Annual Review of Financial Economics

Financial Architecture and Financial Stability

Franklin Allen and Ansgar Walther

Finance Department, Business School, Imperial College London, London SW7 2AZ, United Kingdom; email: f.allen@imperial.ac.uk

Annu. Rev. Financ. Econ. 2021. 13:129–51

First published as a Review in Advance on August 2, 2021

The *Annual Review of Financial Economics* is online at financial.annualreviews.org

https://doi.org/10.1146/annurev-financial-110217-022851

JEL codes: G01, G21, G23

Keywords

banks, non-bank intermediaries, financial crises, networks, contagion, regulation

Abstract

This article studies the links between financial stability and the architecture of financial systems. We review the existing literature and provide organizing frameworks for analyzing three empirically important aspects of financial architecture: the rise of nonbank financial intermediaries, the regulatory response to these structural changes, and the emergence of complex interbank networks. One of our main new results is a necessary and sufficient condition for whether nonbank intermediaries are immune to runs in an extended version of the Diamond–Dybvig model.

1. INTRODUCTION

1.1. Motivation and Overview

The global financial crisis of 2007–2009 (GFC) and the Eurozone crisis of 2010–2012 have renewed interest in financial stability within academic and policy circles. One dimension of novel economic thinking is the study of different and evolving financial architectures. Particularly salient in the GFC was the role of shadow banks and other types of nonbank intermediaries, whose activities in the United States had grown rapidly in the years preceding the crash. Interbank networks with complex networks of exposure came into focus after the failure of Lehman Brothers.

In this article, we survey the literature that relates these themes to financial stability. We also aim to provide new frameworks that are geared toward organizing the main economic insights. In the latter part of this introduction, we begin with a brief review of the significant international differences in financial architectures as well as some historical perspectives on the link between financial architecture and stability. We organize the remainder of the article into three parts.

First, we discuss the role of nonbank intermediaries in financial stability. We focus our review on the concern that nonbank intermediaries, such as shadow banks and mutual funds, may be subject to self-fulfilling runs by investors, in a similar sense to Diamond & Dybvig's (1983) model of bank runs. To organize ideas, we develop a simple theoretical framework that contains the core ideas of the Diamond–Dybvig model but also allows for a wide range of nonbank intermediation contracts, such as mutual funds or money market funds. We derive a necessary and sufficient condition for intermediaries to be run-proof, in a sense that has a natural definition. The theory highlights a special role for mutual funds with a floating net asset value (NAV): A floating NAV run in its purest form is run-proof but also sits on a knife-edge, whereby small frictions can bring back vulnerability. We survey the growing theoretical and empirical literatures on nonbank runs in the context of this central idea. In particular, extensive empirical research has shown that real-world business models deviate from the floating NAV ideal and can therefore become unstable.

In the second part of the article, we consider the question of how a diverse financial architecture should be regulated. Shadow banks, mutual funds, and other nonbank intermediaries are not subject to the same prudential regulations as banks, but recent research has shown that they can impose similar externalities on the rest of the economy in a crisis. If regulators could easily measure these effects and avoid the emergence of new unregulated intermediaries, then the best policy response would be simple and follows Pigou's polluter pays principle: Any institution or activity imposing externalities should be brought under the regulatory umbrella. We focus instead on second-best regulation, in a world where not all institutions or activities can be regulated. In that setting, a trade-off exists between imposing strict regulations on the regulated sector, on the one hand, and being more lenient to prevent excessive leakage to the unregulated sector, on the other hand. We provide an overview of the related theoretical literature and argue that leakage elasticities across regulated and unregulated institutions are key statistics that determine how regulation should be conducted in a second-best world. We also review recent empirical work that has carefully measured these statistics and can inform policy makers.

The third and final part of our review focuses on interbank networks. The early theoretical literature following Allen & Gale (2000) focused on contagion between banks—whereby shocks to one financial institution can lead to distress in others—as an equilibrium phenomenon, with endogenous financial contracts. After their paper, new methods were used to yield more general characterizations of vulnerabilities and contagion. We discuss one instructive example at length. Some papers in the last few years have also attempted a synthesis, namely, general characterizations with endogenous contracts. We close the article by reviewing the related empirical literature and the measures of vulnerability that have been extracted from various data sets on interbank connections.

1.2. International Comparisons and Historical Review

Financial architecture in terms of the role of financial markets and banks differs significantly across countries. **Figure 1*a*** provides evidence of this in terms of total value of financial claims for banks and stock and private bond markets. It can be seen that the United States is very different from the other countries shown. In the United States, the stock market is significantly larger and the

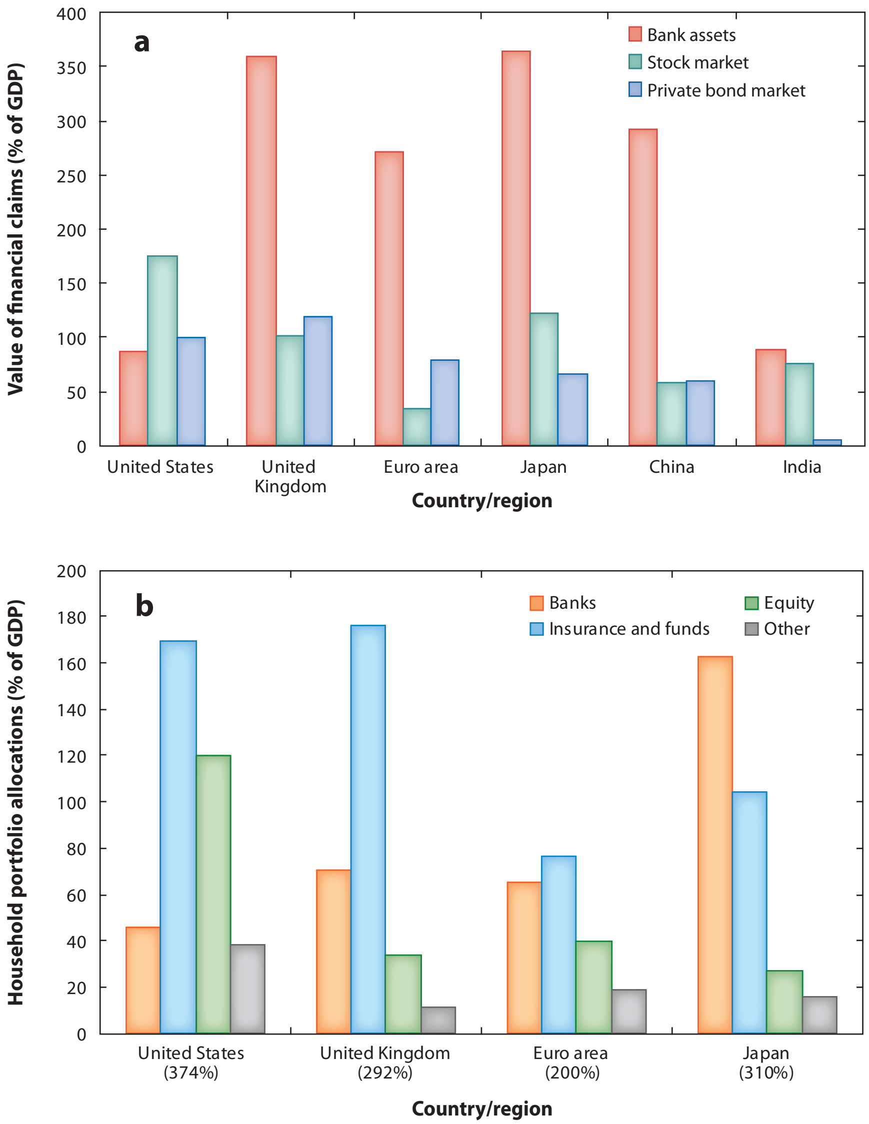

Figure 1

(*a*) Value of financial claims in 2019, percentage of GDP. (*b*) Household portfolio allocations 2000–2019 and total investments in parentheses, percentage of GDP. Panel *a* data are from the World Bank and national authorities. Panel *b* data are from the Bank of Japan, EUROSTAT, Federal Reserve Board, and the IMF. We are grateful to Michael Chui for obtaining the data for these figures.

banking sector significantly smaller than elsewhere (except for India, which has a smaller financial sector). The private sector bond market is also important, being slightly larger in terms of asset values than the banks. It has a market-based financial system. The United Kingdom is also different, with a very large banking sector and an important stock market and private bond market. One of the reasons for this large banking sector is that London is a large international financial center. In the Euro area, banks are significantly more important than the stock market and private bond market. In fact, the stock market is the smallest relative to GDP compared to any of the other places. It has a bank-based financial system. Japan is also bank based but has a stock market that is more important than the Euro area. China is also very much bank based, although over time the stock market and private bond market are steadily becoming more significant. India has a much smaller financial system than the other places, but interestingly, the value of the stock markets is only slightly smaller than that of bank assets.

Figure 1*b* documents the differences in an alternative way, focusing on household portfolio allocations. The United States is again significantly different. Much more equity is held directly by households and much less in banks than elsewhere. Insurance company policies and mutual and pension funds are also very important. The United Kingdom is similar to the United States but with much less in terms of direct ownership of equity. Also, the total amount invested is significantly less, with 292% of GDP in the United Kingdom versus 374% in the United States. The Euro area is also an outlier. It is much smaller in terms of the total resources of 200% of GDP households have in terms of financial investments. The composition has much less in funds and equity than the United States and the United Kingdom. Japan lies between the United States and United Kingdom at 310% of GDP in terms of the total of financial assets held by households. Investors there also have significantly more in banks than in the United States, United Kingdom, or Euro area. These figures, and our literature review below, primarily focus on characteristics of private financial institutions. However, government-owned institutions are important in some countries (e.g., the Japanese government owns a majority stake in the Japan Postal Bank, which is included in the data).

In addition to the differences in terms of the amount financial institutions and markets use in different countries, interbank markets also operate in significantly different ways. Allen et al. (2021) point out that in the United States and Japan interbank markets are small, while in France and Germany they are large, with the United Kingdom being in between. These differences are illustrated in **Figures 2*a*** and **2*b***. **Figure 2*a*** shows bank assets in the different countries, while **Figure 2*b*** shows bank liabilities. It can be seen that the assets and liabilities roughly match in most countries, particularly in France and Germany.

The GFC and the Eurozone crisis of 2010–2012 have underlined the importance of financial stability. However, as Reinhart & Rogoff (2009) have emphasized, financial instability has been an important factor in the operation of financial systems for many centuries.

Prior to the twentieth century, banking panics occurred frequently. Kindleberger (1993, p. 264) points out that in Western Europe financial crises occurred at roughly 10-year intervals over the last 400 years. Over time, one of the most important roles of central banks came to be ensuring financial stability in the banking system. The Bank of England played a particularly important role in the development of effective stabilization policies in the eighteenth and nineteenth centuries. There were no major banking panics in the United Kingdom between the Overend Gurney crisis of 1866 and the collapse of Northern Rock in 2007 in the early stage of the GFC.

As Reinhart & Rogoff (2009) point out, it is not just banks that have been important as a source of financial instability. Historically, another important aspect of financial architecture has been the exchange rate system and the potential it creates for currency crises and sovereign debt

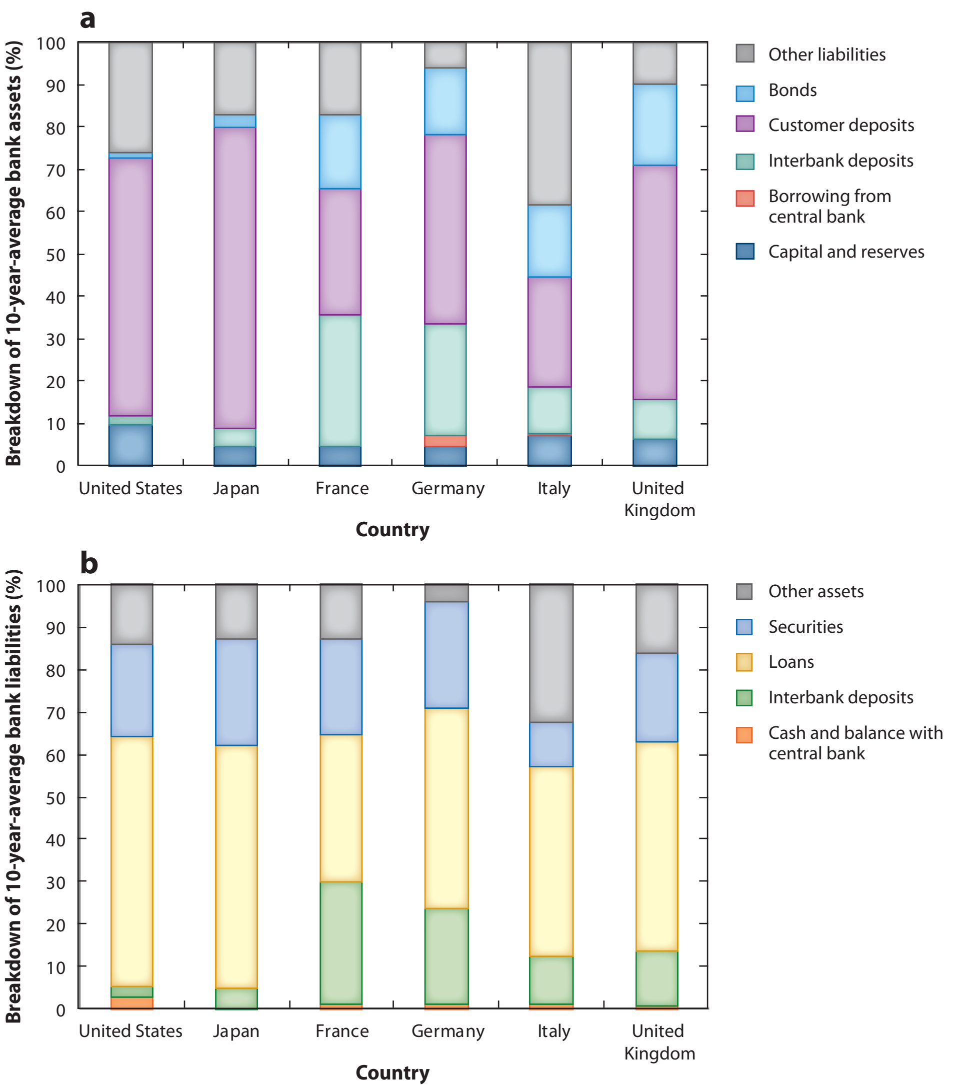

Figure 2

(*a*) Breakdown of 10-year-average bank assets, 2000–2009. (*b*) Breakdown of 10-year-average bank liabilities, 2000–2009. For both panels, data for France, Germany, Italy, the United Kingdom, and the United States are from OECD Statistics (**http://stats.oecd.org/Index.aspx?DataSetCode=BPF1#**), and data for Japan are from the Japanese Bankers Association (**https://www.zenginkyo.or.jp/en/stats/year2-01/**).

default. Stock market crashes have also contributed to financial instability, as with the Great Crash of 1929 in the United States. In this article, we focus on the literature on banks versus nonbank intermediaries and the nature of banking networks, as this has been the area where the theoretical and empirical literatures have developed the most recently.

The next section focuses on runs on nonbank intermediaries, while Section 3 considers financial regulation, architecture, and stability. Section 4 is concerned with banking networks and contagion, and Section 5 contains concluding remarks.

2. RUNS ON NONBANK FINANCIAL INTERMEDIARIES

A striking trend since the GFC has been the growth of nonbank financial intermediaries, in particular, of asset management firms. The data in **Figure 3**, compiled by the Investment Company Institute, show that the global assets of open-ended funds more than doubled between 2009 and 2020 and now account for approximately €50 trillion. This trend reflects the growing complexity of the global financial system, in which asset managers and levered shadow banking institutions conduct a large share of intermediation.

In this section, we consider two broad sets of questions. First, we consider whether nonbank financial intermediaries are likely to suffer from runs. To organize the existing work on this topic, we present a generalized version of the Diamond & Dybvig (1983) model. In this model, we derive necessary and sufficient conditions for vulnerability to runs when intermediaries (e.g., asset managers or money market mutual funds) face sudden withdrawals of short-term funding. This result informs our review of the growing empirical literature on nonbank runs. In practice, a trade-off also exists between guaranteeing financial stability, on the one hand, and avoiding excessive regulation or financial repression that could hamper economic growth, on the other hand. Loayza et al. (2018) provide an excellent review of the related empirical evidence and conclude that strong evidence is in favor of a growth-stability trade-off, the implications of which may be different in advanced and middle-income economies. Given limited space in this article, our formal analysis mostly emphasizes the conditions that ensure stability in a system with nonbank intermediaries.

Second, in Section 2.2, we discuss an active theoretical and empirical literature on the problem of leakage in the financial system, whereby more stringent prudential policies in the regulated segment of the system lead to increased activity by unregulated (or less strictly regulated) intermediaries.

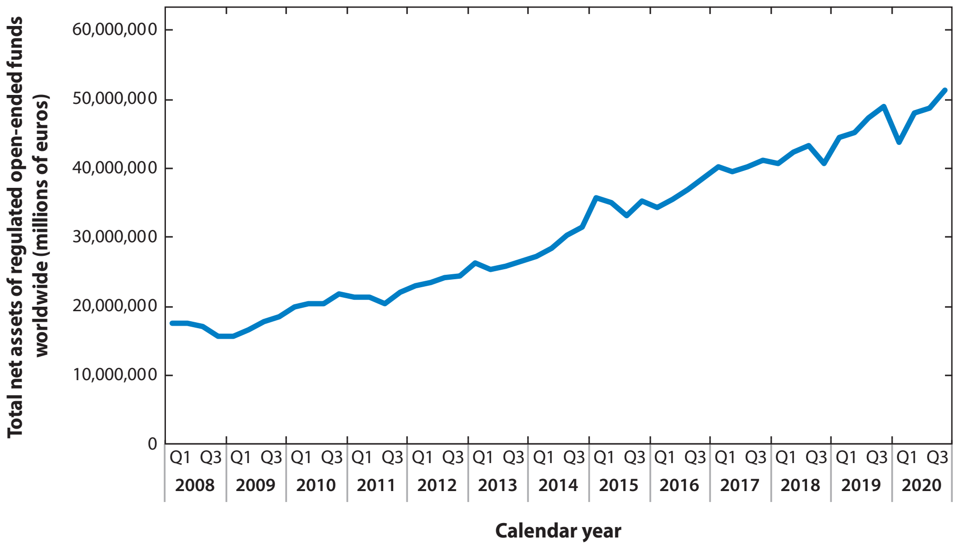

Figure 3

Global growth of nonbank financial intermediation. Data from the International Investment Funds Association (see the Worldwide Public Tables, available at **https://iifa.ca/page/industry_statistics**).

2.1. A Model of Nonbank Runs

We analyze a canonical model that suggests a simple test for stability of nonbank intermediaries with arbitrary architectures. This model is complementary to a growing theoretical literature that studies the stability of particular business models of nonbank intermediaries, such as mutual funds, in more detail (see, e.g., Chen, Goldstein & Jiang 2010; Morris, Shin & Shim 2017; Zeng 2017; Grochulski & Zhang 2019 and references therein). We discuss the relationship of our model to the wider literature at the end of Section 2.1.

2.1.1. Model environment. The environment in this section is akin to the classic model in Diamond & Dybvig (1983). In their analysis, the key economics arise from strategic complementarities among investors: If all other depositors are withdrawing from a bank, any individual has a strong incentive to also withdraw before the bank runs out of cash. As such, their model demonstrates that illiquid investment paired with demandable deposits can trigger instability and bank runs. We consider a similar setting, with illiquid assets, but allow demandable contracts to take a more flexible shape, so that we can analyze the stability of banks and nonbank intermediaries in a unified framework.

The model has three dates indexed by $t = 0, 1, 2$ and a single consumption good. A unit measure of households is born at date 0 with an endowment of one unit of consumption. At date 1, each household learns whether it is an early or a late consumer. With probability λ, each household is an early consumer and enjoys utility $U(c_1)$ from consumption at date 1 only. With probability $1 - \lambda$, it is a late consumer with utility $U(c_2)$ from consumption at date 2.

At date 0, households can invest in a long-term asset. Each unit invested yields R units of consumption at date 2. Long-term assets can be liquidated at date 1 only by selling them to outsider investors. We do not explicitly model the behavior of outsiders but represent them by their demand curve for long assets. When households sell S units of the long assets in total at date 1, the liquidation value of long assets at that date is $P = p(S)$, where $p(.)$ is a decreasing function of sales and satisfies $p(S) \leq R$ for all S. This formulation allows us to capture different kinds of illiquidity in long-term investments. For example, if $p(S) \equiv r$, then the long asset has a fixed liquidation value, as is commonly assumed in the literature on banking.

If $p(S) = R - \beta \cdot S$, then the parameter β measures the depth of the market into which tradeable assets can be sold and is analogous to Kyle's lambda (Kyle 1985). The latter case allows a better description of nonbank intermediaries like mutual funds, who hold assets that can be traded in markets of varying liquidity. Our formalism can therefore capture the creation of price pressure when there are economy-wide early sales. Idiosyncratic sales by individual consumers, by contrast, do not have any price effects because consumers are infinitesimal relative to the aggregate sale S.

Our focus is on a key ingredient of financial architecture, namely, the design of financial intermediaries that issue demandable claims, which allow customers to withdraw funds early. Different kinds of intermediaries handle this problem very differently in the real world—for example, banks tend to offer fixed deposit claims, money market funds offer demandable claims with fixed NAV, and mutual funds allow their NAV to float. We argue that the devil is in the detail: The design of demandable claims is an important determinant of financial stability. We also review additional theoretical literature that highlights other important distinctions between different kinds of intermediary, such as the fact that banks tend to be more leveraged than funds.

2.1.2. Efficient allocation. We first consider constrained efficient allocations. This term refers to allocations chosen by a benevolent social planner who can dictate consumption and investment decisions to households, but who cannot force outsiders to buy assets at a price that is not on their demand curve. The constrained planning problem is:

$$\max_{c_1,c_2,S,P} \lambda U(c_1) + (1-\lambda)\, U(c_2)\,,$$

subject to the budget constraints

$$\begin{aligned}\lambda c_1 &= PS \\ (1-\lambda)\, c_2 &= R\,(1-S)\end{aligned}$$

and the pricing constraint

$$P = p(S).$$

Substituting the budget constraints into the objective and taking the first-order condition for S gives us

$$U'(c_1)P + \mu \frac{\partial p}{\partial S} = U'(c_2)R,$$

where μ denotes the Lagrange multiplier on the pricing constraint. The first-order condition for P is

$$U'(c_1)S = \mu.$$

Combining, we obtain a condition that (along with resource constraints) pins down the efficient allocation $(c_1^\star, c_2^\star)$:

$$U'(c_1^\star)\left(1-\eta_p\right) = \frac{R}{P} U'(c_2^\star),$$

where

$$\eta_p = \left|\frac{\partial \log p\,(S)}{\partial \log S}\right|$$

is the elasticity of outsider's demand. This expression captures the classic trade-off between asset liquidity and consumption insurance. Indeed, when the early liquidation payoff P is a constant, then outsiders' demand is inelastic and the first-order condition above reduces to the standard risk-sharing equation $U'(c_1) = \frac{R}{P} U'(c_2)$: The marginal rate of substitution between periods 1 and 2 is set equal to the rate of return $\frac{R}{P}$ on the long asset between dates 1 and 2. When the demand is more elastic, then the planner distorts c_1 downward because he internalizes the adverse price impact of early liquidation.

2.1.3. Financial architecture with intermediaries. One way for households to implement the efficient allocation is by setting up an intermediary, which operates as follows. The intermediary collects everybody's endowment at date 0 and invests it in the long asset. It then promises consumption c_1 to any household that asks to withdraw its claim early (at date 1) and consumption c_2 to any household that withdraws late (at date 2). At date 1, the intermediary liquidates an appropriate measure S of the underlying long asset for $p(S)$ per unit to honor its promise to early withdrawers. An intermediary is feasible if (c_1, c_2, S) satisfy the budget constraints in the previous section. The efficient allocation is feasible by definition. Note also that the efficient allocation always satisfies

$$c_1^\star \leq c_2^\star.$$

Hence, the intermediary solution is incentive compatible: Early households will withdraw c_1 at date 1, while late households find it optimal to refrain from mimicking early ones and therefore wait to withdraw c_2 at date 2.

The only assertion we have made so far is that an intermediary implements the efficient allocation $(c_1^\star, c_2^\star)$ along the path of some subgame perfect equilibrium. In that equilibrium, all households withdraw in line with their consumption needs, so that the measure of early withdrawals is always exactly λ. However, a practical implementation of this intermediary must also specify what happens off the equilibrium path, that is, the consumption $(c_1(n), c_2(n))$ that is allocated to early and late withdrawers when a general measure $n \neq \lambda$ of households choose to withdraw early. The rigorous version of our assertion is therefore that a feasible intermediation contract $(c_1(n), c_2(n))$ exists such that, on the one hand, $c_1(\lambda) = c_1^\star$ and $c_2(\lambda) = c_2^\star$, and on the other hand, $n = \lambda$ households withdraw early in a subgame perfect equilibrium.

Different models of financial intermediation in the real world share the demandability of claims but diverge when specifying what happens for different n. Bank deposits, for example, offer fixed demandable claims $c_t(n) = D_t$. Shares in an investment fund, by contrast, can typically be liquidated by placing a sell order at date t, which results in a liquidation value per share that is proportional to the fund's NAV at the end of date t. The contract between the fund and investors specifies how NAV is calculated. The standard model in mutual funds is to calculate a floating NAV, which is the value of the fund's total assets at date t after any necessary liquidations have been made. An alternative, common among money market funds before the GFC, is to have a fixed NAV normalized to \$1 per share. The fixed-NAV model makes the contractual arrangements very similar to a banking contract, unless excessive liquidations or bankruptcies force the fund to "break the buck" (i.e., to adjust its NAV below \$1).

Floating-NAV funds are of institutional interest, but as we detail below, they also provide a useful benchmark for financial stability. Therefore, it is worthwhile to spell out what floating NAV means algebraically in our model. If n households withdraw early, the floating NAV of a fund at date 1 is

$$F(n) = p(S(n))S(n) + (1 - S(n))R,$$

where

$$S(n) = \min\left\{S | p\left(S\right) S = nc_1\left(n\right)\right\}$$

is the minimum liquidation required to service withdrawals. So far, we have two equations and four unknowns $(F(n), S(n), c_1(n), c_2(n))$. A floating-NAV fund closes the system by requiring that the payout per share at date 1 is exactly the floating NAV,

$$c_1(n) = F(n),$$

and that the payout per share at date 2 satisfies the resource constraint

$$(1 - n)\, c_2(n) = (1 - S\left(n\right))\, R.$$

It is easy to see that this set of equations implies a unique floating-NAV contract $(c_1(n), c_2(n))$, which moreover has the special property that $c_1(n) = c_2(n)$. This property will be important in the next section, which discusses financial stability.

2.1.4. Financial stability with intermediaries. The fundamental problem of financial stability that we focus on is as follows: While it is possible for intermediaries such as banks to implement the efficient allocation in a subgame perfect equilibrium with $n = \lambda$, it is also possible that another subgame perfect equilibrium with $n \neq \lambda$ exists. For instance, in the Diamond–Dybvig model, the intermediary offers bank deposit claims D_t, where $p(S)S < D_1$ for all $S \in [0, 1]$. If all households decide to withdraw early, then the bank cannot raise enough money to repay them by liquidating assets and will therefore be bankrupt at date 1. Early withdrawers then liquidate the bank and share its remaining value, which is

$$V_1 = \max_S p(S)\,S,$$

while late withdrawers obtain nothing. Hence, if everybody withdraws early, it is optimal to follow, and a bank run with $n = 1$ is another equilibrium. The same issue applies to a fixed-NAV fund.

How, in general, is stability determined? We adopt a robust definition of stability that generalizes our earlier incentive compatibility condition. We say that an intermediary contract $(c_1(n), c_2(n))$ is run-proof if it satisfies

$$c_1(n) \leq c_2(n)$$

for all $n \in [0, 1]$. This condition rules out classic bank runs in which everybody withdraws ($n = 1$) but also less extreme cases with partial inefficient liquidation ($0 < n < 1$).

It turns out that there is a simple necessary and sufficient condition for run-proofness:

Proposition. An intermediary contract $(c_1(n), c_2(n))$ is run-proof if and only if the demandable claim $c_1(n)$ satisfies

$$c_1(n) \leq F(n),$$

where $F(n)$ is the demandable claim of a floating-NAV mutual fund.

The good news is that in the ideal case defined above, a floating-NAV fund is run-proof. This ideal case also provides a natural bound for all other intermediaries. Banks without deposit insurance and money market mutual funds that offer a fixed NAV, for example, are vulnerable to runs because they do not always satisfy our condition for run-proofness.

The bad news is that floating NAV is a knife-edge case. There are a few ways in which small deviations can cause runs on mutual funds. We consider two concrete examples. First, we have assumed that investors are certain that the floating-NAV model is used and that they fully understand it. Suppose instead that investors believe that the NAV at date 1 is fixed at D_1 with probability $\epsilon > 0$ and floating with probability $1 - \epsilon$. Assume that $p(S)S < D_1$ for all $S \in [0, 1]$, so that the fund cannot pay back the fixed NAV if everybody withdraws early. Then if everybody is expected to withdraw early, the net payoff from early withdrawal is

$$c_2(1) - c_1(1) = \begin{cases} 0, & \text{w.pr. } 1 - \epsilon \\ V_1, & \text{w.pr. } \epsilon \end{cases}.$$

The expected incentive to withdraw is $\epsilon V_1 > 0$; therefore, there is a fund run equilibrium where everyone withdraws, even for infinitesimally small ϵ. Chen, Goldstein & Jiang (2010) and Morris, Shin & Shim (2017) both consider deviations from the ideal floating-NAV case along these lines, where early withdrawers are promised a fixed claim.

Second, we have made a stark assumption about asset valuation: At date 1, the fund values its remaining long assets at exactly their date 2 value, which is R per unit. In reality, this continuation value would itself correspond to a market price. If the price effects of forced sales liquidation linger over time, then investors again have potential incentives to run. For simplicity, consider a reduced-form version in which early withdrawals simply reduce the average late return by ϵ per unit, so that the actual asset value at date 2 is

$$R = R - \epsilon S.$$

This formulation can capture lingering price impacts. Zeng (2017) provides another foundation for this assumption, which is based on fund managers' incentive to rebuild their cash balances after outflows. In this model, early withdrawals lead to a predictable decline in NAV. In our reduced form version, claims from late withdrawers satisfy the new budget constraint

$$(1-n)\, c_2(n) = (1 - S(n))\,(R - \epsilon S)\,.$$

Then it is again easy to see that a run equilibrium exists for all $\epsilon > 0$.

Our analysis highlights a simple point: Even an infinitesimal deviation from the ideal case can generate fragility in funds. Having a traditional bank is not a necessary condition for generating runs in the spirit of Diamond & Dybvig (1983).

Further insights on mutual fund runs can be derived by selecting a unique equilibrium of the coordination game among investors. The most popular selection technique is to introduce small deviations from common knowledge in a global game. This criterion implies that investors tend to play the bad equilibrium (i.e., a run on the intermediary) whenever they observe bad news about the fundamental strength of the intermediary. In this context, Morris, Shin & Shim (2017) find that the fund's cash management policy in anticipation of withdrawals is key. If the fund uses cash to smooth out withdrawals, this dampens the strategic complementarity arising from costly liquidations (in our notation, the elasticity η_p is low when the intermediary services redemptions with cash). If, by contrast, the fund hoards cash in anticipation of further outflows, then the strategic complementarity is exacerbated. Also using global games, Chen, Goldstein & Jiang (2010) derive empirical predictions that link the vulnerability to the liquidity of assets and the concentration of fund ownership.

2.1.5. Relationship to the broader theoretical literature. A rich literature develops the theory of classical bank runs. Allen & Gale (2009) provide a more detailed review. On the one hand, papers subsequent to Diamond & Dybvig (1983) considered other architecture aspects such as the information and liquidity structures (e.g., Gorton 1985, 1988; Chari & Jagannathan 1988; Jacklin & Bhattacharya 1988; Allen & Gale 1998; Rochet & Vives 2004; Diamond & Rajan 2005; Goldstein & Pauzner 2005). This classic banking literature focuses on single banks rather than the banking system (e.g., Bryant 1980; Diamond & Dybvig 1983)—but even here architecture mattered, as Cone (1983) and Jacklin (1987) point out that depositors must have restricted access to financial markets. On the other hand, the literature on bank runs advances two additional, complementary perspectives. First, runs and crises can be driven by real events and downturns in the business cycle, in addition to sunspots and self-fulfilling beliefs (e.g., Gorton 1988; Allen & Gale 1998). Second, bank runs give rise to some interesting dynamics (e.g., He & Xiong 2012; Zhong & Zhou 2021). Since the risk of runs is no longer confined to the traditional banking system, these views also have the potential to shed light on the nature of fragility in nonbank intermediaries, although we are unable to review them in detail here.

We have focused on runs on nonbank intermediaries in the broad terms set by Diamond & Dybvig (1983). A growing, complementary literature makes the point that nonbank intermediaries can be fragile even without runs. This literature typically uses macroeconomic models with financial frictions (for a detailed review, see Brunnermeier, Eisenbach & Sannikov 2012). A core idea is that nonbank intermediaries can face binding financial constraints and also increased costs of financing themselves, if they have a limited ability to pledge (safe) assets as collateral in downturns. This effect forces them to sell assets into depressed markets to cover funding shortfalls, which then amplifies the downturn (see, e.g., Gennaioli, Shleifer & Vishny 2013; Luck & Schempp 2014; Moreira & Savov 2017; Ordonez 2018; Martinez-Miera & Repullo 2019). Also related to the business cycle view of financial instability is a recent line of research that studies the impact of large insurance companies on systemic risk (e.g., Koijen & Yogo 2021).

2.2. Empirical Evidence on the Stability of Nonbank Intermediaries

A long literature examines bank runs in the data, and we do not survey this literature in detail here (see, e.g., Gorton 1988; Calomiris & Mason 1997; Iyer & Puri 2012; Iyer, Puri & Ryan 2016 and references therein).

In the GFC, runs affected not just traditional banks but also shadow banks that had financed themselves heavily with securities resembling demandable debt. Schmidt, Timmermann & Wermers (2016) study runs on money market funds that occurred in September 2008 after a major fund (Reserve Primary Fund) "broke the buck." Studying different share classes within the same fund, they argue that runs were more pronounced in funds with more sophisticated investors, which is consistent with the global games solution to a bank run game. Kacperczyk & Schnabl (2013) link these runs to increased incentives for money market funds to take risk in the years before the crisis. Another common business model before 2008 was to finance portfolios of securitized, often housing-related assets, with short-term debt such as repurchase agreements (repo) or asset-backed commercial paper (ABCP). Since short-term debt had to be frequently rolled over, these vehicles were vulnerable to bank runs. Runs on repo and ABCP programs are documented and compared in the work by Gorton & Metrick (2012), Covitz, Liang & Suarez (2013), and Krishnamurthy, Nagel & Orlov (2014), among others. Adrian & Ashcraft (2016) provide a more comprehensive review of the postcrisis literature on shadow banking.

In the rest of this section, we focus on the more contentious question of whether runs and fire sales also affect mutual funds. If a mutual fund is vulnerable to runs, then one might expect that the signals observed by investors, such as the fund's past performance, determine whether or not there is a run. Thus, many theories of runs would imply that fund performance and outflows are negatively correlated. Of course, this correlation does not constitute evidence of runs. For instance, performance-based flows might be an optimal arrangement because they provide incentives for fund managers or because investors learn about managers' skills over time. However, differences in the strength of this correlation across different funds can aid our understanding of runs.

A well-known stylized fact, established for example by Chevalier & Ellison (1997), is that flows respond to performance but that this relationship is nonlinear. For the average fund, positive performance generates significant inflows, while the relation between negative performance and outflows is muted. Thus, flows are convex as a function of performance. At first glance, this seems inconsistent with runs. To test more directly for runs, Chen, Goldstein & Jiang (2010) focus on funds that hold illiquid assets. The relationship between flows and performance is closer to linear for these funds, which is consistent with these funds being more vulnerable to runs. Going further, Goldstein, Jiang & Ng (2017) show that the relationship becomes concave when one focuses on corporate bond funds, which operate in particularly illiquid asset markets. Frazzini & Lamont (2008) argue that mutual fund flows can be interpreted as driven by investor sentiments rather than by fundamental values, consistently with a sunspots view of runs. Taken together, this evidence does suggest performance-related outflows for funds in general, but it also suggests that there are run-related outflows for funds that operate in illiquid asset markets.

A complementary empirical literature tests a core assumption driving fund runs, namely, that funds cannot liquidate their underlying assets at fair value. This looks like a stronger assumption for mutual funds than for banks, because the underlying assets of funds are traded in open markets. Coval & Stafford (2007) define fire sales as situations in which outflow-driven sales by equity funds are concentrated in a small number of securities. These situations are associated with price declines of approximately 8%, which (in contrast to episodes in which funds sell voluntarily) revert when the selling pressure subsides. Edmans, Goldstein & Jiang (2012) find similar effects, while Gompers & Metrick (2001) use the relative prices of small and large company stocks to demonstrate the price impact of large institutional investors. Overall, the evidence suggesting that a mutual fund may be illiquid is quite consistent. Thus, if a run on a large mutual fund or a simultaneous run on many small ones occurs, funds would have to liquidate at a substantial discount, which generates fragility.

As discussed above, whether funds follow a cash-smoothing or a cash-hoarding strategy in the face of outflows is an important ingredient of run incentives. The evidence so far is that both strategies are at play, and in practice, these strategies are not mutually exclusive. Chernenko & Sunderam (2016) study U.S. equity and bond funds in the early twenty-first century. They find that 23–33 cents of every dollar of outflows is accommodated by funds' cash cushions. This finding suggests smoothing and that the effects are stronger in funds that face illiquid conditions in the underlying asset markets. By contrast, Morris, Shin & Shim (2017) find evidence of significant cash hoarding in a sample of global bond funds between 2013 and 2016. They propose an empirical decomposition of fund asset sales into sales driven by investor flows and additional discretionary sales. The latter are suggestive of hoarding and amount to 10 cents for every dollar of flow-driven sales. This evidence is reviewed more carefully by Goldstein (2017). Overall, the mixed findings imply that cash management strategies in practice are a mixture of hoarding and smoothing.

Three further insights from recent empirical work are worth emphasizing. First, the fragility of funds is not confined to open-ended mutual funds but also appears to affect exchange traded funds (e.g., Ben-David, Franzoni & Moussawi 2018). Second, evidence shows that the details of contract design, which we discussed in our model above, are important for stability in practice. For example, Jin et al. (2021) show that a small adjustment to the price at which withdrawals are redeemed, known as swing pricing in open-ended mutual funds, served to eliminate first-mover advantages in UK corporate bond funds. Third, the economic importance of fire sales and fund runs is underlined by their real effects, such as on corporate investment. Significant real effects have been demonstrated by Edmans, Goldstein & Jiang (2012) as well as Hau & Lai (2013) and Dessaint et al. (2019).

3. FINANCIAL REGULATION, ARCHITECTURE, AND STABILITY

In this section, we review a growing theoretical and empirical literature that studies the link between financial policy, financial architecture, and financial stability. Do traditional financial stability policies that focus on traditional banks remain fit for purpose when the financial architecture changes? In what directions should those policies, in principle, be adjusted to account for changes in architecture? These are natural and important questions given the recent rise of nonbank financial intermediaries (see, e.g., **Figure 3**) and given the potential worries about the fragility of nonbanks that we have discussed in the previous section.

A particularly important concern, on which we focus in this review, is the issue of leakage from regulated to unregulated activities. Indeed, many commentators have attributed at least some of the recent rise of shadow banking in the United States (e.g., Buchak et al. 2018a), as well as in China (e.g., Hachem 2018), to the tighter regulatory requirements that the respective regulators have imposed on traditional banks.

It has become customary to refer to this effect, whereby tougher regulation of traditional banks increases the market shares of nonbank intermediaries, as an unintended consequence of financial regulation. This label is somewhat misleading, however, since several strands of recent research have cast a lot of light on the consequences of tougher regulation and how regulators should address them once they are well understood.

3.1. Theories of Financial Regulation with Nonbank Intermediaries

The theory of financial regulation in the presence of leakages has been studied from several different but complementary perspectives. Plantin (2015) was one of the first authors to study the consequences of financial regulation, and the paper shows that tougher capital requirements on the regulated segment of the financial system can increase overall risk because it fuels a rise in unregulated intermediation. Bengui & Bianchi (2018) and Begenau & Landvoigt (2021) explore these effects quantitatively in the context of macroeconomic models.

An important result of this analysis is that leakages are significant, including in realistically calibrated models. However, it is not clear whether this problem calls for weaker or stronger interventions in the regulated segment. One common intuition is that regulators should ease off regulation on traditional banks to avoid pushing activity into the unregulated segment of the system; another intuition is that regulators should crack down on regulated institutions when there is unregulated risk-taking, because only tougher regulation can keep the system as a whole safe. Some subtle differences between different models in this literature could be driving the different conclusions. For example, leakage in some papers occurs because banks themselves engage in unregulated off-balance-sheet activities, so that leakages occur within intermediaries. Other models focus on activities shifting across intermediaries to different, unregulated institutions (see also Chrétien & Lyonnet 2019).

Davila & Walther (2020) show that the design of optimal regulation in each of these cases is driven by the same underlying principles. In this section, we give a brief illustration of their approach, which also guides our review of the relevant evidence in Section 3.2.

Consider an economy where there are two segments of financial intermediary indexed by $i \in \{1, 2\}$, two dates $t \in \{0, 1\}$, and an uncertain state of the world s, which is revealed at date 1. Both types of intermediaries select their volume of risky lending k^i. Davila & Walther (2020) derive microfoundations for the choices of leverage in accordance with some canonical theories in banking and corporate finance. The model can further be extended to include other choice variables, such as the magnitude and allocation of risky investment in different asset classes, or the choice of leverage. However, an important point is that the formulas determining optimal policy do not depend on those details, so we omit them here for brevity.

Regulatory intervention is motivated by the possibility of an ex post bailout (or other distortive policy, such as monetary easing) in states of the world when one or several types of intermediary are troubled. The possibility of a bailout introduces a social cost that intermediaries do not internalize. For simplicity, we write the externality in this review in reduced form as $\Delta(k^1, k^2)$ and note that it depends on the joint activities of both types of intermediary. To address the externality, a regulator can impose a corrective (Pigouvian) tax τ_k^i per unit of leverage on the investment choices of intermediary in segment i. The textbook Pigou principle (i.e., the polluter pays) would suggest that corrective taxes on intermediaries of type i should be set according to

$$\tau_k^i = \frac{\partial \Delta}{\partial k^i} \equiv \delta_k^i.$$

This is the well-known policy whereby corrective taxes are set equal to the marginal Pigouvian distortion δ_k^i, which measures the difference between marginal private and social costs at equilibrium. In other words, the first-best policy imposes restrictions until private and social marginal incentives are aligned.

A more challenging scenario is one in which the regulator faces constraints due to the presence of unregulated intermediaries. For instance, consider the case where intermediaries in segment 2 are unregulated or only partially regulated. To model this, assume that the tax τ_k^2 is fixed at an exogenous level (e.g., at zero, in the case where segment 2 is completely unregulated) but that the regulator is still free to choose any level of taxes τ_k^1 in the regulated segment. In this second-best policy problem, Davila & Walther's (2020) results yield the following necessary condition for optimal policy:

$$\left(\tau_k^1 - \delta_k^1\right) + \left(\tau_k^2 - \delta_k^2\right)\frac{dk^2}{dk^1} = 0.$$

The first term in this equation measures the Pigouvian wedge $(\tau_k^1 - \delta_k^1)$ between private and social marginal costs of leverage in the regulated segment. Indeed, if this wedge is zero, then $\tau_k^1 = \delta_k^1$, so that the then-intermediaries of type 1 are regulated in accordance with the Pigou principle. The second term is the product of the Pigouvian wedge, e.g., $(\tau_k^2 - \delta_k^2)$, between private and social marginal costs in the unregulated segment. Crucially, it is multiplied by the leakage elasticity $\frac{dk^2}{dk^1}$, which quantifies the total impact of a tax reform that increases k^1 on k^2 in general equilibrium (i.e., taking into account all changes in prices and interest rates that the reform brings about).

To understand the formula for second-best policy, consider some concrete examples. On the one hand, if no connection at all between regulated and unregulated intermediaries exists, then the leakage elasticity is zero and the Pigou principle should apply in the regulated sector. On the other hand, the most common case discussed in the literature (e.g., Plantin 2015; Bengui & Bianchi 2018; Begenau & Landvoigt 2021) is when (*a*) the leakage elasticity is positive with $\frac{dk^2}{dk^1} > 0$, meaning that the activities of regulated and unregulated intermediaries are gross substitutes, and (*b*) the Pigouvian wedge in the unregulated segment satisfies $\tau_k^2 - \delta_k^2 < 0$, meaning that social marginal costs of unregulated intermediaries' activities exceed the private costs. In this case, it is easy to see that the optimal policy for regulated intermediaries implies a negative Pigouvian wedge: $\tau_k^1 - \delta_k^1 < 0$. This result implies that the optimal policy is sub-Pigouvian, i.e., that the regulator should stop short of the Pigou principle, imposing less stringent policies than would be needed to align private and social marginal costs. In this sense, under-regulation in one segment of the financial system spills over into optimal under-regulation in other segments.

However, it is important to note that even when negative wedges are optimal, it is possible that the absolute level of the intervention τ_b^1 grows larger when there is imperfect regulation. This is because heightened activity in the unregulated sector can raise the expected marginal social costs δ_b^1 of regulated activities at equilibrium. This scenario applies, for example, when risky lending by shadow banks increases the level of bailout required to restore stability in bad states of the world and thereby also raises the marginal (fiscal) cost of helping traditional, regulated banks. Therefore, the second-best analysis helps to clarify that easing off (in the sense of Pigouvian wedges) and cracking down (in the sense of the stringency of intervention) are both valid intuitions that should guide policy.

A central implication of these arguments is that Pigouvian wedges and leakage elasticities are jointly sufficient statistics for the optimal second-best policy in the presence of nonbank, unregulated intermediaries. Indeed, Davila & Walther (2020) point out this is the case not only in the simple model we have considered here but also in a wider range of settings where there are unregulated activities within a bank (e.g., off-balance-sheet vehicles) or one-size-fits-all policies (e.g., activity-based regulation in short-term borrowing markets, as recently adopted by the Financial Stability Oversight Council in the United States which is discussed in more detail by Kress 2017). This result sheds new light on the active empirical literature that studies shadow banks and leakages, which we review in the next section.

3.2. Empirical Evidence on Regulation and Nonbank Intermediaries

A growing empirical literature casts light on the impact of regulation on financial architecture. In particular, we review a number of papers that estimate leakage elasticities between regulated and unregulated activities. As we saw above, these are crucial statistics for well-calibrated second-best regulation in the presence of unregulated intermediation activities.

Buchak et al. (2018a) study the rise of shadow banking and financial technology (FinTech) lenders in the U.S. mortgage market, arguing that both increased regulatory burdens and technological improvements have contributed to the decline of traditional banks' market share. In particular, they exploit geographical variation across U.S. lenders to estimate leakage elasticities between commercial banks and unregulated intermediaries in mortgage originations. They find robust evidence of substitutability between unregulated and regulated lenders. For example, their estimates suggest that a one percentage point increase in commercial banks' capital requirement increases shadow bank lending volumes by 3.4% and increases shadow banks' market share by a similar 3.3%. Interestingly, the symmetry between the estimated responses of lending volumes and market shares suggests that the total volume of lending remains roughly constant in response to capital requirements—implying near-perfect substitution between traditional and shadow banks. In a related paper, Buchak et al. (2018b) confirm similar numbers in an estimated structural model of regulated and unregulated lending.

Irani et al. (2021) estimate leakage elasticities in the market for syndicated (corporate) lending in the United States. Their results also suggest substitutability. Their study is set against the backdrop of new, tougher capital requirements that were about to be introduced under Basel III. Banks in their sample are heterogeneous in the shortfall of current capital from future required levels. They show that banks with greater shortfalls reduced their retention of syndicated loans, and that this gap was filled partly by nonbank lenders.

This body of work points convincingly to the stylized fact that the activities of regulated and unregulated intermediaries are gross substitutes. Two further recent papers are consistent with this conclusion. First, Xiao (2020) structurally estimates the responses of the market share of traditional banks and shadow banks to monetary policy innovations. Importantly, and in line with other results above, this research finds that the market share of shadow banks increases at times when the market share of traditional banks falls (in this case, when the central bank raises interest rates). Second, Tang (2019) finds that another emergent class of nonbank intermediaries, namely peer-to-peer lending platforms, act as a natural substitute for bank credit.

According to the theory we reviewed above, this pattern of substitutability implies that regulation of traditional banks ought to be sub-Pigouvian, i.e., to stop short of the traditional first-best goal of aligning social and private marginal costs, as long as we believe that risky lending by unregulated intermediaries imposes negative externalities (i.e., distortions $\delta_k^2 > 0$ in our notation above) on the economy.

Many recently developed methodologies suggest that the externalities imposed by risky intermediation are indeed negative across the board (see, e.g., Acharya, Engle & Richardson 2012; Miles, Yang & Marcheggiano 2013; Adrian & Brunnermeier 2016). However, it is not obvious that unregulated intermediation is always undesirable. One relevant case study by Allen et al. (2019) shows that entrusted loans, a common type of shadow lending in China, expands in times when credit is tight overall. Instead of banks lending directly to risky firms, they lend to well-capitalized firms that in turn lend to risky firms. The effect of this is to put an extra buffer of equity in between the risky firm and the bank. This finding suggests that a shift in activity from regulated banks to alternative shadow channels may make the system as a whole more stable in bad times. In that case, shadow lenders can impose a positive externality (i.e., distortions $\delta_k^2 < 0$ in our notation above) on the economy, which means that regulation should be super-Pigouvian and encourage some leakage to the shadow sector.

In summary, a wide body of theory and empirical evidence have helped to make the study of unintended consequences of financial regulation more rigorous. The theoretical equations help us to evaluate whether policy should become tougher or weaker in the presence of unregulated activity. Based on the available evidence, we would conclude that one has to evaluate those equations on a case-by-case basis. Most evidence from the United States suggests that policy should be marginally weaker than the classical Pigou principle suggests, to avoid too much substitution into shadow banking. Conversely, some of the evidence from China suggests that it might sometimes be worth encouraging such substitution, for example, by imposing more stringent restrictions on the regulated segment of the system.

4. BANKING NETWORKS AND CONTAGION

The previous two sections focus on the stability and prevalence of nonbank intermediaries. Another aspect of financial architecture that drives overall financial stability is the structure of networks between banks—for example, the structure of interbank market exposures. Indeed, the issue of financial architecture and stability first came to the fore in the early literature on contagion (e.g., Allen & Gale 2000; Freixas, Parigi & Rochet 2000; Dasgupta 2004). In this section, we review this literature and the more detailed analysis of network structures that has followed it. We then give an explicit theoretical example, based on recent work by Acemoglu, Ozdaglar & Tahbaz-Salehi (2015), that highlights some of the key aspects of banking networks that determine financial stability.

The theoretical literature on banking networks and contagion has evolved in three strands. First, Allen & Gale (2000) propose one of the first formal studies of contagion in banking networks. In their model, the exposures between banks are endogenous and determined in equilibrium. Banks choose to hold claims on each other due to imperfectly correlated regional liquidity shocks. A key result is that contagion, whereby relatively small shocks to one institution cause distress among other institutions, is possible as an equilibrium phenomenon and that complete networks are more robust than concentrated ones. Dasgupta (2004) extends this analysis to dynamic interactions.

Second, many papers have studied the vulnerability of banking networks. This literature typically takes a network of exposures as exogenously given and characterizes the types of shocks (e.g., localized losses) that the network can withstand without generating contagion. In early work in this vein, Freixas, Parigi & Rochet (2000) discuss chain reactions in an interbank market and Eisenberg & Noe (2001) show that even unsystematic negative shocks can reduce the value of financial firms that are exposed to each other via a clearing mechanism. This literature was revived after the GFC. Gai & Kapadia (2010) point out an important property of financial networks that they call robust-yet-fragile: A network that can withstand shocks in good times is not necessarily also a network that avoids vulnerabilities in bad times. Elliott, Golub & Jackson (2014) study the effects of financial integration (more exposure to counterparties) and diversification (number of counterparties per bank). The nature of robustness and fragility is characterized starkly by Acemoglu, Ozdaglar & Tahbaz-Salehi (2015), the work on which we base our detailed examples below.

Two recent papers connect the literature on vulnerability to other aspects of financial stability: Greenwood, Landier & Thesmar (2015) argue that fire sales can lead to spillovers across banks. This connects the contagion literature to the wider literature on fire sales and amplification in competitive equilibrium models, which we reviewed in Section 3. Zhou (2018) connects contagion via exposures to panic-driven financial crises and argues that panics can exacerbate the adverse effects in interbank networks in a crisis.

A third, recent line of research has combined the study of vulnerability in banking networks with endogenous network formation. Babus (2016) shows that the risk of contagion can be significantly reduced if a possibility of mutual insurance exists between banks ex ante. Farboodi, Jarosch & Shimer (2017) argue that the amount of risk generated by banks can nonetheless be excessive from a social perspective. Farboodi, Jarosch & Shimer (2017) study a model in which banks at the front of intermediation shares (i.e., close to the end borrower) get a larger share of the surplus that lending generates. This creates an incentive for banks with risky investment opportunities to connect with each other, to the point where this behavior generates potential contagion and reduces social surplus. Acemoglu et al. (2021) study endogenous contracting between banks and show that the anticipation of contagion is sufficient to generate credit freezes, in which banks cease lending to each other. We now move on to a more concrete example of vulnerability and discuss how it relates to the available empirical evidence.

4.1. Contagion of Banking Crises

A simple example, adapted from Acemoglu, Ozdaglar & Tahbaz-Salehi (2015), illustrates the potential for contagion in interconnected banking systems. Take a system with three banks, called B1, B2, and B3, with the following balance sheets: On the asset side, B1 has investments in the real economy that yield a random return z. On the liability side, B1 owes a total of y to banks B2 and B3 on the interbank market. The assets of B2 and B3, in addition to their claims on B1, are safe reserves worth s in total. Finally, all three banks owe v to other senior creditors outside the network, such as holders of repurchase agreements or depositors. Assume that $v < S$, so that total reserves are enough to cover one bank's deposits.

Figure 4*a* shows a concentrated bank network in which all interbank loans come from B2, as it specializes in interbank loans, while B1 holds only safe reserves. In this case, B1 is bankrupt if its return $z < v + y$. This scenario is contagious and leads to B2's failure if the value of its (junior) interbank claim is $z - v < v$. Thus, we get a joint failure of two banks if returns are bad enough so that $z < \min\{v + y, 2v\}$, but the third bank is always safe.

Figure 4*b* shows an interconnected interbank network with the same total exposure, in which both B2 and B3 lend to B1. Here, the same arguments show that we get contagion, and the failure of all three banks, if $z < \min\{v + y, 3v - s\}$. Since we have $s > v$, this threshold is always lower than the joint failure threshold in the concentrated case. Thus, the interconnected network is more robust. It takes larger losses to cause contagion, but this network is also more fragile—when contagion hits, all banks are affected.

The results in the work by Acemoglu, Ozdaglar & Tahbaz-Salehi (2015) generalize this logic. For small losses, a concentrated ring network, in which each bank borrows from and lends to one other bank, is the least stable, while the complete network, in which each bank is exposed to all others, is the most stable. For large losses, neither the ring nor the complete network do well; instead, stable architectures require pockets of banks that are insulated from others.

4.2. Empirical Work on Bank Networks

An early literature considered the relationship between bank networks and financial stability. Furfine (2003) considered the U.S. banking system; Upper & Worms (2004) analyzed Germany; Boss et al. (2004) looked at Austria; Degryse & Nguyen (2007) analyzed Belgium; and Cocco, Gomes & Martins (2009) considered Portugal. Iyer & Peydró (2011) conduct a case study of interbank linkages resulting from a large bank failure in India due to fraud. Upper (2011) contains a survey of this literature. The main conclusion is that contagion is usually not a serious risk,

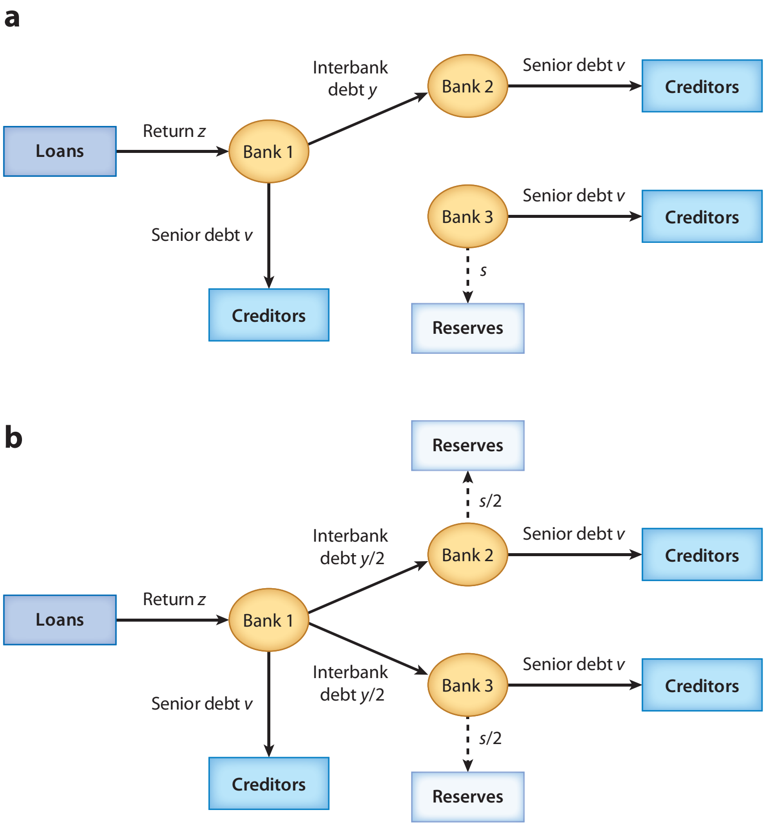

Figure 4

Examples of contagion. Panel *a* shows a concentrated bank network, where bank 1 borrows only from bank 2 in order to make loans. Panel *b* shows an interconnected network, where bank 1 borrows from both bank 2 and bank 3, holding constant the overall amount of interbank borrowing. As described in the text, the concentrated network is vulnerable to small losses, but the interconnected network involves the possible failure of all three banks for large losses. Definitions of the variables: z, the random return on loans made by bank 1; y, interbank debt; v, senior debt; and s, reserves. Solid arrows represent flows associated with bank contracts, and dashed arrows represent investments in safe reserves, such as government securities. It is assumed $v < s$.

provided there are not significant price movements in response to the turmoil. If there are, as Cifuentes, Ferrucci & Shin (2005) have suggested, then contagion effects can be significant.

The GFC resulted in many empirical papers on networks. Afonso, Kovner & Schoar (2013) show that substantial heterogeneity exists in the structure of trading relationships in the U.S. overnight interbank lending market. Billio et al. (2012) develop econometric measures of connectedness between different sectors of the financial system. They use these measures to argue that hedge funds, banks, broker dealers, and insurance companies have become more connected over time and that this connectedness has likely increased systemic risk. Craig & von Peter (2014) show that banks do not lend to each other directly. Instead, they form a core–periphery network by transacting with money center banks at the core. Gofman (2017) estimates a model of interbank lending in the United States and investigates the effect of regulating banks' size and interconnectedness. He shows that restricting interconnectedness improves financial stability. Allen et al. (2021) use a network analysis to argue that the differences in interbank market usage can be explained by the trust of market participants in the stability of the country's banking sector and counterparties, proxied by the history of banking crises and failures.

5. CONCLUDING REMARKS

This article has focused on two aspects of the relationship between financial architecture and financial stability that have been particularly important in the recent literature. First, we focused on the institutional structure of the financial system. While the traditional financial stability literature primarily studied banks, much of the recent literature has looked at other types of financial intermediary. Second, we focused on the network structure of the interbank market and how this structure affects financial stability.

Other aspects of financial architecture and financial stability are the exchange rate system and market crashes. While historically important, these have been less prevalent in recent years. Central bank swap lines have reduced the probability of currency crises, and central bank liquidity interventions have been used to support asset prices and counter financial market crashes.

Going forward, FinTech is likely to considerably change the architecture of the financial system. Allen, Gu & Jagtiani (2021) survey the different aspects of FinTech that will underlie such changes. The effect of these on financial stability remains an important area for future research.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are grateful to an anonymous referee for helpful comments that significantly improved the article.

LITERATURE CITED

Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2015. Systemic risk and stability in financial networks. *Am. Econ. Rev.* 105:564–608

Acemoglu D, Ozdaglar A, Siderius J, Tahbaz-Salehi A. 2021. Systemic credit freezes in financial lending networks. *Math. Financ. Econ.* 15:185–232

Acharya A, Engle R, Richardson M. 2012. Capital shortfall: a new approach to ranking and regulating systemic risks. *Am. Econ. Rev.* 102:59–64

Adrian T, Ashcraft AB. 2016. Shadow banking: a review of the literature. In *Banking Crises: Perspectives from the New Palgrave Dictionary of Economics*. London: Palgrave Macmillan

Adrian T, Brunnermeier MK. 2016. CoVaR. *Am. Econ. Rev.* 106:1705–41

Afonso G, Kovner A, Schoar A. 2013. *Trading partners in the interbank lending market*. Staff Rep. 620, Fed. Reserve Bank of New York

Allen F, Covi G, Gu X, Kowalewski O, Montagna M. 2021. *The interbank market puzzle*. SSRN Work. Pap. 3118363

Allen F, Gale D. 1998. Optimal financial crises. *J. Finance* 53:1245–84

Allen F, Gale D. 2000. Financial contagion. *J. Political Econ.* 108:1–33

Allen F, Gale D. 2009. *Understanding Financial Crises*. Oxford, UK: Oxford Univ. Press

Allen F, Gu X, Jagtiani J. 2021. A survey of fintech research and policy discussion. *Rev. Corp. Finance* 1(3–4):259–339

Allen F, Qian Y, Tu G, Yu F. 2019. Entrusted loans: a close look at China's shadow banking system. *J. Financ. Econ.* 133:18–41

Babus A. 2016. The formation of financial networks. *RAND J. Econ.* 47:239–72

Begenau J, Landvoigt T. 2021. *Financial regulation in a quantitative model of the modern banking system*. SSRN Work. Pap. 2748206

Ben-David I, Franzoni F, Moussawi R. 2018. Do ETFs increase volatility? *J. Finance* 73:2471–535

Bengui J, Bianchi J. 2018. *Macroprudential policy with leakages*. NBER Work. Pap. 25048

Billio M, Getmansky M, Lo AW, Pelizzon L. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J. Financ. Econ.* 104:535–59

Boss M, Elsinger H, Summer M, Thurner S. 2004. Network topology of the interbank market. *Quant. Finance* 4:677–84

Brunnermeier MK, Eisenbach TM, Sannikov Y. 2012. *Macroeconomics with financial frictions: a survey*. NBER Work. Pap. 18102

Bryant J. 1980. A model of reserves, bank runs, and deposit insurance. *J. Bank. Finance* 4:335–44

Buchak G, Matvos G, Piskorski T, Seru A. 2018a. Fintech, regulatory arbitrage, and the rise of shadow banks. *J. Financ. Econ.* 130:453–83

Buchak G, Matvos G, Piskorski T, Seru A. 2018b. *The limits of shadow banks*. NBER Work. Pap. 25149

Calomiris CW, Mason JR. 1997. Contagion and bank failures during the Great Depression: the June 1932 Chicago banking panic. *Am. Econ. Rev.* 87:863–83

Chari VV, Jagannathan R. 1988. Banking panics, information, and rational expectations equilibrium. *J. Finance* 43:749–61

Chen Q, Goldstein I, Jiang W. 2010. Payoff complementarities and financial fragility: evidence from mutual fund outflows. *J. Financ. Econ.* 97:239–62

Chernenko S, Sunderam A. 2016. *Liquidity transformation in asset management: evidence from the cash holdings of mutual funds*. NBER Work. Pap. 22391

Chevalier J, Ellison G. 1997. Risk taking by mutual funds as a response to incentives. *J. Political Econ.* 105:1167–200

Chrétien E, Lyonnet V. 2019. *Traditional and shadow banks*. Work. Pap. 2019-11, Ohio State Univ., Charles A. Dice Cent. Res. Financ. Econ., Columbus, OH

Cocco JF, Gomes FJ, Martins NC. 2009. Lending relationships in the interbank market. *J. Financ. Intermed.* 18:24–48

Cone K. 1983. *Regulation of depository institutions*. PhD Diss., Stanford Univ., Stanford, CA

Coval J, Stafford E. 2007. Asset fire sales (and purchases) in equity markets. *J. Financ. Econ.* 86:479–512

Covitz D, Liang N, Suarez G. 2013. The evolution of a financial crisis: collapse of the asset-backed commercial paper market. *J. Finance* 68:815–51

Cifuentes R, Ferrucci G, Shin HS. 2005. Liquidity risk and contagion. *J. Eur. Econ. Assoc.* 3:556–66

Craig B, von Peter G. 2014. Interbank tiering and money center banks. *J. Financ. Intermed.* 23:322–47

Davila E, Walther A. 2020. Does size matter? Bailouts with large and small banks. *J. Financ. Econ.* 136:1–22

Dasgupta A. 2004. Financial contagion through capital connections: a model of the origin and spread of bank panics. *J. Eur. Econ. Assoc.* 2:1049–84

Degryse H, Nguyen G. 2007. Interbank exposures: an empirical examination of contagion risk in the Belgian banking system. *Int. J. Cent. Bank.* 3:123–71

Dessaint O, Foucault T, Frésard L, Matray A. 2019. Noisy stock prices and corporate investment. *Rev. Financ. Stud.* 32:2625–72

Diamond DW, Dybvig P. 1983. Bank runs, deposit insurance, and liquidity. *J. Political Econ.* 91:401–19

Diamond DW, Rajan RG. 2005. Liquidity shortages and banking crises. *J. Finance* 60:615–47

Edmans A, Goldstein I, Jiang W. 2012. The real effects of financial markets: the impact of prices on takeovers. *J. Finance* 67:933–71

Eisenberg L, Noe TH. 2001. Systemic risk in financial systems. *Manag. Sci.* 47:205–336

Elliott M, Golub B, Jackson MO. 2014. Financial networks and contagion. *Am. Econ. Rev.* 104:3115–53

Farboodi M, Jarosch G, Shimer R. 2017. *The emergence of market structure*. NBER Work. Pap. 23234

Frazzini A, Lamont O. 2008. Dumb money: mutual fund flows and the cross-section of stock returns. *J. Financ. Econ.* 88:299–322

Freixas X, Parigi BM, Rochet JC. 2000. Systemic risk, interbank relations, and liquidity provision by the central bank. *J. Money Credit Bank.* 32:611–38

Furfine CH. 2003. Interbank exposures: quantifying the risk of contagion. *J. Money Credit Bank.* 35:111–28
Gai P, Kapadia S. 2010. Contagion in financial networks. *Proc. R. Soc. A* 466:2401–23
Gennaioli N, Shleifer A, Vishny RW. 2013. A model of shadow banking. *J. Finance* 68:1331–63
Gofman M. 2017. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *J. Financ. Econ.* 124:113–46
Goldstein I. 2017. Comment on "Redemption risk and cash hoarding by asset managers" by Morris, Shim, and Shin. *J. Monet. Econ.* 89:88–91
Goldstein I, Jiang H, Ng D. 2017. Investor flows and fragility in corporate bond funds. *J. Financ. Econ.* 126:592–613
Goldstein I, Pauzner A. 2005. Demand-deposit contracts and the probability of bank runs. *J. Finance* 60:1293–327
Gompers P, Metrick A. 2001. Institutional investors and equity prices. *Q. J. Econ.* 116:229–59
Gorton G. 1985. Bank suspension of convertibility. *J. Monet. Econ.* 15:177–193
Gorton G. 1988. Banking panics and business cycles. *Oxf. Econ. Pap.* 40:751–81
Gorton G, Metrick A. 2012. Securitized banking and the run on repo. *J. Financ. Econ.* 104:425–51
Greenwood R, Landier A, Thesmar D. 2015. Vulnerable banks. *J. Financ. Econ.* 115:471–85
Grochulski B, Zhang Y. 2019. Optimal liquidity policy with shadow banking. *Econ. Theory* 68:967–1015
Hachem K. 2018. Shadow banking in China. *Annu. Rev. Financ. Econ.* 10:287–308
Hau H, Lai S. 2013. Real effects of stock underpricing. *J. Financ. Econ.* 108:392–408
He Z, Xiong W. 2012. Dynamic debt runs. *Rev. Financ. Stud.* 25:1799–843
Irani RM, Iyer R, Meisenzahl RR, Peydró JL. 2021. The rise of shadow banking: evidence from capital regulation. *Rev. Financ. Stud.* 34(5):2181–235
Iyer R, Peydró JL. 2011. Interbank contagion at work: evidence from a natural experiment. *Rev. Financ. Stud.* 24:1337–77
Iyer R, Puri M. 2012. Understanding bank runs: the importance of depositor-bank relationships and networks. *Am. Econ. Rev.* 102:1414–45
Iyer R, Puri M, Ryan N. 2016. A tale of two runs: depositor responses to bank solvency risk. *J. Finance* 71:2687–726
Jacklin CJ. 1987. Demand deposits, trading restrictions, and risk sharing. In *Contractual Arrangements for Intertemporal Trade*. Minneapolis, MN: Univ. Minn. Press
Jacklin CJ, Bhattacharya S. 1988. Distinguishing panics and information-based bank runs: welfare and policy implications. *J. Political Econ.* 96:568–92
Jin D, Kacperczyk M, Kahraman B, Suntheim F. 2021. Swing pricing and fragility in open-end mutual funds. *Rev. Financ. Stud.* In press. **https://doi.org/10.1093/rfs/hhab022**
Kacperczyk M, Schnabl P. 2013. How safe are money market funds? *Q. J. Econ.* 128:1073–122
Kindleberger C. 1993. *A Financial History of Western Europe*. Oxford, UK: Oxford Univ. Press
Koijen R, Yogo M. 2021. *The fragility of market risk insurance*. NBER Work. Pap. 24182
Kress J. 2017. The case against activity-based financial regulation. *The CLS Blue Sky Blog*, Nov. 16. **https://clsbluesky.law.columbia.edu/2017/11/16/the-case-against-activity-based-financial-regulation/**
Krishnamurthy A, Nagel S, Orlov D. 2014. Sizing up repo. *J. Finance* 69:2381–417
Kyle AS. 1985. Continuous auctions and insider trading. *Econometrica* 53:1315–35
Loayza N, Ouaza A, Rancière R. 2018. Financial development, growth, and crisis: Is there a trade-off? In *The Handbook of Finance and Development*, ed. T Beck, R Levine, pp. 289–311. Cheltenham, UK: Edward Elgar
Luck S, Schempp P. 2014. *Banks, shadow banking, and fragility*. Work. Pap. 1726, Eur. Cent. Bank
Martinez-Miera D, Repullo R. 2019. *Markets, banks, and shadow banks.* Work. Pap. 2234, Eur. Centr. Bank
Miles D, Yang J, Marcheggiano G. 2013. Optimal bank capital. *Econ. J.* 123:1–37
Moreira A, Savov A. 2017. The macroeconomics of shadow banking. *J. Finance* 72:2381–432
Morris S, Shin I, Shin H. 2017. Redemption risk and cash hoarding by asset managers. *J. Monet. Econ.* 89:71–87
Ordonez G. 2018. Sustainable shadow banking. *Am. Econ. J. Macroecon.* 10:33–56
Plantin G. 2015. Shadow banking and bank capital regulation. *Rev. Financ. Stud.* 28:146–75
Reinhart CM, Rogoff KS. 2009. The aftermath of financial crises. *Am. Econ. Rev.* 99:466–72
Rochet JC, Vives X. 2004. Coordination failures and the lender of last resort: Was Bagehot right after all? *J. Eur. Econ. Assoc.* 2:1116–47

Schmidt L, Timmermann A, Wermers R. 2016. Runs on money market mutual funds. *Am. Econ. Rev.* 106:2625–57

Tang H. 2019. Peer-to-peer lenders versus banks: Substitutes or complements? *Am. Econ. Rev.* 32:1900–38

Upper C. 2011. Simulation methods to assess the danger of contagion in interbank markets. *J. Financ. Stab.* 7:111–25

Upper C, Worms A. 2004. Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *Eur. Econ. Rev.* 48:827–49

Xiao K. 2020. Monetary transmission through shadow banks. *Rev. Financ. Stud.* 33:2379–420

Zeng Y. 2017. *A dynamic theory of mutual fund runs and liquidity management*. Work. Pap., Univ. Washington, Seattle

Zhong H, Zhou Z. 2021. *Dynamic coordination and bankruptcy regulations*. SSRN Work. Pap. 3355429

Zhou Z. 2018. *Systemic bank panics in financial networks*. Unpublished manuscript, PBC Sch. Finance, Tsinghua Univ., Beijing