Note that the weighted sum of consumption in the two periods for the agent that turns out to be impatient is simply consumption in the first period. That is, we have

$$u(c_1, c_2) = \begin{cases} u(c_1) & \text{with probability } \lambda \ \text{ (impatient)} \\ u(c_1 + c_2) & \text{with probability } 1 - \lambda \ \text{ (patient)} \end{cases}$$

The model also has a technology, where an investment of 1 unit of output in a project at $T = 0$ yields R units of output in $T = 2$, with $R>1$. In period $T = 1$, however, the investment project can be terminated "prematurely," but such a liquidation is inefficient: the liquidation value is merely 1, which is taken to be significantly smaller than R. The invested unit may also be partly liquidated. It is clear, given the stated assumptions, that in the absence of any short-term liquidity needs, it would be optimal for all agents never to liquidate, so that they can consume R in $T = 2$.

[13] In fact, Diamond and Dybvig assume that patient agents only care about period 2 consumption but can store the period 1 consumption good at zero net return, hence making their utility depend on the sum of the consumption goods received in the two periods. Such a formulation is thus equivalent to the one entertained here.

Suppose each agent acts in isolation. Upon having invested their total endowment in the investment project, at $T = 1$, a fraction λ of agents will be hit by a liquidity shock and terminate the project; in this case, they will consume $c_1 = 1$. A fraction $1 - \lambda$ will, however, choose to continue the project and instead consume $c_2 = R$. Then, the expected utility of a typical agent will be

$$\lambda u(1) + (1 - \lambda)u(R).$$

In the autarky solution, each agent will bear the entire risk of being hit by a liquidity shock, and there is no risk sharing. When relative risk aversion is larger than 1, an agent would be better off sharing this risk: i.e., on the margin, a given agent would be better off consuming less in the good state (when there is no liquidity shock) and more in the bad state (when a liquidity shock hits).

To illustrate this, consider a benevolent social planner, who invests the total endowment in the long-term investment project at $T = 0$; liquidates a fraction $1 - x$ of the project and distributes the proceeds to impatient agents at $T = 1$; and distributes the investment proceeds from the remaining part of the project to patient agents at $T = 2$. The social planner will maximize the expected utility of a representative agent, i.e.,

$$\lambda u(c_1) + (1 - \lambda)\, u(c_2),$$

subject to the resource constraints

$$\lambda c_1 = 1 - x$$

and

$$(1 - \lambda)c_2 = xR.$$

Here, x is thus the fraction of the total investments of all agents. Substituting for x in the resource constraints, the social planner's problem simplifies to choosing c_1 and c_2 to maximize

$$\lambda\, u(c_1) + (1 - \lambda)\, u(c_2),$$

subject to

$$\lambda\, c_1 + (1 - \lambda)\left({}^{c_2}/_{R}\right) = 1.$$

The first-order conditions imply that the optimal consumption choices c_1^* and c_2^* satisfy

$$u'(c_1^*) \,/\, u'(c_2^*) = R.$$

To see that the autarky solution where $c_1 = 1$ and $c_2 = R$ is not optimal, note that $-cu''(c)/u'(c) > 1$ implies that the function $cu'(c)$ is decreasing.[14] Hence,

$$Ru'(R) < 1 \cdot u'(1),$$

or

$$u'(1)/u'(R) > R.$$

Since $u''(c) < 0$, the optimal solution will have $c_1^* > 1$ and $c_2^* < R$. Finally, note that R being above one and $u''(c) < 0$ imply that the solution has $c_1^* < c_2^*$.

In other words, there is a role for insurance: *ex ante*, agents would be willing to reduce their expected long-term consumption if it allowed them to increase their short-term consumption when they are subject to a liquidity shock. However, since Diamond and Dybvig assumed that whether an agent gets hit by a liquidity shock or not is private information and cannot be contracted upon *ex ante*, the idiosyncratic liquidity shocks are not insurable in a competitive market setting.[15]

Absent an insurance contract, agents who turn out to value long-term consumption have no interest in providing insurance to those facing liquidity problems.These agents may even pretend that they have experienced a shock in order to get paid themselves. As a result, agents who turn out to be impatient will simply have to liquidate their investment early to meet their liquidity needs and there is too little risk sharing in the economy.

Banks perform maturity transformation

Diamond and Dybvig then explained how financial intermediaries – such as banks – can restore efficient risk sharing. In other words, they can offer an arrangement that reproduces the efficient outcome described above.

Instead of investing directly in the technology themselves, agents deposit their savings at the bank, which in turn invests everyone's deposits. Thus, while agents cannot themselves achieve risk sharing,

[14] If $f(c) = cu'(c)$, then $f'(c) = cu''(c) + u'(c)$. Setting $f'(c) < 0$ implies $-cu''(c)/u'(c) < 1$.

[15] The insurance contract would be incentive-compatible, i.e., it would involve truth-telling for any given agent, if this agent knew that all other agents were truth-telling and did not engage in any side transactions. These presumptions are, in general, not met in this model. Since such an insurance contract would compensate agents who ended up being impatient, patient agents would have an incentive to lie and claim they were impatient as well.

a contract where agents deposit their endowment at a bank, and the bank in turn allows an agent to withdraw her deposits whenever she likes, can implement the optimal solution.

Under the conditions stated on the utility function, if an agent withdraws deposits at $T = 1$, the bank offers a gross return r_1, i.e., $r_1 = c_1^*$. If the agent instead keeps her deposits until $T = 2$, the bank offers a gross return r_2 such that $r_2 = c_2^*$, where $1 < r_1 < r_2 < R$, as shown above.

As long as the fraction of agents hit by a liquidity shock is known – which would be the case if the probability of the idiosyncratic shock is known and there are sufficiently many depositors – the financial intermediary solution gives the same outcome as the first-best insurance contract. The bank only holds enough liquid assets on their balance sheet (or alternatively, liquidate a fraction of the long-term assets) to meet the short-term liquidity needs of agents, while keeping the rest invested long-term. Impatient agents withdraw because they really need the funds. Patient agents could withdraw their funds too, but choose not to, as $r_2 > r_1$.

Banks enhance welfare by "creating liquidity" here: they allow a higher return on early withdrawals – make deposits "more liquid" – by allowing a larger fraction of the long-term project to be liquidated, compared to the autarky solution. Note that we still have that $c_1^* < c_2^*$ as long as $R > 1$; i.e., full insurance is not optimal, since this would involve too much inefficient liquidation of the project.

Bank runs

The Diamond-Dybvig model emphasizes how the banking contract solves the problem of making random, short-term withdrawal needs consistent with investing longer-term at a high rate of return and thereby obtaining high short-run liquidity when agents need to liquidate early. However, the model also shows how the banking contract crucially has a vulnerability feature, allowing panics to occur and, as a result, bank failures. This, Diamond and Dybvig also argued, provides a rationale for financial regulation, such as deposit insurance.

The key to the vulnerability is the assumption that the agent's type – whether they are patient or impatient – is private information. Under this assumption, Diamond and Dybvig showed, the model with a banking contract exhibits multiple Nash equilibria in pure strategies. In the good equilibrium, only agents experiencing short-term liquidity shocks withdraw their funds at $T = 1$, thus revealing their type truthfully, and the intended (optimal) allocation is implemented. In the bad equilibrium, however, a bank run occurs at $T = 1$.

In the bad equilibrium, some agents misrepresent their types, which is possible since the bank must grant any withdrawal requests: they cannot tell the patient and impatient agents apart. In this scenario,

patient agents also withdraw their funds. They do so under the belief that sufficiently many misrepresent their patience that the bank will be fully liquidated. Indeed, the more agents that are patient start withdrawing their deposits at $T = 1$, the less long-term investment will be left, and since the bank does not have enough resources at that time to pay everyone, nothing is eventually left for agents who wait until $T = 2$: the belief is self-fulfilling, so it is in everyone's interest to withdraw early.

The sequential service constraint

The "first-come, first-serve" nature of demand deposits, often denoted as the sequential service constraint, is an important element of the model. Sequential service means that the bank honors every depositor who demands repayment in full as long as there are funds left in the bank. Diamond and Dybvig argued that the sequential service constraint is a natural way to capture the continuous liquidity services that banks offer, where depositors can deposit and withdraw their savings at different random times, in a discrete-time model. A key insight from their model is thus that "it is precisely the 'transformation' of illiquid claims into liquid claims that is responsible both for the liquidity service provided by banks and for their susceptibility of runs" (Diamond and Dybvig, 1983, p. 409).

Solutions to the bank run problem

Diamond and Dybvig showed that one solution to the bank run problem is suspension of convertibility, where banks freeze deposits after a certain fraction of deposits have been withdrawn. Indeed, before the introduction of deposit insurance in the U.S., this was a common response of banks experiencing runs; it was typically implemented not at the level of an individual bank but for a collection of banks, aimed at preventing broader panics. Another possibility is federal deposit insurance, i.e., a government regulation that guarantees that deposits are paid back using government funds. Indeed, at the height of the Great Depression in 1933, as the suspension mechanism appeared not to work, the Federal Deposit Insurance Corporation was founded.

In Diamond and Dybvig's model, deposit insurance and convertibility suspension both operate equally well when the bank knows the actual liquidity needs of its customers for certain, in this case represented by knowing the value of λ. If this value is uncertain, however, so that a truthful revelation of types would lead to a random number of early withdrawers, then a full, unconstrained optimum would be more complicated. The reason is that it would necessarily have returns that depend on the realization of λ, which the bank does not know; to be concrete, how could the bank know what to pay the very first customer who shows up to withdraw? In addition, it is not clear how to use suspension of convertibility optimally for the same reasons: at what point should suspension occur?

For this situation, Diamond and Dybvig argued that deposit insurance, where the government uses efficiently levied *ex-post* taxation to fund promises to repay all deposits in full in case the bank lacks funds, is a good alternative mechanism. They argued in particular that the government can be more appropriate for establishing credibility than could a private, say, insurance entity. Lastly, they discuss how deposit insurance can be implemented by another policy, where the central bank acts as a "lender of last resort," thus purchasing bank assets at a price above their liquidation value.

Closely related research

The model of Diamond and Dybvig shares some features with earlier work by Bryant (1980). As in Diamond and Dybvig's model, Bryant modeled the role of intermediaries to provide liquidity for savers with unexpected consumption needs and explored the role of bank runs and deposit insurance. In contrast to Diamond and Dybvig, Bryant considered a pure endowment economy, building on Samuelson's (1958) pure consumption model, where the role of the bank is to provide intermediate loans between consumers who receive their endowment early versus late.

Since there is no real investment in Bryant's model, it does not consider the maturity transformation role of banks. Also, bank runs emerge from some patient consumers receiving negative private information about future bank fundamentals, and there is no coordination problem between creditors. As a result, deposit insurance does not prevent bank runs, and its desirability depends on the exogenous cost of the government for providing such insurance relative to the insurance value to consumers.

The core Diamond and Dybvig (1983) model is quite stylized. Its purpose is to demonstrate (i) how maturity transformation, a fundamental purpose of financial intermediation, fulfills an important and nontrivial insurance role for consumers; and yet (ii) that it is precisely the maturity transformation, which makes banks more illiquid as a result of creating liquidity for depositors, that makes banks fragile and subject to runs. It is possible to generalize many of the assumptions, such as allowing a richer preference structure.

Researchers have also proposed models with multiple technologies. Assume, for example, that there is a short-run, low-return technology alongside a long-run, high-return technology.[16] Then optimal banking arrangements can feature more intense use of the long-run technology than in the autarky solution, a result which has a different flavor than the result in Diamond and Dybvig's basic model. There, less resources are actually held to period-2 maturity in the optimal banking contract than under

[16] See e.g. Bencivinga and Smith (1991), Wallace (1996), and Allen and Gale (1998, 2000) for examples of models along these lines.

autarky, precisely in order to increase the return on the short-term withdrawals and hence improve insurance.

Yet another natural extension is to allow there to be several technologies that differ in riskiness, thereby allowing moral hazard to be studied; earlier work (e.g., Kareken and Wallace, 1978) suggests that deposit insurance may lead to excessive risk-taking by banks. While Diamond and Dybvig (1983) endorsed deposit insurance as the most effective way to control the risk of runs, they also acknowledged that it can have harmful *ex ante* effects.

Later, Diamond and Dybvig (1986) discussed such issues and argue that banks therefore should not be allowed to use uninsured deposits to enter lines of business that are not related to liquidity provision, such as speculative trading. They also proposed that deposit insurance premiums should be tied to the actual risks banks are taking and recommend higher surcharges if banks experience large loan losses or raise their deposit rates significantly.[17] A subsequent literature has developed extensions of the Diamond-Dybvig model to analyze how deposit insurance can be combined with other types of regulation, such as risk-based insurance premiums, capital requirements, and deposit rate ceilings, in order to both avoid bank runs and ex ante moral hazard.[18]

Another important follow-up paper is from Jacklin (1987), who considers the introduction of financial markets in the Diamond and Dybvig model. In their original model, all the risk-sharing between impatient and patient agents are carried out through the financial intermediary. If there was also a market where agents could trade securities, there is an additional opportunity for risk-sharing: instead of an agent withdrawing r_1 when hit by a liquidity shock in period 1, she could instead go to the financial market and sell a claim on her future payoff r_2 to a patient agent. Jacklin showed that the existence of such a market will undo the insurance mechanism that the intermediary provides, and this undermines the ability of the bank to create liquidity.

To understand why, assume that the return that would be realized in the market between periods 1 and 2 r_m. If $r_m > r_2/r_1$, then all patient agents would find it optimal to pretend to be impatient, withdraw r_1 in period 1 and buy claims in the market, which would return $r_1 r_m > r_1(r_2/r_1) = r_2$, the return on long-term

[17] Dybvig (1993) elaborates further on these points.

[18] Hazlett (1995) and Cooper and Ross (2002) include risky investment choice into the Diamond-Dybvig model to incorporate the moral hazard induced by deposit insurance. The latter paper shows how the first-best allocation can be restored by combining deposit insurance with additional bank capital requirements. A number of papers also show that banking competition can increase moral hazard problems, including Chan et al. (2000), who argue that risk-based deposit insurance fees alone cannot solve the moral hazard problem in a competitive banking industry, and Hellmann et al. (2000), who show how a combination of capital requirements and deposit rate controls together can mitigate moral hazard in competitive environments.

deposits. If $r_m \leq r_2/r_1$, then no one would choose to deposit their savings in the bank. Instead, all agents would directly invest in the long-term asset yielding R in period 2.

If an agent becomes patient, she is better off this way. If she becomes impatient, she would then borrow in the market at rate r_m against their R accruing in period 2. For every unit of investment, the amount they could borrow (and consume) would be $R/r_m \geq R/(r_2/r_1) > r_2/(r_2/r_1) = r_1$, which uses the fact that $1<r_1<r_2<R$. Thus, also if she becomes impatient she would be better off under autarky than with saving through the bank.

In fact, Jacklin showed that no risk-sharing equilibrium exists when agents have access to such a financial market, leading to a reduction in welfare. The findings by Jacklin (1987) led to a subsequent literature on the interplay between banks and markets, which we discuss in more detail in Section 4.1.

The sequential service constraint, which is key for generating the bank run equilibrium, is exogenously assumed by Diamond and Dybvig as a reduced form of capturing liquidity demand. Wallace (1988) provided a model that endogenizes this intuition.

Wallace argued that the appropriate interpretation of the sequential service constraint is that depositors are spatially separated at the time of withdrawal and, furthermore, that the deposit insurance scheme proposed by Diamond and Dybvig, for the case where λ is random, is not consistent with spatial separation. A subsequent literature endogenized the sequential service constraint as a way of disciplining banks. We discuss this literature, particularly the contributions of Calomiris and Kahn (1991) and Diamond and Rajan (2001), below in Section 3.2.

Diamond and Dybvig's discussion of the role of banks in maturity transformation, along with the need for policies such as suspension of convertibility or deposit insurance, has been highly influential. For example, it has made clear how regulation that prevents runs by preventing maturity transformation – such as a 100% reserve requirement – can generate welfare losses. Such a policy, which in the Diamond and Dybvig model can be interpreted as forcing r_1 to equal one, would not allow the bank to fulfill its fundamental insurance role. As financial markets evolve and new institutions appear, regulatory policy needs to analyze how the different roles of these markets – including maturity transformation – keep shifting among market players. As we will describe in Section 5, bank-like institutions not covered by deposit insurance played a critical role during the Great Financial Crisis.

3.3 Delegated monitoring

One of the questions arising from the Diamond-Dybvig model is why "long-term projects," i.e., long-term loans to firms and households, are illiquid. In their model, the illiquidity comes from the particular

production technology they assume. But there are more fundamental reasons, relating to the lending process, for why the bank's assets are worth more when held to maturity compared to if they have to be liquidated or sold prematurely.

Diamond (1984) provided one important reason for this: banks collect valuable information from their borrowers, which makes their loans more valuable within the banking relationship compared to if they are sold to outsiders.[19] His paper can also be seen as formalization of the concept *Cost of Credit Intermediation* (*CCI*), which Bernanke (1983) emphasized is important for understanding the role of banks in the Great Depression.

Delegated monitoring, debt, and diversification

Diamond (1984) argued that delegated monitoring is a central economic role of financial intermediaries: investors delegate their investment decisions to financial institutions, who invest in multiple investment projects/borrowers on their behalf. In their signaling model of financing, Leland and Pyle (1977) also argued informally that financial intermediaries have economies of scale in information production, and that intermediated finance can thus reduce the cost of asymmetric information. Diamond (1984) provided the first formal model of this mechanism and has become a seminal paper in the financial intermediation literature.[20]

A key insight is that diversification across many loans makes it possible for banks to finance risky projects through close-to-riskless debt, thus providing another mechanism for banks to funnel less risky and more liquid savings to riskier and less liquid productive investment. The ability to create liquidity in this way is unique to the lending bank, due to the monitoring of borrowers it performs on behalf of savers. This makes banking relationships valuable, and provides an explanation for why bank loans are illiquid if banks are forced to sell them to outsiders who do not have this information.

Liquidity according to Diamond (1984) is not, however, to be interpreted with the same formalism as defined by Diamond and Dybvig (1983), and consequently there is no scope for discussing bank runs. Rather, the focus is on showing that banks optimally perform monitoring of risky projects and are able to promise depositors a riskless, high return.

[19]Although published in 1984, the paper was originally the third chapter of Diamond's 1980 Yale University dissertation "*Essays in Information and Financial Intermediation*." As such, it is considered the first truly micro-founded theory of financial intermediation.

[20] An analogy would be that the bank is lending to a farmer who buys seeds to put in the ground, and once planted it takes time before the plants are ready to harvest. Before this time, not much value can be recovered from these seeds.

Model setup

As in Diamond and Dybvig (1983), Diamond (1984) provided a model of financial intermediation from first principles. He did so by formulating a general contracting problem, and then showing that a properly designed financial intermediary is the optimal solution to this problem.

The model considers the problem of an entrepreneur, who needs to finance a productive investment project, and a problem of savers, who have funds to invest. There are two periods, now ($T = 0$) and the future ($T = 1$). The project requires an investment now, normalized to 1, and pays a risky payoff y in the future. Agents are assumed to be risk-neutral and the economy's discount rate is $R > 1$. The project is assumed to be productive, so that $E(y) > R + K$ (we will define K in a moment).

Diamond introduced a financial friction: only the entrepreneur observes the payoff y of the project. This implies that the entrepreneur can always claim that the project failed and keep the payoff y to herself. If such misrepresentation cannot be prevented somehow, savers will not expect to receive their required return R and will not provide financing, despite the project being *ex-ante* productive.

Debt as an optimal financial contract

To solve this problem, Diamond assumed that savers can write a financial contract that imposes a punishment on the entrepreneur if she does not repay R to savers.[21] In this context the optimal financial contract will be debt: the entrepreneur promises to pay back a fixed face value D to investors, and if the payment is less than D, a bankruptcy penalty is imposed on the entrepreneur. This incentivizes the entrepreneur to repay if she can, to avoid the bankruptcy penalty, and D is set to the lowest value needed for savers to break even (i.e., so that the expected repayment is R). While this is an optimal contract, since it minimizes the states of the world where bankruptcy is incurred, it still imposes an inefficient bankruptcy cost on the entrepreneur in the states when the project's payoffs turn out to be low, i.e., when *ex post* $y < D$.

[21] The first research to highlight the role of debt as an incentive device was Jensen and Meckling (1976). Diamond (1984) is one of the first models to derive debt as the optimal solution to a formal contracting problem, together with Townsend (1979), Grossman and Hart (1982), and Gale and Hellwig (1985). To make his model tractable, Diamond assumes that the bankruptcy penalty is non-pecuniary, so that the entrepreneur can be punished despite having limited liability. Such non-pecuniary bankruptcy cost can be thought of as a loss of reputation or legal repercussions that provide disutility to the entrepreneur. The financial contracting literature have used other ways to model bankruptcy cost, e.g., that investors can verify payoffs ex post at a cost (Townsend, 1979), or that savers can (inefficiently) liquidate the entrepreneur's assets or force the entrepreneur to post collateral which would be seized in bankruptcy (Hart and Moore, 1989, 1995). The optimal financial contract turns out to be a debt contract in all of these situations, and the intuition is the same in all of them: bankruptcy costs are necessary for disciplining the entrepreneur to repay the loan, and debt contracts minimize the number of states when costly bankruptcy will occur.

Inefficient duplication of monitoring

Diamond then introduced monitoring, a second tool through which lenders can make sure they are repaid: by paying a cost K in $T = 0$, the lender gets to observe realized cash flows y in $T = 1$. This assumption captures the notion that a lender can monitor the borrower by making a thorough credit evaluation, scrutinizing the firm's financials, and by having regular meetings with management to ensure that the entrepreneur does not misbehave. While monitoring is not free and requires time and resources, the benefit is that the risk of costly bankruptcy down the road is reduced.

To capture this notion, Diamond assumed that monitoring costs K are lower than the expected bankruptcy costs that the entrepreneur would face under the optimal debt contract. Monitoring becomes less efficient the more lenders the entrepreneur has, however, since each of them has to pay a monitoring cost of K.

Realistically, most investment projects (think about a factory or a house) are large relative to the amount of savings of any given individual, so that several savers will have to team up to finance one investment project. But this will quickly make the monitoring by individuals infeasible for large projects, since the monitoring cost is multiplied whenever a new saver contributes to the project. With m savers, if every saver has $1/m$ to invest, total monitoring costs will be mK, and when $mK > E(y)$, these costs exceed the expected value of the project. In this case, the only financing alternative is lending by individuals without monitoring, which leads to costly bankruptcy in some states of the world.

Intermediaries as delegated monitors

To avoid the duplication of costs, the m savers could lend their funds to a financial intermediary, who in turn could monitor the entrepreneur on their behalf. But this, in principle, just moves the financing problem up one level: the individual savers still do not observe the project payoff y, so now we need to ensure that the intermediary truthfully reports the project payoffs and repays the savers.

We are back to the same problem as before: it would be too expensive for every saver to monitor the intermediary, and the only feasible contract between savers and the intermediary would be a debt contract, which means that the intermediary will suffer the costs of bankruptcy whenever the risky project payoffs turn out to be low. Thus, when the intermediary only lends to one project, nothing has been gained from delegated monitoring compared to the savers lending directly to the entrepreneur.

In reality, financial intermediaries make loans to multiple firms (projects). A large bank would use deposits from thousands of savers and lends them out to thousands of entrepreneurs. The key insight from Diamond's model is that diversifying the asset side of the balance sheet (by making many loans)

as well as the liability side (pooling savings from many savers) is necessary for financial intermediation to work.

Assume that the intermediary takes deposits from Nm savers to lend to N different projects, which each has a risky payoff y. If the projects are (sufficiently) uncorrelated, then as N becomes sufficiently large, the average payoff on each project will converge to a deterministic payoff $E(y) > R + K$. Thanks to diversification, there is no longer any risk that the intermediary goes bankrupt simply due to bad luck. As long as the intermediary performs its monitoring properly, it will always be able to cover its monitoring cost and give savers their required return, so deposits become risk-free. And the intermediary has the incentive to honor its obligations in order to avoid costly bankruptcy. In this way, the intermediary is able to create safe deposits from a diversified portfolio of risky loans thanks to delegated monitoring.

Closely related research

While the argument so far has relied on many simplifying assumptions, such as project payoffs being uncorrelated, the basic insight is remarkably robust. Relaxing some of these assumptions also yields new insights on financial intermediation. For example, Diamond (1984) showed that if project payoffs are correlated, a financial intermediary should hedge away systematic risks in the market to the extent possible and focus on taking on idiosyncratic credit risks that will be diversified away in a large loan portfolio. To the extent some systematic risks cannot be hedged perfectly, there will be a role for keeping an equity capital buffer in the intermediary in addition to deposits; this way, the risk of bankruptcy in an economic downturn goes down. Too much equity in the bank is not optimal either, however, since the risk of bankruptcy is what incentivizes the bank to properly perform its delegated monitoring function in the first place.

Another implication is that the loans the intermediary makes are illiquid, in the sense that the bank cannot sell them in the market at fair value. Since it is only the bank that has performed the monitoring and can observe *y*, any buyers of these loans in the market would themselves have to pay the cost of monitoring and must be compensated through a lower purchasing price. While these issues were discussed by Diamond (1984), they were developed in detail in subsequent work by Diamond and Rajan (2001), which in turn builds on a contribution by Calomiris and Kahn (1991).

Calomiris and Kahn (1991) pointed out that the role of debt as a disciplining device proposed by Diamond (1984) is incomplete. First, delegated monitoring is possible thanks to the ability of depositors to impose a non-pecuniary penalty on the banker if the bank cannot fulfil its obligations. This raises the question of how the imposition of such a penalty is actually enforced, given that any individual depositor is small and uninformed. Second, Diamond (1984) derived debt as the optimal depositor

contract, but was silent on the maturity of this debt and whether it should be in the form demandable deposits that satisfy a sequential service constraint.

Calomiris and Kahn showed that demandable deposits is an efficient way to discipline a bank when monitoring is costly. The right to take her money out of the bank if she becomes suspicious that realized returns are low makes it in the depositor's interest to keep an eye on the bank. If enough depositors agree with this negative assessment of the bank's future, a bank run will be set in motion that will eventually make the bank fail. Since bank failures are costly to the banker (not necessarily due to non-pecuniary costs but also because of lost future income) the threat of a run will serve as a commitment device.

Diamond and Rajan (2001) combined the models of Diamond (1984) and Calomiris and Kahn (1991) to provide a new theory of banking. Specifically, a bank serves as a delegated agent, who is lending to a borrower on behalf of some less well-informed lenders. The intermediary stands out in its ability to monitor borrowers and collect funds from the loans in the event that a borrower defaults. Thus, the bank has an advantage in both screening borrowers *ex ante* and realizing loan repayments *ex post*.

It is efficient for society to have the repayment specialists involved in the collection of bad credit because these specialists can offer the most attractive rates to depositors; any other entity that made an identical set of loans would collect less than the banks on the loans that default. But this again raises the problem of monitoring the monitor: how does the bank commit itself to working as hard as possible to maximize recovery from borrowers in the event of trouble? If depositors would threaten to liquidate the bank, the most they could count on to receive from the bank would be a lower value, which implies that the bank could hold up the depositors to extract the full value of their monitoring efforts. In other words, the monitoring advantage of banks make the loans illiquid, in that they would have a much lower value if transferred to an outside party in the private market.

Diamond and Rajan solved this conundrum by noting the disciplinary role that demand deposits can play, following Calomiris and Kahn (1990). If the bank offers its customers debt that will be serviced on a first-come, first-serve basis, the bank is setting itself up for failure if the customers believe that the bank is shirking or holding out on them in any other way. Thus, in addition to showing how banks can create liquid savings from illiquid assets as in Diamond-Dybvig, they explained why savings should be in the form of demand deposits to ensure that the bank has the incentives to create liquidity. When the bank has the right quantity of deposits outstanding, any attempt by the banker to extort a rent from depositors by threatening to withdraw will be met by a run, which disintermediates the banker and drives his rents to zero. Notice that in this application of the logic of Calomiris and Kahn is directly tied to the specialized skills and role of the banker, so the same story would not apply to non-financial firms.

Diamond and Rajan developed several extensions of this basic model to analyze the impact of different regulations and government policies on financial stability, which we review in Section 4.3.

3.4 Financial intermediation during the Great Depression

In 1983, Bernanke published a study on the role of bank failures during the U.S. Great Depression of the early 1930s. He provides historical documentary evidence supplemented by data to put forth the argument that financial intermediaries perform a valuable service by channeling savings to productive investments. Moreover, he argues, by disrupting these services, the banking panics in the early 1930s was precisely what generated such a long and deep recession at the time.

From the perspective of the contributions by Diamond and Dybvig, Bernanke's work can be seen as providing evidence supporting their models. Specifically, he provides evidence that bank runs can lead to financial crises (as in Diamond and Dybvig, 1983), which in turn leads to prolonged periods of disruption of credit intermediation, consistent with bank failures destroying the valuable screening and monitoring services banks perform (as in Diamond, 1984).

The insights from Bernanke (1983) were, however, also very important in their own right and led to a deeper understanding of the importance of leverage among banks, firms, and households in exacerbating economic shocks. Moreover, they have generated significant and separate theoretical and quantitative literatures not directly relevant to banking, but relevant to the role of credit in the macroeconomy. We briefly discuss these literatures in later sections.

The dominant explanation at the time for why the Great Depression was so deep and prolonged was due to Friedman and Schwartz (1963). They argued that the waves of banking crises in 1930–1933 substantially reduced the money supply and the money multiplier. The failure of the Fed to offset this decline in money supply in turn led to deflation and a contraction in economic activity.

Bernanke (1983) proposed a new (and in his view complementary) explanation of why the financial crisis affected output. According to this view, the services that the financial intermediation sector provides, including "nontrivial market-making and information gathering," are crucial for connecting lenders to borrowers. The bank failures of 1930–1933 hampered the financial sector's ability to perform these services, resulting in an increase in the real costs of intermediation. Consequently, borrowers – particularly households, farmers, and small businesses – found credit to be expensive or unavailable, which had a prolonged negative effect on aggregate demand. Bernanke combines examination of historical sources, statistical analysis, and (at the time) recent theoretical insights to build this argument.

To be clear, Bernanke's analysis does not engage in the discussion of what caused the initial economic downturn in the late 1920s that subsequently escalated into the Great Depression, and this was not the focus of Friedman and Schwartz either. Similarly, when we discuss the Great Recession below, the core issue is not about its origins but on the mechanisms by which the recession played out.

The impact of financial panics on output

While a recession had begun in the U.S. in 1929, the downturn was comparable in magnitude to earlier recessions that had relatively quick recoveries (such as the recession of 1920–1922). As shown by Friedman and Schwartz, there was a "change of character of the contraction" (Friedman and Schwartz, 1963, p. 311) as the banking crisis began in 1930. By the end of 1933, because of failures and mergers, the number of banks operating had been reduced by almost one half compared to the number that existed in 1929 (although the fall was somewhat smaller in terms of assets, since mostly small banks failed).

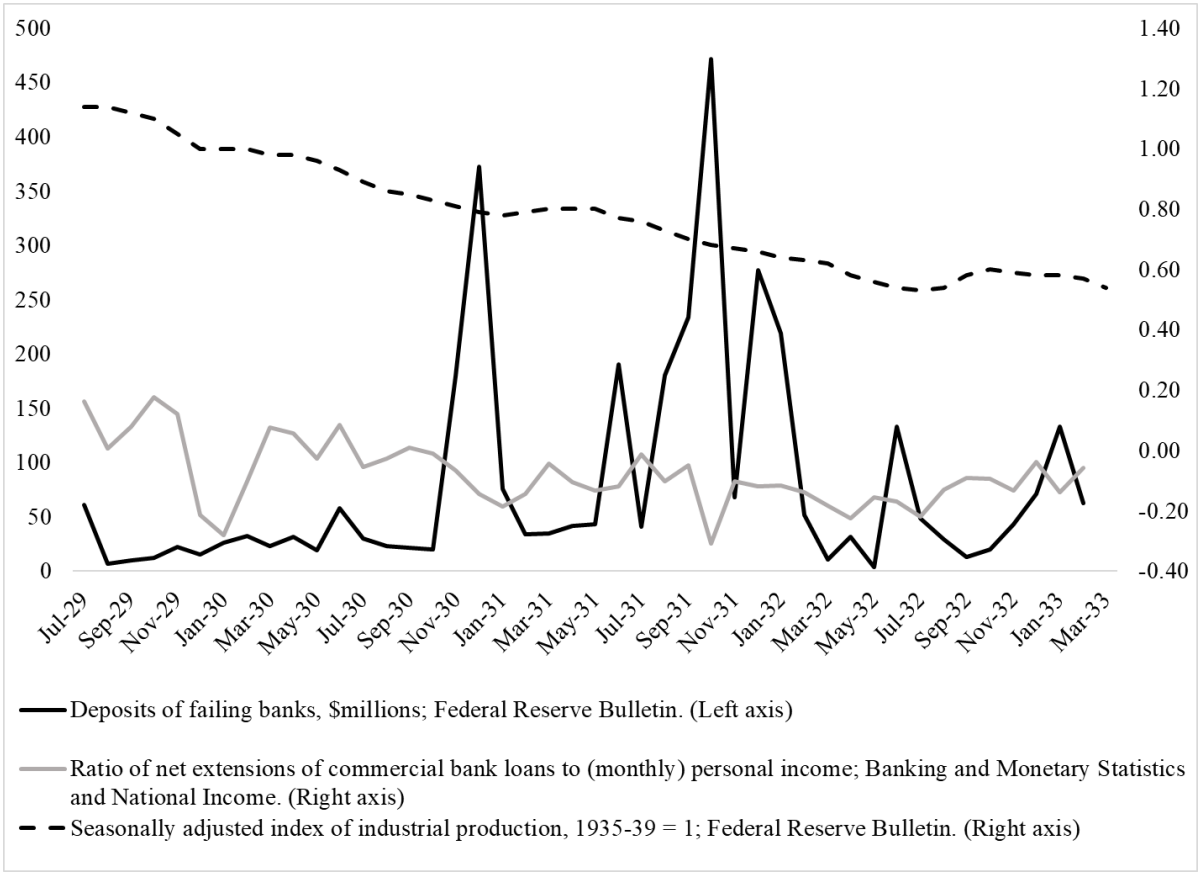

Figure 4: Selected macroeconomic data, July 1928 to March 1933. (Figure based on Table 1, Bernanke, 1983, p. 262.)

Figure 4, based on Bernanke (1983, Table 1), presents data showing that bank failures were accompanied by a large contraction of credit and by a dramatic reduction in industrial output, which

dropped by 46% between January 1930 and March 1933. The financial panic ended in early March 1933, when the government (led by the newly elected president Franklin D. Roosevelt) forced all banks to close their businesses for one week to stop the panic (the so-called National Banking Holiday). Later that year, Congress passed the National Banking Act of 1933, which among other measures introduced federal deposit insurance for U.S. banks.

Bernanke (1983) first discussed the importance of self-fulfilling panics as an explanation for the surge in bank failures, citing the 1981 working paper version of Diamond and Dybvig (1983). Before the introduction of federal deposit insurance, bank runs had in fact been relatively frequent. Large-scale panics had usually been contained, however, through suspension of convertibility, typically initiated and coordinated by loose organizations of urban banks called clearinghouses. According to Friedman and Schwartz (and reiterated by Bernanke), the creation of the Fed in 1913 upset this institutional arrangement because the Fed was believed to have taken over the responsibility for fighting bank runs.

The Cost of Credit Intermediation (CCI)

Next, Bernanke argued that the reason the financial panic accelerated and prolonged the recession was not simply because it led to a drop in the money supply, which Friedman and Schwartz had argued, but also – and maybe more importantly – because it reduced the ability of financial intermediaries to supply credit to firms, farms, and households.

At the time Bernanke wrote his paper, the finance literature had long been dominated by the efficient-markets paradigm, which assumes that markets are complete and frictionless, which implies that both financial structure (Miller and Modigliani, 1958) and financial intermediation (Fama, 1980) are irrelevant. This paradigm, for example, holds that whether a firm is financed by new debt, equity, or retained earnings, is immaterial for the working of the firm: there can be risks in a firm's productive endeavors but risks are not associated with the inability to borrow, since the availability of financing is determined entirely by non-financial factors.

In the late 1970s, however, this paradigm was challenged. In particular, several research papers showed how various frictions, such as imperfect information and agency problems, lead to market arrangements where both the financial structure and the availability of financing were central.[22] Building on this emerging literature, Bernanke defined the *CCI* as the "cost of channeling funds from the ultimate savers/lenders into the hands of good borrowers.

[22] Important early papers on financial frictions and its implications include Jensen and Meckling (1976); Jaffee and Thomas (1976), Myers (1977); Townsend (1979); Leland and Pyle (1979); and Stiglitz and Weiss (1981).

The *CCI* includes screening, monitoring, and accounting costs, as well as the expected losses inflicted by bad borrowers. Banks presumably choose operating procedures that minimize the *CCI*. This is done by developing expertise at evaluating potential borrowers; establishing long-term relationships with customers; and offering loan conditions that encourage potential borrowers to self-select in a favorable way" (Bernanke, 1983, p. 263). He then argued:

> "Fear of runs led to large withdrawals of deposits, precautionary increases in reserve-deposit ratios, and an increased desire by banks for very liquid or re-discountable assets. These factors, plus the actual failures, forced a contraction of the banking system's role in the intermediation of credit. [T]he rapid switch away from the banks (given the banks' accumulated expertise, information, and customer relationships) no doubt impaired financial efficiency and raised the *CCI*." (Bernanke, 1983, p. 264)

In addition to emphasizing direct effects on credit supply when financial intermediaries cannot perform their screening and monitoring services (as in Diamond, 1984), Bernanke also pointed to an indirect financial channel working through demand, building on work by Fisher (1933). An increase in *CCI* increases the effective cost of credit and can even make credit unavailable for some potential borrowers. While the increase in *CCI* might be less important for large, cash-rich firms, it becomes binding for bank-dependent borrowers, including farms, small firms, and households, which cut back on consumption and investment. The resulting decrease in demand dampens economic activity and leads to lower prices and deflationary pressures.

Fisher (1933) argued that during the Great Depression, deflation was particularly damaging to already leveraged firms and households, since their outstanding debt was nominal, and they thus became even more leveraged in real terms. In turn, this caused an increase in insolvency and financial distress, which led lenders to liquidate borrower assets, further depressing prices and exacerbating the feedback loop.[23]

While Fisher's arguments had some influence among policymakers at the time, they were less influential among academics, who argued that debt deflation simply led to redistribution from

[23] Fisher (1933, p. 342) summarized his argument as follows: "Then we may deduce the following chain of consequences in nine links: (1) Debt liquidation leads to distress selling and to (2) Contraction of deposit currency, as bank loans are paid off, and to a slowing down of velocity of circulation. This contraction of deposits and of their velocity, precipitated by distress selling, causes (3) A fall in the level of prices, in other words, a swelling of the dollar. Assuming, as above stated, that this fall of prices is not interfered with by reflation or otherwise, there must be (4) A still greater fall in the net worths of business, precipitating bankruptcies and (5) A like fall in profits, which in a 'capitalistic,' that is, a private-profit society, leads the concerns which are running at a loss to make (6) A reduction in output, in trade and in employment of labor. These losses, bankruptcies, and unemployment, lead to (7) Pessimism and loss of confidence, which in turn lead to (8) Hoarding and slowing down still more the velocity of circulation. The above eight changes cause (9) Complicated disturbances in the rates of interest, in particular, a fall in the nominal, or money, rates and a rise in the real, or commodity, rates of interest." The fire-sale mechanism of financial propagation was later formalized in a highly influential paper by Kiyotaki and Moore (1997), discussed in section 4.3.

borrowers to lenders, which should have no major macroeconomic effects (Bernanke, 1995).[24] However, Bernanke (1983) pointed to the importance of heterogeneity in leverage (and in the associated difficulties in financing) across the population of firms, something that the redistribution argument ignores.

The seriousness of the problem in the Great Depression was not only deflation, which had occurred in previous recessions without resulting in mass insolvency, but the fact that debt deflation disproportionately hit small and bank-dependent borrowers, such as households, farms, and small businesses, which had increased their leverage significantly in the years before of the Great Depression. As a result, debt deflation had a particularly large effect in terms of reducing consumer demand, distorting capital allocation across firms, and causing further losses to the financial intermediaries lending to households and small businesses.

Evidence on the CCI and on aggregate output

Bernanke (1983) presented several pieces of evidence to corroborate the *CCI* channel. He shows that bank failures were followed not only by a decrease in bank credit but also by a widening of credit spreads. He also showed that the credit contraction was particularly harmful for borrowers that were more bank-dependent, such as small firms, farmers, and households, while large firms with access to public equity and bond markets seem to have been much less credit constrained.

The main statistical analyses are time-series regressions of industrial output on unanticipated changes in money and credit supply, following the methodology of Barro (1978). Using monthly data during 1921–1941, Bernanke found that monetary shocks alone, while significant, can only explain a modest part of output fluctuations, while the inclusion of his *CCI* proxies (deposits of failed banks and liabilities of failing businesses) yields considerable additional explanatory power and can explain most of the output drop during the Great Depression.

Based on this evidence, Bernanke (1983) argued that the bank failures during the Great Depression caused the reduction in credit, which in turn reduced economic activity. The counterargument would be that output dropped due to a non-financial shock (e.g., a decrease in consumer sentiment or productivity), which led to reduced demand for credit from firms and consumers. In his regression analysis relating output to bank failures, he included several lags of output to control for this alternative channel.

[24] Notable exceptions arguing for the importance of the debt-deflation channel were Minsky (1964), Kindleberger (1973), and Mishkin (1978).

In addition, Bernanke provided several other pieces of evidence to dismiss this alternative explanation. He reported narratives from contemporary sources that banks started cutting their credit supply dramatically as banks started failing.[25] He also built on Friedman and Schwartz (1963), who identified specific events that were important sources of the 1930–1933 bank runs, and argued that these events were unlikely to be connected to U.S. industrial output.[26]

The persistence of the Great Depression

After having provided evidence that the *CCI* channel can explain the depth of the Great Depression, Bernanke (1983) argued that it can also better explain its persistence compared to alternative theories.[27] Existing theory had a difficult time explaining why pure monetary effects, relying on gradual diffusion of information or sticky prices and wages, would be long-lasting. Bernanke argued that it is easier to reconcile theoretically why the credit channel can have persistent effects, as its duration should depend on the amount of time it takes to revive broken channels of credit flow (e.g., recapitalize banks and/or create new banking relationships) and restructure insolvent debtors, both of which would plausibly be difficult and slow processes.

Bernanke corroborated this argument with historical data and narratives. While the financial crisis culminated with the government-imposed "bank holiday" in March 1933, Bernanke showed that the contraction of credit supply lasted several more years, particularly for small firms and households (mortgages). He also suggested that the recovery would have been even slower if it had not been for government intervention and assistance, such as the establishment in 1934 of the Federal Deposit Insurance Corporation, the Federal Savings and Loan Insurance Corporation, and the Home Owners' Loan Corporation.

Closely related research

Bernanke's work was novel in convincingly showing, with a battery of qualitative and quantitative evidence, that financial intermediaries play a crucial role in the real economy, especially during financial crises. The particular focus on the Great Depression was helpful as a rich case study: he could

[25] An example cited by Bernanke is a 1932 industry survey. It reported: "During 1930, the shrinkage of commercial loans no more than reflected business recession. During 1931 and the first half of 1932 (the period studied), it unquestionably represented pressure by banks on customers for repayment of loans and refusal by banks to grant new loans" (Bernanke, 1983, p. 265).

[26] "These include the revelation of scandal at the Bank of the United States (a private bank, which in December 1930 became the largest bank to fail up to that time); the collapse of the Kreditanstalt in Austria and the ensuing financial panics in central Europe; Britain's going off gold; the exposure of huge pyramiding schemes in the United States and Europe; and others" (Bernanke, 1983, p. 272).

[27] In subsequent work, Bernanke (1995) argues that cross-country evidence provides support for additional factors also contributing to the persistence of the Great Depression, such as lack of exchange rate adjustment due to the gold standard (Eichengreen, 1992; Bernanke and James, 1991) and stickiness in wages (Eichengreen and Sachs, 1985; Bernanke and Carey, 1994).

point to channels through which bank failures – which occurred *en masse* during this period – can have devastating consequences for economic welfare not just through the implied reduction in the money supply, but through a reduction in the ability of households, farms, and firms to obtain credit. Today, researchers refer to "the credit channel" as a catchall for the mechanisms Bernanke pointed to. Although commercial banks were the object of study for Bernanke, the credit channel can of course operate through other financial institutions, both intermediaries and other institutions. Thus, many researchers have built theoretical models to enrich the notion of the credit channel; we briefly touch on some of this work below.

Yet others have looked at more evidence. Some of this work, in line with Bernanke (1983), has been to look at pre–WWII data – both further studies of the Great Depression and data from the 19th century. Some researchers, for example, have argued that the simple time-series regressions in Bernanke (1983) did not allow a proper control for unobserved shocks and that the results were not robust to the inclusion of additional time-series variables (Temin, 1989; Rockoff, 1993; Cole and Ohanian, 2000). However, Calomiris and Mason (2003) overcame much of this criticism by using panel data of bank lending and economic activity at the state- and county-level, with results very much in line with Bernanke (1983).[28]

An even larger number of studies focused on more recent recessions. Our coverage below of the Great Recession should therefore be viewed both as another case study of a significant financial crisis where bank-like institutions played a key role and as a way of introducing some of the more recent empirical work on the credit channel.

While Bernanke emphasized the difficulty of reviving broken lending relationships as an explanation for why the recession was so prolonged, his evidence on this mechanism is indirect and suggestive. Subsequent research, particularly work studying the Global Financial Crisis, has been able to provide direct evidence on this mechanism by utilizing microdata on bank-borrower relationships. We describe some of this work in more detail in Section 5. Moreover, Cohen et al. (2021) showed how relationship lending can be identified even absent such microdata by looking at the responsiveness of loan rates to bank funding costs. Using this methodology, they provided evidence that broken credit relationships indeed propagated the real effects of bank failures during the Great Depression.

Finally, as we also discuss below, many researchers have studied the role of the credit channel also during "normal times," or as the economy goes through more mild recessions. In all of these follow-up analyses, a central motivation has been to examine the role of government policy. Bernanke's own insights in particular indicate that government-sponsored programs that support a well-functioning

[28] Also, Bernanke and James (1991) analyze cross-country data on the Great Depression and argue that national differences in vulnerability to banking crises had more to do with institutional and policy differences than macroeconomic conditions.

financial sector, both in normal times by providing, for example, deposit insurance and during crises by injecting credit into the banking system, can be more beneficial than the research community had envisioned prior to the work of Bernanke (1983).

4. Extensions and the subsequent literature

In this section, we discuss some important extensions and subsequent research related to the laureates' contributions. This discussion is not intended to be a comprehensive literature review. The related literature is extensive, and many important contributions therefore cannot be covered.

4.1 Maturity transformation

The simple Diamond-Dybvig model has proved remarkably robust to generalizations. When the model is enriched, the main logic prevails. In some cases, however, the generalizations also produce major additional insights.

Bank runs and sunspot equilibria

In Diamond and Dybvig (1983), bank runs are self-fulfilling (or "sunspot") equilibria, unrelated to the fundamentals of the bank and the economy.[29] Some researchers viewed the implication that bank runs were a multiple-equilibrium phenomenon unrelated to fundamentals as a weakness of the Diamond and Dybvig model; empirically, banking panics have indeed been shown to correlate strongly with the business cycle (Gorton, 1988; Calomiris and Gorton, 1991). Consequently, subsequent literature explored the implications of introducing fundamental shocks into the Diamond-Dybvig framework, sometimes reaching different conclusions regarding the existence of multiple equilibria and the design of optimal policy.[30]

Morris and Shin (2000), Goldstein and Pauzner (2004), and Rochet and Vives (2004) show that if there is some uncertainty in the payoff of the long-term assets, and if investors receive imperfectly correlated signals about these payoffs, the Diamond-Dybvig model has a unique equilibrium where bank runs only occur when the expected payoffs are sufficiently low.[31] When expected payoffs are higher than a threshold, the unique equilibrium is the "good" risk-sharing equilibrium.

[29] The term sunspot equilibrium was coined by Cass and Shell (1983), who establish the existence of multiple equilibria in general equilibrium models with incomplete markets.
[30] See, e.g., Jacklin and Bhattacharya (1988), Chari and Jagannathan (1988), Hellwig (1994), and Allen and Gale (1998).
[31] In this modified setup, a unique equilibrium can be determined using the global games concept of Carlsson and Van Damme (1983).

Although these three models connect bank runs to fundamentals, bank run equilibria are still self-fulfilling (the bank would be solvent and depositors satisfied if nobody were running), and sharp crises can result from small changes in fundamentals. As a result, the fundamental policy implications from Diamond-Dybvig are still valid, i.e., that deposit insurance and lenders of last resort play an important role for avoiding costly bank runs.

The spread of banking crises

In the stylized Diamond-Dybvig economy, the optimal mechanism is to have one bank in the economy. Subsequent work has analyzed how a failure in one bank can spread to others and eventually grow into a systemic banking crisis, as happened during the Great Depression.

Allen and Gale (2000) consider a Diamond-Dybvig model with two assets, the illiquid long-term asset and a liquid asset.[32] The liquid asset yields a lower return than the illiquid asset over the long term, but a higher return in the short term. Banks hold just enough of their deposits in the liquid asset to cover expected withdrawals from impatient consumers and invest the remainder productive illiquid asset.[33] In addition, they assume that the economy consists of a number or regions, where the fraction of impatient consumers in each region fluctuates randomly. If the aggregate demand for liquidity across all regions is constant, optimal insurance can still be achieved through an interbank market in deposits. A bank that experiences unexpectedly many impatient consumers borrows liquidity from other banks in regions with unexpectedly few withdrawals, and pays the lending banks back next period using the returns from the long-lived asset.

If aggregate demand for liquidity fluctuates, however, the interbank market can instead lead to a systemic crisis, where a liquidity shock in one region can spread by contagion to other regions. In this case, a bank that experiences unexpectedly many impatient consumers cannot borrow enough from other banks and will have to liquidate some of the long-term assets. This may cause a run by impatient consumers, where all the bank's long-term assets have to be liquidated prematurely. The banks who lent to the crisis region may then expect not to be repaid enough to cover the obligations to their own patient consumers next period, leading to runs in their own regions as well and resulting in a systemic crisis. Allen and Gale (2000) show that the resilience of the system depends on the structure of the

[32] An earlier model of systemic crises in a Diamond-Dybvig economy is Smith (1991). In his model, there are many banks and a "reserve bank" that provides liquidity if a given bank has a shortage. He shows under what circumstances a run on the reserve bank can occur, resulting in a suspension of convertibility, and the effect of different regulatory regimes.

[33] The model is an extension of Allen and Gale (1998), which introduces aggregate shocks into a two-asset Diamond-Dybvig economy. When an aggregate shock hits, the bank's liquid assets do not suffice to cover withdrawals from impatient consumers, and it must liquidate part of its long-term assets. If the shock is large enough, patient consumers will anticipate a low long-run payoff and run on the bank. Allen and Gale show how the right type of central bank intervention can restore avoid runs and restore efficiency.

interbank market: in their setup, the more regions that are connected to each other, the less likely that a financial crisis in one region will accelerate into a systemic crisis. The subsequent literature has shown that this result does not necessarily generalize, and has explored how the likelihood of contagion, as well as optimal policy responses, depend on the structure of interbank networks.[34]

Banks and markets

Another important literature that grew out of Diamond and Dybvig (1983) considered the interaction of banks and markets. Jacklin (1987, Section 3.1) showed that trading in securities generally cannot achieve the same level of risk-sharing as a Diamond-Dybvig intermediary can, and that allowing for trading claims in markets actually undermines the liquidity creation of the bank.

Extending this argument, Diamond (1997) analyzed the interaction between banks and markets where he got around Jacklin's critique by assuming that not all savers have access to markets.[35] He showed that banks and markets together actually create more liquidity than each of these institutions alone. First, bank deposits offer an option to obtain funds on short notice at a lower opportunity cost compared to markets. Second, banks improve the liquidity of markets, because long-term assets can be sold at higher market prices than would prevail without banks.[36]

While these extensions may seem of primarily theoretical interest, they have been instrumental for building a deeper understanding of financial regulation more broadly. Consider the question of whether banks are naturally induced to hold the right amount of liquidity in the good equilibrium. In the canonical Diamond-Dybvig model the answer is yes, but as realism is added, the picture becomes more nuanced and interesting.

Following Jacklin (1987), Farhi, Golosov, and Tsyvinski (2009) assumed that agents may be able to engage in anonymous financial trade with each other and not only with a single intermediary. Under these circumstances, it can be shown that there may be too little liquidity provided in the economy. The reason is that each agent fails to internalize that their own decision to hold illiquid assets creates a scarcity of liquid assets for those who are unfortunate enough to have a short-term liquidity need. This leads the interest rate on liquid borrowing to be inefficiently high in equilibrium, absent regulation. A

[34] See, e.g., Freixas et al (2000), Allen et al. (2012), and Acemoglu et al. (2015).
[35] While this may seem like an arbitrary assumption, the limited participation of individuals in financial markets is a well-documented phenomenon. For example Mankiw and Zeldes (1991) document that only 24% of U.S. households owned equities in 1984. Even though participation rates have increases since then, Guiso et al (2003) shows that even in the US and Sweden, the two countries with the highest level of stockholding, about 50% of households do not invest in shares, and many more in countries like Italy and Germany.
[36] Allen and Gale (2004) examine the importance of different assumptions regarding market completeness and contracting completeness in general equilibrium. The show that bank runs can be optimal in equilibrium when contracts are incomplete and markets are complete.

straightforward way of reducing the cost of liquidity is to force some actors to hold more liquid assets than they desire. For example, banks can be required to hold liquid reserves – a requirement that is commonplace in financial regulation in most countries.

Banks and money

Going back to Hume, economists have argued that money plays a critical role in facilitating transactions. Because bank deposits offer a substitute for money, banks have frequently been discussed in the context of monetary economics. Extensions of the model of Diamond and Dybvig (1983) have explored the role of banks in creating money-like assets (or "inside money," as opposed to "outside money" issued by the central bank). While this literature is still developing, several insights have already emerged.

In an important contribution, Gorton and Pennacchi (1990) showed that liquidity needs of agents are best met using "informationally insensitive" securities. As per Diamond and Dybvig (1983), agents face the risk of liquidity shocks, but in addition they assume that there is asymmetric information about asset returns across different agents.[37] Similar to Jacklin (1987), Gorton and Pennacchi also allowed for agents trading in markets. They then showed how banks (and similar financial intermediaries) can play a unique role in creating informationally insensitive demand deposits. These demand deposits provide liquidity as long as they are risk-free, since no party is concerned about the value of the claim, and essentially makes them qualify as money. Deposits are informationally insensitive in the sense that there is no adverse selection when they are traded, and no agent has incentives to produce private information about the value of deposits as long as they are risk-free.

Building on the work by Gorton and Penacchi (1990) and the discussion by Holmström (2014), Dang et al. (2017) explored the role of banks versus markets in creating liquidity, and which types of investments will be funded from each of these sources. They showed that a key condition for banks to be able to create liquid money-like claims is that investors agree on its value. This in turn implies that external investors must be unable to acquire private information about the bank's loan portfolio: it is optimal for bank loan portfolios to be "opaque."

Along the same lines, Dang et al. (2020a) showed how financial panics can occur following a negative shock that leads investors to start doubting that bank debt is truly risk-free, which gives them an incentive to start collecting information and causes adverse selection which decreases liquidity. Dang

[37] Gorton and Pennacchi's (1990) addition of asymmetric information built on the then-emerging market microstructure literature, such as Grossman and Stiglitz (1980) and Kyle (1985).

et al. (2020b) argued that this mechanism is consistent with the runs on shadow banks during the Global Financial Crisis in 2007–2009 (see Section 5).

4.2 Delegated monitoring

We now turn to extensions of Diamond (1984), some of which have been the result of collaboration between Diamond and Raghuram Rajan (2000, 2001, and 2005). Together, they have extended the original model in several directions to investigate the role of banks as delegated agents and to derive implications for optimal financial regulation.

Monitored and unmonitored lending

As Bernanke (1983) showed, the credit crunch in the Great Depression disproportionately affected bank-dependent borrowers – small firms, farmers, and households – while large firms with access to public bond markets were much less hurt (a pattern that has been shown to hold during other financial crises as well; see Section 5). A subsequent empirical literature subsequently documents the importance of bank relationships for certain types of borrowers but not for others (see, e.g., Petersen and Rajan, 1994, 1995; Berger et al., 1995 and 2005). This raises the question of which borrowers will finance themselves through banks versus markets.

While Diamond (1984) derived delegated monitoring as the optimal financial setup, subsequent literature has modelled the coexistence of banks and bond markets. Diamond (1989) explored how firms can use reputational capital to obtain financing when there is no monitoring technology available for lenders. He showed how firms can build a reputation over time, which alleviates the incentive conflicts between borrowers and lenders and lowers financing costs.

Building on this model, Diamond (1991) provided a model where firms can borrow from banks that monitor or from arms-length lenders that do not. A key result is that borrowers with credit ratings (i.e., observable credit quality or reputation) towards the middle of the spectrum will rely on bank financing. Firms with higher credit ratings will borrow from arms-length lenders, since they will be disciplined by their large reputational consequences of defaulting, while monitoring is not profitable for the lowest-rated firms, who have no choice but to turn to arms-length lenders and try to build a reputation over time. [38] [39]

[38] Other influential theories on monitored vs. arms-length financing include Sharpe (1990) and Rajan (1992).

[39] In other work Diamond has explored the interaction of monitoring and debt structure. Diamond (1993a,b) shows that an ideal mix of short-term senior debt and long-term junior debt balances credible threats of liquidation against inefficient liquidation in a way that minimizes the firm's total cost of capital. Diamond (1993b) argues that active monitors should hold short-term senior debt. These predictions have been examined by James (1995, 1996) who finds that, for firms in financial distress, banks tend to not make any concessions unless public debt

Given the important role of bank monitoring, it is puzzling that the secondary market for bank loans has grown significantly in recent decades. It would seem that this very activity would undermine the incentive for banks to screen borrowers *ex ante* and to monitor *ex post*. Gorton and Pennachi (1995) examined whether banks have incentive-compatible arrangements that could explain loan sales. They find that banks keep a portion of the cash flows on their balance sheet which is consistent with maintaining incentives. In principle this could preserve the incentive to monitor and screen, but it is puzzling that banks sell a large fraction of loans, and also transfer the credit risk of their loans to third parties via credit default swaps. Subsequent work analyzing the implications of loan sales on monitoring includes research from Gorton and Winton (2003) and Parlour and Plantin (2008).

Government policy and financial stability

In Section 3.2, we outlined the model of Diamond and Rajan (2001), which combines the forces from Diamond (1984) and Calomiris and Kahn (1991) to show how the combination delegated monitoring and demand deposits makes it possible for banks to create liquidity. In subsequent work, Diamond and Rajan have extended their basic model to address issues around optimal financial regulation and macroeconomic aspects of banking.

Diamond and Rajan (2000) extended the basic model to analyze the role of bank capital requirements. When the outcomes of the investment projects to which the bank lends are uncertain, and this uncertainty is not contractible and cannot be diversified away, depositors may inefficiently run on the bank when project payoffs are low. To avoid this, the bank can replace some of the deposit financing with equity or long-term debt, or "capital." Capital be more expensive than deposits, however, since it does not provide the same discipline to the bank and thus leaves the bank with some rents. The optimal bank capital structure then trades off the higher cost of bank capital against the risk of bank runs in bad states of the world.

Diamond and Rajan (2005) analyzed a version of this model with a competitive banking sector and aggregate uncertainty, in that some of investment projects to which banks lend can be delayed with some probability, leading to a temporary shortage of liquidity in the economy. They showed that when liquidity shocks are sufficiently large, failures can spread across banks and cause a systemic crisis, even absent any interbank links (in contrast to Allen and Gale, 2000). Different policy responses vary in their ability to alleviate a crisis. For example, a when a crisis is the result of liquidity rather than solvency problems, a bailout of banks can worsen the crisis by further prolonging an aggregate liquidity shortage.

holders do, and that banks typically make fewer concessions than public debt holders. This evidence is consistent with banks holding more senior claims and that banks are tougher negotiators.

The optimal policy analysis is further developed by Diamond and Rajan (2006, 2012). Diamond and Rajan (2012) showed how unconstrained bailouts of failing banks also undermines the disciplinary role of deposits, and can lead to excessive liquidity creation by banks, which increases the risk of crises. They argued that the optimal policy response is instead a lender of last resort policy, where the central bank charges a higher real interest rate to banks in good times to offset distortions from reducing rates in adverse times. Diamond and Rajan (2006) introduced monetary policy into the model, and show that expansionary monetary policy can counteract the risk that a real liquidity shortage develops into a systemic crisis by deflating the value of nominal deposits.

4.3 The financial propagation mechanism

In explaining how the financial factors propagated the Great Depression, Bernanke (1983) emphasized two channels: a fall in credit supply, due to stressed and failing banks cutting back their lending, and lower credit demand, as a higher *CCI* led firms and households to demand less debt. The second channel was also emphasized by Fisher (1933), but only informally. In the late 1970s, however, a number of papers written by microeconomics and finance researchers had demonstrated how informational asymmetries and agency problems could give rise to financial frictions.[40] In a short piece Bernanke wrote two years before his landmark 1983 paper (Bernanke, 1981), he argued that these financial frictions should have an impact on aggregate demand and should be incorporated into macroeconomic analysis.

While Bernanke (1983) did not provide any formal macroeconomic model, he did so a few years later in work with Mark Gertler.[41] Bernanke and Gertler (1989) incorporated Townsend's (1979) model of external finance constraints into a general-equilibrium macroeconomic model. The financial friction stems from an asymmetric information problem, where the entrepreneurs privately observe project outcomes, while lenders must pay a monitoring cost to do so.

This monitoring cost makes it more expensive for entrepreneurs to finance investment by borrowing from lenders (external financing) compared to using their own funds (internal financing). As a

[40] The earliest papers include Jensen and Meckling (1976), Jaffee and Russell (1976), Townsend (1979), Myers (1979), and Stiglitz and Weiss (1981).

[41] A few years before, Williamson (1987) presented what was likely the first dynamic macroeconomic model with endogenous financial frictions, also using a model with costly monitoring as in Townsend (1979) and Diamond (1984). As in Diamond (1984), lending is optimally delegated to a financial intermediary in Williamson's model; the intermediary monitors all projects on behalf of savers (i.e., depositors). Monitoring is carried out ex post as in Townsend (1979) (rather than ex ante, as in Diamond, 1984) and will only occur when the borrower defaults. If the likelihood of default is higher, the expected monitoring cost of the intermediary is also higher. Since the intermediary has to break even on the loan including expected monitoring costs, some projects which would have been financed in the absence of asymmetric information will not be financed. Since expected monitoring costs are higher in economic downturns when more projects default, more entrepreneurs will be credit-constrained in economic downturns, which generates a financial propagation mechanism.

consequence, monitoring costs, which are a deadweight cost in the economy that reduces the resources left for consumption and investment, are higher when entrepreneurs' net worth (i.e., their available savings) is lower.

Bernanke and Gertler (1989) showed that this creates a financial multiplier that exacerbates business-cycle fluctuations: a negative (positive) shock to productivity is associated with higher (lower) monitoring deadweight costs and larger decreases (larger increases) in investment than would be the case without financial frictions.[42] In an influential extension, Bernanke, Gertler, and Gilchrist (1999) embedded a financial accelerator into a dynamic stochastic general-equilibrium model with nominal rigidities (see Woodford, 2003) that allowed for interactions between financial frictions and monetary policy. Variants of this model became widely adopted by central banks around the world for prediction and monetary policy evaluation.

In an influential paper, Kiyotaki and Moore (1997) modeled another important financial amplification mechanism, working through asset prices and collateral. [43] They started from a different microeconomic model of financial frictions, one due to Hart and Moore (1994), where lenders cannot force borrowers to pay back their loans unless their debt is secured by some collateral, which can be seized by lenders upon default. The value of this collateral will then influence the amount of funding a lender will provide.

Kiyotaki and Moore (1997) incorporated this mechanism into a dynamic-general equilibrium framework where capital is both a factor of production and serves as collateral for borrowing. When a negative shock hits, the net worth of borrowers is reduced and they have to cut back on borrowing and investment, as in Bernanke and Gertler's work. In addition, the reduced demand for capital makes the equilibrium price of capital drop. A lower price of capital leads to a lower collateral value, which reduces the debt capacity of borrowers even further, leading to further drop in investment, collateral values and so forth.

Such "fire-sale" dynamics had been shown to be important in the Great Depression by Fisher (1933) and was again emphasized in Bernanke (1983), but Kiyotaki and Moore (1997) provided the first dynamic equilibrium model of such effects. In addition to explaining the depth of a downturn, their model also shows how a prolonged increase in asset prices can cause a "credit boom," which makes the economy more vulnerable once an economic downturn materializes; this mechanism had been

[42] The model implies smaller firms should account for a larger decline in economic activity in a downturn compared to larger firms, since they suffer more from asymmetric information problems. This prediction has been verified empirically by Bernanke, Gertler, and Gilchrist (1996) and others.
[43] A similar mechanism had previously been demonstrated in a static model by Shleifer and Vishny (1992).

emphasized in Minsky (1964) and Kindleberger (1973) and is believed to have played an important role for explaining the build-up in leverage during the years preceding the Great Recession.[44]

These models of financial amplification do not incorporate any credit supply channel, however, as savers were assumed to lend directly to firms, giving no role for financial intermediaries and panics.[45] Thus, while quantitative dynamic macro models such as Bernanke, Gertler, and Gilchrist (1999) are able to capture financial accelerator effects under more "normal" business cycles, they have difficulties in generating the large and rapid drops in real activity that can occur when credit demand and credit supply effects interact during a financial crisis, such as in the Great Depression and the Great Recession.[46] Subsequent research has demonstrated that financial amplification effects become substantially larger when intermediaries are introduced in these types of models, as bank-dependent borrowers become financially constrained not only due to the reduction in their own borrowing capacity, but also because of a decrease in the supply of bank credit as banks become more stressed.[47]

Bernanke and co-authors also have explored the interaction of financial frictions with monetary policy empirically and proposed a credit channel for monetary transmission, as a complement to traditional explanations based on sticky wages and prices (see Bernanke and Gertler, 1995, for a review).[48] According to this view, when the central bank raises interest rates, it depresses economic activity by increasing the cost of external financing for borrowers. This would occur in two ways: by lowering the net worth of borrowers through higher interest payments and lower asset prices (the "balance sheet" channel) and by increasing the funding cost of banks, leading to a reduction in bank lending (the "bank lending channel"). The credit channel for monetary transmission is currently an active research area and is regularly brought up in monetary policy discussions, especially since the Great Recession.[49]

---

[44] See Brunnermeier (2010) and Adrian and Shin for empirical evidence, and Brunnermeier and Pedersen (2009), Geanakoplos (2010), and Diamond and Rajan (2011) for examples of theoretical work using collateral prices to explain the dynamics during the Global Financial Crisis and the Great Recession.

[45] Bernanke and Gertler (1989) assume that the outcome from monitoring is observable to everyone in the economy; thus Diamond's (1984) problem of "monitoring the monitor" does not occur and there is no role for financial intermediaries.

[46] See Kocherlakota (2000).

[47] Holmström and Tirole (1997) develop a static general-equilibrium model incorporating both financial intermediaries and market financing (as in Diamond 1991a). As in Diamond (1984, 1991), intermediaries monitor borrowers at a cost and there is a problem with monitoring the monitor, which in their model is solved by the intermediaries investing its own wealth. They assume that the outcomes of different investment projects are perfectly correlated, which rules out perfectly solving the "monitoring the monitor" problem through diversification, as in Diamond (1984). After the Great Recession, work in this area has intensified, aiming to develop quantitative dynamic macroeconomic models with financial intermediaries. Research in this direction include He and Krishnamurthy (2012, 2013), Brunnermeier and Sannikov (2014), Gertler and Kiyotaki (2015), Gertler et al. (2016, 2020),

[48] Central papers in this literature include Bernanke and Blinder (1988, 1992), Gertler and Gilchrist (1994), Kashyap et al (1994), and Kashyap and Stein (2000). For more recent evidence on the bank lending channel, see Jimenez et al. (2012) and Drechsler et al. (2017).

[49] See, e.g., https://www.federalreserve.gov/newsevents/speech/quarles20190530a.htm

5. Discussion of recent crises and policy responses in light of the laureates' work

In this section, we use the Global Financial Crisis and the ensuing Great Recession as a case study to show that the laureates' research from the 1980s is highly relevant for understanding recent crises and for developing policy tools that can help prevent financial market disruptions from creating long-lasting adverse effects on the economy. These tools are regularly in use by policymakers worldwide and played an important role also during the onset of the COVID-19 pandemic.

5.1 The Global Financial Crisis

The Global Financial Crisis of 2007–2009 was the most severe global economic crisis since the Great Depression. The crisis culminated with the bankruptcy of Lehman Brothers on September 15, 2008, which led to a panic in U.S. financial markets that spread rapidly worldwide, leading to the failure and/or government rescue of several large financial institutions. The financial market panic had a dramatic effect on the real economy and gave rise to what is today commonly referred to as the Great Recession.[50]

The years leading up to the Global Financial Crisis were characterized by a booming economy in the U.S., particularly in the real estate sector, where house prices had increased rapidly. This, together with concurrent increases in other types of household debt (such as auto loans and credit cards), had resulted in very high levels of household indebtedness (Mian and Sufi, 2013). While most loans were traditionally provided by regulated banks, an increasing component of credit intermediation and maturity transformation had migrated to the unregulated market-based intermediaries, the so-called shadow-banking sector (Poszar et al., 2010). Traditional banks had increasingly moved to an "originate and distribute" model, where they repackaged the loans they made and sold them to other types of investors in a process called securitization (Brunnermeier, 2009). Securitization replaced the intermediation model with a long intermediation chain.[51]

Similar to traditional banks, the various intermediaries in the shadow-banking sector financed themselves largely with short-term debt. Rather than being provided by savers in the form of deposits, however, the financing was provided by debt markets in the form of wholesale funding, such as Asset-Backed Commercial Paper (ABCP) and repurchase agreements (repos). The combination of high debt levels among both households and financial intermediaries, together with the reliance on short-term

[50] See Brunnermeier (2009) and Gertler and Gilchrist (2018) for detailed research overviews on the causes and consequences of the Global Financial Crisis and the Great Recession.
[51] See Adrien and Shin (2010), Figure 2 and 6 for a description.

debt that had to be continuously rolled over, made the financial system highly vulnerable, for the fundamental reason described in Diamond and Dybvig (1983).

When loans were securitized, the credit evaluation of borrowers deteriorated (Keys et al., 2010 and 2012). In 2006, defaults on household mortgages started rising, particularly in the subprime segment, and U.S. house prices started falling. As mortgage credit losses increased, bank solvency worsened, but initially only slowly. The first signs of a financial panic occurred in August 2007, when French bank BNP Paribas announced that it was unable to value its subprime mortgage assets. This put wholesale funding markets under pressure. In March 2008, the investment bank Bear Stearns came close to failing due to its various exposures to mortgage-backed securities and found it increasingly difficult to roll over its short-term debt. To avoid bankruptcy, Bear Stearns was acquired by JPMorgan Chase in a government-assisted rescue. Over the next couple of months, an increasing number of financial intermediaries came under pressure, including the government-sponsored mortgage intermediaries Fannie Mae and Freddy Mac, who were put under federal conservatorship.

One effect of the long intermediation chains in the shadow-banking sector was that financial institutions in Europe and other parts of the world became exposed to the risk of U.S. mortgage and loan markets. In late July 2007, the German Industriekreditbank IKB announced that it had incurred significant losses on its large exposures to the subprime market. A few weeks later, another German bank, Sachsen Landesbank, was hit. In September 2007, the U.K. mortgage lender Northern Rock, which had financed its operations in the commercial paper and interbank markets, found itself facing a liquidity crisis and had to be saved by the U.K. government. In October and November, several large international banks announced that they had incurred significant impairment losses on their exposures to the subprime sector.

The financial panic culminated with the Lehman Brothers bankruptcy and the government rescue of insurer AIG in September of 2008, after which short-term debt markets came close to a stand-still. While the panic in short-term funding markets was quelled over the next few months, concurrent with massive government and regulatory intervention, the panic led to deteriorating bank solvency and significant reductions of lending volumes, which did not recover until 2010. Real economic activity slowed significantly both in the U.S. and other parts of the world, reflected in large decreases in GDP, industrial production, and employment.

5.2 The financial panic had a large effect on economic activity

The Global Financial Crisis and the ensuing Great Recession illustrate the dramatic effect that financial panics can have on the real economy. As in the Great Depression, the financial panic lead to a rapid

increase in the cost of credit intermediation, which ultimately affected the real activity resulting in the Great Recession. In particular, the demand for durables fell sharply (Gertler and Gilchrist, 2018).

Bernanke (2018) provides evidence of the differential impact various financial channels had on the severity of the Great Recession. In particular, he considers the relative importance of two different mechanisms: i) panics, due to lost confidence in financial institutions, reducing credit supply and ii) worsening of household balance sheets, resulting in deleveraging and less household spending. Bernanke uses factor analysis to reduce a large number of financial data series into four indexes of financial health: two capturing the first mechanism and two the second. His empirical analysis suggests that the two panic factors have a significantly larger explanatory power for the deterioration in real activity than the balance sheet factors. In other words, Bernanke concludes that the main driving force for the economic downturn was the financial panics.

5.3 The panic was triggered by runs on short-term debt

A main message of Diamond and Dybvig (1983) is that absent deposit insurance or access to a lender of last resort, maturity transformation is vulnerable to panics. Extensions of their model show that such self-fulfilling panics can be triggered by a negative shock to fundamentals (Allen and Gale, 1998, 2000; Goldstein and Pauzner, 2005). In line with this, Bernanke (2018) shows that the fundamental shock that started the crisis was the slump in the housing sector, but it did not lead to a dramatic economic downturn until panic arose in the financial markets. As in Diamond and Dybvig (1983), this panic led to self-perpetuating runs. [52] While a few banks experienced traditional deposit runs (e.g., Northern Rock, Washington Mutual, and Wachovia Bank), most of the runs occurred in the shadow-banking sector.[53] These runs triggered the failures of several large investment banks during the financial crisis, including Bear Stearns, Lehman Brothers, and Merrill Lynch. However, since commercial banks were also increasingly relying on short-term wholesale funding to complement (insured) core deposits, traditional banks were also affected by the panic, and they responded by reducing the supply of bank credit.

Cornett et al. (2011) show that banks that relied more on insured deposits for their funding continued to lend during the financial crisis. Banks that did not, and that had more illiquid assets on their balance sheet, started cutting lending and replacing loans with liquid assets (very similar to what Bernanke, 1983, documented for the Great Depression).[54] Acharya and Merrouche (2012) show that the

[52] See Goldstein (2013) for a survey of the empirical literature on bank runs and deposit insurance.
[53] Covitz et al. (2013) describe runs on asset-backed commercial paper, Schmidt et al. (2016) discuss the run on money-market mutual funds, and Gorton and Metrick (2009) show that the Lehman failure was followed by a run on the repo market.
[54] The connection between reliance to wholesale funding and bank credit supply was also found by Irani and Meisenzahl (2014) and Dagher and Kazimov (2015).

unwillingness of large wholesale-funded banks to part with their liquidity contributed to a freeze in interbank markets worldwide,[55] and this affected banks outside of the U.S., including even those with no exposure to U.S. mortgage markets (see, e.g., Iyer et al., 2014). Traditional banks were also exposed to a different type of run, not on their deposits but on the credit lines they had provided to firms. During the Global Financial Crisis, corporations started drawing down their credit lines and kept the liquidity as cash on their own balance sheets, because they were concerned that their credit line would be cancelled if their bank failed. Ivashina and Scharfstein (2010) show that runs on credit lines contributed to affected banks cutting their lending to firms and households.

The runs experienced during the Global Financial Crisis had several features in common. First, they involved shadow banks that engaged in maturity transformation, where the intermediaries promised their funding providers short-term liquidity and then invested the funds borrowed into longer-term, less liquid securities. Second, shadow banks were not subject to the same regulation as traditional banks and were for this reason not covered by deposit insurance. Moreover, even for regulated banks, deposit insurance did not apply to their wholesale and interbank funding and could not prevent a run on credit lines. Third, while the origin of the runs was the losses on U.S. subprime mortgages, they eventually led to runs on intermediaries with very little exposure to the original fundamental shock (i.e., the drop in U.S. housing prices and the mortgage losses that followed), consistent with a Diamond and Dybvig (1983) financial panic. For example, both securitization vehicles holding loans to non-financial corporations and non-U.S. banks relying on wholesale funding but with no or at best indirect exposure to the U.S. subprime market faced runs.

5.4 Disruption in credit intermediation hurt the real economy

Parallel to the facts Bernanke documented for the Great Depression, the 2007–2009 panics in financial markets led to a contraction in credit supply, which was followed by a sharp drop in real economic activity. While this suggests that the financial market disruption caused or deepened the Great Recession, this correlation could be the result of some other unobserved non-financial shock to demand or productivity. Using microdata (and identification methods pioneered by the 2021 Prize in Economic Sciences laureates Angrist, Card, and Imbens), however, several researchers have provided causal evidence on the effect of the Global Financial Crisis on real economic activity. For example, Duchin et al. (2010) and Almeida et al. (2011) use variation in non-financial firms' liabilities for identification to show that the drop in credit supply that followed the financial panic led firms to cut investment in the crisis.

[55] Interbank markets allow banks with excess liquidity at the end of a day to lend it to banks needing liquidity, typically overnight.

The fact that a stressed bank cuts its credit supply would be less damaging if borrowers could easily switch to other sources of financing. Bernanke (1983) and Diamond (1984) argue that the unique information the bank obtains about the borrower through their screening and monitoring makes it difficult for firms to replace their bank with a new lender.

Several papers have use microdata on bank-firm relationships to isolate the effect of reductions in loan supply from reduced loan demand. When firms have several bank relationships, Khwaja and Mian (2008) show that loan demand effects can be controlled for, allowing causal identification of the effect of a loan supply shock on firm performance, which has been extensively used in the subsequent literature.

Using U.S. microdata matching bank loans to borrowers, Chodorow-Reich (2014) use the dispersion in lender health following the Lehman Brothers crisis to show that firms with weaker lenders borrowed less, paid higher rates when they borrowed, and reduced employment more than other firms. The strongest employment effects were at small and medium-sized firms, and these firms are more dependent on bank financing.[56] Along these lines, Adrian et al. (2012) and Becker and Ivashina (2014) provide evidence suggesting that large firms with access to bond markets were much less credit constrained, as public debt markets recovered more quickly than did bank lending after the Global Financial Crisis.

Further causal evidence of the credit channel comes from research showing that credit-supply effects extended to countries outside of the U.S. that were not directly exposed to the bust in U.S. real estate markets. Puri et al. (2011) shows that the effect of losses in the German Landesbanks on their holdings of U.S. mortgage-backed securities led to a contraction in household credit by German savings banks, who were the owners of the Landesbanks.

Also for Germany, Huber (2018) considers the effect of the lending cuts of one of Germany's largest banks, Commerzbank. The bank suffered significant trading losses on their holdings of international securities, which were economically unrelated to the fundamentals of their German domestic market. These trading losses depleted Commerzbank regulatory capital and forced the bank to cut its lending to German borrowers. Huber shows that this resulted in persistent adverse effects on output, employment, and productivity in firms and regions where the Commerzbank had a relatively larger market share

[56] In the Great Recession, many leveraged borrowers ended up breaching their loan covenants because of a drop in their profits. Using a bank-borrower data set, Chodorow-Reich and Falato (2022) found that firms borrowing from a stressed banks were more likely to experience a reduction in their loan commitments following a covenant valuation, and that this channel account for most of the cross-sectional variation in credit supply to firms over this period.

before the crisis. Similar evidence of spillovers from the U.S. financial panic on credit supply outside of the U.S. have been found for many other countries around the world.[57]

The failure of shadow banks did not only affect credit supply indirectly through the spillovers on commercial banks, but also directly affected access to credit in the real economy. Using a data set linking every U.S. car sale to an associated supplier of auto credit, Benmelech et al. (2017) show that the collapse of the asset-backed commercial paper market led to a decrease in car sales. Studying the subsequent European sovereign debt crisis in 2011, Chernenko and Sunderam (2014) show that the outflows from money-market mutual funds exposed to Eurozone banks led to worse access to credit for non-European borrowers who were relying on these funds for financing. Their evidence suggests that like traditional banks, shadow banks also develop valuable lending relationships that are hard to replace.

Asian countries went through a severe financial crisis in the late 1990s. In the countries most affected by the Asian financial crisis, there was a long period of post-crisis deleveraging when private credit fell by some 40% to 50% of GDP. Asian governments responded to the crisis by implementing regulatory and prudential policies that resulted in more reliance on deposit funding, higher bank asset quality, and less reliance on external short-term financing leading up to the Global Financial Crisis (Jeasakul et al., 2014). While many Asian banks, particularly in Japan, had exposures to U.S. subprime assets, their losses and write-downs were much more modest than those of U.S. and European banks. Moreover, because they were less reliant on wholesale funding, they did not face a debilitating funding squeeze. Despite entering the crisis with a healthier financial sector, Asian economies shrunk during the Great Recession, but they contracted less, and rebounded more quickly, than did the U.S. and European countries.[58]

5.5 Policy responses

The panic in short-term debt markets following the demise of Lehman Brothers in 2008 prompted the Fed and the U.S. Department of Treasury (U.S. Treasury) to intervene, acting as a "lender of last resort" not only for commercial banks but also extending support to important shadow banks (such as large investment banks).[59] To support the flow of credit to households and firms, the Fed launched a series of programs targeting different financial intermediation functions. The Fed's programs included: the Term Auction Facility (TAF), which provided longer-term funding for depository institutions that was

[57] See, e.g., Aiyar (2012) for the UK, Iyer et al. (2014) for Portugal, Jensen and Johannesen (2017) for Denmark, and de Haas and Van Horen (2012) using data on syndicated lenders from 59 countries.
[58] Comparing Asia to the rest of the world, Jeasakul et al. (2014) find that the depth of the output decline was smaller by 2.8 percentage points, the recovery to the 2008Q3 output levels was more than 3 quarters quicker, and the cumulative output loss was lower by 11% of annualized 2008Q3 GDP.
[59] Ben Bernanke was the chairman of the Federal Reserve between 2006 and 2014, and as such one of the main U.S. policymakers during the Global Financial Crisis. For his perspective on government interventions during the crisis, see Bernanke, Geithner, and Paulson (2019).

intended to help circumvent the stigma of the discount window; the Term Securities Lending Facility (TSLF), which allowed primary dealers to exchange illiquid assets for Treasuries; and the Primary Dealer Credit Facility (PDCF), where the Fed lent to primary dealers against collateral accepted in the repo market.

In an effort to ensure that mortgage financing would remain available to creditworthy borrowers, the U.S. Treasury made an initial pledge of $200 billion (which was later increased to $400 billion) when Fannie Mae and Freddie Mac were placed into conservatorship. The Fed also intervened by conducting large purchases of agency mortgage-backed securities with aim to reduce the cost of mortgages by expanding central bank intermediation to offset the contraction in credit provision by financial intermediaries.

Finally, the U.S. Treasury launched the Troubled Asset Relief Program (TARP), which involved purchasing $250 billion of preferred equity in the nine largest U.S. commercial banks (Gertler and Gilchrist, 2018). These interventions, coupled with a temporary public guarantee on the debt of these banks, helped stabilize the markets, especially the short-term funding market, and many observers believe they helped restore investor and creditor confidence in the solvency and viability of financial institutions.

Similarly, the European Central Bank (ECB) took a number of steps to support the smooth functioning of the euro area interbank markets in response to the Global Financial Crisis. These non-standard measures became known as enhanced credit support, and helped secure the flow of credit to households and firms. They focused primarily on commercial banks, as these are the main source of funding for households and business in the euro area (about 70% of the funding comes from banks). The ECB policy interventions included:[60] "the full accommodation of banks' liquidity requests at fixed interest rates; the expansion of the list of assets eligible as collateral; the lengthening of the maturities of refinancing operations, up to one year; the provision of liquidity in foreign currencies, notably the U.S. dollar; and, finally, outright purchases of euro-denominated covered bonds issued in the euro area." Individual European governments also adopted measures to support their financial markets, thus safeguarding the stability of the European financial system. The measures included increasing deposit insurance ceilings, guarantees for bank liabilities, and bank recapitalizations.

Central banks around the world took similar steps using a battery of policies to avoid runs and to keep credit flowing to households and firms. The Bank of Japan, for example, purchased stocks held by banks and provided subordinated loans to banks with the objective of ensuring the stability of the

[60] Gertrude Tumpel-Gugerell, Member of the Executive Board of the ECB, Speech at Bank of New York Mellon Headquarter, New York, October 16, 2009 (https://www.ecb.europa.eu/press/key/date/2009/html/sp091016_1.en.html).

financial system.[61] Furthermore, the Bank of Japan took a number of steps to facilitate corporate financing, including increasing the frequency and size of commercial paper repurchase agreement operations, expanding the range of asset-backed commercial paper and corporate debt eligible as collateral, and conducting outright purchases of commercial paper and corporate bonds.

Thus, even though Asian banks, as we discussed above, were less exposed to subprime and regulation and prudential policy and Japan had more limited private credit expansion, in light of the global financial turmoil, Japanese policymakers recognized the need to intervene to safeguard the financial system and the flow of credit in order to avoid a deep and prolonged recession. In other words, they too understood the lessons from the research of this year's laureates, Bernanke, Diamond, and Dybvig.

5.6 Banking regulation

To be sure, supporting financial institutions during a financial crisis comes at a cost. That taxpayers bear this cost may be particularly controversial if the crisis is considered to have arisen from excessive risk taking in the banking sector. Governments bailing out banks that have taken excessive risks produces incentives for risk-taking at the expense of taxpayers (moral hazard). Requirements that a bank keep a certain capital ratio (equity capital divided by assets) are supposed to mitigate excessive risk taking and ensure that sufficient equity capital is available to support banks' lending activities also in bad times, but policymakers are aware that these requirements may increase credit costs during normal times.

The capital requirements in place at the time clearly failed to prevent the Global Financial Crisis, arguably because they did not sufficiently mitigate excessive risk-taking. The international financial regulation since the Global Financial Crisis has made capital requirements more stringent, including the introduction of counter-cyclical capital buffers and "absolute" caps on bank leverage that complement the risk-based ones that were already in place. In addition to capital requirements, post–Global Financial Crisis regulation has introduced new liquidity requirements that depends on the illiquidity of the bank's assets and the extent of maturity transformation that it undertakes.[62]

Moreover, existing bank capital requirements prior to the Global Financial Crisis addressed risk-taking by an individual bank. If a bank's capital ratio falls below the required level, it can restores it by raising more equity or shrinking its balance sheet by reducing lending. However, if several large banks' capital ratios fall below the required level at the same time, and they respond by shrinking their balance sheets, a credit crunch will ensue. If banks try to sell assets urgently (so called fire sales), asset values market-wide may fall even further, leading to additional reduction of lending activity by banks seeking to shrink

[61] See https://www.boj.or.jp/en/mopo/outline/cfc.htm/.
[62] Diamond and Kashyap (2016) extend the Diamond-Dybvig model to analyze optimal liquidity regulation.

the balance sheet (deleveraging). The result is a harmful reduction in credit supplied to households and firms, curtailing real activity such as consumer spending and business activities. Hence, it became clear that bank regulation solely focusing on risk-taking by individual banks was inadequate.

The Global Financial Crisis also made it clear that it was insufficient to focus prudential policies simply on traditional banks when much of the solvency problems were in the shadow-banking system. Policymakers needed updated tools to address the evolving landscape of financial intermediation – "a 'macroprudential' approach [that] recognizes general equilibrium effects, and seeks to safeguard the financial system as a whole" (Hansen, Kashyap, and Stein, 2011, p. 3). Thus, *macroprudential policy* has now become widespread language to cover this broad range of policy and regulatory interventions. These are aimed at striking a balance between the negative consequences of disrupted credit networks emphasized in Bernanke (1983) and the various costs of intervention.

The regulatory response to the Global Financial Crisis came in the form of the July 2010 Wall Street Reform and Consumer Protection Act, or "Dodd–Frank," in the U.S., and the recommendations made by the Basel Committee on Banking Supervision in September 2010, the "Basel III" process. The former focuses on consumer protection, regulation of over-the-counter derivatives, and resolution authority, while the latter addresses some of the deficiencies in the pre–Global Financial Crisis bank capital requirements.[63] Banking regulators around the world now also require large (systemically important) banks to conduct stress tests, analyses to determine whether a bank has enough capital to absorb losses during stressful conditions while meeting obligations to creditors and counterparties, and continuing to be able to lend to households and businesses. In other words, the stress tests help ensure that large banks can support the economy during economic downturns.

5.7 The COVID-19 pandemic

Also during our most recent economic crisis, caused by the COVID-19 pandemic, it is clear that policymakers understand that it is imperative to safeguard the viability of the financial system to minimize disruptions to credit supply. As first shown by Bernanke (1983) and emphasized in the present document, the Great Depression started as a "normal" recession but developed into something much worse due to disruptions in the financial system, particularly bank failures. The steps taken by federal, state, and local officials to mitigate the spread of the virus during the pandemic led to a sudden and very deep reduction in economic activity. Through a battery of measures, central banks all over the world ensured that credit continued to flow to households and firms, preventing financial market disruptions from intensifying the economic damage. The Fed intervened directly in the markets for corporate and

[63] Hansen, Kashyap, and Stein (2011) propose several ways in which bank capital requirements can be improved in light of academic research, and evaluate how the Basel III process corresponds to their proposals.

municipal bonds to ensure that key players could raise funds to pay workers and avoid bankruptcy. The Fed also provided unlimited liquidity to financial institutions so they could meet credit drawdowns and make new loans to businesses and households. These measures were aimed to help firms survive the crisis and resume hiring and production once the pandemic receded.

In Europe, the ECB responded to the COVID-19 crisis by dramatically increasing its purchases of government bonds, regional and local authorities' bonds, corporate bonds, asset-backed securities, and covered bonds under its existing programs, and significantly expanded the scope of bond-buying activity by launching the €750 billion Pandemic Emergency Purchase Program (PEPP), which also covered commercial paper issued by non-financial corporations. Furthermore, the ECB incentivized banks to lend by expanding its targeted long-term refinancing operations. These offered banks cheap (with negative interest rate, banks actually get paid to borrow money), long-term loans with additional incentives to use the funds to lend to euro area consumers and businesses. Finally, the ECB launched several other initiatives to encourage banks to lend to consumers, business, and other banks, including temporarily relaxing capital requirements; relaxing the rules around the classification of non-performing loans; easing collateral restrictions; providing support for bank funding and money markets; and established international swap lines.[64]

Similar massive interventions were undertaken by central banks outside of the Eurosystem. For example The Bank of Japan engaged in special funds-supplying operations to provide loans to financial institutions with the goal of facilitating lending to small and medium-sized firms. The Bank's stated objective was to do "its utmost to ensure smooth financing of the private sector and maintaining stability in financial market through appropriate market operations, given the impact of the outbreak of COVID-19 on economic activity."[65]

While the root cause for the crises we describe in this section differs, they share a common feature: policymakers around the world realize the importance of maintaining market participants' faith in the ability of not just traditional banks, but also in other financial intermediaries such as shadow banks, to channel savings toward investment without disruption. This year's laureates, Bernanke, Diamond, and Dybvig, have taught us that exactly these types of policy responses are crucial ingredients for preventing bank runs and for preserving valuable credit relationships

[64] Belz, Sage, Jeffrey Cheng, David Wessel, Daniel Gros, and Angela Capolongo, June 4, 2020, What's the ECB doing in response to the COVID-19 crisis? (https://www.brookings.edu/research/whats-the-ecb-doing-in-response-to-the-covid-19-crisis/).
[65] https://www.boj.or.jp/en/mopo/measures/mkt_ope/ope_v/index.htm/.

6. Conclusion

Banks and bank-like institutions have existed for thousands of years. Today they are active in every country around the world. Banks obviously perform important functions, but they have also been at the epicenter of some of history's most devastating economic crises such as the Great Depression. Nevertheless, it was not until the work of this year's laureates, Ben S. Bernanke, Douglas W. Diamond, and Philip H. Dybvig, that we had a comprehensive theory of why banks exist in the form we observe, what role they play in the economy, why they are fragile, and an empirical account of how devastating and long-lasting the consequences of massive bank failures can be.

Diamond and Dybvig (1983) showed that an institution using demand deposits to finance long-term lending is perfectly suited to satisfy the conflicting needs of savers and borrowers. The former need liquid assets to satisfy random spending needs, while the latter need long-term commitments to be able to finance investments that cannot be prematurely liquidated without large costs. Banks do this by transforming illiquid assets into liquid assets.

The theory of Diamond and Dybvig (1983) also implies that maturity transformation naturally is a fragile business. A rumor that a bank is about to fail can lead to a bank run, where the expectations that other people will demand their deposits will lead all savers to run to the bank to withdraw their funds. Even healthy banks may get into trouble if bank runs become widespread. However, the theory also implies that deposit insurance and central banks promising to stand in as lender of last resort can be a remedy for this fragility.

Diamond (1984) showed that the way banks are constructed is key for their ability to act as delegated monitors. In practice, small lenders could not themselves undertake the monitoring of all final users of their savings – it needs to be delegated. But who should then monitor the bank? Diamond (1984) showed that a debt contract between lenders and the bank, along with diversification, provides the bank with the right incentive to monitor.

According to the theories of Diamond and Dybvig, banks are middlemen between savers and borrowers. But this situation does not impose costs on society. On the contrary, maturity transformation and borrower monitoring are socially productive activities that reduce the cost of credit and minimize wasteful bankruptcy costs. Thus, the economy works better with banks than without them, provided their inherent fragility can be managed.

Bernanke analyzed the Great Depression. In his seminal work on the subject, Bernanke (1983 showed that the key mechanism behind the depth and in particular the length of the depression was bank failures

and fear of bank runs. Banks could not fulfill the important tasks described theoretically by Diamond and Dybvig. The consequence was the largest economic crisis in modern history.

The monitoring task described by Diamond (1984) requires knowledge about the borrower. This informational capital takes time to build, it is difficult to transfer to other banks and thus often gets destroyed in a bank failure. This, according to Bernanke, explains why the Great Depression and other financial crises have been so protracted.

The research from the 1980s for which this year's Prize in Economic Sciences is awarded obviously does not provide us with final policy recommendations. Deposit insurance does not always work as intended. It can lead to perverse incentives for banks and their owners to gamble to take the profit if things go well and let taxpayers pay the bill if not. Runs on new financial intermediaries, engaging in profitable maturity transformation like banks, but operating outside of bank regulation, were arguably key for the financial crisis 2007–2009 leading to the Great Recession. When central banks act as lenders of last resort, this can lead to large and unintended wealth redistribution and have negative moral hazard effects on banks who may increase reckless lending, potentially leading to future crises.

How to regulate the financial market so that it can perform its important function of channeling savings to productive investments, without from time to time causing financial crises, is a question that is actively debated to this day. The same is true about what policies are most effective in preventing a threatening crisis from developing. However, based on the foundational work of the laureates and all research that has followed, society is now better equipped to handle financial crises.

References

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, 2015, Systemic risk and stability in financial networks, *American Economic Review* 105(2), 564–608.

Acharya, Viral, Philipp Schnabl, and Gustavo Suarez, 2013, Securitization without risk transfer, *Journal of Financial Economics* 107, 515–536.

Adrian, Tobias, and Hyun Shin, 2010, The changing nature of financial intermediation and the Financial Crisis of 2007-09, *Annual Review of Economics* 2, 603–618.

Adrian, Tobias, and Hyun Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19, 418–437.

Adrian, Tobias, Paolo Colla, and Hyun Shin, 2012, Which financial frictions? Parsing the evidence from the financial crisis of 2007–09, *NBER Macroeconomics Annual* 27, 159–214.

Aiyer, Shekhar, 2012, From Financial Crisis to Great Recession: The role of globalized banks, *American Economic Review Papers & Proceedings,* 102(3), 225–230.

Allen, Franklin, and Douglas Gale, 1997, Financial markets, intermediaries, and intertemporal smoothing, *Journal of Political Economy* 105(3), 523–546.

Allen, Franklin, and Douglas Gale, 1998, Optimal financial crises, *Journal of Finance* 53(4), 1245–1284.

Allen, Franklin, and Douglas Gale, 2000, Financial contagion, *Journal of Political Economy* 108, 1–33.

Allen, Franklin, Ana Babus, and Elena Carletti, 2012, Asset commonality, debt maturity and systemic risk, *Journal of Financial Economics* 104(3), 519–534.

Almeida, Heitor, 2012, Corporate debt maturity and the real effects of the 2007 credit crisis, *Critical Finance Review* 1(1), 3–58.

Bagehot, Walter, 1873, *Lombard Street: A description of the money market.* John Wiley and Sons (1999 edition).

Bank of International Settlements, 2021, Assessing the impact of Basel III: Evidence from macroeconomic models: literature review and simulations, Basel Committee on Banking Supervision Working Paper 38, April 2021.

Barth, James, Philip F. Bartholomew, and Michael G. Bradley, 1990, Determinants of thrift resolution costs, *Journal of Finance* 45, 731–754.

Barro, Robert, 1978, Unanticipated money, output, and the price level in the United States," *Journal of Political Economy* 86, 549–580.

Becker, Bo, and Victoria Ivashina, 2014, Cyclicality of credit supply: Firm-level evidence, *Journal of Monetary Economics* 62, 76–93.

Bencivinga, Valerie, and Bruce Smith, 1991, Financial intermediation and endogenous growth, *Review of Economic Studies* 58(2), 195–209.

Benston, George, 1985, *An Analysis of Causes of Savings and Loan Association Failures*, Monograph Series in Finance and Economics; New York University Graduate School of Business, Salomon Brothers Center for the Study of Financial Institutions, Monograph 1985–4/5.

Berger, Allen, Nathan Miller, Mitchell Petersen, Raghuram Rajan, and Jeremy Stein, 2005, Does function follow organizational form? Evidence from the lending practices of small and large banks, *Journal of Financial Economics* 76, 237–269.

Berger, Allen, and Raluca Roman, 2017, Did saving Wall Street really save Main Street? The real effect of TARP on local economic conditions, *Journal of Financial and Quantitative Analysis* 52(5), 1827–1867.

Berlin, Mitchell, and Loretta Mester, 1999, Deposits and relationship lending, *Review of Financial Studies* 12:3 (Fall), 579–607.

Bernanke, Ben S., 1981, Bankruptcy, liquidity, and recession, *American Economic Review Papers & Proceedings* 71(2), 155–159.

Bernanke, Ben S., 1983, Nonmonetary effects of the financial crisis in the propagation of the Great Depression, *American Economic Review* 73, 257–276.

Bernanke, Ben S., 1995, The macroeconomics of the Great Depression: A comparative approach, *Journal of Money, Credit and Banking* 27(1), 1–28.

Bernanke, Ben S., 2018, The real effects of disrupted credit: Evidence from the Global Financial Crisis, *Brookings Papers on Economic Activity* Fall 2018, 251–322.

Bernanke, Ben, and Alan Blinder, 1988, Credit, money, and aggregate demand, *American Economic Review Papers & Proceedings* 78(2), 435–439.

Bernanke, Ben, and Alan Blinder, 1992, The Federal Funds Rate and the channels of monetary transmission, *American Economic Review* 82(4), 901–921.

Bernanke, Ben S., and Kevin Carey, 1996, Nominal wage stickiness and aggregate supply in the Great Depression, *Quarterly Journal of Economics* 111(3), 853–883.

Bernanke, Ben S., Timonthy F. Geithner, and Henry M. Paulson Jr, 2019. *Firefighting: The financial crisis and its lessons.* Penguin Books.

Bernanke, Ben S., and Mark Gertler, 1989, Agency costs, net worth, and business fluctuations, *American Economic Review* 79:1, 14–31.

Bernanke, Ben S., and Mark Gertler, 1990, Financial fragility and economic performance, *Quarterly Journal of Economics* 105, 87–114.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, 1996, The financial accelerator and the flight to quality, *Review of Economics and Statistics* 78(1), 1–15.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, 1999, The financial accelerator in a quantitative business cycle framework. In: John Taylor and Michael Woodford (eds), *Handbook of Macroeconomics, Volume 1*, (Chapter 21, 1341–1393), Princeton University Press.

Bernanke, Ben, and Harold James, 1991, The Gold Standard, deflation, and Financial Crisis in the Great Depression: An international comparison. In R. G. Hubbard, ed., *Financial Markets and Financial Crises*, Chicago: University of Chicago Press.

Bolt, Jutta and Jan Luiten van Zanden, 2020, Maddison style estimates of the evolution of the world economy. A new 2020 update.

Bordo, Michael, and Barry Eichengreen, 1999, Is our international financial environment unusually crisis prone? In David Gruen and Luke Gower, eds., *Capital Flows and the International Financial System* (Sydney: Reserve Bank of Australia), 18–75.

Bordo, Michael, and Barry Eichengreen, 2003, Crises now and then: What lessons from the last era of financial globalization?, in Paul Mizen (ed.) *Monetary History, Exchange Rates and Financial Markets: Essays in Honor of Charles Goodhart*. Vol 2, Cheltenham; Edward Elgar.

Brainard, William, and James Tobin, 1963, Financial intermediaries and the effectiveness of money controls, *American Economic Review Papers & Proceedings* 53(2), 383–400.

Brunnermeier, Marcus, 2009, Deciphering the liquidity and credit crunch of 2007–2009, *Journal of Economic Perspectives* 23(1), 77–100.

Brunnermeier, Marcus, and Lasse Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22(6), 2201–2238.

Brunnermeier, Marcus, and Yuli Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104(2), 379–421.

Bryant, John, 1980, A model of reserves, bank runs, and deposit insurance, *Journal of Banking and Finance* 4, 335–344.

Buser, Steven, Andrew Chen, and Edward Kane, 1981, Federal deposit insurance, regulatory policy, and optimal bank capital, *Journal of Finance* 35, 51–60.

Calomiris, Charles W., 1990, Is deposit insurance necessary? A historical perspective, *Journal of Economic History* 50, 283–295.

Calomiris, Charles W., 1993, Financial factors in the Great Depression, *Journal of Economic Perspectives* 7(2), 61–85.

Calomiris, Charles W., and Gary Gorton, 1991, The origins of banking panics: Models, facts, and bank regulation, in R. Glenn Hubbard, ed.: *Financial Markets and Financial Crises* (University of Chicago Press, Chicago), 109–173.

Calomiris, Charles W., and Gary Kahn, 1991, The role of demandable debt in structuring optimal banking arrangements, *American Economic Review* 81(3), 497–513.

Calomiris, Charles W., and Joseph R. Mason, 2003, Consequences of bank distress during the Great Depression, *American Economic Review* 93(3), 937–947.

Cantillo Simon, Miguel, and Julian Wright, 2000, How do firms choose their lender? *Review of Financial Studies* 13, 155–189.

Capra, Frank, and James Stewart, *It's a Wonderful Life*. Los Angeles, CA: Liberty Films, 1946.

Cass, David, and Karl Shell, 1983, Do sunspots matter? *Journal of Political Economy* 91, 193–228.

Chan, Yuk-Shee, Stuart Greenbaum, and Anjan Thakor, 1992, Is fairly priced deposit insurance possible? *Journal of Finance* 47(1), 227–245.

Chari, V.V., and Ravi Jagannathan, 1988, Banking panics, information, and rational expectations equilibrium, *Journal of Finance* 43(3), 749–761.

Chodorow-Reich, Gabriel, 2014, The employment effects of credit market disruptions: Firm-level evidence from the 2008–2009 financial crisis*, Quarterly Journal of Economics* 129, 1–59.

Chodorow-Reich, Gabriel, and Antonio Falato, 2022, The loan covenant channel: How bank health transmits to the real economy, *Journal of Finance* 77(1), 85–128.

Cohen, Jon, Kinda Hachem, and Gary Richardson, 2021, Relationship lending and the Great Depression, *Review of Economics and Statistics* 103(3), 505–520.

Cole, Harold L. and Lee E. Ohanian, 2000, Re-Examining the contributions of money and banking shocks to the U.S. Great Depression, in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER macroeconomics annual 2000*. Cambridge, MA: MIT Press, 183–226

Cooper, Russell and Thomas Ross, 2002, Bank runs: Deposit insurance and capital requirements, *International Economic Review* 43(1), 55–72.

Cornett, Marcia M., Jamie J. McNutt, Philip E. Strahan, and Hassan Tehranian, 2011, Liquidity risk management and credit supply in the financial crisis, *Journal of Financial Economics* 101(2), 297–312.

Dang, Tri Vi, Gary Gorton, and Bengt Holmström, 2017, Banks as secret keepers, *American Economic Review* 107(4), 1005–1029.

Dang, Tri Vi, Gary Gorton, and Bengt Holmström, 2020a, Ignorance, debt, and financial crises, working paper, Yale School of Management.

Dang, Tri Vi, Gary Gorton, and Bengt Holmström, 2020b, The information view of financial crises, *Annual Review of Financial Economics* 12, 39–65.

Davila, Eduardo, and Anto Korinek, 2018, Pecuniary externalities in economies with financial frictions, *Review of Economic Studies* 85, 352–395.

DeGennaro, R. P., L. Lang, and J. B. Thomson, 1993, Troubled savings and loan institutions: Turnaround strategies under insolvency, *Financial Management* (Autumn), 163–175.

de Haas, Ralph, and Neeltje Van Horen, 2012, International shock transmission after the Lehman Brothers collapse: Evidence from syndicated lending, *American Economic Review Papers & Proceedings* 102(3), 321–237.

Despres, Emile, Albert Hart, Milton Friedman, Paul Samuelson, and Donald Wallace, 1950, The problem of economic instability, *American Economic Review* 40(4), 501–503+505–538.

Diamond, Douglas W., 1980, Essays in Information and Financial Intermediation, PhD Dissertation, Yale University, May.

Diamond, Douglas W., 1984, Financial intermediation and delegated monitoring, *Review of Economic Studies* 51, 393–414.

Diamond, Douglas W., 1985, Optimal release of information by firms, *Journal of Finance* 40, 1071–1094.

Diamond, Douglas W., 1989, Reputation acquisition in debt markets, *Journal of Political Economy* 97, 828–862.

Diamond, Douglas W., 1991a, Monitoring and reputation: The choice between bank loans and directly placed debt, *Journal of Political Economy* 99, 698–721.

Diamond, Douglas W., 1991b, Debt maturity structure and liquidity risk, *Quarterly Journal of Economics* 106, 709–737.

Diamond, Douglas W., 1993a, Seniority and maturity of debt contracts, *Journal of Financial Economics* 33:3 (June), 341–368.

Diamond, Douglas W., 1993b, Bank loan maturity and priority when borrowers can refinance. In Mayer, Colin , and Xavier Vives, eds., *Capital Markets and Financial Intermediation* (Cambridge University Press, Cambridge, New York, and Melbourne), 46–68.

Diamond, Douglas W., 1997, Liquidity, banks, and markets, *Journal of Political Economy* 105, 928–956.

Diamond, Douglas W., and Phillip H. Dybvig, 1983, Bank runs, deposit insurance, and liquidity, *Journal of Political Economy* 91, 401–419.

Diamond, Douglas W., and Anil K. Kashyap, 2016, Liquidity requirements, liquidity choice and financial stability, in *Handbook of Macroeconomics*, Vol 2B, John B. Taylor and Harald Uhlig (eds.), North Holland.

Diamond, Douglas W., and Raghuram G. Rajan, 2000, A theory of bank capital, *Journal of Finance* 55, 2431–2465.

Diamond, Douglas W., and Raghuram G. Rajan, 2001, Liquidity risk, liquidity creation and financial fragility: A theory of banking, *Journal of Political Economy* 109, 287–327.

Diamond, Douglas W., and Raghuram G. Rajan, 2005, Liquidity shortages and banking crises, *Journal of Finance* 60, 615–647.

Diamond, Douglas W., and Raghuram G. Rajan, 2006, Money in a theory of banking, *American Economic Review* 96, 30–53.

Diamond, Douglas W., and Raghuram G. Rajan, 2011, Fear of fire sales, illiquidity seeking, and credit freezes, *Quarterly Journal of Economics* 126, 557–591.

Diamond, Douglas W., and Raghuram G. Rajan, 2012, Illiquid banks, financial stability, and interest rate policy, *Journal of Political Economy* 120, 552–591 .

Diamond, Douglas W., and Robert E., Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277–311.

Diamond, Douglas W., and Robert E. Verrecchia, 1991, Disclosure, liquidity, and the cost of capital, *Journal of Finance* 46, 1325–1359.

Donaldson, R. Glenn, 1992a, Costly liquidation, interbank trade, bank runs and panics, *Journal of Financial Intermediatio*n 2, 59–82.

Donaldson, R. Glenn, 1992b, Sources of panics: Evidence from the weekly data, *Journal of Monetary Economics* 30, 277–305.

Donaldson, R. Glenn, 1993, Financing banking crises: Lessons from the panic of 1907, *Journal of Monetary Economics* 31, 69–95.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2017, The deposits channel of monetary policy, *Quarterly Journal of Economics* 132(4), 1819–1876.

Duchin, Ran, Oguzhan Ozbas, and Berk Sensoy, 2010, Costly external finance, corporate investment, and the subprime mortgage credit crisis, *Journal of Financial Economics* 97, 418–435.

Duffie, Darrell, and Peter DeMarzo, 1999, A liquidity-based model of security design, *Econometrica* 67(1), 65–99.

Dybvig, Philip H., 1993, Remarks on banking and deposit insurance, *Federal Reserve Bank of Saint Louis Review* January/February 1993, 21–24.

Dybvig, Philip H., and G.D. Jaynes, 1980a, Output supply, employment, and intra-industry wage dispersion, Cowles Foundation Discussion Paper Series No. 546, March 6, 1980.

Dybvig, Philip H., and G.D. Jaynes, 1980b, Microfoundations of wage rigidity, Yale working paper.

Dybvig, Philip H., and Chester Spatt, 1983, Adoption externalities as public goods, *Journal of Public Economics* 20, 231–247.

Edgeworth, Francis Ysidro, 1888, The mathematical theory of banking, *Journal of the Royal Statistical Society* 51(1), 113–127.

Egan, Mark, Hortacsu, Ali, and Gregor Matvos , 2017, Deposit competition and financial fragility: Evidence from the US banking sector, *American Economic Review* 107, 169–216.

Eichengreen, Barry, 1992, *Golden Fetters: The Gold Standard and the Great Depression, 1919–1939.* Oxford University Press.

Eichengreen, Barry, and Jeffrey Sachs, 1985, Exchange rates and economic recovery in the 1930s, *Journal of Economic History* XLV, 925–46.

Farhi, Emmanuel, Golosov, Mike, and Aleh Tsyvinski, 2009, A theory of liquidity an regulation of financial intermediation, *Review of Economic Studies* 76, 973–992.

Farhi, Emmanuel, and Ivan Werning, 2016, A theory of macroprudential policies in the presence of nominal rigidities, *Econometrica* 84, 1645–1704.

Fisher, Irving, The Debt-Deflation Theory of Great Depressions, *Econometrica* 1, 337–57.

Freixas, Xavier, Bruno Parigi, and Jean-Charles Rochet, 2000, Systemic risk, interbank relations, and liquidity provision by the central bank, *Journal of Money, Credit, and Banking* 32(3), 611–638.

French, Kenneth R., Baily, Martin N., Campbell, John Y., Cochrane, John H., Diamond, Douglas W., Duffie, Darrell, Kashyap, Anil K., Mishkin, Frederic S., Rajan, Raghuram G., Scharfstein, David S., Shiller, Robert J., Shin, Hyun Song, Slaughter, Matthew J., Stein, Jeremy C. and René M. Stulz, 2010, *Squam Lake Report*, Princeton University Press.

Friedman, Milton, and Anna Schwartz, 1963, *A Monetary History of the United States 1867–1960,* Princeton University Press.

Gale, Douglas, and Martin Hellwig, 1985, Incentive-compatible debt contracts: The one-period problem, *Review of Economic Studies* 52:4, 647–663.

Geanakoplos, John, 2010, The leverage cycle, *NBER Macroeconomics Annual* 24, 1–66.

Gersbach, Hans, and Jean Rochet, 2012, Aggregate investment externalities and macroprudential regulation, *Journal of Money, Credit and Banking* 44, 73–109.

Gertler, Mark, and Simon Gilchrist, 1994, Monetary policy, business cycles, and the behavior of small manufacturing firms, *Quarterly Journal of Economics* 109(2), 309–340.

Gertler, Mark, and Simon Gilchrist, 2018, What happened: Financial factors in the great recession, *Journal of Economic Perspectives* 32, 3–30.

Gertler, Mark, Kiyotaki, Nobuhiro, and Andrea Prestipino, 2020, Macroprudential policy, *American Economics Association Papers and Proceedings* 110, 463–469.

Goldstein, Itay, 2013, Empirical literature on financial crises: Fundamentals vs. panic, Chapter 36 in eds., The Evidence and Impact of Financial Globalization, (Elsevier), 523–533.

Goldstein, Itay, and Ady Pauzner, 2005a, Contagious speculative self-fulfilling financial crises due to diversification of investment portfolios, *Journal of Economic Theory* 119, 151–183.

Goldstein, Itay, and Ady Pauzner, 2005b, Demand-deposit contracts and the probability of bank runs, *Journal of Finance* 60, 1293–1327.

Goodhart, Charles, 2010, The changing role of central banks, Bank of International Settlements Working Paper No. 326.

Gorton, Gary, 1988, Banking panics and business cycles, *Oxford Economic Papers* 40, 751–781.

Gorton, Gary, and Andrew Metrick, 2009, Securitized banking and the run on repo, NBER Working Paper No. 15223.

Gorton, Gary,and George Pennachi, 1990, Financial intermediaries and liquidity creation, *Journal of Finance* 45, 49–72.

Gorton, Gary, and George Pennachi , 1995, Bank and loan sales: Marketing non-marketable assets, *Journal of Monetary Economics* 35, 389–411.

Gorton, Gary, and Andrew Winton, 2003, Chapter 8 in financial intermediation, in Milton Harris, George M. Constantinides, and Rene M. Stulz, eds.: *Handbook of the Economics of Finance: Corporate Finance*, 431–552 (Elsevier, Amsterdam, The Netherlands).

Goodhart, Charles. A. E.,1975, *Money, Information, and Uncertainty* Ch. 6, MacMillan.

Grossman, Richard, 1992, Deposit insurance, regulation, and moral hazard in the thrift industry: Evidence from the 1930's, *American Economic Review* 82, 800–822.

Grossman, Sanford, and Oliver Hart, 1982 , Corporate financial structure and managerial incentives, in *The Economics of Information and Uncertainty*, McCall, J.J. (ed.), University of Chicago Press.

Gurley, John, and Edward Shaw, 1960, *Money in a theory of finance.* The Brookings Institution.

Hazlett, Denise, 1995, Deposit insurance and regulation in a Diamond-Dybvig banking model with a risky technology, *Economic Theory* 9(3), 453–470.

He, Zhiguo, and Arvind Krishnamurthy, 2012, A model of capital and crises, *Review of Economic Studies* 79, 735–777.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103(2), 732–770.

He, Zhiguo, and Wei Xiong, 2012, Dynamic debt runs, *Review of Financial Studies* 25(6), 1799–1843.

Hellmann, Thomas, Kevin Murdock, and Joseph Stiglitz, 2000, Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review* 90(1), 147–165.

Hellwig, Martin, 1994, Liquidity provision, banking, and the allocation of interest rate risk, *European Economic Review* 38, 1363–1389.

Holmström, Bengt, 2015, Understanding the role of debt in the financial system, Working Paper no. 479, Bank of International Settlements.

Holmström, Bengt, and Jean Tirole, 1997, Financial intermediation, loanable funds, and the real sector, *Quarterly Journal of Economics* 112(3), 663–691.

Ivashina, Victoria, and David Scharfstein, 2010, Bank lending during the financial crisis of 2008, *Review of Financial Studies* 97, 319–338.

Iyer, Rajkamal, and Manju Puri, 2012, Understanding bank runs: The importance of depositor-bank relationships and networks, *American Economic Review* 102, 1414–1445.

Iyer, Rajkamal, Lopes, Samuel, Peydró, José-Luis, and Antoinette Schoar, 2014, Interbank liquidity crunch and the firm credit crunch: Evidence from the 2007–2009 crisis, *Review of Financial Studies* 27, 347–372.

Jacklin, Charles, 1987, Demand deposits, trading restrictions, and risk-sharing, in Edward Prescott and Neil Wallace eds., *Contractural Arrangements for Intertemporal Trade* (University of Minnesota Press; Minneapolis), 26–47.

Jacklin, Charles, and Sudipto Bhattacharya, 1988, Distinguishing panics and information-based bank runs: Welfare and policy implications, *Journal of Political Economy* 96(3), 568–592.

Jaffee, Dwight, and Thomas Russell, 1976, Imperfect information and credit rationing, *Quarterly Journal of Economics* 90, 651–66.

James, Christopher, 1995, When do banks trade equity? An analysis of bank loan restructurings and the role of public debt, *Review of Financial Studies* 8, 1209–1234.

James, Christopher, 1996, Bank debt restructurings and the composition of exchange offers in financial distress, *Journal of Finance* 51, 711–727.

Jaynes, G.D., 1976, *Essays in the Economics of Imperfect Information*, PhD Dissertation, University of Illinois – Urbana-Champaign.

Jaynes, G.D., 1978, Equilibria in monopolistically competitive insurance markets, *Journal of Economic Theory* 19, 394–422. (Received March 4, 1977).

Jeasakul, Phakawa, Lim, Cheng Hoon, and Erick Lundback, 2014, Why was Asia resilient? Lessons from the past and for the future, IMF Working Paper 14/38.

Jensen, Michael C., and William H. Meckling, 1976, Theory of the firm: managerial behavior, agency costs and ownership structure*, Journal of Financial Economics* 3, 305–360.

Jensen, Thais Lærkholm, and Niels Johannesen, 2017, The consumption effects of the 2007 & 2008 Financial Crisis: Evidence from households in Denmark, *American Economic Review* 107(11) 3386–3414.

Jimenez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina, 2012, Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications, *American Economic Review* 102(5), 2301–2326.

Kacperczyk, Martin, and Philipp Schnabl, 2013, How safe are money-market funds? *Quarterly Journal of Economics* 128(3), 1073–1122.

Kashyap, Anil, Raghuram Rajan, and Jeremy Stein, 2001, Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking, *Journal of Finance* 57(1), 33–73.

Kashyap, Anil, and Jeremy Stein, 1995, The impact of monetary policy on bank balance sheets, *Carnegie–Rochester Conference Series on Public Policy* 42, 151–195.

Kashyap, Anil, Jeremy Stein, and David Wilcox, 1993, Monetary policy and credit conditions: Evidence from the composition of external finance, *American Economic Review* 83(1), 78–98.

Kaufman, George G., 1973, *Money The Financial System and The Economy*, Ch. 4, Rand McNally.

Kelley, Morgan and Cormac O Grada, 2000, Market contagion: Evidence from the panics of 1854 and 1857, *American Economic Review* 90(5), 1110–1124.

Keynes, John Maynard, 1936, *The General Theory of Employment, Interest, and Money.* New York: Harcourt, Brace.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, Did securitization lead to lax screening? Evidence from subprime loans, *Quarterly Journal of Economics* 125, 307–362.

Keys, Benjamin, Amit Seru, and Vikrant Vig, 2012, Lender screening and the role of securitization: Evidence from prime and subprime mortgage markets, *Review of Financial Studies* 25(7), 2071–2108.

Khwaja, Asim I., and Atif Mian, 2008, Tracing the impact of bank liquidity shocks: Evidence from an emerging market, *American Economic Review* 98, 1413–1442.

Kindleberger, Charles P., 1973, *The World in Depression 1929–39*. University of California Press.

Kindleberger, Charles P., 1993, *A Financial History of Western Europe,* 2nd Edition. Oxford University Press.

Kiyotaki, Nobuhiro, and John Moore, 1997, Credit Cycles, *Journal of Political Economy* 105(2), 211–48.

Kocherlakota, Narayana, 2000, Creating business cycles through credit constraints, *Federal Reserve Bank of Minneapolis Quarterly Review* 24(3), 2–10.

Leland, Hayne E., and David H. Pyle, 1977, Informational asymmetries, financial structure, and financial intermediation, *Journal of Finance* 32, 371–387.

Lindgren, Carl-Johan, Gilliam Garcia, and Matthew Saal, 1996, *Bank Soundness and Macroeconomic Policy* (Washington: International Monetary Fund).

Marcus, Alan, 1984, Deregulation and bank financial policy, *Journal of Banking and Finance* 8, 557–565.

Marcus, Alan, and Israel Shaked, 1984, The valuation of FDIC deposit insurance using option pricing estimates, *Journal of Money, Credit and Banking* 16, 446–460.

Martinez-Peria, M.S., and S.I. Schmukler, 2001, Do depositors punish banks for bad behavior? Market discipline, deposit insurance, and banking crises, *Journal of Finance* 56, 1029–1051.

Mian, Atif, and Amir Sufi, 2016, Who bears the cost of recessions? The role of house prices and household debt. Chapter 5 of *Handbook of Macroeconomics* (Elsevier), vol 2A, 255–296.

Minsky, Hyman P., 1964, Longer waves in financial relations: Financial factors in the more severe fepressions, *American Economic Review* 54, 324–35.

Modigliani, Franco, and Merton H. Miller, 1958, The cost of capital, corporation finance and the theory of investment, *American Economic Revi*ew 48, 261–297.

Miller, Merton, and Franco Modigliani, 1961, Dividend policy, growth, and the valuation of shares, *Journal of Business* 34(4), 411–433.

Morris, Stephen, and Hyun Song Shin, 1998, Unique equilibrium in a model of self-fulfilling currency attacks, *American Economic Review* 88, 587–597.

Morris, Stephen, and Hyun Song Shin, 2000, Rethinking multiple equilibria in macroeconomic modeling, *NBER Macroeconomics Annual*.

O Grada, Cormac, and Eugene N. White, 2003, The panics of 1854 and 1857: A view from the emigrant industrial savings bank, *Journal of Economic History* 63, 213–240.

Parlour, Christine, and Guillaume Plantin, 2008, Loan sales and relationship banking, *Journal of Finance* 58, 1291–1314.

Patinkin, Don, 1961, Financial intermediaries and the logical structure of monetary theory: A review, *American Economic Review* 51(1), 95–116.

Pennacchi, George, 1987, A reexamination of the over- (or under-) pricing of deposit insurance, *Journal of Money, Credit and Banking* 19, 340–360.

Petersen, Mitchell, and Raghuram G. Rajan, 1994, The benefits of lending relationships: Evidence from small business data, *Journal of Finance* 49, 3–37.

Petersen, Mitchell, and Raghuram G. Rajan, 1995, The effect of credit market competition on lending relationships, *Quarterly Journal of Economics* 110, 407–443.

Poszar, Zoltan, Tobias Adrian, Adam Ashcraft, and Hayley Boesky, 2010, Shadow banking, Federal Reserve Bank of New York Staff Report No. 458.

Rajan, Raghuram, 1992, Insiders and outsiders: The choice between informed and arms-length debt, *Journal of Finance* 47(4), 1367–1400.

Reinhart, Carmen, and Kenneth Rogoff, 2009, *This time is different: Eight centuries of financial folly,* Princeton University Press.

Reinhart, Carmen, and Kenneth Rogoff, 2014, Recovery from financial crisis: Evidence from 100 Episodes, *American Economic Review* 104(5), 50–55.

Rockoff, Hugh, 1993, The meaning of money in the Great Depression. National Bureau of Economic Research (Cambridge, MA) Historical Paper No. 52.

Rothermund, Dietmar, 1996, *The Global Impact of the Great Depression 1929–1939*. London and New York: Routledge.

Sargent, Thomas, and Neil Wallace, 1982, The real bills doctrine versus the quantity theory: A reconsideration, *Journal of Political Economy* 90(6), 1212–1236.

Schmidt, Lawrence, Allan Timmermann, and Russ Wermers, 2018, Runs on money market mutual funds, *American Economic Review* 106(9), 2625–2657.

Schumacher, L., 2000, Bank runs and currency run in a system without a safety net: Argentina and the 'tequila' shock, *Journal of Monetary Economics* 46, 257–277.

Sharpe, Steven, 1990, Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships, *Journal of Finance* 45(4), 1069–1087.

Smith, Bruce, 1991, Bank panics, suspensions, and geography: Some notes on the "contagion fear" in banking, *Economic Inquiry* 29(2), 230–248.

Stiglitz, Joseph E. and Weiss, Andrew, 1981, Credit rationing in markets with imperfect information, *American Economic Review* 71, 393–410.

Temin, Peter, 1989, *Lessons from the Great Depression*. Cambridge, MA: MIT Press.

Townsend, Robert, 1979, Optimal contracts and competitive markets with costly state verification, *Journal of Economic Theory* 21, 265–293.

Wallace, Neil, 1988, Another attempt to explain an illiquid banking system: The Diamond-Dybvig model with sequential service taken seriously, *Federal Reserve Bank Minneapolis Quarterly Review* 12, 2–16.

Wallace, Neil, 1996, Narrow banking meets the Diamond-Dybvig model, *Federal Reserve Bank Minneapolis Quarterly Review* 20(1), 3–13.

Wang, Chunyang, 2013, Bailouts and bank runs: Theory and evidence from TARP, *European Economic Review* 64, 169–180.

Wheelock, David, 1992a, Regulation and bank failures: New evidence from the agricultural collapse of the 1920`s, *Journal of Economic History* 52, 806–825.

Wheelock, David, 1992b, Deposit insurance and bank failures: New evidence from the 1920s, *Economic Inquiry* 30, 530–543.

Wheelock, David, and Subal Kumbhakar, 1995, Which banks choose deposit insurance? Evidence of adverse selection and moral hazard in a voluntary insurance system, *Journal of Money, Credit and Banking* 27, 186–201.

Wheelock, David, and Paul Wilson, 1995, Explaining bank failures: Deposit insurance, regulation, and efficiency, *Review of Economics and Statistics* 77, 689–700.

Wicksell, Knut, 1898, *Geldzins and Güterpreise*, Gustav Fisher publishers, Jena.

Williamson, Stephen, 1986, Costly monitoring, financial intermediation, and equilibrium credit rationing, *Journal of Monetary Economics* 18, 159–179.

Williamson, Stephen, 1987, Financial intermediation, business failures, and real business cycles, *Journal of Political Economy* 95(6), 1196–1216.

Winton, Andrew, 1995a, Delegated monitoring and bank structure in a finite economy, *Journal of Financial Intermediation* 4:2 (April), 158–187.

Winton, Andrew, 1995b, Costly state verification and multiple investors: The role of seniority, *Review of Financial Studies* 8:1 (Spring), 91–123.

Woodford, Michael, 2003. *Interest and Prices: Foundation of a Theory of Monetary Policy.* Princeton University Press.

Yellen, Janet, 2013, Interconnectedness and Systemic Risk: Lessons from the Financial Crisis and Policy Implications, Speech by Vice Chair of the Federal Reserve Board of Governors, https://www.federalreserve.gov/newsevents/speech/yellen20130104a.htm