NBER WORKING PAPER SERIES

SHADOW BANKING AND THE FOUR PILLARS OF TRADITIONAL FINANCIAL INTERMEDIATION

Emmanuel Farhi
Jean Tirole

Working Paper 23930
http://www.nber.org/papers/w23930

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017, Revised February 2020

The authors are grateful to Aimé Bierdel, Mathias Dewatripont, Veronica Guerrieri, Frédéric Malherbe, Paul-Henri Moisson, John Moore, Francisco Nadal de Simone, Guillermo Ordonez, Guillaume Plantin, Jeremy Stein, Joseph Stiglitz, John Vickers, four referees, and participants at the Shanghai Frontiers of Macroeconomics conference, LACEA-LAMES Buenos Aires conference, 2019 ASSA meetings, Bocconi, Central Banks of England, France, Luxembourg and Uruguay, Harvard, NY Fed, and ULB for helpful comments. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n°669217 - ERC MARKLIM). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Shadow Banking and the Four Pillars of Traditional Financial Intermediation
Emmanuel Farhi and Jean Tirole
NBER Working Paper No. 23930
October 2017, Revised February 2020
JEL No. E0,G0

ABSTRACT

Traditional banking is built on four pillars: SME lending, insured deposit taking, access to lender of last resort, and prudential supervision. This paper unveils the logic of the quadrilogy by showing that it emerges naturally as an equilibrium outcome in a game between banks and the government. A key insight is that regulation and public insurance services (LOLR, deposit insurance) are complementary. The model also shows how prudential regulation must adjust to the emergence of shadow banking, and rationalizes structural remedies to counter financial contagion: ring-fencing between regulated and shadow banking and the sharing of liquidity in centralized platforms.

Emmanuel Farhi
Harvard University
Department of Economics
Littauer Center
Cambridge, MA 02138
and NBER
emmanuel.farhi@gmail.com

Jean Tirole
Toulouse School of Economics
1 Esplanade de l'Universite
31080 Toulouse Cedex 6
FRANCE
jean.tirole@tse-fr.eu

1 Introduction

Traditional banking is built on four pillars: the commercial or retail bank is prudentially supervised and in exchange gets access to lender of last resort and to deposit insurance. It caters to "special depositors", who want a liquid and safe vehicle for their savings, and to "special borrowers", the small and medium enterprises that need close oversight to secure financing. These two activities are the "core functions", the successful delivery of which governments attach great social value to. Other investors and borrowers have access and resort to financial markets. Other financial institutions traditionally have been left unregulated and could not claim access to deposit insurance and public liquidity.

This institutional framework is not a matter of course. The access to the "public insurance services"- the discount window and other liquidity facilities on the one hand and cheap deposits on the other - could be priced and offered to the financial system as a whole. Besides, this conventional definition of retail banking is called into question by recent developments. Many shadow financial institutions (money market mutual funds and investment banks) gained access to public liquidity facilities during the 2008 crisis. Another challenge to the conventional wisdom is the observation that in recent years shadow banks have gained much market share in retail banks' classical territories, the core functions. This seems in the West to be due to a migration of activities in reaction to tighter prudential standards. But they also have grown in importance in India and other emerging markets. Shadow banks in China lend to small and medium enterprises and cater to retail depositors through wealth management funds. Should we reconsider the conventional SME lending/insured deposit taking/regulation/lender of last resort quadrilogy?

To start answering this question, we unveil the logic of the conventional wisdom. While our work is normative, we build on the usual hazards facing financial systems: over-leverage and bailouts, threat of migration to shadow banking, cross exposures between retail and shadow banking sectors. These modeling ingredients are there for good reason. Moral hazard in the form of over-leverage (or more broadly risk taking), combined with the possibility of bailouts, creates an externality of unregulated banking on public finances. This externality may vindicate costly supervision. Cross-exposures (motivated by imperfect risk correlation and mutual insurance on our model) raise the possibility of financial contagion, which allows us to investigate the desirability of ring-fencing and of central counterparty clearing houses (CCPs). Mechanism design calls for optimally devising institutions that best mitigate the combined hazards.

This paper's central insight is that there are basic complementarities between regulation and the other components of the quadrilogy. Through its monopoly privilege on taxation -i.e. its access to future earnings-, the government has a special ability to create liquidity and therefore to insure banks and/or individuals when private markets are unable or unwilling to do so. However, deposit insurance (DI) and lender of last resort (LOLR) services are costly for the officials (political opprobrium) or for society (as they require the government to raise funds even in financial straits). Banking supervision lowers the cost of these put options on taxpayer money to the extent that it monitors leverage (in our model) or more generally reduces banking moral hazard. SME lending magnifies the benefit of regulation, as the fear of industrial disruption may trigger ex-ante-unwanted banking bailouts.[1]

Overall, the broad normative picture is one in which core banking clients- small depositors and SMEs-, who are most affected by a banking failure and therefore are politically sensitive, are served by a regulated entity and benefit (directly for depositors and indirectly for SMEs) from extended insurance from the government. The attractive pricing of this insurance in turn incentivizes banks to remain in the regulated sector instead of migrating to the shadow banking sector. This picture chimes with the UK legislation, which is cast in terms of the continuity of provision of "core services" – to households and SMEs that lack non-bank alternatives.

Our model, described in Section 2, has two dates. At date 0, the representative bank chooses its leverage, freely so if in the shadow banking sector and in a constrained way if part of the regulated sector. At date 1, the bank receives a random revenue (which in practice reflects fee earnings, asset values or the availability of cheap deposits), and honors (or not) its unsecured debt obligations. It then invests if it can: The bank, which has a specific expertise in lending to the industry (to the SMEs) may be illiquid/unable to finance it at date 1, and furthermore such an event is more likely if the bank is highly levered. Furthermore, if the bank has issued insured deposits at date 0, the government has to make good on these deposits when the bank cannot pay the depositors.

An important distinction here is between bailouts and insurance.[2] Bailouts arise when the government would like to commit not to rescue the bank or its depositors, but cannot refrain from doing so ex post when the bank is illiquid. By contrast, "public insurance

[1] The key feature of bank loans to SMEs is that the government attaches social value to them. In practice SME lending figure prominently as part of a bank's "core functions". The model applies more broadly to systemically important functions such as payment systems, market making, etc.

[2] LOLR is often described as following Bagehot's dictum: To avert panics, central banks should lend early and freely (i.e. without limit), to solvent firms, against good collateral, and at "high rates". In practice, it is very difficult to distinguish illiquidity from insolvency, and LOLR ends up subsidizing financial institutions. We focus on this dimension and model LOLR as a commitment at date 0 to bail out banks at date 1.

services" (LOLR for banks and DI for depositors) are contractual features[3] in which the government finds its own interest ex ante: It thereby monetizes its unique ability to provide liquidity in extreme events against other benefits, such as bringing the bank into the regulated sector through an access to public liquidity and to cheap deposits. Section 3 derives the complementarity between regulation and the two forms of liquidity assistance.

The paper's second contribution is to provide a rationale for two structural remedies: ring-fencing and migration of transactions towards CCPs[4]. Ring-fencing and CCPs feature prominently in a number of post-crisis reforms worldwide, and, for the former, in the philosophy of the Glass-Steagall act (in force from 1933 through 1999 in the US) separating regulated commercial and unregulated investment banking. Nonetheless, to the best of our knowledge, these policies have not yet been subject to a formal analysis. To perform such an analysis, we introduce a rationale for cross-exposures among financial institutions: Imperfectly correlated liquidity shocks create scope for desirable liquidity pooling and therefore counterparty risk. We show that the provision of mutual insurance among financial intermediaries is subject to gaming at the taxpayer's expense in which a regulated bank is only partially covered by its insurance counterparty and therefore holds "bogus liquidity" (as was de facto offered by AIG in the CDS market)[5], and "syphons off" its liquidity to a shadow banking entity (which happened when retail banks offered credit lines to the conduits they had created).

Section 4 first assumes that when counterparties are both supervised, the regulator can learn the correlation structure between them (say through joint stress testing). It thereby can prevent the hazards described above; a simple regulation forcing regulated banks to co-insure through mutual lines of credit (which is a form of liquidity regulation) then delivers the second-best welfare level. In contrast, such an understanding is not available if one of the parties lies outside the regulated sphere, and liquidity pooling can then game the supervisory system. Ring-fencing can help prevent such abuses.

Second, we make the opposite polar assumption that, unlike the counterparties, the regulator never learns the correlation structure. Ring-fencing then no longer suffices to

[3]In practice, examples of LOLR along these lines include access to the discount window and other facilities which are reserved to regulated banks. Deposit insurance is also reserved to regulated banks.

[4]CCPs become the buyer to every seller and the seller to every buyer; they thereby ensure the future performance of open contracts. Under the Basel framework, clearing member banks operating through a "qualified CCP" get preferential capital treatment.

[5]The notion of "bogus liquidity" is documented by Yorulmazer (2013) (who does not use this term), who analyzes the correlation between the insurer's default and the bank's shocks and argues that CDSs, which according to Basel regulation, can be counted as hedges and allow banks to free up regulatory capital, have been used to create a false sense of safety due to counterparty risk. There have also been concerns that regulated banks be dependent on investment funds for their short-term funding (see Jin-Nadal de Simone 2016 for evidence on the exposure of major European banks to investment funds).

deliver the second best. Regulated banks can game the liquidity requirements and arrange bogus liquidity lines to each other, knowing that they will be protected by bailouts anyway. They thereby maximize their put on taxpayer money. To restore the second best, the regulator can complement ring-fencing with the requirement that liquidity pooling occur through a CCP rather than bilaterally. This prevents banks from fine-tuning their liquidity provision at the expense of the taxpayer, i.e. from engaging in risk selection.

2 Model

We start by setting up the model in Section 2.1. We then go through a thorough discussion of our assumptions and modeling choices in Section 2.2.

2.1 Description

We first describe the model and later comment on the plausibility of, and robustness to the key assumptions. There are two periods ($t = 0, 1$), a single good, and three groups of players: investors, bankers, and the government (or "regulator"). At date 1, all uncertainty is resolved with the realization of an aggregate state $\omega = (\omega_1, \omega_2)$, where $\omega_1 \in \{G, B\}$ is the fiscal state and $\omega_2 \in \{g, b\}$ is the liquidity state for the banks which are indexed by $i \in [0, 1]$. The fiscal state is verifiable but the liquidity state is not.

- The fiscal state $\omega_1 \in \{G, B\}$ determines the shadow cost $1 + \lambda_{\omega_1}$ of public funds. The deadweight cost of taxation λ_{ω_1} is lower in the good fiscal state G than in the bad fiscal state B, i.e. $0 \leq \lambda_G < \lambda_B$.
- The liquidity state $\omega_2 \in \{g, b\}$ determines the common revenues r_{ω_2} of the banks. The revenues of the banks are higher in the good liquidity state g than in the bad liquidity state b, i.e. $r_g > r_b$. For simplicity we assume that $r_g = 2$ and $r_b = 0$.
- All probabilities are denoted by $p_{\cdot}$. The probability of fiscal state $\omega_1 \in \{G, B\}$ is p_{ω_1}. The conditional probabilities are denoted $p_{\omega_2|\omega_1}$ and the probability of state $\omega = (\omega_1, \omega_2)$ is $p_\omega \equiv p_{\omega_1} p_{\omega_2|\omega_1}$. The probability of a good fiscal state and liquid banks is p_{Gg}, etc.

There is no store of value in the economy and the correlation of banks' liquidity shocks precludes any cross-insurance (we relax this assumption in Section 4). Furthermore, as in Holmström and Tirole (1998), we assume that consumers/investors cannot commit to pay any funds in the future. As a result, while they can save, they can neither borrow nor grant credit lines to bankers. This will imply that only the government, through its exclusive access to taxpayer money, can provide liquidity in bad times. This builds the foundations for the unique ability of the government to offer *lender of last resort (LOLR)* and *deposit insurance (DI)*.

Investors. There are two kinds of investors.[6] "Ordinary" investors/consumers have risk-neutral preferences with no discounting over consumption. Their utility from consumption vector $\{c_0, c_{1,\omega}\}$ is given by $\mathbb{E}_{\omega}[c_0 + c_{1,\omega}]$; these depositors will in equilibrium be uninsured. They have large endowments in every period. "Special depositors" are formalized as in Gennaioli et al (2012, 2013), Stein (2012) and Caballero-Farhi (2018): They are ex-ante risk averse (à la Epstein-Zin)—actually infinitely risk averse—at date 0; they are willing to pay $1 + \theta$ with $\theta > 0$ at date 0 for the certainty of receiving 1 at date 1 (they are willing to pay 1 for 1 beyond that amount). These deposits must be absolutely safe in order to be valued. There is a mass $\mu \leq 1$ of special depositors.

Since bankers experience an aggregate liquidity shock with some probability, they cannot supply safe deposits on their own. They must rely partly on the government. It may therefore be optimal for the government to run a deposit insurance scheme. This deposit insurance scheme promises to make whole every special depositor who has deposited funds in a given bank in case this bank experiences a liquidity shortfall. Bankers can then market this insured product at price $1 + \theta$ at date 0 (the state will have no incentive to constrain this decision as it can always recoup an arbitrary part of the banker's rent from cheap deposits through the date-0 pricing of the deposit insurance service); furthermore, charging the special depositors' willingness to pay for the service, or for that matter any amount exceeding 1, implies that ordinary depositors lose money if they masquerade as special depositors.

Bankers. There is a mass 1 of bankers protected by limited liability. Each banker $i \in [0, 1]$ will need to finance a "project" (understand "SMEs") requiring a unit investment at date 1. The involvement of the banker is indispensable to run his project, which can be motivated by special knowledge or expertise. As there is no store of value in the economy, this date-1 investment must be financed from the bank's date-1 revenue or through a transfer from the state. As stated earlier, the bank's date-1 revenue r_{ω_2} is random with $r_g = 2$ and $r_b = 0$.[7]

[6] Of course, depositors may have several incarnations. They may be special depositors for returns up to 1, and ordinary depositors beyond that level.

[7] We could alternatively assume that the investment need is random.

At date 0, each bank i issues wholesale, uninsured debt with face value $\{d^i_{\omega_1}\}$ to be reimbursed at date 1. This debt can be contingent on the fiscal state ω_1, which is verifiable, but not on the liquidity state ω_2, which is not.

The state-contingent debt is reimbursed at the beginning of date 1, just after revenue accrual. We assume that bankers are protected by limited liability and that unsecured debt repayments cannot exceed revenue and are hence given by $\min\{d^i_{\omega_1}, r^i_{\omega_2}\}$. The justification for the upper bound is either that the government is committed to enforce the bailinability of uninsured debt (in the spirit of recent regulatory reforms) or that creditors have no bargaining power in renegotiations with the government (and thus cannot demand compensation).

This issuance decision captures bank i's risk-taking and will be called "choice of leverage". Bank i may also issue μ^i insured deposits, necessarily with the assent of the government, which as we have seen is indispensable for the creation of safe claims.

In the absence of regulation (shadow banking), the bank freely chooses leverage. We will thereby capture the idea that unmonitored balance sheet decisions may make the bank illiquid and unable to pursue its activities in the absence of liquidity assistance; such illiquidity may occur even in the absence of leverage.

The project's payoff comes in the form of a non-pledgeable, private benefit $\mathfrak{b} > 1$ for the banker. Banker i values consumption c_0 at date 0 and the private benefit from investment according to

$$U \equiv \mathbb{E}_\omega[c^i_0 + \mathfrak{b} x^i_\omega].$$

where $x^i_\omega = 1$ if bank i is able to invest at date 1 and $x^i_\omega = 0$ otherwise. The key feature, shared by all models of liquidity management, is that the surplus associated with continuation is not fully pledgeable to investors and so refinancing problems may emerge.

Government. At date 0 the regulator may monitor (regulate) banks that are not in the pure shadow banking sector. Monitoring bank i means setting[8] its leverage $\{d^i_{\omega_1}\}$ (a form of capital requirement) and costs $c \geq 0$.[9]

Second, the government may raise revenues by taxing consumers at date 1 and use the proceeds for bailouts (of banks and/or their depositors) and to honor its public insurance services. As started earlier, the collection of taxes for these purposes is costly: collecting 1 involves a *cost of public funds* equal to $1 + \lambda_{\omega_1}$, where the deadweight costs of taxation

[8]It in fact suffices to set a (state-contingent) upper bound for the bank's liability.

[9]Some elements of this activity- such as counting wholesale liabilities- are straightforward. But figuring out the implications of maturities, correlations, credit line and derivative exposures and the like can be competence- and time-intensive. We are agnostic as to the level of this cost, except for the fact that the more complex the banks' activities, the higher c is likely to be.

satisfy $\lambda_B > \lambda_G \geq 0$. By contrast, rebating one unit of revenue to consumers does not entail any efficiency benefit.[10]

The government can contract at date 0 on public insurance services (LOLR and DI) for date 1 with either regulated or shadow banks, or both; indeed, we aim at demonstrating the complementarity of regulation and public insurance services. LOLR consists in promising a bank to bring the missing cash if there is too little cash to invest 1 at date 1 when investment is meant to take place (see exact timing below); this service is contingent on the verifiable state, i.e. the fiscal state. Similarly, DI is a promise to special depositors to make up for any shortfall in the deposit contract's reimbursement of 1; and the DI agreement between bank i and the government specifies a volume μ^i of deposits.

Next, we describe the government's preferences. The government puts a welfare weight of 1 on consumers/investors, ordinary or special, implying in particular that special depositors will not be bailed out ex post at date 1 if the commitment to DI has not been extended ex ante at date 0. The government also puts weight on bank stakeholders (to be interpreted as the SMEs that rely on the banking relationship: credit crunch foundations are provided in Section 2.2), and zero or positive weight on bankers, with a total resulting weight[11] β on a project being implemented; the social value β of continuation satisfies

$$1 + \lambda_B > \beta > 1 + \lambda_G.$$

This assumption guarantees that bankers lacking cash at date 1 receive a discretionary bailout in state G but not in state B. It also implies that at date 1 in state B, the government would like to renege on any promise of liquidity assistance that would be part of the LOLR scheme, and is only prevented from doing so by its commitment at date 0.

For future use, we define the expected deadweight losses

$$\Lambda_{\omega_2} \equiv \Sigma_{\omega_1 \in \{G,B\}} p_{\omega_1} p_{\omega_2|\omega_1} \lambda_{\omega_1}$$

associated with one unit of public transfer to banks (or via the bank to depositors), contingent on the liquidity state $\omega_2 \in \{g, b\}$. The average shadow cost is

$$\bar{\lambda} \equiv \Lambda_g + \Lambda_b = p_G \lambda_G + p_B \lambda_B.$$

[10] As will shortly become clear, deposit insurance transfers to special depositors to make them whole has date-0 (but not date-1) efficiency benefits since $\theta > 0$, and we fully take them into account in the analysis.

[11] More generally the weight put on bankers could be positive, as long as it is smaller than 1. In that case, β is the total weight put on bankers and on bank stakeholders.

Payoffs. Let $m^i = 1$ if bank i is monitored (regulated) and $m^i = 0$ if it is not. The equilibrium values of $\{d^i_{\omega_1}\}$ are chosen by the government if bank i is regulated and is determined by the incentives of banker i seeking to maximize his utility if it is not. The total date-0 transfer made by the state to bank i is τ^i_0: it encompasses insurance premia for DI and LOLR and subsidies for abiding by regulatory constraints, and can be positive or negative; to simplify the notation, this we leave this dependence implicit.

Bank i's utility is

$$U_i \equiv \mathbb{E}_\omega[\mathfrak{b}x^i_\omega + \min\{d^i_{\omega_1}, r_{\omega_2}\} + (1+\theta)\mu^i + \tau^i_0]. \tag{1}$$

The term $\mathfrak{b}x^i_\omega$ reflects the benefits of continuation. The last three terms in U_i add up to the banker's total net date-0 income: the term $\min\{d^i_{\omega_1}, r_{\omega_2}\}$ is the date-0 proceeds of debt issuance, the term $(1+\theta)\mu^i$ the date-0 revenue from special deposits, and the term τ^i_0 the date-0 transfer from the government. The expression of U_i ignores the date-0 endowment (and possible activities) of the bank.

Total (government + bank) welfare is

$$W \equiv \mathbb{E}_\omega \left[\int_0^1 \left[(\mathfrak{b}+\beta-1)x^i_\omega + r_{\omega_2} + \theta\mu^i - cm^i \right.\right.$$
$$\left.\left. -\lambda_{\omega_1} \max\{0, x^i_\omega + \mu^i + \min\{d^i_{\omega_1}, r_{\omega_2}\} - r_{\omega_2}\} \right] di \right]. \tag{2}$$

W adds up for each bank i: (a) the social value of investment $(\mathfrak{b}+\beta-1)$ if it takes place; (b) the date-1 revenue r_{ω_2}; (c) the social surplus θ on each unit of deposit; (d) the cost c of monitoring, if any; and (e) the shadow cost of public funds on each public payment at date 1. Note that because of our assumption of transferable utility at date 0, date-0 transfers τ^i_0 does not appear in this expression. Conditional on $\{d^i_{\omega_1}\}$, μ^i, x^i_ω, and m^i, these transfers influence the respective levels of utility of the state and of bankers but not total welfare.

Public policy. We assume that all banks are treated symmetrically through a one-size-fits-all scheme. We will verify later in Section 3.5 that symmetry is optimal under a simple condition. Therefore, we omit the i indices unless otherwise stated. But we will also characterize the changes to the analysis if this condition is not satisfied and show that the optimal scheme may involve otherwise identical banks selecting different options.

The government sets a single regulatory contract $\{\tau_0, k, l, m, \{d_{\omega_1}\}\}$, where: τ_0 is the transfer from the government; $k = 1$ if there is DI and $k = 0$ if not; $l = 1$ is there is LOLR and $l = 0$ if not; $m = 1$ if regulation is required and $m = 0$ is not; and d_{ω_1} are prescribed debt levels (which are only enforced if the bank is regulated).

Abusing notation, we denote by $W(k, l, m, \{d_{\omega_1}\})$ the value of W in (2) when banks' leverage is $\{d_{\omega_1}\}$. In this expression, we have $\mu^i = \mu$ if and only if $k = 1$ (DI). Otherwise we have $\mu^i = 0$.[12] Similarly, we have $x^i_\omega = 1$ if either: $r_{\omega_2} \geq d^i_{\omega_1} + \mu^i + 1$; or $r_{\omega_2} < d^i_{\omega_1} + \mu^i + 1$ and either $l = 1$ (LOLR) or $\omega_1 = G$ (bailout). Otherwise have $x^i_\omega = \max\{0, r_{\omega_2} - d^i_{\omega_1} - \mu^i\}$.

"Pure shadow banking" is always available to the banker as an outside option, and the corresponding payoffs can be replicated by the regulatory contract $k = l = m = \tau_0 = 0$.[13]

Timing. After observing the proposed regulatory policy $\{\tau_0, k, l, m, d_{\omega_1}\}$, the banks choose whether to abide by the regulatory contract or to operate in the shadow banking sector. The timing is summarized in Figure 1. In Section 3, there is no (reason for an) interdependence among banks, and so we can view the banks' decisions of which sector to join as independent.[14]

The government chooses regulatory policy in order to maximize $W(k, l, m, \{d_{\omega_1}\})$ subject to the banks' participation constraint which determines τ_0 (they can choose to be pure shadow banks). So we solve an optimal mechanism design problem.

(i) Government offers regulatory contract $\{\tau_0, k, l, m, \{d_{\omega_1}\}\}$

(ii) Banks either accept the contract or become pure shadow banks.

(iii) Each bank i chooses its leverage $\{d^i_{\omega_1}\}$ if it decides to be a pure shadow bank; and abides by the regulatory contract otherwise.

(i) State ω is realized. Revenue r_{ω_2} materializes.

(ii) Debt repayment $\min\{d^i_{\omega_1}, r_{\omega_2}\}$.

(iii) Investment $x^i_\omega \in \{0, 1\}$.

(iv) Depositor repayment.

(v) Government, if called to the rescue, inspects the bank to ascertain the remaining cash in the bank, and bails it out (state G) or, if the bank has opted in the regulatory contract, honors its LOLR contract (state B) and its DI contract.

Figure 1: Timing for bank's operations.

[12] At the optimum, the mass of special depositors who are serviced by each bank is at a corner, either 0 or μ, and so we only consider these values.

[13] In our model, there is no private demand for supervision, and so were we to allow the private sector to operate the regulation technology, it would not use it unless it were compelled to do so. By contrast, there is a public demand for supervision stemming from the desire to reduce the fiscal costs of bailouts and LOLR.

[14] In Section 4, imperfect correlation will lead to liquidity pooling, and so for instance the benefit of joining the shadow banking sector may depend on whether other banks do so.

2.2 Discussion of modeling choices

A number of assumptions require explanations.

(a) Modeling of special depositors

Our use of Epstein-Zin preferences implies that government support is required to create the liabilities that depositors demand. While this unique ability is by and large descriptive, recent history has taught us that non-banks create quasi-safe assets that appeal to such investors, and that governments come to the rescue of entities (money market mutual funds, life-insurance vehicles) that do not deliver their explicitly or implicitly guaranteed return. Our companion paper (Farhi-Tirole 2020) assumes instead that special depositors' preferences, and the associated demand for safe assets, admit a Von Neumann-Morgenstern representation; special depositors need money at date 1 to accomplish or fulfill specific needs. The state may be tempted to make good on a financial claim that is held primarily by special depositors when the claim fails to deliver. We show that the shadow banking sector may cleverly use financial engineering so as to attract special depositors and create a put on taxpayer money. Another interesting feature is the phenomenon of clientele-dependant valuations: special depositors may in the absence of bailout prefer portfolio 1 to riskier portfolio 2, but nonetheless purchase (outbid ordinary depositors on) portfolio 2 that then becomes safer than portfolio 1 due to the investor bailout triggered by the special-depositor-heavy clientele. Some ordinary depositors benefit from an investor bailout as they mix with special depositors. Despite these differences (and others), the basic insight of a complementarity between supervision and liquidity assistance services holds.

(b) Indispensability of banker for investment

That the banker is indispensable is much stronger than needed for the results. For one thing, disposing of the banker deprives the banker of the private benefit attached with managing the asset; it thus makes it costly for the government to ex-ante ensure banker participation, unless alternative bankers at date 1 have sufficient cash to pay for the full private benefit. For another, in the absence of indispensability, managerial turnover costs would make firing the banker time-inconsistent given the absence of adverse selection in the model.

(c) Absence of store of value

The absence of store of value makes the government the only source of liquidity outside the banking sector. The correlation of banking shocks further implies that inside liquidity (that produced by the banking sector) is insufficient. Like in other models of liquidity shortages, one could introduce stores of value, provided that they are scarce, i.e. cannot allow the banking sector to self-insure against all macro shocks at no cost. Costly stores of value would still make it efficient for the public sector to bring liquidity, at least in tail events.

(d) Narrow banking

Relatedly, an alternative to public insurance services would be narrow banking: The government would issue sufficient debt that banks could hoard at date 0, and banks' leverage (now net of liquidity hoarding) would be regulated so as to preclude any transfer of public funds to the banks at date 1. This alternative however would be suboptimal. Revenues associated with the banking activity can be pledged to special depositors and special borrowers, thereby limiting the need to rely on government funds (which come together with a distortionary cost) to the cases where banking revenues fall short. Narrow banking would require more reliance on government distortionary taxation since special deposits and SME lending would be covered more extensively.

(e) Foundations for β

The key assumption underlying the possibility of bailouts is that the government cares sufficiently about credit flowing to the productive sector. The foundations for this are standard.[15]

(f) Public insurance services

Recall that LOLR means that the government commits at date 0 to enable the date-1 unit investment if the bank has less than 1 at the end of date 1, and similarly that DI means that the state commits at date 0 to make special depositors whole at the end of date 1 if the bank does not have enough funds to do so .

In principle, the LOLR and DI options could be made contingent on the liquidity state ω_2. For expositional simplicity only, we assumed that this aggregate liquidity state is not verifiable. One can imagine that the only verifiable information regarding aggregate liquidity comes in the form of reports by the government. In our model, the government would always prefer at date 1 to slant its report ex post in order to minimize its LOLR and DI liabilities.

This restriction affects only one of our results. Were regulatory contracts contingent

[15] Following Holmström-Tirole (1997), suppose that, at date 0, each bank makes an investment in knowledge/staff so as to be able to invest in, and monitor a mass 1 of firms, each with investment need 1 at date 1 and no net worth. At date 1, firms which have invested succeed or fail (then return 0). Success is guaranteed if none of the bank, the firm's managers and the workers shirks. Otherwise, success accrues with probability 0. Shirking at date 1 brings benefit $\mathfrak{b}$ to the bank, benefit $\mathfrak{b}_f$ to the firm manager and shirking by the workers yields them $\mathfrak{b}_w$. There is no payoff beyond the incentive payoffs $\mathfrak{b}$, $\mathfrak{b}_f$ and $\mathfrak{b}_w$ of these stakeholders. To simplify the firms' revenue in case of success is $\mathfrak{b} + \mathfrak{b}_f + \mathfrak{b}_w$, so that it covers incentive payments. We then get $\beta = \beta_b \mathfrak{b} + \beta_f \mathfrak{b}_f + \beta_w \mathfrak{b}_w$, where β_b, β_f and β_w are the welfare weights for bankers, firms and workers.

on the entire state ω, offering LOLR to regulated and shadow banks would be equally costly for the government: the quadrilogy would become a trilogy, the theory making no prediction on whether LOLR is better targeted at regulated or shadow banks (so it would be offered to both or to none). Were the liquidity state verifiable, the regulator could perfectly detect excess leverage. In state B, the LOLR option could be made contingent on the other banks' themselves being illiquid (there would be no change in state G, as the bank is then bailed out anyway). Furthermore, if there were idiosyncratic as well as aggregate shocks, and even if the aggregate liquidity state were verifiable, then regulation would lower the state's cost of LOLR, and so the quadrilogy would re-emerge.

One can combine the two paradigms by assuming that with some probability $z \in (0, 1)$ the entire state ω is verifiable, while with the complementary probability only the fiscal state ω_1 is verifiable. All results, including the strict complementarity between regulation and LOLR, then hold.[16]

(g) State-contingency of debt

Because we assume that the liquidity state ω_2 is not verifiable, debt contracts d_{ω_1} can only depend de jure on the fiscal state ω_1. However, because of limited liability, debt repayments $\min\{d_{\omega_1}, r_{\omega_2}\}$ are de facto contingent on the liquidity state. Furthermore because $r_b = 0$, it can easily be verified that all the results of the model would go through unchanged if we had instead allowed debt contracts to be de jure contingent on both the fiscal and the liquidity state.

(h) Timing

The bank repays creditors and invests before depositors are repaid and LOLR commitments/bailouts are implemented. This assumption captures, in a stylized way, the fact that banks have complex and continuously evolving balance sheets, and that neither depositors nor the banking supervisor monitor the bank's operations in continuous time. This raises the possibility that bad surprises happen, with too little cash left to pay back depositors or to finance the economy.

(i) Existence of shadow banks

Like most models of shadow banking, our framework posits that the government cannot prevent banking activities from developing outside the regulatory umbrella (or that shadow banks cannot be arbitrarily taxed, as enough taxation would de facto force banks

[16] Another approach goes as follows: Suppose that in state b of widespread illiquidity, some banks have revenue 2 and others revenue 0 depending the realization of a purely idiosyncratic shock. The regulator does not know which banks have revenue 0. Thus the absence of liquidity in a bank in state b may be due to lack of luck or to the syphoning off of revenue 2 through a debt level of 2. The regulator then faces a trade-off between not offering LOLR and offering LOLR and giving rise to moral hazard. So again the provision of LOLR is cheaper when the bank is regulated, and the quadrilogy obtains.

to be regulated). There are different rationales for this (realistic) assumption. First, the government and legislation may not react fast enough to financial innovation that enables corporate entities to offer banking services. Second, the regulator may find some activities/products too complex to be regulated and the benefits of regulation would not be worth the monitoring cost. Third, and following some of the literature, the shadow banking sector may be viewed as offering some breathing room to the industry, even though it may have detrimental consequences as well, such as those modeled here. We could obviously allow the government to tax shadow banks as long as the tax is low enough for shadow banking to remain an attractive outside option.

(j) Date-0 activities

For notational simplicity, we ignore date-0 activities, for instance date-0 lending or investment which could be the cause of the random date-1 return, or else any fixed cost investment. These could be added to the model, but, being separable, would not affect the qualitative results.

(k) Moral hazard

We identify banking moral hazard with the choice of leverage. We could equivalently assume that it relates to a decision affecting the net date-1 return, i.e. either the revenue or, as we did in a previous version of the paper, a reinvestment need. Again, these alternative modeling choices do not affect the results.

(l) Endogenous shadow price of public funds

We consider an exogenous cost of public funds (but have allowed it to be correlated with the banks' liquidity shock, as banks and government might be affected by the same macroeconomic shock). Following the doom-loop literature, though, the cost of public funds might depend on the banks' leverage, as Spain and Ireland learnt the hard way.[17] The analysis can be extended to accommodate such endogeneity of λ at the cost of a more complex analysis (which requires solving for fixed points).

(m) The government's commitment

The paper makes a key distinction between a "bailout" and "LOLR". The former refers to a potentially[18] involuntary recapitalization that follows an observation that the bank does not have the funds to continue; the government ex post bails out the bank because it has an important stake in the bank's continuation, although it might want to commit

[17] The other leg of the doom loop- with the government debt's credibility affecting the banks' balance sheets- by contrast is already captured by the presumed correlation.

[18] "Potentially" refers to the fact that bailouts are desirable in state (Gb) (in which the bank has no cash anyway), and undesirable in state (Gg) (in which the bank does have enough cash, but may have committed it to *investors*).

ex ante not to do so, in order to avoid excess leverage. The latter by contrast is a time-inconsistent intervention that is triggered by an earlier commitment. Our model thus adopts the paradigm of "commitment and renegotiation": parties to a contract can each demand that the contract be enforced but can renegotiate the contract if both benefit from doing so.

The potential undesirability of bailouts raises the question of whether parties who lose when the bank is recapitalized (CDS holders, bank's competitors, taxpayers. . .) should not be able to sue the government for rescuing the bank. Such ability to challenge government decisions would not change anything as long as the government has the bargaining power in renegotiations; but more generally it could enhance the government's ability to commit not to bail out banks. In practice though, the government cannot be sued for this (there are exceptions: other countries may challenge the rescue within the state-aid rules). Such empowerment would presumably confer nuisance power on many economic agents, as the large number of actions taken by national and local governments every day all could be challenged by a multitude of agents with a (possibly very small) stake in the decisions; this would certainly stall any public action. The study of optimal challenges to public policies clearly lies well beyond the scope of this paper, so we content ourselves with taking on board the existing institutions.

(n) Banker's payoff function

We assumed that the banker values only date-0 consumption. If banker i's utility were equal to $\mathbb{E}_\omega[c_0^i + c_{1,\omega}^i + \mathfrak{b}x_\omega^i]$, the only function of leverage would be to pump out cash out of the bank so as to benefit from a bailout (in fiscal state G) or LOLR (in fiscal state B). There would still be a social benefit from regulating the leverage and a complementarity between regulation and public insurance services. Alternatively, there could be an intrinsic, and not only a strategic, demand for leverage. First, the banker might be more impatient than the investors: $\mathbb{E}[c_0^i + \delta c_{1,\omega}^i + \mathfrak{b}x_\omega^i]$, where $\delta < 1$ (equivalently, investors might be willing to pay a premium for the ability to benefit from a scarce savings technology). Second, we could introduce a date-0 banking investment, combined with a limited net worth of the bankers, who then need to borrow. For all these reasons, our assumption on the banker's utility function is not restrictive.

3 Shadow banking and the quadrilogy

We now present the solution of the model laid down in the previous section. We first derive the outcome under pure shadow banking, i.e. when the bank is unregulated and is not offered any public insurance service; this outcome later defines the reservation utility U^{SB} of a representative bank and is associated with welfare W^{SB}.

We then solve the social planning problem of the maximization of social welfare $W(k, l, m, \{d_{\omega_1}\})$. We exploit the fact that utility is transferable between the bank and the government at date 0. The bank's utility is equal to its shadow banking reservation utility, and the government's utility is equal to total welfare minus the bank's reservation utility, $W(k, l, m\{d_{\omega_1}\}) - U^{SB}$. The date-0 transfer τ_0 from the government to the bank, which can be positive or negative, is chosen so as to make the bank indifferent to becoming a shadow bank.

We show that there are *complementarities* between regulation and public liquidity support: the net benefits of DI and LOLR are greater under regulation than in the absence of regulation; equivalently, the presence of liquidity support (DI or LOLR) increase the net benefits of regulation. These complementarities are the signature of *economies of scope* in regulation: regulation facilitates both LOLR and DI.[19]

Because of these complementarities, the traditional banking quadrilogy naturally emerges at the optimum with the coexistence of lending to SMEs, DI, regulation, and LOLR. Public liquidity support (DI and LOLR) may be underpriced in order to prevent migration to the shadow banking sector.[20]

We first derive social welfare under the all the different configurations of (k, l, m). We divide the analysis into two parts: shadow banking ($m = 0$) in Section 3.1 and regulated banking ($m = 1$) in Section 3.2. Then in Section 3.3, we establish the complementarities between regulation and public liquidity support and derive the conditions for the emergence of the traditional banking quadrilogy. In the analysis up to this point, shadow banking influences the optimal arrangement only by serving as an outside option. Finally in Sections 3.4 and 3.5, we analyze natural extensions that can give rise to the actual coexistence of traditional and shadow banking.

[19]In our model, the balance sheet monitoring technology takes the form of a fixed cost. Indeed, the cost of monitoring the balance sheet of a bank to predict shortages of cash, which mostly arises from setting up an informational infrastructure (for reporting, auditing, etc.), is unlikely to change much whether the shortages lead to bailouts of the bank or of its depositors. This naturally gives rise to increasing returns to scope in regulation. Milder versions of this assumption would preserve but soften this conclusion, as would happen for example if we assumed that the cost preventing the abuse of LOLR and DI is larger than the each of the costs but smaller than the sum of the costs of preventing the abuse of these public provision programs.

[20]As Peltzman (1989) argues: "The putative motive for this [government guarantee of deposits] subsidy is to use the banks as the government's agents for providing a cheap, liquid substitute for government money. The quid for this quo is that banks should refrain from using their access to the government guarantee simply to maximize profits." [We are grateful to John Vickers for referring us to Peltzman's work].

3.1 Shadow banking

We start by analyzing the case of shadow banking where banks are not monitored and so bank leverage is unregulated. As described above, a bank's objective is two-fold: invest, and receive money from debt and/or deposits issuance. Investing dominates receiving money as the private benefit exceeds the cost of investment ($\mathfrak{b} > 1$). But the two need not be inconsistent. Indeed, left unsupervised, the bank has an incentive to issue as much as possible in state G so that $d_G = 2$; it thereby collects maximal date-0 revenue without jeopardizing date-1 investment, which is secured by the subsequent bailout. The same is true in state B if the bank benefits from LOLR so that $d_B = 2$. By contrast, in the absence of LOLR, the bank cannot count on a bailout in state B, and is better off limiting its liability to $d_B = 1$, so as to have enough cash to undertake the investment.

(a) Pure shadow banking

Under pure shadow banking (and so without access to public liquidity insurance services), the bank optimally sets $d_G = 2$ for state G. It thereby receives $2p_{Gg}$ at date 0 from investors. Being an empty shell in state G, it is rescued by the government (receives bailout 1 to invest) and obtains benefit $\mathfrak{b}$ from continuation. In state (B, g) by contrast, it cannot count on a bailout. Because $\mathfrak{b} > 1$, the bank prefers to continue, and it therefore issues debt $d_B = 1$. The bank cannot continue in state (B, b) since it has no revenue and receives no bailout. Its utility in the shadow banking sector is therefore

$$U^{SB} = (1 - p_{Bb})\mathfrak{b} + [2p_{Gg} + p_{Bg}].$$

Social welfare further accounts for the cost $1 + \lambda_G$ of the bailout in fiscal state G as well as of the social benefit of continuation β when it occurs:[21]

$$W^{SB} = (1 - p_{Bb})(\mathfrak{b} + \beta - 1) + 2p_g - \lambda_G p_G.$$

This expression for W^{SB} embodies the expected surplus from investment assuming private-sector financing, $(1 - p_{Bb})(\mathfrak{b} + \beta - 1)$, plus the expected date-1 revenue, $2p_g$, minus the deadweight loss of public financing of investment in the good fiscal state, $\lambda_G p_G$.

[21] The net effect of bailouts on welfare can be positive or negative depending on the relative strength of two effects: first, bailouts allow the bank to continue in state (G, b); second, bailouts allows the bank to increase leverage without risking termination in state (G, g). Bailouts are efficient as in Bianchi (2016) when the former effect dominates the latter. In this case, bailouts are a form of implicit public outside liquidity, which remedies the lack of private inside liquidity that banks could rely on to weather liquidity shocks in state (G, g).

		Fiscal state	
		$1+\lambda_G < \beta$	$1+\lambda_B > \beta$
		G	B
Liquidity state	$r = 2$: g	bailout, invest	invest
	$r = 0$: b	bailout, invest	no invest

Pure shadow bank

(b) LOLR

With systematic access to public liquidity, a shadow bank selects to always be an empty shell: $d_G = d_B = 2$. Because it already counted on a bailout in state G, LOLR has an effect only in the B state. The net social benefit of LOLR is

$$W^{SBL} - W^{SB} = p_{Bb}(\mathfrak{b} + \beta - 1) - p_B \lambda_B.$$

	G	B
g		LOLR, invest $-\lambda_B$
b		LOLR, invest $\mathfrak{b} + \beta - (1 + \lambda_B)$

Shadow bank + LOLR (compared with pure SB)

The first term corresponds to the net social benefit of investment, $\mathfrak{b} + \beta - 1$ in state (B, b). The second term is the shadow cost of public funds in state B.

(c) DI

With deposit insurance, the shadow bank (which does not internalize depositor or tax-payer welfare) sets $d_G = 2$ as earlier, and $d_B = 2$ in the presence of LOLR and $d_B = 1$ without LOLR. In both cases, the net social benefit associated with DI is, per unit of deposit, θ, minus the deadweight loss associated with the full provision of the deposit repayment (as the bank always manages to be an empty shell by the time the repayment to depositors is due), $\Lambda_g + \Lambda_b = p_G \lambda_G + p_B \lambda_B = \bar{\lambda}$. And so the net benefit of deposit insurance in the shadow banking sector is

$$W^{SBD} - W^{SB} = \mu(\theta - \bar{\lambda}).$$

	G	B
g	$\mu(\theta - \lambda_G)$	$\mu(\theta - \lambda_B)$
b	$\mu(\theta - \lambda_G)$	$\mu(\theta - \lambda_B)$

Shadow bank + DI (compared with pure SB)

We collect these results in a proposition.

Proposition 1. *(optimum in the absence of regulation)*
The optimal arrangement given no regulation ($m = 0$) features

(i) LOLR ($l = 1$) and continuation in all states if and only if

$$W^{SBL} > W^{SB} \iff \mathfrak{b} + \beta - 1 > \lambda_B \frac{p_B}{p_{Bb}}, \tag{3}$$

and otherwise no LOLR ($l = 0$) and continuation in all states except state B when there are no revenues at date 1.

(ii) DI and servicing of special depositors ($k = 1$) if and only if

$$W^{SBD} > W^{SB} \iff \theta > \bar{\lambda}, \tag{4}$$

and otherwise no DI and no servicing of special depositors ($k = 0$).

There is no direct complementarity between LOLR and DI: The net benefit of LOLR and DI combined is $[W^{SBL} - W^{SB}] + [W^{SBD} - W^{SB}]$. The conditions (3) and (4) for LOLR and DI can thus be applied independently. For example, the optimal arrangement features LOLR and DI if (3) and (4) hold, LOLR but no DI if (3) does but (4) does not, etc. As we will see, the possibility of regulation creates an indirect complementarity (natural co-variation) between these two forms of public liquidity support.

3.2 Regulated banking

We now analyze the case of regulated banking where banks are monitored and so bank leverage is regulated.

(a) Pure regulated banking (no insurance services)

A regulated bank is optimally constrained to set $d_G = 1$ so that it continues by itself in state (G, g) and is saved by the government in state (G, b). The bank still sets $d_B = 1$. The efficiency gains of regulation come from the fact that the deadweight loss associated with bailouts in state (G, g) is avoided. The net benefit of regulation in the absence of LOLR or DI is

$$W^{RB} - W^{SB} = p_{Gg}\lambda_G - c.$$

	G	B
g	no bailout λ_G	
b		

Regulated bank (compared with pure SB)

(b) LOLR

In the shadow banking sector, LOLR generates moral hazard in the form of an increase in leverage to $d_B = 2$. With regulation, this increase in leverage is prevented and the bank is constrained to set $d_B = 1$. The efficiency gains of regulation come from the fact that the deadweight loss associated with LOLR in state (B, g) is avoided. The net social benefit of LOLR given regulation is

$$W^{RBL} - W^{RB} = [W^{SBL} - W^{SB}] + p_{Bg}\lambda_B = p_{Bb}[\mathfrak{b} + \beta - (1 + \lambda_B)].$$

	G	B
g		
b		LOLR, invest $\mathfrak{b} + \beta - (1 + \lambda_B)$

Regulated bank + LOLR (compared with pure RB)

(c) DI

In the shadow banking deposit insurance is triggered in both states whether or not the bank has a revenue. With regulation, deposit insurance is triggered only when the bank does not have a revenue. Bank leverage is constrained at $d_G = d_B = 1 - \mu$ so as to be able to finance one unit of investment and to repay μ special depositors when its revenue are equal to 2.[22]

	G	B
g	$\mu\theta$	$\mu\theta$
b	$\mu(\theta - \lambda_G)$	$\mu(\theta - \lambda_B)$

Regulated bank + DI (compared with pure RB)

The net benefit of deposit insurance given regulation is

$$[W^{RBD} - W^{RB}] = [W^{SBD} - W^{SB}] + \mu\Lambda_g = \mu(\theta - \Lambda_b).$$

Proposition 2. *(optimum given regulation)*

The optimal arrangement given regulation ($m = 1$) features

(i) LOLR ($l = 1$) and continuation in all states if and only if

$$W^{RBL} > W^{RB} \iff \mathfrak{b} + \beta > 1 + \lambda_B, \tag{5}$$

and otherwise no LOLR ($l = 0$) and continuation in all states except state B when there are no revenues at date 1.

(ii) DI and servicing of special depositors ($k = 1$) if and only if

$$W^{RBD} > W^{RB} \iff \theta > \Lambda_b \tag{6}$$

and otherwise no DI and no servicing of special depositors ($k = 0$).

Like under shadow banking, there is no direct complementarity between LOLR and DI. The conditions (5) and (6) for LOLR and DI can be applied independently. For example,

[22] As long as the banks are all regulated, there are multiple ways of achieving the same DI outcome. For example $\int \mu^i di = \mu$ is consistent with $\mu_i = \mu$ for all i; each bank's debt is then constrained to not exceed the free cash flow left once investment and deposits are covered: $d^i_{\omega_1} \leq 1 - \mu$. Alternatively, one can specialize banks, i.e. allocate all deposits to a fraction μ of them (which then have $\mu^i = 1$), and require that each of these banks issue no debt ($d^i_{\omega_1} = 0$), while the non-deposit-taking ones can issue debt ($d^i_{\omega_1} \leq 1$). These two options are equivalent if it is optimal to regulate all banks and then our maintained "one-size-fits-all" treatment makes us choose the former. However, as we will observe in Section 3.5, this treatment is not warranted if it is optimal to regulate deposit-taking banks but to not to regulate non-deposit-taking banks; the second option is then uniquely optimal, and we will therefore accommodate it.

the optimal arrangement features LOLR and DI if (5) and (6) hold, LOLR but no DI if (5) does but (6) does not, etc. But there is an indirect complementarity (natural co-variation) between these two forms of public liquidity support.

3.3 Complementarity of regulation and public liquidity services

Leveraging the analysis in Sections 3.1 and 3.2, we can write down the policy problem as the maximization of social welfare over all the different possible arrangements:

$$\max_{\{k,l,m\}} \mathcal{W}(k,l,m),$$

where

$$\begin{aligned}\mathcal{W}(k,l,m) = W^{SB} &+ [W^{SBL} - W^{SB}]l + [W^{SBD} - W^{SB}]k \\ &+ [W^{RB} - W^{SB}]m + [(W^{RBL} - W^{RB}) - (W^{SBL} - W^{SB})]ml \\ &+ [(W^{RBL} - W^{RB}) - (W^{SBD} - W^{SB})]mk. \end{aligned} \tag{7}$$

The right-hand side of this expression features both linear terms in l, k, m and quadratic terms in ml and mk. The following corollary confirms that the coefficients of the quadratic terms are positive, which indicates complementarities between regulation and public liquidity support: the net benefits of DI and LOLR are greater under regulation than in the absence of regulation; equivalently, the presence of liquidity support (DI or LOLR) increase the net benefits of regulation.

Corollary 1. *(complementarity) Public insurance services and regulation are complements:*

$$W^{RBL} - W^{RB} = W^{SBL} - W^{SB} + p_{Bg}\lambda_B > W^{SBL} - W^{SB}$$

and

$$W^{RBD} - W^{RB} = W^{SBD} - W^{SB} + \mu\Lambda_g > W^{SBD} - W^{SB}.$$

The complementarity between regulation and LOLR and that between regulation and DI are the signature of economies of scope in regulation: regulation facilitates both LOLR and DI.[23] This complementarity is at the heart of the quadrilogy: the coexistence of lending to SMEs, deposit taking DI, regulation, and LOLR.

Armed with these results, we can now characterize the overall optimal arrangement.

[23]This rationale for the co-existence of lending and deposit-taking is distinct from the one articulated by Kashyap et al. (2002). They also emphasize economies of scope but arising from a different mechanism: the need for a pool of safe and liquid assets for these two imperfectly correlated activities.

Proposition 3. *(traditional banking system)*

(i) Regulation is optimal if and only if the maximization of $\mathcal{W}(k,l,m)$ *yields* $m = 1$*, i.e. iff* $c \leq c^*$*, for some regulation cost threshold* $c^* \geq p_{Gg}\lambda_G$*, where* $p_{Gg}\lambda_G$ *is the benefit of avoided bailouts.*

(ii) A traditional banking system with regulation, LOLR and DI is optimal, while LOLR and DI would not be granted to shadow banks if and only if

$$0 < p_{Bb}[\mathfrak{b} + \beta - (1 + \lambda_B)] < p_B\lambda_B \tag{8}$$

$$\Lambda_b < \theta < \bar{\lambda} = \Lambda_g + \Lambda_b \tag{9}$$

$$c \leq c^* = p_{Gg}\lambda_G + p_{Bb}[\mathfrak{b} + \beta - (1 + \lambda_G)] + \mu(\theta - \Lambda_b). \tag{10}$$

3.4 Observed heterogeneity and the coexistence of the two sectors

Assume now that banks differ in how hard they are to monitor. It may be that some activities, such as SME and mortgage lending or plain-vanilla interest-rate and exchange-rate derivatives, are sufficiently well-understood to be reasonably supervised by the government, while others involve very complex instruments such as bespoke derivatives, that either are poorly understood by the government or are extremely time-consuming to monitor and assess. One can imagine that there is a distribution of banks, each associated with a pattern of banking activities and characterized by its monitoring cost.[24] The monitoring cost is c for a fraction α of banks and $+\infty$ for the remaining fraction. The latter banks are necessarily in the shadow banking sector, while the former have a choice.[25] So this extended version boils down to the previous one for $\alpha = 1$. A bank's value of the monitoring cost is observable by the supervisor.

The heterogeneity of activities creates the possibility of co-existence of regulated and shadow banks in equilibrium. Regulated banks may enjoy LOLR and DI, while shadow banks may enjoy neither LOLR nor DI. And special depositors are serviced by regulated banks because doing so is more efficient since it economizes on public funds and on the associated deadweight costs of taxation.

[24] Actually the banks can be ex-ante identical. What matters is that activities differ in their surveillance cost. One of the strengths of our modeling is that we do not presume that some banks cannot do certain things.

[25] For example, securitized assets, which are held to a large extent by shadow banks, might be intrinsically harder to understand and monitor due to their complexity.

Corollary 2. *(heterogeneity) Assume that conditions (8) through (10) hold. Then, there are two sectors in the economy. A fraction α of the banks are regulated and enjoy LOLR and DI ($k = l = m = 1$); the remaining fraction $1 - \alpha$ is in the shadow banking sector and has no public liquidity support ($k = l = m = 0$); special depositors are serviced by regulated banks (and all are serviced if $\alpha \geq \mu$).*

3.5 One-size-fits-all or a menu of options?

The fact that all banks are ex-ante identical suggests that a single option is optimal. This is so only subject to an extra condition. Intuitively, the allocation of deposits does not matter, provided that $\mu^i \leq 1$ for all i so that a bank with revenue 2 can both cover its investment and service its depositors if its leverage is low enough (nil if $\mu^i = 1$), as long as: (a) all banks are supervised; and (b) DI pricing makes them indifferent to taking more or less insured deposits (which is always doable). However, when $\mu < 1$, it conceivably may be desirable to specialize banks and split them into two groups: those which serve special depositors and are monitored, and those which have no special depositors and are part of the shadow banking sector.

Proposition 4. *((sub)optimality of menus)*

(i) Suppose that $W^{RB} + \max\{W^{RBL} - W^{RB}, 0\} \geq W^{SB} + \max\{W^{SBL} - W^{SB}, 0\}$ (monitoring is optimal even in the absence of special depositors). Then the optimal regulation can be implemented through a single, one-size-fits-all contract (with e.g. $\mu^i = \mu$).

(ii) By contrast, if $W^{RB} + \max\{W^{RBL} - W^{RB}, 0\} < W^{SB} + \max\{W^{SBL} - W^{SB}, 0\}$ (implying that monitoring is suboptimal in the absence of special depositors), $W^{RB} + \max\{W^{RBL} - W^{RB}, 0\} + \max\{W^{RBD} - W^{RB}, 0\} > W^{SB} + \max\{W^{SBL} - W^{SB}, 0\} + \max\{W^{SBD} - W^{SB}, 0\}$ (implying that it is optimal to service all special depositors and that it is more efficient to do so in the regulated banking sector), and $\mu < 1$ (implying that special depositors can be serviced by a fraction of the banks), then the optimal regulation consists in a menu of two options (equally attractive to banks): one with deposit insurance in the regulated sector (with $\mu^i = 1$), and one without deposit insurance in the shadow banking sector (with $\mu^i = 0$).

4 Contagion, ring-fencing, and CCPs

So far liquidity shocks were perfectly correlated and so there was no rationale for liquidity pooling and therefore for counterparty risk. In practice, liquidity pooling occurs through credit default swaps, interest and FX swaps, lines of credit, guarantees, money market lending, and other varieties of financial instruments. To capture liquidity pooling and its regulatory consequences, we allow liquidity shocks to be imperfectly correlated. We assume that banks are able to recognize the patterns of correlation. The regulator cannot assess the correlation when one of the two counterparties is in the shadow banking sector: Figuring out the correlation requires at the very least the supervision of both balance sheets. When both counterparties are regulated, we look at the polar cases in which the regulator learns (say, through joint stress tests) or does not learn the pattern of correlation of the two institutions. Imperfect regulatory knowledge will create opportunities for gaming which can be thwarted by structural remedies: *ring-fencing* the regulated sector from the shadow banking sector; *or setting up a central counterparty clearing house (CCP)*.

We start by setting up the model in Section 4.1. We discuss our modeling choices in Section 4.2. We then analyze optimal liquidity sharing in Section 4.3. In Section 4.4, we ask when optimal liquidity sharing can be implemented, by examining the opportunities for regulatory gaming and how this can be contained via ring-fencing and CCPs. Finally, in Section 4.5, we consider an extension where activities in the regulated and shadow banking sectors are naturally imperfectly correlated, which creates a cost of liquidity segregation, and work out its implications for the optimal arrangement of liquidity sharing, public liquidity support, and regulation.

4.1 Setup

Suppose, for expositional simplicity only, that there are no special depositors ($\mu = 0$). The model is otherwise a generalization of the perfect-correlation model. The fiscal states G and B are as earlier. The possible realizations of the aggregate liquidity state are now $\omega_2 \in \{g, b, m^H, m^T\}$. States g and b are as before: all banks receive revenue $r_g = 2$ and $r_b = 0$, respectively, and so there is no cross-insurance opportunity. By contrast, in states m^H and m^T (which are intermediate in terms of total liquidity), half of the banks receive revenue 2 and the other half receive revenue 0. More precisely, there are two equal-size "groups" or "types"[26] of banks: H (heads) and T (tails). The realization of a coin toss (heads or tails, equally likely) occurs at the beginning of date 1. In sub-state m^H, banks in group H receive revenue 2, while banks in group T receive no revenue; and conversely in sub-state m^T.

As earlier, the fiscal state $\omega_1 \in \{G, B\}$ is verifiable. The liquidity state ω_2 is partly

[26] Belonging to group H or T is the only parameter distinguishing banks at date 0; so we can legitimately refer to it as the bank's "type".

verifiable, in a way that will make the model of Section 3 nested in this one: the verifiable partition is $\{\{g, b\}, m^H, m^T\}$. We assume that, unlike g and b, m^H and m^T are verifiable and can be contracted on at date 0. In particular, bankers can write debt contracts promising repayment $d^i_{\omega_1}$ up to limited liability in states ω with $\omega_2 \in \{g, b\}$ exactly as before, as well as debt contracts promising repayments d^i_{ω} up to limited liability in states ω with $\omega_2 \in \{m^H, m^T\}$. Similarly, the government can offer separate LOLR contracts guaranteeing reinvestment indistinctly in states (B, g) and (B, b) exactly as before, as well in states (B, m^H) and (B, m^T).

At date 0, banks do know which group they, and their potential counterparties belong to. Counterparties observe each other's leverage.[27] By contrast, and as announced earlier, we will assume that the regulator never observes the type of a shadow bank; and we will consider the two polar cases in which the regulator knows or does not know the type of a supervised bank.

When $p_{m^H} = p_{m^T} = 0$, the model is strictly isomorphic to our baseline model, in which there is no scope for liquidity sharing. When $p_{m^H} = p_{m^T} > 0$, there are opportunities for liquidity sharing between different types of banks in the "insurance opportunity" states m^H and m^T. A transfer from a cash-rich bank to a cash-poor one enables investment in both, and thereby economizes on public funds that would go to the cash-poor bank in a bailout (state G) or a LOLR operation (state B). Two banks are said to be "natural counterparties" if one belongs to group H and the other to group T. They are "correlated-risk counterparties" if they belong to the same group. The analysis of states g and b is the same as in Section 3. So the focus of this section will be entirely on the insurance-opportunity states m^H and m^T.

We introduce the following financial contracts. A bank can sign state-contingent liquidity-sharing contracts at date 0, over and above the debt contracts that we have already introduced.[28] We will focus on contracts in which one bank promises to transfer 1 in state m^H and the other in exchange promises to transfer 1 in state m^T;[29] we will look at

[27] If we allowed for multiple counterparties (which is not needed here: one counterparty is sufficient to exhaust insurance opportunities), we would further require that a bank observes the overall date-0 balance sheet, including other liquidity-sharing contracts, so as to preserve informational symmetry among counterparties.

[28] The rationale for cross pledging is different from that in Diamond (1984). In that paper, the benefits come from economies of scope in incentive provision. In our model instead, it is a form of risk sharing which does not help remedy an internal agency problem. Instead, our focus is on the agency problem of banks in cahoots to avoid sharing liquidity.

[29] At date 1, the grantee can threaten to take the grantor into a bankruptcy proceeding if the latter does not abide by its commitment; but if the grantor has no revenue, there is no bankruptcy proceeding (to ensure that, one may envision a small cost of bankruptcy proceedings). It does not matter whether banks can observe or not each other's revenue because they know each other's type and therefore can infer each other's revenue. If a bank with revenue does not honor its due payment as a grantor, then the grantee can

whether such contracts suffice to implement the optimum, and when they do not, at how centralized liquidity pooling and dispatch enables society to reach this optimum. Such mutual insurance contracts are best thought of as swaps, transferring mechanically cash between the two banks in a state-contingent manner.[30]

The modified timing is described in Figure 2 (recall that there are no special depositors), with the novel part highlighted in bold.

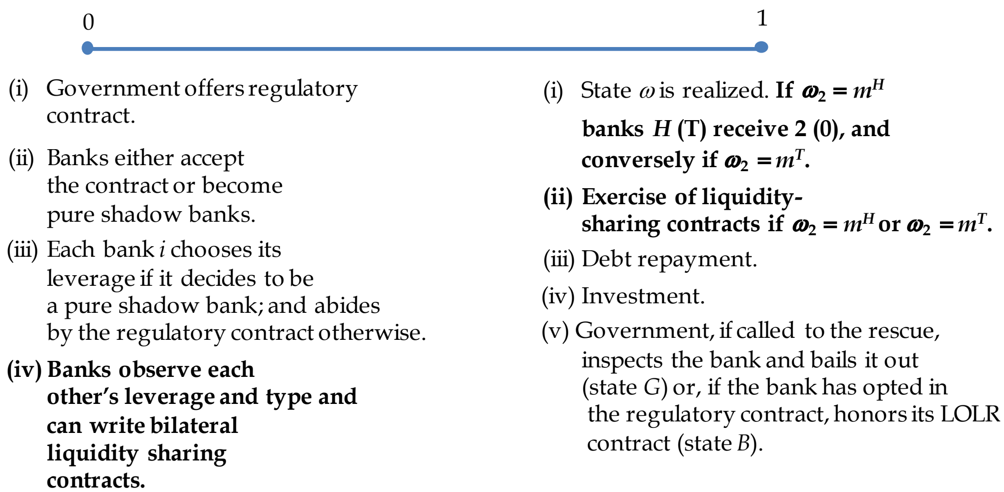

Figure 2: Timing under bilateral hedging.

Our contemplated policy interventions will impose constraints on the choice of counterparties in liquidity sharing contracts at stage (iv) of date 0. Ring fencing is the requirement that the counterparty of a regulated bank itself be in the regulated sector. A CCP requirement will refer to a contract on an exchange rather than a bilateral one (see below for more detail).

4.2 Discussion of modeling choices

(a) Definition of ring fencing

In the public debate, "ring fencing" mostly refers to the insulation of a relatively safe retail bank performing traditional activities from a riskier and harder-to-monitor investment bank that is part of the same financial entity. Indeed, the Vickers rule in the UK leaves the investment bank unregulated and just prevents it from "polluting" the regulated entity.

get the payment in court.

[30]Had we introduced pledgeable income, we could have alternatively interpreted the contracts as mutual credit lines, which the two parties commit to at date 0 (indeed, the signature of a committed credit line is that this credit line would not necessarily be granted ex post).

Our insights apply just the same when the unregulated entity is part of the same group and when it is independent from the retail bank.

(b) One-way hedging contracts

Insurance contracts need not be bilateral; there could be stand-alone contracts, like CDSs, in which case multiple contracts would be needed so as to achieve the required insurance. For example, a bank of group H (the grantee) can sign a liquidity-sharing contract with a counterparty bank of group T (the grantor), whereby the grantor pays 1 to the grantee in state m^T.

(c) Risk profile as a choice

In this version of the model, each bank is exogenously assigned a type (H or T). Alternatively, we could have assumed that each bank chooses its risk profile, namely whether to be an H or a T bank. Anticipating on the results, the logic behind the insight that the regulated sector must be ringfenced from the shadow sector remains the same under this moral-hazard version of the model, because the profile choice of a shadow bank reacts to the possibility of matching with a regulated bank of the same group. The insight about using CCPs also carries over: Assuming that constituting a portfolio of type H is more expensive (as expensive, cheaper) than building a portfolio of type T if a majority of (half, a minority of) banks choose the H option, the equilibrium in which banks divide equally between the H and T choices is still an equilibrium under ring-fencing and use of a CCP, while assortative matching still prevails when the regulator is unable to assert the banks' types even within the regulated sector.

(d) Ring-fencing vs. location of the bailinable debt

Recent reforms aiming at a substantial enlargement of the scope of bailinability (in the past *de facto*, although not *de jure*, circumscribed to equity) to most unsecured bank liabilities stress that the bailinable debt of banks should be held outside of the banking sector. This location of the bailinable debt differs in spirit from the idea of ring-fencing, and is more related to the avoidance of double counting (the old concern about "double gearing"): bailouts are not reduced by the bailinability of a bank A liability if it held by bank B, which due to the loss of money must in turn be bailed in (or reduce credits to SMEs). The idea therefore is that retail banks' distress spill over to the shadow banking sector, while ring-fencing aim at preventing the reverse.

(e) Verifiability of intermediate liquidity state

We have assumed that liquidity states m^H and m^T are verifiable and can be contracted on at date 0. This allows bankers to sign targeted debt and liquidity sharing contracts for these states. This also allows the government to avoid inefficiently extending LOLR support in these states. Basically, these assumptions decouple private and public liquidity decisions in the intermediate liquidity states m^H and m^T from those in states g and b that we have analyzed in the previous section. They allow enough flexibility to deliver the efficient pooling of liquidity under adequate regulation and monitoring, as well as the gaming of bailouts under inadequate regulation and monitoring, which are precisely the phenomena that we want to analyze in this section.

Because expositional simplicity requires the basic model to be a special case of this one, we keep the assumption that liquidity states g and b are contractually indistinguishable. However, assuming that only $\{m^H, m^T\}$ (and not m^H and m^T) is verifiable would have been equally consistent with this nesting logic. In that alternative formulation, the liquidity contracts would be based on coarser information and take the form of mutual credit lines in state $\{\omega_1, \omega_2\}$ with $\omega_2 \in \{m^H, m^T\}$, to be reimbursed at the end of date 1 if there is any cash left in the bank then. Anticipating a bit, the same gaming possibility would arise: when the regulator is unaware of correlations, two correlated-risk counterparties could write a "cross-insurance" contract that is not one, and leaves them with probability $1/2$ with no cash in state $\{G, \omega_2\}$ with $\omega_2 \in \{m^H, m^T\}$, triggering a bailout. This alternative treatment is more intricate, though, which explains our modeling choice.

4.3 Socially optimal liquidity sharing

Consider the liquidity states m^H and m^T. The socially efficient arrangement consists in: (a) the sharing of liquidity between two natural counterparties in the two states H and T; and (b) the absence of leverage in these states: $d^i_{\omega_1 m^H} = d^i_{\omega_1 m^T} = 0$ for $\omega_1 \in \{G, B\}$. In state m, each bank ends up with income 1, allowing it to cover its investment need without any government money. When banks that are natural insurance counterparties engage in such liquidity sharing, the occurrence of bank distress is minimized, allowing the government to economize on taxpayer money (leaving scope only for a bailout in state (G, b) in the absence of regulation and a possible LOLR rescue in state (B, b)). Thus, to reach maximal welfare, the regulator must not only induce banks to join the regulated sector and perhaps give them access to LOLR, but also ensure that the banks are hedged through proper risk-transfer schemes. This raises the issue of potential gaming of the hedging function.[31]

[31] It is interesting in this respect to look at the case of pure shadow banks. Pure shadow banks share liquidity in the bad, but not in the good fiscal state (or equivalently, for the good fiscal state they contract with a correlated-risk counterparty). This yields

$$U^{SB} = (1 - p_{Bb})\mathfrak{b} + 2p_{Gg} + p_{Gm} + p_{Bg}$$

Proposition 5. *(optimal risk sharing) Social welfare is maximized by*

(i) in states ω with $\omega_2 \in \{g, b\}$: the same allocation (i.e., the same leverage d_{ω_1} and the same public policy and liquidity assistance $\{k, l, m\}$) as when $p_{m^H} = p_{m^T} = 0$,

(ii) in states ω with $\omega_2 \in \{m^H, m^T\}$: setting $d_{\omega_1 m^H} = d_{\omega_1 m^T} = 0$ and arranging liquidity-sharing contracts among natural counterparties (i.e. matching a type H with a type T).

4.4 When is optimal liquidity sharing implementable?

Let us assume that the implementation of the optimum requires regulation. It may or may not involve LOLR for states (B, g) and (B, b). By contrast, it never involves LOLR for states (B, m^H) and (B, m^T). Recall also that we have assumed away special depositors for simplicity so that DI is an empty policy.

In order to implement optimal liquidity sharing, and thereby the full optimum, the regulator needs to regulate not only leverage but also the existence of liquidity sharing contracts by mandating that each bank enter a pair of bilateral liquidity sharing arrangements for states ω with $\omega_2 \in \{m^H, m^T\}$. However, this is not enough as banks face perverse incentives in the choice of their counterparty.

Ring-fencing and within-regulated-sector correlation monitoring

Consider first fiscal state B. In states (B, m^H) and (B, m^T), the banks want to be covered as there will be no bailout and no LOLR. So, for these states, natural counterparties spontaneously engage in an optimal liquidity-sharing arrangement and specify debt $d_{Bm^H} = d_{Bm^T} = 0$.

More interesting is fiscal state G. Suppose first that the regulator imposes to each regulated bank to limit its leverage to $d_{Gm^H} = d_{Gm^T} = 0$ and to enter a bilateral liquidity-sharing arrangement. As long as each bank picks a natural counterparty in the regulated sector, optimal risk sharing is implemented. But the banks have incentives to game this requirement and engage in bogus liquidity provision with correlated counterparties (see below). If the government is not careful along dimensions that we will make clear, banks will succeed in evading the insurance imperative, thereby preventing the implementation of the optimum.

---

and

$$W^{SB} = U^{SB} + (1 - p_{Bb})\beta - p_G(1 + \lambda_G).$$

Consider first the case where the government does not impose that the counterparties be in the regulated sector. It is then unable to monitor the correlation of a shadow bank counterparty with a regulated bank. A regulated bank has an incentive to sign a liquidity-sharing contract with a correlated-risk counterparty (i.e. of the same type) in the shadow banking sector for states (G, m^H) and (G, m^T). Consider for example the case where the regulated bank is of type H and picks a counterparty in the shadow banking sector of type H. The shadow bank increases its leverage to $d_{Gm^H} = 3$.[32] Their contract specifies a transfer of 1 to the shadow bank in state m^H and to the regulated bank in state m^T. The regulated bank has the "right contract with the wrong bank".

In state m^H, the regulated bank has revenue 2, does not pay anything to creditors (it is constrained to $d_{Gm^H} = 0$) and pays 1 to the shadow bank which combines this payment with its revenue of 2 to pay down its debt of 3. This represents *liquidity syphoning* from the regulated sector to the shadow banking sector. In state m^T, the shadow bank has no revenue and defaults on the payment of 1 that it must make to the regulated bank; both banks are then bailed out. This represents *bogus liquidity provision* by the shadow bank to the regulated bank. It increases the reliance on bailouts of the regulated bank.[33] By forming such a coalition and engaging in bogus liquidity provision, the banks generate a joint surplus of p_{Gm^H} at the taxpayer's expense which they can split at date 0 through appropriate transfers between themselves.[34] This shows that ring-fencing the regulated and shadow banking sectors is a necessary condition for the implementation of the optimum.[35]

The conclusion is therefore the following:

[32]As earlier, bailinability of unsecured claims implies that uninsured investors cannot receive more than there is in cash at the bank.

[33]The model exhibits a form of "liquidity syphoning" in that liquidity flows from the regulated sector to the shadow banking sector without any flow in the other direction. Liquidity syphoning and bogus liquidity are therefore two sides of the same coin. Our timing does not allow for transfers from the government to a regulated bank to directly leak to a shadow bank. However, it allows for an indirect form of leaking since liquidity syphoning and bogus liquidity increase the cost of bailouts in the regulated sector. In this sense, public liquidity is indirectly syphoned from the regulated sector to the shadow banking sector. Extensions of the model could capture direct public liquidity syphoning.

[34]The joint surplus is computed in reference to the counterfactual where the two banks are regulated and engage in contracts with natural counterparties, and where regulated banks are prevented from engaging in contracts with shadow banks through ring-fencing. Under this counterfactual, the utility of a regulated bank is equal to that of a shadow bank.

[35]Our purpose here is simply to point at the necessity of ring-fencing, and not to compute the exact deadweight loss associated with allowing liquidity sharing between regulated and shadow banks. Indeed, suppose that all banks are regulated; a bank that deviates and becomes a shadow bank can create competition among regulated banks to be its counterparty. Then all the private surplus generated by bogus liquidity and liquidity syphoning accrues to the shadow bank. And so the reservation utility of all banks if all banks are to be kept in the regulated sector is higher than when ring-fencing is prohibited. Again we do not solve for the optimal policy in the absence of ring-fencing.

Proposition 6. *(ring-fencing and correlation monitoring)*

(i) In the absence of ring-fencing, regulated banks enter bilateral liquidity-sharing contracts with correlated-risk shadow banks, leading to bogus liquidity and liquidity syphoning in states (G, m^H) and (G, m^T).

(ii) The optimum can be implemented via bilateral liquidity-sharing contracts between banks if (a) both parties to the hedging contract are in the regulated sector, and if (b) the regulator is able to monitor their correlation (though joint stress testing, say), so as to check that they are natural counterparties.

This leaves open the question, to which we turn next, of how the optimum can be implemented when instead the regulator cannot assess correlations even when the two banks both belong to the regulated sector.

Ring-fencing and absence of correlation monitoring: the role of CCPs.

Consider now the case where the government imposes ring-fencing by stipulating that regulated banks cannot sign liquidity-sharing contracts with shadow banks, but that it does not observes type (H or T) even in the regulated sector, and so cannot monitor correlations within the regulated sector.

Our assumptions that leverage is regulated and that taking money out of the bank for private purposes would constitute an abuse of corporate assets leave little scope for moral hazard under regulation in our model. To reintroduce some incentive to game the cross-insurance mechanism, we assume that excess cash in the bank can be wasted and produce (arbitrarily small) private benefits for the banker. Namely, assume that bankers at date 1 can consume free cash flow[36] and that they value a unit of free cash flow available at the end of date 1 at $\varepsilon < 1$, where ε can be arbitrarily small, but strictly positive. This relaxes the extreme assumption in our baseline model that bankers value only continuation but not cash at date 1 and does not alter our previous analysis.

A regulated bank then has an incentive to sign a liquidity-sharing contract with a correlated-risk counterparty in the regulated banking sector in state (G, m).[37] In the absence of LOLR, the only difference with the case considered above is that the counterparty

[36] At the end of period 1, say, provided that there was no inspection.

[37] This only proves that the absence of monitoring of such contracts leads to gaming. The regulator might impose that the same liquidity-sharing contract apply to both $\{G, m^H\}$ and $\{G, m^T\}$ on the one hand and to $\{B, m^H\}$ and $\{B, m^T\}$ on the other hand so as to impose a cost on such gaming: Choosing a correlated-risk counterparty then would lead to no investment with probability $1/2$ in the latter states. However this would not discourage correlated-risk counterparties to match as long as $p_{Bm^H}/p_{Gm^H} = p_{Bm^T}/p_{Gm^T}$ is not too large.

cannot increase its leverage because it is regulated. Consider for example the case of a regulated bank of type H signing a liquidity-sharing contract with a regulated bank of type H in states (G, m^H) and (G, m^T). In state (G, m^H), each bank has revenue 2; one of them pays 1 to the other, and is left with 1, which is then used for investment as d_{Gm^H} is constrained to be 0. The other bank has a surplus of 2 after investment. In state (G, m^T) both banks default on the payment due to the other because of the absence of revenues (another case of bogus liquidity) and both banks are bailed out. By forming such a coalition and engaging in bogus liquidity provision, the banks generate a joint surplus of $2p_{Gm^H}\epsilon$ at the taxpayer's expense.[38] This shows that under bilateral contracts, monitoring correlations in the regulated sector is a necessary condition for the implementation of the optimum.

An alternative arrangement to implement optimal liquidity sharing is to set up a CCP. Banks in the regulated sector are forced to participate in the CCP and banks in the shadow banking sector are banned from participating (ring-fencing). More precisely, banks are forced to enter a bilateral liquidity-sharing contract with the CCP. The CCP centralizes liquidity sharing. At date 1, member banks are called upon to provide 1. Those which do not contribute and demand liquidity assistance are audited and receive 1 if they indeed are missing cash. In equilibrium half of the regulated banks are of type H while the other half is of type T, so the CCP is indeed able to perform the intermediation. Shadow banks are banned from signing liquidity-sharing contracts with the CCP and with regulated banks.

This arrangement guarantees the efficient distribution of liquidity within the regulated sector and eliminates bogus liquidity by preventing banks from fine-tuning their liquidity provision at the expense of the taxpayer. The key is that the CCP removes the counterparty risk that banks are endogenously generating by picking correlated counterparties. Should shadow banks be prevented from participating in the CCP by imposing ring-fencing blocks any syphoning of liquidity to the shadow banking sector and any bogus liquidity provision by the shadow banking sector? Using the same institutional arrangement as described above, shadow banks could indeed join the CCP scheme provided that they can be audited if they refuse to contribute to the CCP. In practice, though, the inspection might be harder than for a regulated bank, as no data has been accumulated by the regulator about the shadow bank prior to the liquidity sharing event.

The conclusion is therefore the following:

[38] As above, the joint surplus is computed in reference to the counterfactual where the two banks are regulated and engage in contracts with natural counterparties, and where regulated banks are prevented from engaging in contracts with shadow banks through ring-fencing.

Proposition 7. *(CCPs)*

(i) With ring-fencing but in the absence of correlation monitoring in the regulated sector, regulated banks enter bilateral liquidity-sharing contracts with correlated-risk regulated banks in which there is bogus liquidity in states (G, m^H) and (G, m^T).

(ii) If correlations within the regulated sector cannot be observed, an alternative implementation can be used to reach the optimum: mandating participation in a CCP in the regulated sector.

4.5 Observed heterogeneity and coexistence of the two sectors

The arguments above apply in a model where a representative fraction $1 - \alpha$ of banks have a monitoring cost equal to ∞. Banks with infinite monitoring cost operate in the shadow banking sector.

We continue to denote by (m, l) the configuration applying to banks with finite monitoring cost c. Under the same assumptions as above, the optimal arrangement has the following features: banks with finite monitoring costs are monitored in the regulated sector, benefit from LOLR, and pool their liquidity among natural insurance counterparties in aggregate states ω with $\omega_2 \in \{m^H, m^T\}$; and banks with infinite monitoring costs are not monitored in the shadow banking sector, do not benefit from LOLR, and pool their liquidity among natural insurance counterparties only in aggregate state (B, m^H) and (B, m^T).

As in Proposition 6, the underlying liquidity arrangement can be implemented by mandating regulated banks to enter bilateral liquidity-sharing contracts between natural counterparties within the regulated sector. This requires both ring-fencing and either the monitoring of correlations or the use of CCPs within the regulated sector.

4.6 Shadow banking as diversification and the costs of liquidity segregation

As in Section 3.4, one can imagine that banks are heterogeneous in their activities and so logically are not subject to exactly the same shocks. Heterogeneous activities also can help justify the existence of a shadow banking sector, which so far was a pure nuisance for the social planner (shadow banking defined the banks' reservation values and augmented their rents). In this case, shadow banking is socially useful, but is still a constraint on what the regulator can achieve. Ring-fencing on the other hand limits liquidity pooling and so therefore now comes with a meaningful tradeoff.

We accordingly modify the setup to point out a potential cost of segregating liquidity across sectors. Consider the model with a fraction $1-\alpha$ of banks with an infinite regulation cost. The government knows at date 0 who are these natural shadow banks. We modify the stochastic structure of the economy, only in liquidity state m. We assume that the ex-ante type of a bank is perfectly correlated with its monitoring cost.[39]

This means that there is no scope for liquidity sharing in state m neither within the regulated sector nor within the shadow banking sector. Liquidity sharing can only be implemented across the regulated and shadow banking sectors via cross exposures. When $\alpha < 1/2$, banks with finite monitoring costs are on the short side of the liquidity market. In this case, there is no change in the outside option of operating in the shadow banking sector. Mandating liquidity sharing for regulated banks via bilateral liquidity sharing with shadow banks increases welfare by reducing the fiscal cost of bailouts in both sectors.[40] When $\alpha > 1/2$, there is a tradeoff: Mandating liquidity sharing via bilateral liquidity-sharing contracts (with potential ex-ante transfers) across sectors reduces welfare by increasing the outside option of operating in the shadow banking sector on the one hand, but increases welfare by reducing the fiscal cost of bailouts in both sectors.[41]

5 Relationship to the literature

There are widely different views, both among economists and in the policy debate, about the social merits of shadow banking. The most positive view states that regulatory constraints stifle innovation, limit lending and distort markets; shadow banking then offers some breathing room and undoes a state failure. See for example Ordoñez (2018) for an elaboration of this point in a model where banks are asymmetrically informed about their investment opportunities, and where migration into the shadow banking sector provides a way for the banks with the best opportunities to pursue them by avoiding blunt regulation.[42]

[39] Implicitly, this assumption de facto ensures that banks with low monitoring costs can operate both in the regulated and in the shadow banking sectors while banks with high monitoring costs can only operate (at finite cost) in the shadow banking sector.

[40] Note that this requires checking that shadow banks do not grant multiple credit lines, which might be challenging in practice given that the balance sheets of shadow banks are not regulated.

[41] The government can always adjust transfers to implement the equilibrium where banks with a finite regulation cost do not migrate to the shadow banking sector. To go beyond implementation and guarantee unique implementation, the government may need to allow transfers to depend on the mass of banks in the shadow banking sector.

[42] Feve et al. (2019) provide some evidence of such migration in response to higher capital requirements. See also Buchak et al. (2018) who study the rise of fintech and non-fintech shadow banks in the residential lending market and find that financial technology innovation can account for about 35% of shadow bank

Different strands of the academic literature articulate a more negative view. One branch of the literature stresses regulatory arbitrage: Shadow banking is then a (perhaps unavoidable) nuisance. The regulatory arbitrage view includes two possible subviews. In the first, retail banks evade capital requirements by providing liquidity support off-balance-sheet to shadow banks; Acharya et al (2013) find evidence that such regulatory arbitrage was a key motive behind setting up ABCP conduits, as losses from conduits remained with retail banks[43]. The underpricing of this absence of effective risk transfer was corrected by Basel 3, which put the corresponding exposures back on the retail bank's balance sheet. The second subview, spelled out for example in Acharya-Richardson (2009) and Claessens et al. (2012), involves capital requirement "evasion" by shadow banks, which face no capital adequacy requirement and yet receive public assistance. Shadow banks cut regulatory corners and have their cake and eat it too: They are free of constraints in normal times, and are bailed out if tail risk materializes.[44] Perhaps consistent with this view, Buchak et al. (2018) also finds that the migration to shadow banking induced by the increasing regulatory burden faced by traditional banks account for 55% of shadow bank growth over the same period.

Another branch of the literature stresses behavioral factors: Shadow banks exploit neglected risk. Gennaioli et al (2012, 2013, 2015) assume that investors overweigh a favorable scenario upon good news and similarly overreact when bad news occur. Shadow intermediaries create false substitutes for truly safe bonds. Financial crises can be triggered by the repricing of risk following the sudden realization of the true risks embedded in these pseudo-safe assets. In Farhi-Tirole (2020), shadow banks can create relatively (but not entirely) safe assets via financial engineering to attract special depositors but without exploiting the behavioral biases of the latter.

Finally, a last branch of the literature emphasizes comparative advantage.[45] For example, in Hanson et al (2015), households are willing to pay a premium for safe assets, as in Stein (2012).[46] Safe assets can be created in two ways; in the regulated sector through deposit insurance offered by the state in exchange of costly capital requirements; by an

growth over the period 2007-2015.

[43]See also Gorton-Metrick (2010) and Pozsar et al (2013).

[44]In the context of these two subviews, Farhi-Tirole (2012, 2018) and Di-Iasio-Pierobon (2012) emphasize strategic complementarities in regulatory arbitrage arising from a security in numbers due to the fact that bailouts are imperfectly targeted.

[45]See e.g. Perotti (2014) for an early policy discussion.

[46]The demand for safe assets also figures prominently in Diamond (2019)'s theory of segmentation. In Diamond, firms tranche their liabilities so as to create relatively safe assets (debt), which are then held by banks. Banks transform these assets into really safe assets (deposits) through an equity add-on. In our model, only the state can create safe assets, but it finds it cheaper to do so if banks themselves hold relatively safe assets. The state then optimally piggybacks on the banks' balance sheets to do so.

early exit option and the costly liquidation of assets in the shadow sector. In equilibrium, shadow banks therefore hold relatively liquid assets. The paper does not analyze optimal regulation, but identifies an externality in the unregulated sector, due to fire sales. This externality creates a tendency for the shadow banking sector to be too large compared to the regulated sector. Chrétien-Lyonnet (2019) pursue this logic by assuming that rather than outside investors, it is banks in the regulated sector that purchase the assets that are liquidated by shadow banks, and that they do so using cheap insured deposits. They study the resulting interactions between the two sectors. Relatedly, Gertler et al. (2016) build a model in which wholesale shadow banks borrow from regulated retail banks which in turn raise deposits from households. In their model, the relative size of the two sectors is determined by a tradeoff between assumed comparative advantages of wholesale banks in managing assets and of retail banks in overcoming agency frictions in fund borrowing. In a different vein, Moreira-Savov (2017) emphasize the coexistence of money (securitization products that are safe and liquid all the time) and shadow money (securitization products that are safe and liquid most of the time). In their model, compared to money, shadow money economizes on collateral but is more fragile. Periods of low uncertainty are associated with expansions in shadow money and economic booms, which come to an end when uncertainty increases, shadow money collapses, and the economy tanks.

Our model incorporates elements of these different branches of the literature. At its core is a problem of regulatory arbitrage, along the lines of the two corresponding subviews mentioned above: Shadow banks avoid the capital requirements of the regulated sector and yet receive some public support in the form of bailouts; banks in the regulated sector must also be prevented by regulation from extending liquidity support to shadow banks. An extension of our model (see Sections 3 and 4) also incorporates a notion of comparative advantage: Some activities are simply too costly to regulate, perhaps because they are too complex, and so they are better performed by the shadow banking sector. Moreover, to the extent that the risks of the shadow banking sector are not perfectly correlated to those of the regulated sector, allowing for the two sectors not only to co-exist, but also to share some risks, is desirable (see Section 4).

Few papers study optimal regulation in the presence of a shadow banking sector. Beguenau-Landvoigt (2018) solve for optimal capital requirements in a quantitative model where banks can migrate to the shadow banking sector in the presence of exogenous bailouts occurring with a higher probability in the regulated sector than in the shadow banking sector. The idea that regulation must account for the possibility of migration of banking activities can be found in earlier papers.[47] For example, Grochulski-Zhang (2014)

[47] See e.g. Hanson et al. (2011) for an early policy discussion.

analyze a model à la Diamond-Dybvig (1983), where regulation is motivated by a pecuniary externality arising from the possibility of private re-trades among banks as in Farhi et al. (2009), and introduce shadow banking as a nuisance in the form of a participation constraint which limits the scope of regulation. Similarly, Plantin (2015) sets up a model where a bank engages in excessive risk-taking and evades regulatory risk-monitoring through securitization and the granting of lines of credit to the resulting conduits. In his regulatory-evasion model, shadow banking is therefore a nuisance, and he shows that tightening capital requirements may spur a surge in shadow banking activity and reduce welfare. Harris et al. (2014) emphasize a different perverse effect of tighter regulation, namely that increased capital requirements can actually induce risk shifting in the regulated banking sector because of bailouts and because the competition of shadow banks is more intense for safe positive net-present-value projects than for risky negative-present value projects. In a different vein, Bengui-Bianchi (2014) analyze the optimal design of capital controls in a small open economy with pecuniary externalities when some possibility of evasion exists. In their model, tighter capital controls curb risk-taking in the regulated sector, increase it in the unregulated sector, and are overall desirable.

Elements of the quadrilogy are embodied in general equilibrium models, such as Begenau (2019)'s calibrated estimation of the optimal capital adequacy requirement. In the latter, banks issue risk-free deposits, supply credit and engage in risk-taking.[48] Her main point is that stricter capital requirements may increase the supply of credit: while the direct, partial equilibrium effect is a reduction in credit, an increase in capital requirement reduces deposits and, through a substitution effect, the cost of unsecured borrowing (labelled "equity" in her framework), raising the possibility of a credit expansion.

Our theory is unique in explaining the complementarities between regulation, LOLR, and DI, and in showing how the optimal deployment of these attributes endogenously gives rise to a regulated banking sector associated with the aforementioned attributes and a shadow banking sector devoid of them. Relative to the existing literature, our paper also makes forays into two new areas: the complementarity between the four classic markers of traditional banking, and the use of ring-fencing and CCPs, adding two further markers. Finally, our paper emphasizes and distinguishes between bank bailouts and investor bailouts.

[48]Government subsidies are modeled in a reduced-form fashion as transfers that increase in the size, leverage and losses of the banking sector.

6 Conclusion

We studied the optimal regulation of banks when supervision reduces moral hazard and the riskiness of balance sheets and financial intermediaries can migrate to shadow banking in response to regulatory requirements. We did not posit that shadow or retail banks had a comparative advantage, and rather derived differences in their behavior from equilibrium considerations.

Our main insight is the complementarity between regulation and the forms of insurance provided by the state: LOLR to banks and deposit insurance to depositors. Insurance is costly and supervision helps reduce the risk that its promises are called upon. Our analysis makes room for both bank and depositor implicit and explicit guarantees. Second, we provide the first formal rationale for ring-fencing and for incentivizing the migration of transactions towards CCPs. To this purpose, we showed how imperfect regulatory information may lead to gaming of hedging among financial intermediaries, resulting in banks being only partially covered as they hoard bogus liquidity and in public liquidity being syphoned off to the shadow sector. Overall the picture emerging from the analysis is an hexalogy: prudential supervision of banking goes hand in hand with servicing special borrowers (SMEs) and special lenders (retail depositors), LOLR, deposit insurance, incentivized migration to CCPs and ring-fencing.[49]

There are many alleys for future research. For example, our model has logically led to a focus on public supervision as the externalities were on public finances. In practice, monitoring is performed both by the public sector (banking supervision) and by the private sector (holders of shares and other bailinable securities, rating agencies), and in both cases it is potentially subject to moral hazard and capture. So it may be useful to look at the respective stakes, and to derive the optimal pattern of monitoring in richer environments. In our core model in Section 3, there are no private incentives to monitor because there is no way for banks to dilute creditors. A previous version of this paper considered a different model with moral hazard in the choice of riskiness of bank projects as opposed to moral hazard in leverage. We showed that there were private incentives to monitor, but that private monitoring incentives were likely weaker than public monitoring incentives because of fiscal externalities.

Another issue relates to universal banks and the choice between regulating institutions vs. regulating activities. We have not analyzed the question of banks involved in different

[49]These insights are likely to apply more broadly to systemically important functions such as payment systems, market making, etc. In particular, the payment system is based on trust that itself relies on the existence of relatively safe claims, obtained either from collateral-taking or from the government's presence as a regulator and lender of last resort.

activities, some traditionally thought of as belonging to the regulated sector and some traditionally thought of as belonging to the shadow banking sector. Our model suggests an argument for regulating institutions rather than activities to the extent that liquidity can be reallocated more easily inside universal banks than through arms-length financial transactions. Opening-up the black box of financial institutions and tackling the question of firm boundaries is an important area where future research will be needed.[50]

Needless to say, the sharp picture obtained in the paper is only meant to stress natural covariations. Reality as always is more complex than the model suggests. The unique features associated with the traditional banking sector themselves impose costs, leading to a finer overall picture. We hope that future work will sharpen this analysis.

[50] First attempts at studying the costs and benefits of separation of traditional lending and investment banking and therefore the merits of universal banking are provided by Shy and Stenbacka (2017) and Vickers (2017).

References

Acharya, V., and M. Richardson, eds., (2009) *Restoring Financial Stability: How to Repair a Failed System*, John Wiley and Sons.

Acharya, V., Schnabl, P., and G. Suarez (2013) "Securitization without Risk Transfer," *Journal of Financial Economics*, 107(3): 515–536.

Beguenau, J., and T. Landvoigt (2018) "Financial Regulation in a Quantitative Model of the Modern Banking System," Working Paper 3558, Stanford Graduate School of Business.

Beguenau, J. (2019) "Capital Requirements, Risk Choice, and Liquidity Provision in a Business-Cycle Model," forthcoming *Journal of Financial Economics*.

Bengui, J., and J. Bianchi (2014) "Capital Flow Management when Capital Controls Leak," Minneapolis Fed Working Paper.

Berg, T., and J. Gider (2017) "What Explains the Difference in Leverage between Banks and Non-Banks?" *Journal of Financial and Quantitative Analysis* 52(6): 2677–2702

Bernheim, D. Peleg, B. and M. Whinston (1987) "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, 42(1): 1–12.

Buchak, G., Matvos, G., Piskorski, T., and A. Seru (2018) "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks," *Journal of Financial Economics*, 130(3): 453-483.

Caballero, R.J., and E. Farhi (2018) "The Safety Trap," *Review of Economic Studies*, 85: 223-274.

Chrétien, E., and V. Lyonnet (2019) "Traditional and Shadow Banks," Working Paper 2019-11, Fisher College of Business.

Claessens, S., Z. Pozsar, L. Ratnovski and M. Singh (2012) "Shadow Banking: Economics and Policy," IMF Staff Discussion Note, 12/12, http://dx.doi.org/10.2139/ssrn.2187661.

Di Iasio, G., and F. Pierobon (2012) "Shadow Banking, Sovereign Risk and Collective Moral Hazard," http://dx.doi.org/10.2139/ssrn.2060839.

Diamond, D. W., and P.H. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91(3): 401–419.

Diamond, D. W. (1984) "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51(3): 393–414.

Diamond, W. (2019) "Safety Transformation and the Structure of the Financial System," Research Papers Series, The Wharton School, University of Pennsylvania.

Farhi, E., Golosov, M., and A. Tsyvinski (2009) "A Theory of Liquidity and Regulation of Financial Intermediation," *Review of Economic Studies*, 76(3): 973–992.

Farhi, E., and J. Tirole (2012) "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts," *American Economic Review*, 102(1): 60–93.

Farhi, E., and J. Tirole (2018) "Deadly Embrace: Sovereign and Financial Balance Sheets Doom Loops," *Review of Economic Studies*, 85(3): 1781–1823.

Farhi, E., and J. Tirole (2020) "Clientele-Dependent Asset Pricing," forthcoming mimeo.

Feve, P., Moura, A., and O. Pierrard (2019) "Financial Regulation and Shadow Banking: A Small-Scale DSGE Perspective," *Journal of Economic Dynamics and Control*, 101: 130–144.

Gennaioli, N., Shleifer, A., and R. Vishny (2012) "Neglected Risks, Financial Innovation, and Financial Fragility," *Journal of Financial Economics*, 104(3): 452–68.

Gennaioli, N., Shleifer, A., and R. Vishny (2013) "A Model of Shadow Banking," *The Journal of Finance*, 68: 1331–1363.

Gennaioli, N., Shleifer, A., and R. Vishny (2015) "Neglected Risks: The Psychology of Financial Crises," *American Economic Review: Papers & Proceedings*, 105(5): 310–314.

Gertler, M., Kiyotaki, N and A. Prestipino (2016) "Wholesale Banking and Bank Runs in Macroeconomic Modelling of Financial Crises," *Handbook of Macroeconomics*, edited by John B. Taylor and Harald Uhlig, 2: 1345–1425, chapter 16.

Gorton, G., and A. Metrick (2010) "Regulating the Shadow Banking System," *Brookings Papers on Economic Activity*, 41: 261–312.

Grochulski, B., and Zhang, Y. (2015) "Optimal Liquidity Regulation with Shadow Banking," FRB Richmond Working Paper No. 15-12, http://dx.doi.org/10.2139/ssrn.2673540.

Hanson, S., Kashyap, A.K., and J. Stein (2011) "A Macroprudential Approach to Financial Regulation," *Journal of Economic Perspectives*, 25(1): 3–28.

Hanson, S., Shleifer, A., Stein, J. and R. Vishny (2015) "Banks as Patient Fixed-Income Investors," *Journal of Financial Economics*, 117: 449–469.

Harris, M., Opp, C. and M. Opp (2014) "Higher Capital Requirements, Safer Banks? Macroprudential Regulation in a Competitive Financial System," Wharton Working Paper (revised 2017).

Holmström, B., and J. Tirole (1997) "Financial Intermediation, Loanable Funds and the Real Sector," *Quarterly Journal of Economics*, 112(3), 663–691.

Holmström, B., and J. Tirole (1998) "Private and Public Supply of Liquidity," *Journal of Political Economy*, 106(1): 1–40.

Jin, X., and F. Nadal de Simone (2016) "Tracking Changes on the Intensity of Financial Sector?s Systemic Risk," BCL WP 102.

Kashyap, A.K., Rajan, R., and J.C. Stein (2002) "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking," *Journal of Finance*, 57(1): 33–73.

Moreira, A., and A. Savov (2017) "The Macroeconomics of Shadow Banking," *Journal of Finance*, 72(6): 2381–2432.

Ordoñez, G. (2018) "Sustainable Shadow Banking," *AEJ: Macroeconomics*, 10(1): 33–56.

Peltzman, S. (1989) "The Economic Theory of Regulation after a Decade of Deregulation," *Brookings Papers: Microeconomics*, 1989: 1–59.

Perotti, E. (2014) “The Roots of Shadow Banking,” in Claessens, Stijn et al. (eds), *Shadow Banking*, Federal Reserve Bank of Chicago International Banking Series.

Plantin, G. (2015) “Shadow Banking and Bank Capital Regulation,” *Review of Financial Studies*, 28(1): 146–175.

Pozsar, Z., T. Adrian, A. Ashcraft and H. Boesky (2013) “Shadow Banking,” *Economic Policy Review*, 19(2): 1–16.

Shin, H. (2012) “Global Banking Glut and Loan Risk Premium,” *IMF Economic Review*, 60: 155–192.

Shy, O., and R. Stenbacka (2017) “Ring-Fencing, Lending Competition, and Taxpayer Exposure,” available at SSRN: https://ssrn.com/abstract=2874316 or http://dx.doi.org/10.2139/ssrn.2874316.

Stein, J. (2012) “Monetary Policy as Financial Stability Regulation,” *Quarterly Journal of Economics*, 127: 57–95.

Vickers, J. (2017) “Economics of Structural Reform of Banks,” Thys Visser Lecture 3.

Yorulmazer, T. (2013) “Has Financial Innovation Made the World Riskier? CDS, Regulatory Arbitrage and Systemic Risk,” discussion paper.