Shadow Banking and the Four Pillars of Traditional Financial Intermediation*

Emmanuel Farhi† and Jean Tirole‡

December 11, 2018

Abstract

Traditional banking is built on four pillars: SME lending, deposit taking, access to lender of last resort and deposit insurance, and prudential supervision. This paper unveils the logic of the quadrilogy by putting core services to "special depositors and borrowers" at the heart of the analysis, and makes room for bank and depositor implicit and explicit guarantees. It analyzes how prudential regulation must adjust to the emergence of shadow banking. The model also rationalizes structural remedies to counter syphoning and financial contagion: ring-fencing between regulated and shadow banking and the sharing of liquidity in centralized platforms.

Keywords: Retail and shadow banks, lender of last resort, deposit insurance, supervision, migration, ring-fencing, CCPs.

JEL numbers: E44, E58, G21, G28.

*The authors are grateful to Aimé Bierdel, Mathias Dewatripont, Veronica Guerrieri, Frédéric Malherbe, Paul-Henri Moisson, John Moore, Francisco Nadal de Simone, Guillermo Ordonez, Jeremy Stein, Joseph Stiglitz, John Vickers, four referees, and participants at the 3rd Bank of England-CCBS macro-finance conference, Banque de France-TSE Prize conference, Shanghai Frontiers of Macroeconomics with Applications in China conference, LACEA-LAMES Buenos Aires conference, Bocconi, Central Bank of Luxembourg, 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).

†Harvard University and NBER.

‡TSE and IAST.

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 definition of retail banking (and by default of shadow banking) with exceptions has received little examination. Yet, the access to the "public insurance services"- the discount window and other liquidity facilities on the one hand and cheap deposits on the other - can be priced and could be 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 shadow banks have in recent years 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/deposit taking and deposit insurance/regulation/lender of last resort quadrilogy?

To start answering this question, we unveil the logic of the conventional wisdom, which has traditionally been taken for granted. While our work is therefore intrinsically normative, we build on experience to identify the hazards facing financial systems: over-leverage, threat of migration to shadow banking, cross exposures between retail and shadow banking sectors, bailouts. 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 first argues 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 state 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 state to raise funds even in financial straits). Regulation 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. 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 state. The attractive pricing of this insurance in turn incentivizes banks to remain in the regulated sector instead of migrating to the shadow banking sector. For example, 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 has three 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 debt obligation. 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. Finally, date 2 as usual stands for "the future".

In the tradition of prudential control of capital adequacy, we define regulation as the ability of the regulator to restrict leverage. Our model's rationale for such regulation is that unregulated banks choose a high leverage, and that this leverage exerts a negative externality on public finances in two ways. First, the bank's inability to lend to the industry ex post justifies a bailout even though the state puts no welfare weight on banks themselves in our model. Second, if the bank has issued demand deposits at date 0, the state has to make good on these deposits when the bank cannot pay the depositors.

An important distinction here is between bailouts and insurance.[1] Bailouts arise when

[1] LOLR is often described as following Bagehot's dictum: To avert panics, central banks should lend

the state would like to commit not to rescue the bank or its depositors, but cannot refrain from doing so when the bank is illiquid. By contrast, "public insurance services" (LOLR for banks and DI for depositors) are contractual features[2] in which the state finds its own interest: 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. The first key result of the paper is that this quid-pro-quo is optimal due to a complementarity between prudential regulation and the provision of public insurance services. Such services are costly and prudential supervision reduces their cost.

Date-1 rescues involve both a benefit and a cost. The benefit for the state is that it keeps the real economy going (in our model, the state does not care about the bankers per se). This is where SME lending enters the picture: Absent SME lending, the state would have no incentive to bail out banks, and there would be no moral hazard and no market failure.[3] The cost is that bailouts create date-0 over-leverage.

The paper's second contribution is to develop a rationale for the two concepts of 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 in which either a regulated bank is only partially covered by its insurance counterparty and therefore holds

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. We could capture the notion that LOLR is only deployed against good collateral by introducing a new dimension on moral hazard which cannot be eliminated by regulation, whereby banks can take an unobservable action that increases or preserves the value of their collateral. In order to provide incentives for banks to take this action, the implicit LOLR contract would require committing to bail out banks only if the value of their collateral is high enough.

[2] 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.

[3] 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.

[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.

"bogus liquidity" (as was de facto offered by AIG in the CDS market[5]) or public liquidity is "syphoned off" to benefit a shadow banking entity (which happened when retail banks offered credit lines to the conduits they had created).

We first assume 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 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 or LOLR 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.

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.[6]

[5]The notion of "bogus liquidity" is documented by Yorulmazer (2013), 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).

[6]Feve-Pierrard (2017) provide some evidence of such migration in response to higher capital requirements. See also Buchak et al. (2017) 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 growth over the period 2007-2015.

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[7]. 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.[8] Perhaps consistent with this view, Buchak et al. (2017) 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 (2019), 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.[9] For example, in Hanson et al (2015), households are willing to pay a premium for safe assets, as in Stein (2012).[10] 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 early exit option and the costly liquidation of assets in the shadow sector. In equilibrium,

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

[8] 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.

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

[10] The demand for safe assets also figures prominently in Diamond (2017)'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.

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 (2017) 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 (2017) 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.[11] For example, Grochulski-Zhang (2014) analyze a model à la Diamond-Dybvig (1983), where regulation is motivated by a pecu-

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

niary 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.

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.

2 Model

There are three periods indexed by $t = 0, 1, 2$, a single good, and three classes of players: investors, bankers, and the government (or "state", or "regulator"). At date 1, all uncertainty is resolved with the realization of an aggregate state (ω, χ), where $\omega \in \{G, B\}$ is the fiscal state and $\chi \in \{NI, WI\}$ is the liquidity state for the banks. The fiscal state is good ($\omega = G$) with probability p_G or bad ($\omega = B$) with probability p_B. The liquidity state can be "No Illiquidity" ($\chi = NI$) or "Widespread Illiquidity" ($\chi = WI$); contingent on the realization of the fiscal state $\omega \in \{G, B\}$, the probability of the former is x_ω and that of the latter $1 - x_\omega$. We describe the meaning of these states below. There is no store of value in the economy.

Investors. There are two kinds of investors.[12] "Ordinary" investors/consumers have risk-neutral preferences with no discounting over consumption. Their utility is given by $\mathbb{E}[c_0^I + c_1^I + c_2^I]$. They have large endowments in every period. "Special depositors" are formalized as in Gennaioli et al (2012, 2013), Stein (2012) and Caballero-Farhi (2017): 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.[13] Importantly, these deposits must be absolutely safe in order to be valued.[14] There is a mass $\mu \leq 1$ of special depositors.

Since bankers experience aggregate liquidity shocks 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 will 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.

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)*.

Bankers. There is a mass 1 of bankers protected by limited liability. Each banker 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

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

[13]See also Malherbe-McMahon (2017) for a model with risk-averse households where only bank equity and deposits are traded, and where deposit insurance leads to risk-shifting.

[14]The exact microfoundations are as follows. Special depositors live for two periods, at dates 0 and 1. Their utility at date 0 is given by a form of recursive utility given by $c_0^{SD} + (1+\theta)\min_0\{\min\{c_1^{SD}, 1\}\}$, where the notation $\min_0$ denotes a minimum over states of the world and $\theta > 0$. Their utility at date 1 is given by $\min\{c_1^{SD}, 1\}$. Special depositors are infinitely risk averse but have an infinite intertemporal elasticity of substitution up to a satiation point at which it becomes zero. See the conclusion for a discussion of alternative modeling of special depositors.

by special knowledge or expertise.[15] 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. The bank's date-1 revenue is random. It depends on the realization of the liquidity state χ and is denoted by r_χ, with $r_\chi = 2$ if $\chi = NI$ and $r_\chi = 0$ if $\chi = WI$.[16]

At date 0, each bank i sets its leverage $\{d_{(\omega,\chi)i}\}$. This debt, to be reimbursed at date 1, is contingent on both ω and χ and must be positive $d_{(\omega,\chi)i} \geq 0$. The debt is honored (or not) at date 1 once revenue is realized, and before the investment decision. We assume that creditors have no bargaining power ex post and are therefore negotiated down to their outside option in bankruptcy. The actual repayment in state (ω, χ) is therefore $\min\{d_{(\omega,\chi)i}, r_\chi\}$. Without loss of generality, we thus assume throughout the paper that $0 \leq d_{(\omega,NI)i} \leq r_{NI} = 2$ and $d_{(\omega,WI)i} = r_{WI} = 0$ for $\omega \in \{G, B\}$.

In the absence of regulation (shadow banking), the banker freely chooses leverage. We thus capture the idea that unmonitored balance sheet decisions may make the bank illiquid and unable to pursue its activities in the absence of a bailout; of course such illiquidity may also occur even in the absence of leverage (as long as $x_\omega < 1$).

The project's payoff comes in the form of a private benefit $b > 1$ for the banker. Bankers value consumption at date 0 and the private benefit from investment according to

$$U \equiv \mathbb{E}[c_{0i}^B + bj_i]$$

where $j_i \in \{0, 1\}$ is an indicator for the realization of the project of banker i.[17]

Our assumption that revenue shocks are perfectly correlated across bankers ensures that bankers cannot obtain insurance from financial claims on other bankers, and that the government cannot economize on taxpayer money by forcing banks to co-insure. We relax this assumption in Section 4.

[15] 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.

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

[17] We could add some date-2 pledgeable income and an incentive payment for the banker, but that does not make any qualitative difference to the analysis. That is, we could alternatively posit that the bankers need to be incentivized through a monetary incentive scheme so as to exert some date-1 effort or to not take a private benefit at that date. 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 state monitors banks that accept to be in the regulated sector. Monitoring bank i means controlling its leverage $\{d_{(\omega,\chi)i}\}$ (a form of capital requirement) and costs $c \geq 0$.[18]

Second, the state 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. We assume that the collection of taxes for these purposes is costly: collecting 1 involves a *cost of public funds* equal to $1 + \lambda_\omega$, where the deadweight costs of taxation satisfy $\lambda_B > \lambda_G \geq 0$. By contrast rebating one unit of revenue to consumers does not entail any efficiency benefit.[19]

The state contracts at date 0 on public insurance services for date 1. At this stage, it can do so 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. Similarly, DI is a promise to make up for any shortfall in the deposit contract's reimbursement of 1.[20]

Next we describe the government's preferences. It 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: See footnote 22 below), but no weight on bankers, with a resulting weight[21] on the realization of a project of β with

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

This assumption guarantees that bankers hit by a revenue shock receive a discretionary bailout in state G but not in state B. This also means that ex post at date 1 in state B, the

[18]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.

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

[20]SME lending is coupled with deposit taking because the revenues associated with the banking activity can be pledged to special depositors, thereby limiting the need to rely on government funds which come together with a distortionary cost to the cases where SME revenues fall short and DI needs to be activated. Narrow banking would require more reliance on government funds since special deposits would be covered more extensively through distortionary taxation.

[21]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.

government would like to renege on any promise of bailout that would be part of the LOLR scheme, and is only prevented from doing so by its commitment at date 0. This social welfare function can be given credit-crunch foundations.[22]

Let us define the expected deadweight losses associated with one unit of public transfer to banks (or via the bank to depositors), contingent on bank revenue being equal to 2 and 0 respectively:

$$\Lambda_2 \equiv \Sigma_\omega p_\omega x_\omega \lambda_\omega$$

and

$$\Lambda_0 \equiv \Sigma_\omega p_\omega (1 - x_\omega) \lambda_\omega.$$

Payoffs. We denote by $j_i(\omega, \chi, \mu_i, d_{(\omega,\chi)i})$ bank i's investment if the state is (ω, χ), by $d_{(\omega,\chi)i}$ its state-contingent debt for that state, and by μ_i its volume of special deposits. We write $m_i = 1$ if bank i is regulated and $m_i = 0$ if it is not. The total date-0 transfer made by the state to bank i is τ_{0i}, which encompasses premia for DI and LOLR and can be positive or negative.

Different configurations of policies (DI, LOLR, regulation, and transfers) give rise to different equilibrium values for μ_i, $j_i(\cdot)$, m_i, and $\tau_{0,i}$. The equilibrium values of $\{d_{(\omega,\chi)i}\}$ are chosen by the state if bank i is regulated and is determined by the incentives of banker i seeking to maximize his utility if it is not. We will analyze these dependences in detail in Section 3. Here we only describe payoffs given μ_i, $j_i(\cdot)$, m_i, $\tau_{0,i}$, and the state-contingent debts $\{d_{(\omega,\chi)i}\}$.

It will be convenient to define the net funding gap of a bank as a function of investment $\hat{j}$, special deposits $\hat{\mu}$, debt $\hat{d}$, and revenues $\hat{r}$:

$$nf(\hat{j}, \hat{\mu}, \hat{d}, \hat{r}) \equiv \hat{j} + \hat{\mu} + \hat{d} - \hat{r}.$$

It will also be useful to decompose it into its positive and negative parts:

$$nf(\hat{j}, \hat{\mu}, \hat{d}, \hat{r}) = nf^+(\hat{j}, \hat{\mu}, \hat{d}, \hat{r}) + nf^-(\hat{j}, \hat{\mu}, \hat{d}, \hat{r}),$$

[22] 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 a mass 1 of firms, each with investment need 0 or 1 at date 1 and no net worth. The bank monitors firms (or shirks) at both dates 0 and 1. At date 1, the firms monitored at date 0 by the bank need 1 unit of cash each. At date 2, firms succeed or fail (then return 0). Success is guaranteed if none of the managers and the workers shirks. Otherwise, success accrues with probability 0. Shirking at date 1 brings benefit b_f to the firm manager and shirking by the workers yields them b_w. There is no payoff beyond the incentive payoffs $b_f + b_w$ of these stakeholders. We then get $\beta = \beta_f b_f + \beta_w b_w$, where β_f and β_w are the welfare weights for firms and workers.

where $nf^+(\hat{j},\hat{\mu},\hat{d},\hat{r}) \equiv \max\{0, nf(\hat{j},\hat{\mu},\hat{d},\hat{r})\}$ and $nf^-(\hat{j},\hat{\mu},\hat{d},\hat{r}) \equiv \min\{0, nf(\hat{j},\hat{\mu},\hat{d},\hat{r})\}$.

The net funding gap of bank i at date 1 in state (ω,χ) is $nf(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi)$. When the net funding gap is positive, the bank necessarily receives a state-contingent date-1 transfer $nf^+(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi)$ from the state in the form of bailouts, LOLR, and DI. This transfer allows the bank to carry out its investment and to make special depositors whole.[23] When the net funding gap is negative, there are leftover funds in the bank $-nf^-(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi)$ at date 1. These leftover funds can neither be consumed nor taxed. They represent a complete loss and do not serve any other purpose. It is always preferable to soak them up via an increase in debt and a corresponding increase in revenues at date 0. As a result, in equilibrium, they will always be zero. But we cannot impose that they are zero off equilibrium.

The state's utility is

$$V \equiv \int [\mathbb{E}[\beta j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) \\ - (1+\lambda_\omega) nf^+(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi) - \tau_{0i} - cm_i]] di.$$

The term $\beta j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i})$ reflects the stakeholders' benefits of continuation for bank i, the term $(1+\lambda_\omega) nf^+(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi)$ the costs of the date-1 transfer to that bank, the term τ_{0i} the cost of the date-0 transfer to that bank, and the term cm_i the cost of regulating it.

Bank i's utility is

$$U_i \equiv \mathbb{E}[b j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + (1+\theta)\mu_i + d_{(\omega,\chi)i} + \tau_{0i}].$$

The term $b j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i})$ reflects the benefits of continuation, the term $d_{(\omega,\chi)i}$ the date-1 debt payment that can be pledged at date 0, the term $(1+\theta)\mu_i$ the date-0 revenues from special deposits, and the term τ_{0i} the date-0 transfer from the state. This ignores the date-0 endowment (and possible activities) of the bank. It will be convenient to rewrite it as

$$U_i = \mathbb{E}[(b-1) j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \theta\mu_i + r_\chi + nf(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi) + \tau_{0i}].$$

This expression reflects, in each state (ω,χ), the net benefit $(b-1) j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i})$ from reinvestment for the bank, the risk-free discount $\theta\mu_i$ on insured deposits, the date-1 revenue r_χ, the date-1 transfer from the state $nf^+(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi)$, and the loss from date-1 leftover funds $nf^-(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi)$, as well as the date-0

[23] Because public funds are costly at date 1, it is never optimal to set a higher transfer.

transfer from the state τ_{0i}.

Aggregating over all banks yields

$$U \equiv \int U_i di = \int [\mathbb{E}[(b-1)j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \theta\mu_i + r_\chi \\ + nf(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi) + \tau_{0i}]]di.$$

Finally, total (state + bank) welfare is

$$W \equiv U + V = \int [\mathbb{E}[(b+\beta-1)j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \theta\mu_i + r_\chi \\ + nf^-(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi) - \lambda_\omega nf^+(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi) - cm_i]]di.$$

Note that because of our assumption of transferable utility at date 0, date-0 transfers $\tau_{0,i}$ does not appear in this expression. Conditional on $\{d_{(\omega,\chi)i}\}$, μ_i, $j_i(\cdot)$, and m_i, these transfers influence the respective levels of utility of the state and of bankers but not total welfare.

As already mentioned, in equilibrium, $\{d_{(\omega,\chi)i}\}$ is always chosen (by the state or by banker i depending on whether bank i is regulated or not) so that there no leftover funds at date 1, i.e. $nf^-(j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}),\mu_i,d_{(\omega,\chi)i},r_\chi) = 0$. Imposing these condition leads to simplifications in some of the expressions above. For example, we get

$$U_i = \mathbb{E}[(b-1)j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \theta\mu_i + r_\chi \\ + \max\{0, j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \mu_i + d_{(\omega,\chi)i} - r_\chi\} + \tau_{0i}]$$

and

$$W = \int [\mathbb{E}[(b+\beta-1)j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \theta\mu_i + r_\chi \\ - \lambda_\omega \max\{0, j_i(\omega,\chi,\mu_i,d_{(\omega,\chi)i}) + \mu_i + d_{(\omega,\chi)i} - r_\chi\} - cm_i]]di.$$

Public policy. We assume that all banks are treated symmetrically through a one-size-fits-all scheme. We will verify later in Section 3.6 that this is the case under a simple condition. 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. Therefore, from now on, we omit the i indices.

At the optimum, the mass of special depositors that are serviced by each bank is at a corner, either 0 or μ, and so we only consider these values. In the former case, we write $k = 0$ to indicate that there is no DI, and in the latter $k = 1$ to indicate that there is DI. Similarly, we write $m = 1$ if banks are regulated and $m = 0$ if they are not. And finally, we write $l = 1$ if there is LOLR and $l = 0$ otherwise. The state sets a policy $\{\tau_0, k, l, m\}$, where $\tau_0 \in \mathbb{R}$ and $(k, l, m) \in \{0, 1\}^3$. Pure shadow banking corresponds to $k = l = m = 0$ and $\tau_0 = 0$ and is always available to banks as an outside option.[24]

Public insurance services. Recall that LOLR means that the state 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 χ. For expositional simplicity only, we assume that this aggregate liquidity state is not verifiable. One can imagine that the only verifiable information regarding aggregate states comes in the form of reports by the government. In our model, the government would always prefer to slant its report ex post at date 1 in order to minimize its LOLR and DI liabilities. This restriction affects only one of our results. In its absence, offering LOLR to regulated and shadow banks would be equally costly for the state: 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.[25]

Timing. After observing the proposed policy $\{\tau_0, k, l, m\}$, the banks choose to be in the regulated or shadow banking sector. The timing is summarized in Figure 1. In Section

[24] 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.

[25] Suppose that in state WI, 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 WI 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.

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. In Section 4 however, 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. Consequently, we will require the Nash equilibrium among banks in their sectoral choice (for a given public policy) to be coalition-proof in the sense of Bernheim et al. (1987); that is, it must be immune to a self-enforceable deviation by a subset of banks that makes the banks in the "coalition" better off.

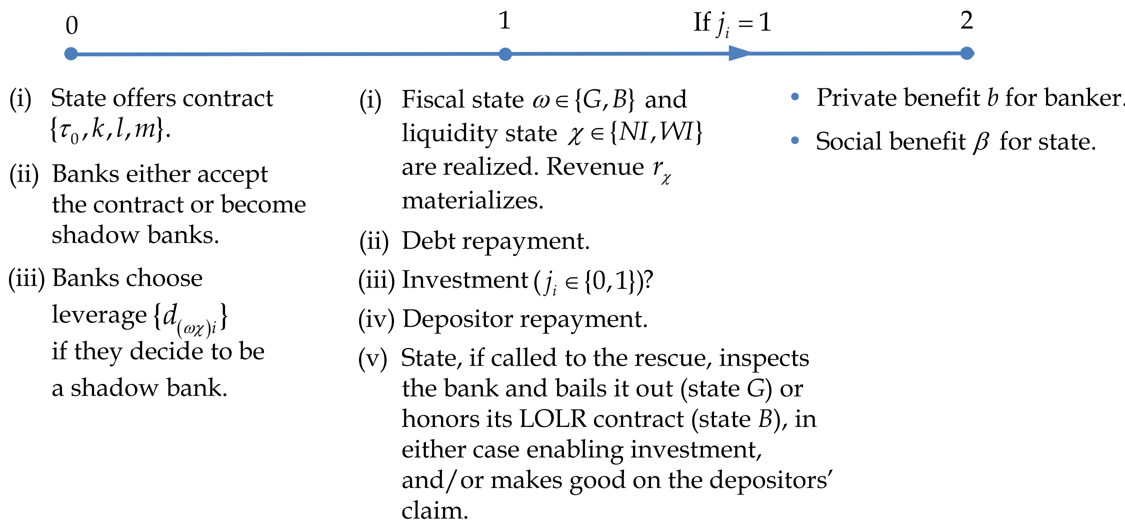

Figure 1: Timing for bank's operations.

Roadmap. We first compute 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}.[26]

We then solve the social planning problem of the maximization of social welfare $W(k, l, m)$ over all possible policies $(k, l, m) \in \{0, 1\}^3$, where $W(k, l, m)$ will denote the maximum achievable welfare under configuration (k, l, m). We denote the maximum by W^*. To do so, we exploit the fact that utility is perfectly transferable between the bank and the state at date 0. The bank's utility for policy configuration (k, l, m) is equal to its shadow banking reservation utility, and the state's utility is equal to total welfare minus the bank's

[26]Implicit in our analysis is the assumption that the state cannot lower the rewards to being a pure shadow bank, and that pure shadow banking therefore acts as an outside option. In reality, the state may also be able to tax shadow banks, at least to some degree, and thereby to reduce their reservation utility. The only difference to our analysis would be a reduced utility of bankers.

reservation utility, $W(k,l,m) - U^{SB}$. The date-0 transfer τ_0 from the state to the bank, which can be positive or negative, is chosen so as to make the bank indifferent to becoming a shadow bank. For example, when traditional banking with regulation, LOLR, and DI, emerges as the optimum, DI and LOLR may be underpriced in order to prevent migration to the shadow banking sector.[27]

Each policy configuration gives rise to a particular volume of special deposits per bank and a particular state-contingent date-1 investment policy function $j(\cdot)$. The date-1 investment policy function is $j(G,\chi,\hat{\mu},\hat{d}) = 1$ because of bailouts, $j(B,\chi,\hat{\mu},\hat{d}) = 1_{\{r_\chi - d_{(B,\chi)} \geq 1\}}$ if there is no LOLR ($l = 0$), and $j(B,\chi,\hat{\mu},\hat{d}) = 1$ if there is LOLR ($l = 1$). The volume of special depositors per bank is 0 if there is no DI ($k = 0$) and μ if there is DI ($k = 1$). State-contingent debts $\{d_{(\omega,\chi)}\}$ are chosen by bankers to maximize their utility if they are not regulated ($m = 0$), and by the state to maximize its utility (or equivalently social welfare) if they are regulated ($m = 1$). Since $d_{(\omega,WI)} = 0$ for $\omega \in \{G,B\}$, we only need to characterize $d_{(\omega,NI)}$ for $\omega \in \{G,B\}$. We now proceed to characterize these choices and the associated value of social welfare $W(k,l,m)$ for all possible policy configurations.

Notes on modeling choices. (a) 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. (b) Furthermore, 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.

3 Shadow banking and the quadrilogy

In this section we analyze the different configurations. We first start by examining the different cases when there is no regulation (shadow banking) in Section 3.1. We then turn to the different cases when there is regulation (regulated banking) in Section 3.2. We then demonstrate the complementarity that endogenously emerges between regulation and LOLR and between regulation and DI in Section 3.3. We continue by characterizing the

[27] 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].

full optimum configuration in Section 3.4 by integrating the analyses of the preceding sections. Finally, we consider two extensions. In Section 3.5, we consider a situation where there is ex-ante heterogeneity between banks, and in Section 3.6, we consider whether ex-post heterogeneity endogenously emerges at the optimum via selection of different menus offered by the state.

3.1 Shadow banking

We start by analyzing the case of shadow banking. As described above, a bank's objective is two-fold: invest and receive money from debt issuance and/or deposits. Investing dominates receiving money as the private benefit exceeds the cost of investment ($b > 1$). But the two need not be inconsistent. Indeed, left unsupervised, the bank has an incentive to issue the maximal possible liability ($d_{(G,NI)} = 2$) in state (G, NI); 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, NI) if the bank benefits from LOLR. By contrast, in the absence of LOLR, the bank cannot count on a bailout in state (B, NI), and is better off limiting its liability to 1, so as to have enough cash to undertake the investment.

Pure shadow banking

Under pure shadow banking without access to public liquidity insurance services, the bank optimally sets $d_{(G,NI)} = 2$ for state (G, NI). It thereby receives $p_G(2x_G)$ 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 b from continuation. In state (B, NI) by contrast, it cannot count on a bailout. Because $b > 1$, the bank prefers to continue, and it therefore issues debt $d_{(B,NI)} = 1$. The bank cannot continue if in state (B, WI) since it has no revenues. Its utility in the shadow banking sector is therefore

$$U^{SB} = p_G(b + 2x_G) + p_B x_B(b + 1).$$

Social welfare further accounts for the cost $1 + \lambda_G$ of the bailout in fiscal state G as well as of the social benefits of continuation β when it occurs:[28]

$$W^{SB} = p_G[b + \beta + 2x_G - (1 + \lambda_G)] + p_B x_B[b + \beta + 1].$$

Value of public liquidity insurance services to the shadow banking sector

We first analyze LOLR and then DI.

LOLR. With systematic access to public liquidity, a shadow bank selects to always be an empty shell: $d_{(G,NI)} = d_{(B,NI)} = 2$. Therefore, LOLR has an effect only in the B state. The net social benefit of LOLR is

$$w^{\text{LOLR,SB}} = p_B(1 - x_B)[b + \beta - (1 + \lambda_B)] - p_B x_B \lambda_B.$$

The first term in $w^{\text{LOLR,SB}}$ corresponds to the net social benefit of investment, $b + \beta - (1 + \lambda_B)$ in state (B, WI). The second term is the extra burden imposed by excess leverage in state (B, NI).

DI. With deposit insurance, the shadow bank (which does not internalize depositor or taxpayer welfare) sets $d_{(G,NI)} = 2$ as earlier, and $d_{(B,NI)} = 2$ in the presence of LOLR and $d_{(B,NI)} = 1$ 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), $\Sigma_\omega p_\omega \lambda_\omega$. And so the net benefit of deposit insurance in the shadow banking sector is

$$w^{\text{DI,SB}} = \mu(\theta - \Lambda_0 - \Lambda_2).$$

The net benefit of LOLR and DI combined is $w^{\text{LOLR,SB}} + w^{\text{DI,SB}}$.

We collect these results in a proposition.

Proposition 1. *(optimum given no regulation)*

The optimal arrangement given no regulation ($m = 0$) features

[28] The net effect of bailouts on welfare is $p_G[(1 - x_G)(b + \beta) - x_G \lambda_G]$, which can be positive or negative depending on the relative strength of two effects: first, bailouts allow the bank to continue in state (G, WI); second, bailouts allows the bank to increase leverage without risking termination in state (G, NI). 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, NI).

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

$$w^{LOLR,SB} > 0 \iff b + \beta > 1 + \lambda_B + \frac{x_B}{1 - x_B}\lambda_B, \tag{1}$$

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

$$w^{DI,SB} > 0 \iff \theta > \Lambda_0 + \Lambda_2, \tag{2}$$

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

There is no direct complementarity between LOLR and DI. The conditions (1) and (2) for LOLR and DI can be applied independently. For example, the optimal arrangement features LOLR and DI if (1) and (2) hold, LOLR but no DI if (1) does but (2) does not, etc. As we will see, there is 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 bank leverage is regulated.

Pure regulated banking

Under shadow banking, the bank sets $d_{(G,NI)} = 2$ and is systematically rescued by the government in state G. With regulation but without LOLR or DI, the bank is optimally constrained to set $d_{(G,NI)} = 1$ so that it continues by itself in state (G, NI) and is saved by the government in state (G, WI). The bank still sets $d_{(B,NI)} = 1$. The efficiency gains of regulation come from the fact that the deadweight loss associated with bailouts in state (G, NI) is avoided. The net benefit of regulation in the absence of LOLR or DI is

$$w^{RB/SB} = p_G x_G \lambda_G - c.$$

Value of public liquidity insurance services to the regulated banking sector

We first analyze LOLR and then DI.

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

$$w^{\text{LOLR,RB}} = w^{\text{LOLR,SB}} + p_B x_B \lambda_B = p_B(1 - x_B)[b + \beta - (1 + \lambda_B)].$$

DI. In the shadow banking deposit insurance is triggered in both states whether or not the bank has revenues. With regulation, deposit insurance is triggered only when the bank does not have revenues. Bank leverage is constrained at $d_{(\omega,NI)} = 1 - \mu$ for $\omega \in \{G, B\}$ so as to be able to finance one unit of investment and to repay μ special depositors when its revenue are equal to 2.[29]

The net benefit of deposit insurance given regulation is

$$w^{\text{DI,RB}} = w^{\text{DI,SB}} + \mu\Lambda_2 = \mu(\theta - \Lambda_0).$$

The net benefit of LOLR and DI combined given regulation is $w^{\text{LOLR,RB}} + w^{\text{DI,RB}}$. Exactly like in shadow banking, there is no direct complementarity between LOLR and DI.

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

$$w^{LOLR,RB} > 0 \iff b + \beta > 1 + \lambda_B, \tag{3}$$

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

[29] 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: $d_{(G,NI)i} = d_{(B,NI)i} = 1 - \mu$. 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 ($d_{(G,NI)i} = d_{(B,NI)i} = 0$), while the non-deposit-taking ones can issue debt ($d_{(G,NI)i} = d_{(B,NI)i} = 1$). 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 3.6, this treatment is not warranted if it is optimal to regulate deposit-taking banks but to not to regulated non-deposit-taking banks; the second option is then uniquely optimal, and we will therefore accommodate it.

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

$$w^{DI,RB} > 0 \iff \theta > \Lambda_0 \tag{4}$$

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

The conditions (3) and (4) for LOLR and DI can 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, 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 planning problem as the maximization of social welfare over all the different possible arrangements

$$W^* = \max_{\{k,l,m\}} W(k,l,m),$$

where

$$W(k,l,m) \equiv W^{SB} + w^{\text{DI,SB}}\mu k + w^{\text{LOLR,SB}} l + w^{RB/SB} m \\ + (w^{\text{DI,RB}} - w^{\text{DI,SB}})\mu mk + (w^{\text{LOLR,RB}} - w^{\text{LOLR,SB}}) ml.$$

The right-hand side of this expression features both linear and quadratic terms in k, l, and m. The quadratic terms are the last two terms $(w^{\text{DI,RB}} - w^{\text{DI,SB}})\mu mk$ $(w^{\text{LOLR,RB}} - w^{\text{LOLR,SB}})ml$. The following corollary confirms that these two 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:*

$$w^{LOLR,RB} = w^{LOLR,SB} + p_B x_B \lambda_B > w^{LOLR,SB}$$

and

$$w^{DI,RB} = w^{DI,SB} + \Lambda_2 > w^{DI,SB}.$$

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.[30] This complementarity is at the heart of the quadrilogy: the coexistence of lending to SMEs, deposit taking DI, regulation, and LOLR.

3.4 Overall optimum

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

Proposition 3. *(traditional banking system)*

(i) There exists a threshold regulation cost $c^* > 0$ such that regulation is optimal if and only if

$$c \leq c^*, \tag{5}$$

where

$$c^* = p_G x_G \lambda_G + \max\{w^{DI,RB}, 0\} + \max\{w^{LOLR,RB}, 0\} - \max\{w^{DI,SB}, 0\} - \max\{w^{LOLR,SB}, 0\}.$$

(ii) A necessary and sufficient condition for the traditional banking system with regulation, LOLR, and DI ($m = l = k = 1$) to be optimal is that (3), (4), *and* (5) *be satisfied.*

The analysis demonstrates the impact of shadow banking on the quadrilogy. The shadow banking sector acts as an outside option which determines the equilibrium rents enjoyed by banks in the form of cheap public liquidity insurance services (LOLR and DI).

3.5 Observed heterogeneity

Assume now that banks differ in how hard they are to monitor. For example, monitoring complex exposures on OTC markets is harder than assessing the risk on plain vanilla rated municipal bonds or even a (diversified) loan portfolio. One can imagine that there is a distribution of banks, each associated with a pattern of banking activities and characterized by its monitoring cost.[31] 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,

[30] 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.

[31] 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.

while the former have a choice.[32] 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.

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.

Corollary 2. *(heterogeneity) Assume that conditions (3) through (5) hold and that* $w_0^{LOLR,SB} \leq 0$ *and* $w^{DI,SB} \leq 0$*. 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 support* ($k = l = m = 0$)*; special depositors are serviced by regulated banks.*

3.6 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 it is un-levered, as long as: (a) all banks are supervised; and (b) they are made 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 $w^{RB/SB} + \max\{w^{LOLR,RB}, 0\} \geq \max\{w^{LOLR,SB}, 0\}$ *(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 $w^{RB/SB} + \max\{w^{LOLR,RB}, 0\} < \max\{w^{LOLR,SB}, 0\}$ *(implying that monitoring is suboptimal in the absence of special depositors),* $w^{RB/SB} + \max\{w^{DI,RB}, 0\} + \max\{w^{LOLR,RB}, 0\} > \max\{w^{DI,SB}, 0\} + \max\{w^{LOLR,SB}, 0\}$ *(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*

[32] 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.

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 (which requires knowing the two types) 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 then analyze optimal liquidity sharing in Section 4.2. In Section 4.3, 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 4.4, 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 an extension of the perfect-correlation basic model. The possible realizations of the aggregate liquidity state are now $\chi \in \{NI, WI, IO\}$, where IO denotes a new "Insurance Opportunity" state. Contingent on the realization of the fiscal state $\omega \in \{G, B\}$, the realizations of the aggregate liquidity state χ are as follows:

- "No Illiquidity" (NI): With probability x_ω, all banks have revenue 2.

- "Widespread Illiquidity" (WI): With probability $1 - x_\omega - y_\omega$, banks have revenue 0.

- "Insurance Opportunity" (IO): With probability y_ω, half of the banks receive revenue 2 and the other half receives 0. More precisely, there are two groups, M and N, of equal size and two sub-states $IO = IOM \cup ION$. The two groups have perfectly negatively correlated revenues in the insurance-opportunity state. In state IOM, banks of group M have revenues 2 and banks of group N have revenues 0. In state IOM, the opposite happens. States IOM and ION each have probability $y_\omega/2$.

When $y_\omega = 0$ for all ω, the model is strictly isomorphic to our baseline model, in which there is no scope for liquidity sharing. When $y_\omega > 0$, there are opportunities for liquidity sharing between the different ex-ante types of banks in the "insurance opportunity" state. Two banks are said to be "natural counterparties" if one belongs to group M and the other to group N. They are "correlated-risk counterparties" if they belong to the same group. The analysis of the no-illiquidity and widespread-illiquidity states are the same as in Section 3. So the focus of this section will be entirely on the insurance-opportunity state.

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.[33]

We will focus on contracts in which one bank promises to transfer 1 in state (ω, IOM) and the other in exchange promises to transfer 1 in state (ω, ION); we will look at 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. 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).

And, as we noted, 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 M (the grantee) can sign a liquidity-sharing contract with a counterparty bank of group N (the grantor), whereby the grantor

[33] 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.

pays 1 to the grantee in state (ω, ION). At date 0, the counterparties observe each other's leverage and liquidity-sharing choices and can organize transfers among themselves to split the associated surplus. 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).[34]

The regulator cannot check whether parties to a liquidity-sharing contract are natural counterparties if at least one of them is a shadow bank and therefore is not monitored. Banks can observe at date 0 the ex-ante type of another bank in the shadow banking sector, but the regulator cannot, in the sense that a bank in the shadow banking sector can portray itself to the regulator to be of any possible ex-ante type (M or N) at date 0; implementing the optimal allocation, as we will show, requires a form of ring-fencing of liquidity which makes sure that banks in the regulated sector do not trade liquidity with banks in the shadow banking sector.

The modified timing is described in Figure 2 (recall that $k = 0$ as there are no special depositors).[35]

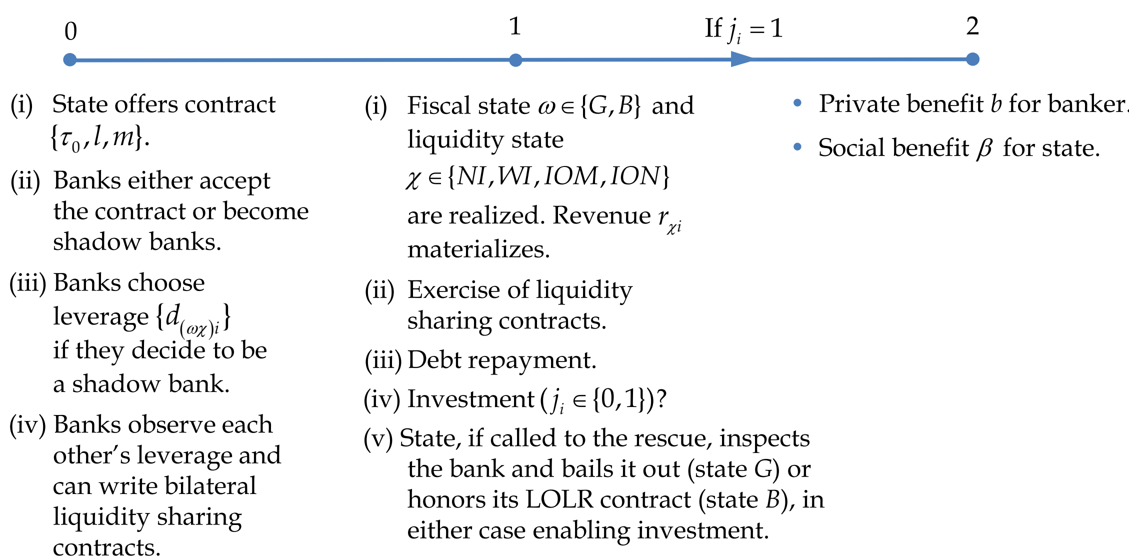

Figure 2: Timing under bilateral hedging.

[34] 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 get the payment in court.

[35] We introduce the notation $r_{\chi i}$ because the revenue of bank i depends on its ex-ante type (M or N) in states $\chi \in \{IOM, ION\}$. For example, $r_{IOM,i} = 2$ if i is of type M and $r_{IOM,i} = 0$ if i is of type N. Bank revenue in states $\chi \in \{NI, WI\}$ is independent of i and so $r_{\chi i} = r_\chi$, where $r_{NI} = 2$ and $r_{WI} = 0$.

4.2 Optimal liquidity sharing

In this section we characterize the outcome under optimal liquidity sharing. Later in Section 4.3, we ask under what conditions this optimal liquidity sharing arrangement can be obtained.

Optimal liquidity sharing

Consider the IO liquidity state. The socially efficient arrangement consists in: the sharing of liquidity between two natural counterparties in the two states (G, IO) and (B, IO); the absence of leverage in these states so that $d_{(\omega,\chi)} = 0$ for $\omega \in \{G, B\}$ and $\chi \in \{IOM, ION\}$. More precisely, a bank of group M signs two pairs of liquidity-sharing contracts (one for each fiscal state) with two banks of group N. The pair of liquidity-sharing contracts for the fiscal state (G, IO) consists in the bank of group M granting one unit of insurance to the bank of group N in state (G, IOM) and the bank of group N granting one unit of insurance to the bank of group M in state (G, ION). The pair of liquidity-sharing contracts for the fiscal state (B, IO) are similar. Then in both states (G, IO) and (B, IO), each bank de facto has 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 state to economize on taxpayer money (for a bailout in state (G, IO) and a possible LOLR rescue in state (B, IO)). 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.

Suppose for now that the government can ensure that banks share liquidity optimally within the regulated sector. We will later analyze what the government needs to do to ensure that this is the case. Then the analysis is analogous to our baseline analysis. There are only some small differences introduced by the introduction of the new state IO, which we briefly outline now. Recall that we have assumed away special depositors for simplicity, so that the outcomes should be compared with the case $\mu = 0$ where DI is an empty policy.

Modifications to the baseline analysis with optimal liquidity sharing

We start with pure shadow banking. 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

$$U^{SB} = p_G\left[b + 2\left(x_G + \frac{y_G}{2}\right)\right] + p_B[x_B(b+1) + y_B b]$$

and

$$W^{SB} = p_G\left[b + \beta + 2\left(x_G + \frac{y_G}{2}\right) - (1+\lambda_G)\right] + p_B[(x_B + y_B)(b+\beta) + x_B].$$

The utility of bankers is higher than in our baseline analysis by an amount $p_G y_G + p_B y_B b$ because they can pledge revenues with probability $1/2$ in state (G, IO) and because they can continue in the state (B, IO).[36] Social welfare is higher than in our baseline analysis by an amount $p_G y_G + p_B y_B(b+\beta)$ for the same reason.

Similarly, the net benefits of LOLR under no regulation are given by

$$w^{\text{LOLR,SB}} = p_B(1 - x_B - y_B)[b + \beta - (1+\lambda_B)] - p_B(x_B + y_B)\lambda_B$$

and the net benefits of LOLR under regulation are given by

$$w^{\text{LOLR,RB}} = w^{\text{LOLR,SB}} + p_B(x_B + y_B)\lambda_B.$$

The difference with our baseline analysis is that x_B has been replaced by $x_B + y_B$, the probability of the state $(B, NI) \cup (B, IO)$ in which the banks privately have the possibility of avoiding illiquidity when the fiscal state is B.

Along the same lines, the net benefits of regulation without LOLR are given by

$$w^{RB/SB} = p_G(x_G + y_G)\lambda_G - c,$$

The difference with our baseline analysis is that x_G has been replaced by $x_G + y_G$, the probability of the state $(G, NI) \cup (G, IO)$ in which the banks privately have the possibility of avoiding illiquidity when the fiscal state is G.

With these new expressions for the net social values, we have

$$\begin{aligned} W = W^{SB} + w^{\text{DI,SB}}\mu k + w^{\text{LOLR,SB}} l + w^{RB/SB} m \\ + (w^{\text{DI,RB}} - w^{\text{DI,SB}})\mu m k + (w^{\text{LOLR,RB}} - w^{\text{LOLR,SB}}) m l. \end{aligned}$$

The rest of the analysis follows along the same line leading to similar results for Proposi-

[36] Only for the purpose of this comparison, and only in order to facilitate the exposition, we assume that the probabilities x_G and x_B are unchanged. This assumption plays no role in the analysis.

tions 1, 2, 3, and 4, as well as Corollary 1.

4.3 When is optimal liquidity sharing implementable?

Let us assume that the optimum involves regulation, with or without LOLR (recall 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, one for states (G, IO) and (B, IO).

However, this is not enough as banks face perverse incentives in the choice of their counterparty. To contain these incentives, the state needs to put in place the following measures: ring-fencing between the regulated sector and the shadow banking sector; and either correlation monitoring within the regulated sector or centralized liquidity pooling. Failure to implement any of these two measures will allow banks to game liquidity regulation by engaging in *liquidity syphoning* and *bogus liquidity* provision. This regulatory evasion will in turn lead to an unravelling of the full optimum. If correlation monitoring is impossible, an alternative is the introduction of a CCP. Banks in the regulated sector are mandated to sign bilateral liquidity-sharing contracts with the CCP and shadow banks are banned from participating. All in all, the regulatory gaming incentives can be countered by appropriate structural remedies.

Ring-fencing and within-regulated-sector correlation monitoring

Since we have assumed away special depositors for simplicity, there are therefore only two cases to consider: with and without LOLR, depending on which configuration is optimal.

No LOLR. The cost and benefit of regulation in the NI and WI liquidity states are the same as in Section 3. So we focus on the IO liquidity state. In fiscal state (B, IO), the banks want to be covered as there will be no bailout (no LOLR). So, for that state, natural counterparties spontaneously engage in an optimal liquidity-sharing arrangement. Only those who have specified zero debt in the event (B, IO) can of course have access to such contracts, as otherwise no-one would want to obtain (bogus) insurance from them; and so all banks are happy to specify debt $d_{(B,IOM)} = d_{(B,ION)} = 0$ for that event.

More interesting is state (G, IO). Suppose first that the regulator imposes to each regulated bank to limit its leverage to $d_{(G,IOM)} = d_{(G,ION)} = 0$ and to enter a bilateral liquidity-sharing arrangement. As long as each bank picks natural counterparties in the regulated sector, optimal risk sharing is implemented. But as we shall see, the banks have incentives to game this requirement and engage in liquidity syphoning and bogus liquidity provision with correlated counterparties. If the state 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.

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 then has an incentive to sign a liquidity-sharing contract with a correlated-risk counterparty (i.e. of the same group) in the shadow banking sector for state (G, IO). Consider for example the case where the regulated bank is in group M and picks a counterparty in the shadow banking sector in group M for this state. The shadow bank increases its leverage to $d_{(G,IOM)} = 3$.[37] Their contract specifies a transfer of 1 to the shadow bank in state (G, IOM) and to the regulated bank in state (G, ION). The regulated bank has the "right contract with the wrong bank".

In state (G, IOM), the regulated bank has revenues 2, pays 1 to the shadow bank which combines this payment with its revenues of 2 to pay down its debt of 3. This represents liquidity syphoning from the regulated sector to the shadow banking sector. In state (G, ION), the shadow bank has no revenues and defaults on the payment of 1 that it must make to the regulated bank, and both banks are then bailed out. This represents bogus liquidity provision by the shadow bank to the regulated bank. By forming such a coalition and engaging in liquidity syphoning and bogus liquidity provision, the banks generate a joint surplus of $p_G y_G/2$ at the taxpayer's expense which they can split at date 0 through appropriate transfers between themselves. This shows that ring-fencing the regulated and shadow banking sectors is a necessary condition for the implementation of the optimum.

LOLR. The analysis with LOLR is similar. The difference is that there are now more gaming possibilities as banks can capitalize on LOLR in fiscal state (B, IO) on top of bailouts in fiscal state (G, IO). For example, in the absence of ring-fencing, a regulated bank can now also enter in a bilateral liquidity-sharing arrangement with a correlated shadow bank in state (B, IO). This coalition generates a surplus at the taxpayer's expense by syphoning liquidity provided to the regulated sector via LOLR and increases

[37] Note that it is important that we do not impose a priori that $0 \leq d_{(\omega,IOM)} \leq 2$ and $0 \leq d_{(\omega,ION)} \leq 2$ for $\omega \in \{G, B\}$. By contrast, we still assume, without loss of generality, that $d_{(\omega,WI)} = 0$ and that $0 \leq d_{(\omega,NI)} \leq 2$ for $\omega \in \{G, B\}$.

the reliance of the regulated bank on LOLR through bogus liquidity.[38] As in the case without LOLR, ring-fencing and correlation monitoring eliminate the problem.

The conclusion is therefore the following, whether or not LOLR is optimal:

Proposition 5. *(ring-fencing and correlation monitoring)*

(i) In the absence of ring-fencing, regulated banks enter bilateral liquidity-sharing contracts with correlated-risk shadow banks in which there is bogus liquidity and liquidity syphoning in state (G, IO), and , in case of LOLR, in state (B, IO) as well.

(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 state imposes ring-fencing by stipulating that regulated banks cannot sign liquidity-sharing contracts with shadow banks, but that it does not monitor correlations within the regulated sector. Technically, we assume that the state knows the type of each bank, but does not condition the treatment of each bank to the identity of its counterparty. This is meant to capture the idea that in practice, banks have myriads of counterparties and that the composition of these counterparties is fast-moving.[39] Monitoring the identities of all these different counterparty pairs is a complex task, which might simply be too costly to be worth executing.

Assume that bankers at date 1 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

[38] Recall that we assume that only state $\omega \in \{G, B\}$ is contractible and so LOLR amounts to a commitment to a bailout in state B. In the absence of this assumption, LOLR would only be provided in state (B, WI) and never in states (B, IO) or (B, NI).

[39] Even if the state incurs the high monitoring cost of matching the bank's type with the types of all its counterparties, an alternative interpretation of the analysis below is that CCPs economize on the cost of monitoring: The state needs only to run a stress test on the bank in isolation to understand whether its net position is sustainable (technically in this model, whether its type is M or N).

incentive to sign a liquidity-sharing contract with a correlated-risk counterparty in the regulated banking sector in state (G, IO). In the absence of LOLR, the only difference with the case considered above is that the counterparty cannot increase its leverage because it is regulated. Consider for example the case of a regulated bank of type M signing a liquidity-sharing contract with a regulated bank of type M in state (G, IO). In state (G, IOM), each bank has revenue 2, pays 1 to the other one and receives 1 from the other one, and is left with 2, which is then consumed. In state (G, ION) 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 $2p_G y_G \epsilon = (p_G y_G/2)(4\epsilon)$ at the taxpayer's expense which they can split at date 0 through appropriate transfers between themselves. This shows that under bilateral contracts, monitoring correlations in the regulated sector is a necessary condition for the implementation of the optimum. A similar analysis and conclusion applies in state (B, IO) when there is LOLR.

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. For example, a bank of group M must pay one to the CCP in state IOM and receives one from the CCP in state ION. 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. Preventing shadow banks 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.

The conclusion is therefore the following, whether or not LOLR is optimal:

Proposition 6. *(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 state (G, IO), and, in case of LOLR, in state (B, IO) as well.

(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 and banning participation from the shadow banking sector in the CCP (ring-fencing).

4.4 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 (G, IO) and (B, IO); 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, IO).

As in Proposition 5, 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 the monitoring of correlations or the use of CCPs within the regulated sector. There are a few interesting differences. For example, it is interesting to note that there are additional costs from relaxing ring-fencing since the associated liquidity syphoning and bogus liquidity decreases welfare not only by increasing the outside option of operating in the shadow banking sector[40], but also by directly increasing the fiscal costs of bailouts and LOLR in the regulated sector.[41] Alternatively, as in Proposition 6, the underlying liquidity arrangement can be implemented with a CCP.

[40] As long as $1-\alpha < 1/2$, we can assume that the banks with infinite monitoring cost reap all the gains from liquidity sharing if the underlying trades are possible since they are on the short side of the liquidity market.

[41] We could already have made this point in the homogenous bank case, but our purpose there was 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 in the homogenous bank case, 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).

4.5 Shadow banking as diversification and the costs of liquidity segregation

Finally, 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). 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 state, while others involve very complex instruments such as bespoke derivatives, that either are poorly understood by the state or are extremely time-consuming to monitor and assess. 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. We modify the stochastic structure of the economy, only in state IO. We assume that the ex-ante type of a bank is perfectly correlated with its monitoring cost.[42]

This means that there is no scope for liquidity sharing in state IO 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 via bilateral liquidity sharing increases welfare by reducing the fiscal cost of bailouts in both sectors and of LOLR in the regulated sector. 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 and of LOLR in the regulated sector.

[42] 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.

5 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.

The first key 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.

Our use of Epstein-Zin preferences implied 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 2019) 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. Second, over-leverage is even costlier than in the absence of special depositors, as it generates investor bailouts on top of bank bailouts. Third, shadow banks enjoy a large rent, as they sell the safe asset at a price that reflects the prospect of an investor bailout by the state. We show that all three problems can be resolved by incentivizing special depositors to remain in the regulated sector. Again deposit insurance and regulation co-vary.

There are many other 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 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 stakes, and to derive the optimal pattern of monitoring in richer environments. In our model, 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 activities, some traditionally thought of as belonging to the regulated sector and some traditionally thought of as belonging to the shadow banking sector. An interesting question is whether regulation should be performed at the activity level or at the institution level. 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.[43]

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.

[43] 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 (2017) "Financial Regulation in a Quantiative Model of the Modern Banking Sytem," mimeo Stanford.

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 (2017) "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks," Chicago Booth Working Paper.

Caballero, R.J., and E. Farhi (2017) "The Safety Trap," forthcoming *Review of Economic Studies*.

Chrétien, E., and V. Lyonnet (2017) "Traditional and Shadow Banks," HEC and INSEE CREST mimeo.

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, W. (2017) "Safety Transformation and the Structure of the Financial System," job Market Paper, Harvard University.

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.

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 (2019) “Safe Assets and Investor Bailouts,” mimeo.

Feve, P., and O. Pierrard (2017) “Financial Regulation and Shadow Banking: A Small-Scale DSGE Perspective”, TSE working paper 17–829.

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.

Malherbe, F., and M. McMahon (2017) "Financial Sector Origins of Economic Growth Delusion," mimeo.

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* 1-41.

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.