Working Paper Series

Optimal Liquidity Regulation with Shadow Banking

WP 15-12R

Borys Grochulski
Federal Reserve Bank of Richmond

Yuzhe Zhang
Texas A&M University

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Optimal liquidity regulation with shadow banking*

Borys Grochulski† Yuzhe Zhang‡

November 3, 2017

Working Paper No. 15-12R

Abstract

We study the impact of shadow banking on optimal liquidity regulation in a Diamond-Dybvig maturity mismatch environment. A pecuniary externality arising out of the banks' access to private retrade renders competitive equilibrium inefficient. A tax on illiquid assets and a subsidy to the liquid asset similar to the payment of interest on reserves (IOR) constitute an optimal liquidity regulation policy in this economy. Shadow banking gives banks an outside option that allows them to escape regulation but also entails a cost. We derive two implications of shadow banking for optimal liquidity regulation policy. First, optimal policy must implement a macroprudential cap on illiquid asset prices that binds only when the return on illiquid assets is high. Second, optimal policy must implement a fire sale of illiquid assets when high demand for liquidity is anticipated. We show how these features can be implemented by adjusting the IOR rate and the illiquid-asset tax rate.

Keywords: maturity mismatch, shadow banking, mechanism design, pecuniary externality, private retrade, liquidity regulation, interest on reserves, illiquid-asset tax
JEL codes: G21, G23, E58

1 Introduction

Beginning in the 1980s and leading up to 2007, a shadow banking system developed as a venue for origination and funding of illiquid bank assets outside of the realm of the government bank regulatory framework.[1] By 2007, the shadow banking sector had become about as large as the traditional, regulated banking sector.[2] The view that shadow banking was a key factor

*The authors are grateful to Javier Bianchi, Dean Corbae, Douglas Diamond, Huberto Ennis, Alan Moreira, Guillaume Plantin, Kieran Walsh, and Ariel Zetlin-Jones for their helpful comments. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

†Federal Reserve Bank of Richmond, borys.grochulski@rich.frb.org.

‡Texas A&M University, zhangeager@tamu.edu.

[1]Pozsar et al. (2012) define shadow banking as intermediation of credit through a wide range of securitization and secured funding techniques such as asset-backed commercial paper (ABCP), asset-backed securities (ABS), collateralized debt obligations (CDOs), and repurchase agreements (repos). The growth of shadow banking beginning in 1980 has been widely documented, see for example Greenwood and Scharfstein (2013).

[2]As measured by the value of outstanding liabilities, see Figure 1 in Adrian and Ashcraft (2012).

contributing to the financial crisis of 2007-2009, and to liquidity problems in particular, is shared by many academics and policymakers.[3] Optimal liquidity regulation of banks, therefore, should recognize that the option to move assets from regulated banks into shadow banks can potentially render bank liquidity regulations ineffective. Yet, this issue is understudied in the literature on optimal bank regulation.

In this paper, we study shadow banking as a costly but unregulated alternative to traditional banking. The question we ask is how the presence of this alternative affects optimal liquidity regulation of banks. As our benchmark, we take the pecuniary-externality-based theory of optimal liquidity regulation of Farhi et al. (2009). We extend this theory by giving banks an outside option: the possibility of moving assets to a shadow banking sector, where origination of illiquid assets is more costly but all activities are free of regulation. We solve the resulting mechanism design problem, where shadow banking introduces an ex ante participation constraint on banks.

We derive two implications of shadow banking for optimal liquidity regulation policy in this model. First, we show that optimal policy must implement a cap on the price of illiquid assets when the return on illiquid assets is high. This cap is necessary to prevent an exodus of banks to the shadow sector. A shadow bank's optimal strategy is to free ride on market liquidity by holding no liquid assets and dumping its illiquid assets on the market in case of an idiosyncratic liquidity shock. If the market price of illiquid assets is low, this strategy does not yield a payoff high enough to tempt banks to join the shadow sector, the participation constraint is slack, and the price of illiquid assets can increase when their return increases. But the payoff to shadow banking grows steeply with the market price of illiquid assets. When this price becomes sufficiently high, the participation constraint binds, and the price must be kept from increasing any further, even if the return on illiquid assets continues to increase.

Second, we show that optimal policy must implement a fire sale of illiquid assets when high demand for liquidity is anticipated. We model high liquidity demand as a high fraction of banks that receive an idiosyncratic shock compelling them to sell their illiquid assets before maturity. High anticipated liquidity demand has a negative wealth effect in the economy due to the need to retain a lot of aggregate liquidity and, thus, invest little in high-yield, illiquid assets. Absent the possibility of shadow banking, or if the shadow banking constraint does not bind, high anticipated liquidity demand has no implication for the equilibrium price of illiquid assets, as their supply and demand both decrease. With the shadow banking constraint binding, however, the price of illiquid assets must drop in a fire-sale manner when high liquidity demand is anticipated, as this is the only way to pass the negative wealth effect on to shadow banks.

Conversely, when anticipated liquidity demand is low, optimal policy implements higher investment in illiquid assets and a higher market price of illiquid assets. In this way, our model shows positive comovement between asset prices and investment, driven by anticipated

[3] Brunnermeier (2009), Gorton and Metrick (2010), Financial Stability Board (2012), Bernanke (2012).

liquidity needs. This relation between liquidity, asset prices, and investment has been an object of interest in the macroeconomic literature with financial frictions.[4]

These implications for optimal liquidity regulation policy, obtained as a solution to a mechanism design problem, depend only on the primitives of the economy and not on the particular implementation of the (constrained) optimal allocation. We discuss two optimal policy regimes. One is the quantity restriction of Farhi et al. (2009), mandating a minimum proportion of liquid assets on the balance sheet of a bank. The other is an alternative implementation mechanism in which asset holdings are taxed or subsidized but no balance-sheet restrictions are imposed. We discuss how these policy tools must respond to changes in the fundamental return on illiquid assets and to the level of aggregate liquidity demand in order to implement the two optimal illiquid asset price responses described above.

Our formal model builds on the classic maturity mismatch problem studied in Diamond and Dybvig (1983), Holmstrom and Tirole (1998), Allen and Gale (2004b), Farhi et al. (2009), and Farhi and Tirole (2012), among others. There are three dates, 0, 1, 2, and a population of ex ante identical banks, each with initial resources e.[5] Banks have the opportunity to invest in a long-term project at date 1 and, subject to an idiosyncratic shock, also at date 2. Each bank maximizes its continuation value, which is strictly increasing and concave in the scale of the long-term project funded by the bank. The long-term projects are bank-specific, i.e., nontraded. The idiosyncratic risk each bank faces is that its long-term investment opportunity may close early, i.e., at date 1.

This structure gives banks Diamond-Dybvig preferences over the timing of the funding of their long-term investments. A bank whose opportunity remains open beyond date 1 loses nothing by postponing investment until date 2, i.e., such a bank is perfectly patient at date 1. A bank whose idiosyncratic opportunity closes at date 1 becomes extremely impatient at that date. Similar to Holmstrom and Tirole (1998) and Farhi and Tirole (2012), our model leaves out the withdrawal behavior of depositors and instead focuses directly on the banks' maturity mismatch problem.

There are two assets banks can use at date 0 to transfer resources forward in time: a liquid asset that matures at date 1 and yields a gross return of 1 (cash or central bank reserves), and an illiquid asset that matures at date 2 and yields a gross return of $R > 1$ at that date (bank loans, e.g., mortgages). After banks find out their patience type at date 1, a competitive market opens in which the illiquid asset is traded for date-1 resources (cash) at price p.[6] The illiquid asset, therefore, is completely illiquid technologically (i.e., cannot be physically turned into date-1 resources) but the presence of a market in which it can be sold gives it a degree of

[4] Kiyotaki and Moore (2012), Shi (2015).

[5] With all agents ex ante identical in our model, we abstract from leverage.

[6] In an equivalent formulation, banks could borrow against the illiquid asset instead of trading it at date 1. We assume assets are traded for the ease of exposition.

market liquidity. How liquid the asset is in the market sense depends on the equilibrium level of p. If $p < R$, the asset is not perfectly liquid, as it trades at date 1 at a liquidity discount. Diamond and Dybvig (1983) show that $p = 1 < R$ in a unique laissez-faire equilibrium.

Following Lorenzoni (2008) and Farhi et al. (2009) among others, we assume that the date-1 market for the illiquid asset is private/anonymous, i.e., it cannot be interfered with by a regulator.[7] In the mechanism design problem, banks report their patience type to a planner, the planner distributes liquid and illiquid assets to banks based on their reports, then banks enter the private retrade market. Access to this market makes the banks' incentive constraint tighter, as private retrade can enhance the banks' value of misrepresenting their patience type. Moreover, the value of this misrepresentation depends on the market price p, which in turn is determined by supply and demand in the retrade market. This dependence creates the so-called pecuniary externality: by taking p as given, an individual bank does not internalize the impact of its actions on the tightness of the incentive constraint faced by other banks. A planner solving the optimal mechanism design problem does internalize this impact. This discrepancy drives a wedge between the market and the planner's preferred allocation in this economy. Due to this wedge, the market allocation is inefficient, which gives rise to a role for regulation.

To this environment we add the shadow banking, which we model as follows. At date 0, banks can choose to move to an unregulated shadow banking sector, where, as in the regulated sector, they invest their resources e in liquid and illiquid assets. The benefit of moving to shadow banking is that shadow banks' activities cannot be regulated. The disadvantage of moving to shadow banking is that shadow banks face a markup $\lambda \geq 0$ on the cost of investing in illiquid assets. This cost is a simple way of modeling the shadow banks' lack of access to the government safety net that is available to formal banks.[8] Ceteris paribus, the safety net, which includes both the explicit (e.g., deposit insurance, discount window) and implicit (i.e., ex post bailouts) forms of support, decreases the formal banks' cost of funding.[9] Since we abstract from leverage in this paper, we model the disadvantage of funding illiquid bank assets in the shadow sector directly as a markup λ on the cost of originating illiquid assets (loans) that shadow banks incur relative to formal banks.[10]

After origination, the illiquid assets held by formal banks and shadow banks are homogeneous. In the spirit of Jacklin (1987) and Kehoe and Levine (1993), shadow banks retain access

[7] One way to think about this assumption is that trading can be moved out of the jurisdiction (to an offshore location) and coordination of regulations across jurisdictions is impossible to achieve.

[8] Adrian and Ashcraft (2012) in fact define shadow banking as intermediation of credit outside of the government safety net.

[9] Using Fitch ratings data, Ueda and Weder di Mauro (2013) estimate the funding cost advantage of banks covered by the expectation of ex post government support at between 60 and 80 basis points. Explicit and priced safety net programs like deposit insurance and access to the discount window can also reduce the banks' cost of funding if bank investors/depositors are risk averse.

[10] In the baseline model, we assume that all shadow banks face the same markup λ. In Appendix B, we allow for heterogeneity in λ.

to the private retrade market for illiquid assets at date 1.

The ex ante participation constraint, which we also refer to as the shadow banking constraint, requires that the ex ante value of remaining in the regulated banking sector be not smaller than the value of becoming a shadow bank. Both these values depend on the equilibrium illiquid asset price p, which in turn depends on the allocation itself. The mechanism design problem with the shadow banking constraint is therefore nonstandard in that the agents' outside option value is endogenous to the mechanism.

To solve this mechanism design problem, we first transform it into one in which the planner indirectly chooses the retrade price p while the rest of the allocation (i.e., the initial and final investment made by banks) is determined by the requirements of resource feasibility and incentive compatibility. This transformation is particularly convenient for the analysis of the ex ante participation constraint. We show that this constraint can be reduced to an upper bound on the set of retrade prices p feasible for the planner. Using the transformed mechanism design problem, we then show that ex ante welfare attained in this economy (i.e., the banks' expected continuation value) is strictly increasing in the retrade price p, up to the first-best price $p^{\text{fb}} > 1$ that captures the (unconstrained-) optimal trade-off between liquidity insurance and the average return on investment. The objective of the planner, therefore, is to increase the retrade price p above the laissez-faire equilibrium price $p = 1$ as far toward the first-best price p^{fb} as possible without violating the banks' participation constraint.

Comparative statics with respect to the fundamental return on illiquid assets, R, show that shadow banking requires a macroprudential suppression of illiquid asset prices when R is high. Absent shadow banking, the first-best price p^{fb} is attainable for any R. With shadow banking, the participation constraint binds at all sufficiently high R, and the constrained-optimal price p^* is strictly below p^{fb}. Similarly, the fire-sale property of the optimal allocation follows from the binding participation constraint in a comparative statics exercise in which the fraction of impatient banks increases.

The optimal allocation in this economy can be implemented as follows. Suppose that at date 0 the government/regulator imposes a proportional tax τ on origination of illiquid assets paired with a proportional subsidy i to investment in liquid assets.[11] Such a subsidy is akin to payment of interest on liquid reserves held by banks at the central bank (IOR). It is intuitive that a subsidy to the liquid asset and a tax on the illiquid asset tilt the asset allocation trade-off faced by the banks in favor of the liquid asset. This tilt increases the supply of liquidity and decreases the supply of the illiquid asset in the retrade market at date 1, thus increasing the market-clearing price p. Both policy rates i and τ are uniquely determined by the optimal allocation. More precisely, they are determined by how much liquidity insurance the optimal allocation provides to banks. The optimal IOR rate i is equal to the net return that impatient

[11] We discuss an alternative implementation with a balance-sheet quantity restriction in Section 8.

banks are able to earn at the optimal allocation.[12] This return is zero in the laissez-faire equilibrium. The tax rate τ is the corresponding discount in the total return that patient banks earn at the optimal allocation relative to the total return of R that patient banks earn in the laissez-faire equilibrium. Jointly, thus, i and τ implement an ex post transfer from patient to impatient banks, which, from the ex ante perspective, amounts to provision of liquidity insurance.

There is a substantial debate in policymaking circles on whether and how bank regulation policy should respond to changes in business cycle conditions.[13] Our comparative statics results have implications for how the IOR rate i and the illiquid-asset tax rate τ should respond to changes in the rate of return R and the fraction of impatient banks π. These implications show the response of optimal policy rates to *current* macroeconomic conditions, which complements the usual exercise of extending the model to allow for ex post shocks to R and π and showing the response to *future* conditions.

The implications are as follows. The optimal policy rates must be low and sensitive to changes in R when R is low, but high and insensitive to changes in R when R is high. Since the return on bank assets is procyclical in U.S. data, we can identify high R with times of economic expansion and low R with recessions.[14] Optimal policy, thus, should be sensitive to the state of the business cycle in recessions with deeper reductions in policy rates in deeper recessions.

In response to high anticipated liquidity demand, measured by the fraction of impatient banks π, the IOR rate i must be decreased due to the negative wealth effect of high liquidity demand. The optimal tax rate τ is increased to the extent allowed by the banks' participation constraint. If this constraint is slack, τ increases and p remains approximately constant (exactly constant if banks have constant relative risk aversion). This increase in τ keeps banks indifferent to investing in liquid and illiquid assets. If the shadow banking constraint binds, however, τ cannot be increased. Instead, the price p must drop. This fire-sale-like price drop not only keeps banks indifferent to investing in both assets, but also decreases the return earned by shadow banks, thus preserving the banks' participation constraint.

In our model, shadow banking is an off-equilibrium event that nevertheless impacts on-equilibrium outcomes. Our results do not depend on shadow banking remaining off equilibrium. In Appendix B, we extend the model allowing for an active shadow sector and show that macroprudential asset price suppression and fire sales in high liquidity demand states remain optimal.

Related literature This paper is primarily related to the large literature on the provi-

[12]This return is analogous to the first-period return on the deposit contract in Diamond and Dybvig (1983).

[13]See, for example, Financial Stability Forum (2009).

[14]Using FDIC data on all insured institutions, https://www.fdic.gov/bank/statistical/, it is easy to check that aggregate Return on Assets and Yield on Earning Assets are positively correlated with GDP growth. Detailed analysis of the data is available upon request.

sion of liquidity under maturity-mismatch conditions: Diamond and Dybvig (1983), Jacklin (1987), Bhattacharia and Gale (1987), Diamond (1997), Allen and Gale (2004a), Allen and Gale (2004b), Farhi et al. (2009), Gale and Yorulmazer (2013), and Geanakoplos and Walsh (2017), among others. Our model adds an outside option, which represents shadow banking, and examines its implications for optimal liquidity regulation policy. Our model provides a novel explanation (derived from the value of the option of shadow banking) for the optimality of macroprudential assets price caps and fire sales in states of high liquidity demand.

Our paper is also related to the literature on inefficiencies arising from pecuniary externalities in general: e.g., Kehoe and Levine (1993), Golosov and Tsyvinski (2007), Lorenzoni (2008), Bianchi (2011), and Di Tella (2014). We show that an outside option can limit the strength of the pecuniary externality and reduce the scope for market intervention. We conjecture that this result does not depend on the details of the Diamond-Dybvig model and can be extended to other environments with pecuniary externalities.

Several recent studies build positive models of shadow banking: Huang (2014), Moreira and Savov (2014), and Ordonez (2015). This literature aims to explain the role of shadow banking as an unregulated, off-balance-sheet asset funding vehicle available to banks. In Ordonez (2015), for example, shadow banking is a value-enhancing response of banks to regulatory constraints on risk-taking that are too tight. Our objective in this paper is not to propose a positive theory of shadow banking. Rather, we model shadow banking simply as an arbitrage-seeking activity and aim to explain the impact of this possibility on the optimal liquidity regulation of banks. Although the positive literature on shadow banking considers several interesting government interventions, it does not characterize general, constrained-optimal regulation policy, which is the focus of our paper.

Plantin (2014) studies optimal capital regulation policy in a model with shadow banking. The role for regulation comes from the banks' failure to internalize the real adjustment costs in the production sector caused by the volatility in the final goods demand, which increases with the riskiness of bank deposits. Our model studies optimal regulation of liquidity, instead of capital, and uses an environment in which the role for regulation comes from a different friction (the pecuniary externality). However, as does Plantin (2014), we model shadow banking as unmonitored, ex post spot-market trade between banks and shadow banks. He allows for adverse selection in this market and finds that it can be beneficial, as it disrupts unmonitored trade and limits the size of shadow banking.[15]

Like we do here, Farhi and Tirole (2017) study optimal regulation of banks with shadow banking serving as an unregulated outside option. Their focus, however, is much broader than ours. Building on the framework of Holmstrom and Tirole (1998), they derive implications

[15] House and Masatlioglu (2015) study the effectiveness of equity injections and asset purchases in an interbank market with adverse selection. Bengui et al. (2015) assume illiquid assets to be completely nontradable and study public provision of liquidity through ex post bailouts of banks covered by an explicit government guarantee.

for access to lender of last resort facilities, bailouts, deposit insurance, asset ring-fencing, and central counterparty regulations. These regulations are implemented at the institution level, i.e., using direct transfers to and supervision of banks. In contrast, we focus on the optimal liquidity regulation policy and show that taxing/subsidizing activities at flat rates can change market prices and correct a specific pecuniary externality, thus implementing optimal provision of liquidity by the market.

We assume no uncertainty about asset quality in the market for illiquid assets. In doing so, we follow the large literature on interbank markets that dates back to Bhattacharia and Gale (1987) and includes, among others, Allen and Gale (2004b), Allen et al. (2009), Freixas et al. (2011), and Gale and Yorulmazer (2013). In contrast to most studies in this literature, however, we do not assume contract or market incompleteness. Instead, we solve a general mechanism design problem with resource, incentive, private retrade, and ex ante participation constraints. Further, because shadow banks retain access to the private retrade market, the value of the outside option is in our problem endogenous to the mechanism. This feature is common with the optimal market intervention mechanism studied in Tirole (2012) in a model with adverse selection.

The paper is organized as follows. Section 2 presents the baseline model. Section 3 discusses competitive equilibrium without intervention. Section 4 defines and solves the mechanism design problem describing constrained-efficient allocations. Section 5 derives comparative statics results for the optimal mechanism. Section 6 studies the implementation with IOR and a tax on illiquid assets. Section 7 derives implications of the comparative statics for the optimal policy rates. Section 8 discusses an alternative implementation with minimum liquidity requirements. Section 9 concludes. All proofs are relegated to Appendix A. Appendix B presents an extended model with active shadow banking.

2 The model

We consider an economy populated by a continuum of ex ante identical banks, each endowed with initial resources $e > 0$. The economy extends over three dates $t = 0, 1, 2$ and has two assets available at date 0: a liquid, low-yield asset s (cash or central bank reserves) and a technologically illiquid, high-yield asset x (bank loans such as mortgages or mortgage-backed securities). In addition, banks have idiosyncratic, long-term investment opportunities available at date 1 and, in some cases, also at date 2. Banks' objective is to maximize their continuation value $V(I)$, where I is the amount of capital invested in the long-term project.

At date 0, each bank has the option to become a shadow bank. If they decide to do business as a bank, they are subject to government regulation. If they decide to do business as a shadow bank, they are free from bank regulation but face an extra cost $\lambda \geq 0$ of originating the illiquid asset x.[16] In equilibrium, remaining as a bank will yield at least as high a value as becoming a shadow bank.[17] The possibility of shadow banking will therefore serve as an out-of-equilibrium outside option that restricts the scope of government regulation that can be imposed on banks in equilibrium.[18]

After the decision to not become shadow banks at date 0, banks make their initial investments s_0 and x_0 in, respectively, the liquid cash asset s (central bank reserves) and the illiquid asset x (bank loans). The cash asset pays a riskless return of 1 at date 1 and nothing at date 2. The illiquid asset pays nothing at date 1 and a riskless return of $R > 1$ at date 2. Note that as of date 1, asset x is technologically illiquid, i.e., resources invested in x cannot be used to fund long-term investment I at date 1.

The opportunity to invest in I is subject to an idiosyncratic Diamond-Dybvig shock $\theta \in \{0, 1\}$ realized at date 1. If $\theta = 1$, the bank can invest in I at either date 1 or 2. If $\theta = 0$, the bank has access to the long-term investment I only at date 1. The shock θ is a liquidity shock: banks of type $\theta = 0$ need liquidity at date 1 in order to invest in I before the specific long-term investment opportunity available to them closes. We will refer to them as impatient banks. Banks with $\theta = 1$ will be called patient. We use these assumptions to model the classic maturity-mismatch problem. The illiquid investment x, on the one hand, produces excess return $R > 1$, but, on the other hand, exposes the bank to liquidity risk at date 1.

After banks find out their type θ, but before the opportunity to invest in I at date 1 closes, a competitive market for the illiquid asset opens, where banks can trade the illiquid asset x for cash at a market price p. The existence of this market allows impatient banks to avoid getting stuck with illiquid assets at date 1, as their holdings of x can always be sold at the market price p. This price, however, can in equilibrium be lower than the asset's face value R.

In sum, given that the outside option of becoming a shadow bank is not more attractive, all banks choose to remain as banks at date 0 and then choose a portfolio $(s_0, x_0) \in \mathbb{R}^2_+$ subject to

$$s_0 + x_0 + \mathcal{T}(s_0, x_0) \leq e, \tag{1}$$

where s_0 is the amount invested in the cash asset, x_0 is the amount invested in the illiquid asset, and $\mathcal{T}(s_0, x_0)$ represents the costs of government regulations for a bank with asset portfolio

[16] As discussed in the introduction, this cost is a simple way of modeling the shadow banks' lack of access to the government safety net. For simplicity, we assume that acquiring liquid assets (cash) does not carry an extra cost for a shadow bank. Such a cost would not make a difference in our analysis because, as shown in Section 4.3.2, shadow banks strictly prefer to hold only the illiquid asset and no cash in their portfolios, even without the cost of acquiring cash.

[17] In case of indifference, we assume banks remain as banks. This assumption is without loss of generality because shadow banking is a weakly inefficient way of holding the illiquid asset due to the deadweight cost $\lambda \geq 0$.

[18] In Appendix B, we present an extension of our model in which banks are heterogeneous and a positive fraction of them (those with the lowest cost λ) become shadow banks in equilibrium. The main conclusions of our analysis continue to hold in this extended model.

(s_0, x_0).[19] At date 1, after they find out their type θ, banks choose their net demands $n(\theta) \geq -x_0$ in the market for the illiquid asset, and the final long-term investments $I_1(\theta) \geq 0$ and $I_2(\theta) \geq 0$ subject to budget constraints

$$I_1(\theta) \leq s_0 - pn(\theta), \tag{2}$$

$$I_2(\theta) \leq (x_0 + n(\theta))R. \tag{3}$$

As these budget constraints show, the government does not impose any regulations/taxes on the secondary market for the illiquid asset. Following Lorenzoni (2008), Farhi et al. (2009), and others, we assume, here as well as in the mechanism design problem of the planner/government, that the secondary market for the illiquid asset is out of reach of government regulation.[20]

The objective of a bank is to maximize the continuation value from the long-term investment I. For a bank of type θ that invests I_1 at date 1 and I_2 at date 2, the total continuation value is $V(I_1 + \theta I_2)$. The value function V is strictly increasing, strictly concave, and satisfies the following assumption.

Assumption 1 (Enough concavity) *$V'(I)I$ is strictly decreasing in I for all $I \geq e$.*

The ex ante expected payoff of a bank, therefore, is

$$\mathbb{E}[V(I_1 + \theta I_2)] \equiv \pi V(I_1(0)) + (1 - \pi)V(I_1(1) + I_2(1)), \tag{4}$$

where $0 < \pi < 1$ is the probability of $\theta = 0$. The bank's problem is to maximize (4) subject to budget constraints (1), (2), and (3).

If a bank decides to become a shadow bank, its objective remains the same but its budget constraint at date 0 changes to

$$s_0 + (1 + \lambda)x_0 \leq e. \tag{5}$$

In this constraint, there are no costs of government taxes/regulations $\mathcal{T}$, but there is a cost λ of doing business as a shadow bank. As shadow banks have access to the same retrade market at date 1, their budget constraints at date 1 remain the same, i.e., (2), (3). Denote by $\tilde{V}_0(p, \lambda)$ the value of becoming a shadow bank, i.e., the maximum of (4), subject to (5), (2), and (3).

We focus on the market provision of liquidity and financial firms' liquidity management and not on other aspects of banking (leverage, fragility to runs, etc). Our banks face the trade-off between liquidity and return. Unless they invest 100 percent in the low-yield cash asset, they face a maturity mismatch problem in that their long-term investment opportunity I may close

[19] Note that these costs can be negative, e.g., when the government pays interest on reserves to the banks.

[20] The reason why the government cannot interfere with the secondary market could be anonymity of trade (i.e., any trades that banks and shadow banks execute in this market are not observable to the government). More generally, monitoring of transactions in this market may be very costly given the possibility of these transactions being moved to a different legal jurisdiction.

before the illiquid asset x matures. The secondary market for illiquid assets lets banks access liquidity but only at a cost (because $p < R$).

3 Competitive equilibrium

In this section, we discuss competitive equilibrium in the laissez-faire (LF) economy, that is, the unregulated economy with $\mathcal{T} = 0$. Since $\lambda \geq 0$, it is immediate that with $\mathcal{T} = 0$ all banks prefer, at least weakly, to remain as banks. Diamond and Dybvig (1983) show that there exists a unique equilibrium in this LF economy.

Formally, banks' choices (s_0, x_0), $(n(\theta), I_1(\theta), I_2(\theta))$, and a price p are a competitive equilibrium if, taking p as given, the choices solve the banks' maximization problem, and the date-1 market for the illiquid asset x clears, i.e.,

$$\mathbb{E}[n] \equiv \pi n(0) + (1 - \pi)n(1) = 0. \tag{6}$$

Theorem 1 *In the economy with $\mathcal{T} = 0$, there exists a unique equilibrium: $p = 1$, $(s_0, x_0) = (\pi e, (1 - \pi)e)$, $n(0) = -x_0$, $n(1) = s_0$, $I_1(0) = e$, $I_1(1) = 0$, $I_2(0) = 0$, $I_2(1) = Re$.*

The equilibrium price $p = 1$ is pinned down by an arbitrage-type argument comparing the liquid-asset investment and a one-period investment in the illiquid asset (that is, invest in the illiquid asset at date 0 and sell it at date 1 with probability one). If p is not 1, one of these two investments dominates the other, and so the optimal investment choice at date 0 is not an interior one, which is inconsistent with clearing of the secondary market for the illiquid asset at date 1.

As shown in Jacklin (1987), the above argument does not depend on the simple market structure we consider here (a spot market for the illiquid asset at date 1). This argument carries over to a general market structure in which banks are allowed to enter and trade all conceivable state-contingent contracts. The intuition here is that it is not possible to contract around an arbitrage opportunity. Thus, $p = 1$ in any, even generally defined, competitive equilibrium.[21]

Note that in the LF equilibrium, the illiquid asset trades at date 1 at a discount relative to its fundamental value R. Due to inelastic supply of the illiquid asset (by the impatient types), a riskless payoff of R is bought by agents who are indifferent between resources at date 1 and 2 (the patient types) at an equilibrium price $p = 1 < R$. The difference between R and 1, thus, represents a liquidity discount at which the riskless one-period bond sells in equilibrium. Next, we ask if this discount is efficient. We define a general mechanism design problem, derive the constrained-optimal price p^*, and show that this price is less than R but more than 1, which

[21]See Section 3.1 in Farhi et al. (2009) for a formal proof of this result.

means that the LF equilibrium liquidity discount is too large and, therefore, the LF equilibrium allocation is inefficient.

4 Constrained-optimal allocations

In order to study the extent of the pecuniary externality and the scope for government regulation in this economy, we characterize in this section constrained-optimal allocations. In the first subsection, we define the optimal mechanism design problem. In the second subsection, we reduce this problem to a simple, single-dimensional maximization problem. In the third subsection, we characterize its solution. In the fourth subsection, we comment on the pecuniary externality that exists in this environment with private retrade.

4.1 Mechanism design problem

In this economy, an allocation is $A = (s_0, x_0, s_1(\theta), x_1(\theta), n(\theta), I_1(\theta), I_2(\theta))$, where (s_0, x_0) are the amounts invested at date 0 in the two assets, $(s_1(\theta), x_1(\theta))$ is a state-contingent asset allocation at date 1, and $(n(\theta), I_1(\theta), I_2(\theta))$ are recommendations for the private actions that agents/banks are to take: n is the recommended trade in the private market for the illiquid asset, and I_t is the recommended investment in the long-term projects at date t.

Definition 1 *Allocation A is incentive-feasible (IF) if*

(i) it is incentive compatible (IC), i.e., there exists a price $p \geq 0$ such that a) for both θ

$$(\theta, n(\theta), I_1(\theta), I_2(\theta)) \in \underset{\tilde{\theta},\tilde{n},\tilde{I}_1,\tilde{I}_2}{\operatorname{argmax}} \quad V(\tilde{I}_1 + \theta\tilde{I}_2) \tag{7}$$

$$s.t. \quad \tilde{I}_1 \leq s_1(\tilde{\theta}) - p\tilde{n}, \tag{8}$$

$$\tilde{I}_2 \leq (x_1(\tilde{\theta}) + \tilde{n})R, \tag{9}$$

and, b) the secondary market clears at p, i.e., $\mathbb{E}[n] = 0$,

(ii) it is resource feasible (RF), i.e.,

$$\begin{aligned} s_0 + x_0 &\leq e, \\ \pi s_1(0) + (1-\pi)s_1(1) &\leq s_0, \\ \pi x_1(0) + (1-\pi)x_1(1) &\leq x_0, \end{aligned}$$

(iii) it satisfies the ex ante participation constraint

$$\mathbb{E}[V(I_1 + \theta I_2)] \geq \tilde{V}_0(p, \lambda), \tag{10}$$

where

$$\tilde{V}_0(p,\lambda) \equiv \max_{\tilde{s}_0,\tilde{x}_0,n(\theta),\tilde{I}_1(\theta),\tilde{I}_2(\theta)} \quad \mathbb{E}[V(\tilde{I}_1+\theta\tilde{I}_2)]$$
$$s.t. \quad \tilde{s}_0+(1+\lambda)\tilde{x}_0 \le e,$$
$$\tilde{I}_1(\theta) \le \tilde{s}_0 - p\tilde{n}(\theta),$$
$$\tilde{I}_2(\theta) \le (\tilde{x}_0+\tilde{n}(\theta))R, \ \text{for } \theta = 0,1,$$

is the ex ante value of becoming a shadow bank.

The incentive compatibility condition (7) requires that, taking the retrade price p as given, banks cannot improve their value by any joint deviation combining a misrepresentation of their type θ with retrading in the private market. This condition is the same as the notion of incentive compatibility with retrade used in Farhi et al. (2009) as well as in other studies in the literature on pecuniary externalities.

The ex ante participation constraint (10) is new to this literature. In particular, the novel element here is the dependence of $\tilde{V}_0$ on p. That is, the value of the banks' outside option is not taken parametrically in our model but rather is endogenous to the mechanism.

Social welfare is given by the ex ante expected value delivered to the representative bank:

$$\mathbb{E}[V(I_1+\theta I_2)]. \tag{11}$$

The mechanism design problem is to find an IF allocation A that maximizes this objective among all IF allocations. Such an allocation will be referred to as a constrained-optimal allocation.

4.2 Reduction of the mechanism design problem

In this subsection, we show that the general mechanism design problem defined above can be reduced to one in which the planner chooses, indirectly, just the secondary-market price for the illiquid asset, p, and the rest of the allocation A is determined by the requirements of incentive-feasibility.

Lemma 1 *It is without loss of generality to restrict attention to allocations in which the planner recommends no trade in the private market, i.e.,* $n(\theta)=0$ *for* $\theta=0,1$.

This result is analogous to Cole and Kocherlakota (2001). It follows because using the state-contingent allocation of assets at date 1, $(s_1(\theta),x_1(\theta))$, the planner can replicate any trades that banks may want to execute in the private market.

Given this lemma, in the remainder of this section we focus on allocations with zero recommended retrade. With $n(\theta) = 0$, budget constraints (8) and (9) imply

$$I_1(\theta) = s_1(\theta), \quad I_2(\theta) = Rx_1(\theta).$$

Also, for a given state-contingent allocation of assets at date 1, $(s_1(\theta), x_1(\theta))_{\theta\in\{0,1\}}$, RF constraints determine the requisite initial investment (s_0, x_0). Thus, under Lemma 1, the full allocation A is determined by the date-1 state-contingent allocation of assets $(s_1(\theta), x_1(\theta))_{\theta\in\{0,1\}}$, with ex ante social welfare (11) simplified to $\mathbb{E}[V(s_1 + \theta Rx_1)]$.

Proposition 1 *An allocation is incentive compatible with price $p > 0$ if and only if*

(i) $s_1(0) + px_1(0) = s_1(1) + px_1(1)$,

(ii) $x_1(0) = 0$,

(iii) $p \begin{cases} \geq R, & \text{if } s_1(1) > 0; \\ \leq R, & \text{if } x_1(1) > 0. \end{cases}$

Condition (i) shows that, as in Allen (1985) and Cole and Kocherlakota (2001), incentive compatibility with retrade implies that the market value of assets allocated at date 1 to those who announce θ, i.e., $s_1(\theta) + px_1(\theta)$, must be the same for both announced types. Otherwise, all banks would report the realization of θ that receives the asset allocation $(s_1(\theta), x_1(\theta))$ with the higher market value. When the value of assets allocated to each announcement is the same, banks have no reason to misrepresent their type.

Condition (ii) is necessary for the recommendation of no private trade to be incentive compatible for the impatient types. Clearly, the impatient banks must receive no illiquid asset at date 1 for otherwise they would prefer to trade in the private market (sell $x_1(0)$ at any price).

Likewise, condition (iii) is necessary to guarantee that the patient types do not go to the private market. Indeed, if the allocation gives patient banks some liquid asset at date 1 and the banks are supposed to not trade in the private market, the return from buying the illiquid asset in that market must be weakly negative. If the allocation gives them a positive allocation of the illiquid asset, the return from selling it in the private market must be weakly negative.

Lemma 2 *At a constrained-optimal allocation: a) the RF constraints hold as equality, b) $s_1(1) = 0$, and c) $p \leq R$.*

This lemma gathers three immediate necessary conditions for efficiency. In particular, condition b) says providing liquidity to the type that does not need it is not optimal, as liquidity is costly to provide. Condition c) says that since the patient types are to postpone their final investment to date 2, they need to earn a nonnegative return between dates 1 and 2.

4.2.1 Reduced mechanism design problem

Proposition 1 and Lemma 2 imply that we can further focus our analysis on a simple class of allocation mechanisms in which the planner indirectly chooses just the retrade price p, while the asset allocation $(s_1(\theta), x_1(\theta))_{\theta\in\{0,1\}}$, and thus the whole allocation A as well, is determined by the requirements of incentive-feasibility.

To see this, note first that with $x_1(0) = s_1(1) = 0$, the present value condition (i) in Proposition 1 reduces to $s_1(0) = px_1(1)$. This shows that for given $s_1(0)$ and $x_1(1)$, there exists a unique retrade price consistent with incentive compatibility. That price is the one at which the cash allocated to the impatient banks and the illiquid assets allocated to the patient banks have the same market value, i.e.,

$$p = s_1(0)/x_1(1). \tag{12}$$

Second, with parts a) and b) of Lemma 2, the resource feasibility conditions imply

$$\pi s_1(0) + (1-\pi)x_1(1) = e. \tag{13}$$

Third, part c) of Lemma 2 requires $p \leq R$.

The mechanism design problem, thus, boils down to the choice of $s_1(0)$ and $x_1(1)$ subject to (12), (13), $p \leq R$, and the banks' ex ante participation constraint (10). Equivalently, we can think of the planner as choosing a price $p \leq R$ with $s_1(0)$ and $x_1(1)$ determined by (12) and (13).

The social welfare function that the planner maximizes (i.e., the banks' continuation value) can be conveniently expressed in terms of just the retrade price p.

Lemma 3 *In the mechanism in which the planner indirectly chooses the retrade price $p \leq R$, the social welfare function is*

$$\mathbb{E}\left[V\left(\frac{p}{\pi p + 1 - \pi}e\left(1 - \theta + \theta\frac{R}{p}\right)\right)\right]. \tag{14}$$

The objective (14) shows the trade-off involved in the setting of the retrade price p. At date 1, all banks earn the return $\frac{p}{\pi p+1-\pi}$ on their initial resources e. Impatient banks invest $I_1 = \frac{p}{\pi p+1-\pi}e$ at that time. Patient banks are able to wait until the illiquid asset matures. They earn an additional return $\frac{R}{p}$ and invest $I_2 = I_1\frac{R}{p} = \frac{R}{\pi p+1-\pi}e$ at date 2. Higher p increases I_1, decreases I_2, and decreases the average return earned in the economy, $\pi\frac{I_1}{e} + (1-\pi)\frac{I_2}{e} = \frac{\pi p+(1-\pi)R}{\pi p+1-\pi}$. The planner, therefore, faces a trade-off between return and insurance. Higher p provides more insurance at the cost of lower average return. By setting $p = R$, the planner can achieve full insurance with $I_1 = I_2 = \frac{R}{\pi R+1-\pi}e$, but the average return, $\frac{R}{\pi R+1-\pi}$, is low. By setting the price $p = 0$, the return on e is maximal, R, (supported by $x_1(1) = e$ and $s_1(0) = 0$), but risk sharing is very poor, as $I_1 = 0$.

The indirect formulation of the mechanism design problem in which the planner chooses p is convenient because the banks' value of the outside option (the option of becoming a shadow bank), $\tilde{V}_0(p, \lambda)$, depends on the allocation A (that is offered to banks) only through the illiquid asset resale price p associated with A.

Lemma 4 *In the mechanism in which the planner indirectly chooses the retrade price $p \leq R$, a shadow bank's objective $\tilde{V}_0(p, \lambda)$ can be expressed as*

$$\mathbb{E}\left[V\left(\max\left\{1, \frac{p}{1+\lambda}\right\} e\left(1-\theta+\theta\frac{R}{p}\right)\right)\right]. \tag{15}$$

We see in (15) that the structure of the payoff for a shadow bank mirrors that of a bank, given in (14). All shadow banks earn the same return between dates 0 and 1, and patient shadow banks earn the additional return $\frac{R}{p}$ by postponing their final investment to date 2. The return all shadow banks earn between dates 0 and 1, $\max\{1, \frac{p}{1+\lambda}\}$, is attained by investing all-cash at date 0 if $p < 1 + \lambda$ or putting all initial resources in the illiquid asset if $p > 1 + \lambda$. If $p = 1 + \lambda$, shadow banks are indifferent with respect to the asset allocation at date 0.

With these formulas for the payoffs to banks and shadow banks, we can express the banks' ex ante participation constraint in the following simple form.

Lemma 5 *In the mechanism in which the planner indirectly chooses the retrade price $p \leq R$, the participation constraint reduces to*

$$\frac{\lambda}{\pi} \geq p - 1 \geq 0. \tag{16}$$

The right inequality in (16) is the constraint imposed by the bank's option to become a shadow bank and invest all-cash. The left inequality in (16) follows from the bank's option to become a shadow bank and invest all-illiquid.

In sum, the mechanism design problem boils down to the choice of a retrade price $p \leq R$ that maximizes (14) subject to (16). This reduction of the full mechanism design problem leads to a simple characterization of all optimal allocations and their dependence on λ, which we provide in the next subsection.

4.3 Characterization of optima

4.3.1 Without the ex ante participation constraint

As a benchmark, let us first solve the planning problem with the RF and IC constraints but disregarding the ex ante participation constraint (16), i.e., as if banks did not have the option to become shadow banks.

Proposition 2 *There exists a unique maximizer* p^{fb} *for the objective function (14). Under Assumption 1, this maximizer satisfies*

$$1 < p^{\text{fb}} < R. \tag{17}$$

The price p^{fb} reflects the optimal trade-off between investment efficiency (the return on e) and insurance.[22] The optimal price falls between 1 and R because banks' relative risk aversion with respect to the liquidity shock, embedded in the concavity of V, is greater than 1 but less than infinity.[23] Under Assumption 1, V has enough concavity to lift the optimal price above 1 but keep it below R.

Note that, despite the fact that buyers in the private market for the illiquid asset (the patient types) are indifferent to the timing of their cash flow, the optimal allocation is consistent with the riskless payoff R being sold at date 1 at a price $p^{\text{fb}} < R$. This price is consistent with equilibrium in the private market due to infinite impatience of the impatient types, or cash-in-the-market pricing, as in Allen and Gale (1994).

Associated with the optimal price p^{fb} is the optimal allocation $(s_1^{\text{fb}}(0), x_1^{\text{fb}}(1))$ of the liquid and illiquid asset at date 1. From (12) and (13) we have

$$(s_1^{\text{fb}}(0), x_1^{\text{fb}}(1)) = \left(\frac{p^{\text{fb}} e}{\pi p^{\text{fb}} + 1 - \pi}, \frac{e}{\pi p^{\text{fb}} + 1 - \pi} \right). \tag{18}$$

The impatient banks attain the final investment $I_1^{\text{fb}}(0) = s_1^{\text{fb}}(0)$ and the patient banks attain $I_2^{\text{fb}}(1) = R x_1^{\text{fb}}(1) = \frac{R}{p^{\text{fb}}} s_1^{\text{fb}}(0) > s_1^{\text{fb}}(0)$. The inital investment in the two assets in this allocation is $s_0^{\text{fb}} = \pi s_1^{\text{fb}}(0)$ and $x_0^{\text{fb}} = (1-\pi) x_1^{\text{fb}}(1)$.

4.3.2 With the ex ante participation constraint

Next, we reintroduce the participation constraint (16). Does the first-best optimum p^{fb} satisfy this condition? Clearly, (17) implies $p^{\text{fb}} - 1 \geq 0$, so the right-hand side of (16) is satisfied. Whether $\frac{\lambda}{\pi} \geq p^{\text{fb}} - 1$ depends on the value of λ.[24] If λ is high enough, p^{fb} will continue to be feasible even in the presence of the ex ante private investment constraint (16). In particular, let

$$\bar{\lambda} \equiv \pi(p^{\text{fb}} - 1). \tag{19}$$

[22] Although p^{fb} is defined as a solution to the second-best problem with private retrade, the notation reflects the fact that p^{fb} also solves the first-best planning problem—the problem in which all information is public, i.e., there are no IC constraints. Without the ex ante participation constraint (10), the IC constraint (7) does not bind, i.e., the first-best solution remains incentive compatible under private information about θ and private retrade in the secondary market for illiquid assets. See Farhi et al. (2009) for a full exposition of this result.

[23] It is easy to check that $p^{\text{fb}} = 1$ if V is logarithmic, and $p^{\text{fb}} = R$ if V is Leontief.

[24] We take π to be an exogenous constant for now. We discuss comparative statics with respect to π in Section 6.2.

This value is the threshold level of λ such that the ex ante participation constraint binds if $\lambda < \bar{\lambda}$ and does not bind if $\lambda \geq \bar{\lambda}$. With $\lambda < \bar{\lambda}$, thus, the first-best price p^{fb} is not incentive-feasible.

The next proposition solves for the constrained optimum with the participation constraint (16). The constrained-optimal retrade price will be denoted by p^*.

Proposition 3 *For all $\lambda \geq \bar{\lambda}$, the ex ante participation constraint does not bind and $p^* = p^{\text{fb}}$. For all $0 \leq \lambda < \bar{\lambda}$, the ex ante constraint binds and $p^* = \frac{\lambda}{\pi} + 1 < p^{\text{fb}}$. In sum,*

$$p^* = \min\left\{p^{\text{fb}}, \frac{\lambda}{\pi} + 1\right\}. \tag{20}$$

The proof follows from the fact that social welfare is increasing in p at all p smaller than p^{fb}. The constrained-optimal retrade price, therefore, is the largest price satisfying $\frac{\lambda}{\pi} \geq p - 1$ but not larger than p^{fb}.

The full constrained-optimal allocation, A^*, is determined by the price p^*. The allocation of the liquid and illiquid asset at date 1 is

$$(s_1^*(0), x_1^*(1)) = \left(\frac{p^* e}{\pi p^* + 1 - \pi}, \frac{e}{\pi p^* + 1 - \pi}\right), \tag{21}$$

and the final investment attained by the impatient and patient banks is, respectively,

$$I_1^*(0) = s_1^*(0) \text{ and } I_2^*(1) = R x_1^*(1) > s_1^*(0). \tag{22}$$

Substituting (20) into (21) and comparing with (18), we can also show how the date-1 allocation of assets in the constrained optimum relates to that in the first best:

$$s_1^*(0) = \min\left\{s_1^{\text{fb}}(0), \left(1 + \frac{1-\pi}{\pi}\frac{\lambda}{1+\lambda}\right) e\right\}, \tag{23}$$

$$x_1^*(1) = \max\left\{x_1^{\text{fb}}(1), \frac{e}{1+\lambda}\right\}. \tag{24}$$

The intuition behind Proposition 3 is as follows. If the ex ante constraint binds, i.e., if $\lambda < \bar{\lambda}$, the amount of insurance against the liquidity shock the planner can provide to banks is constrained by the threat of banks becoming shadow banks and employing the investment strategy described in Jacklin (1987). This strategy delivers the maximum value a shadow bank can attain. In this strategy, the shadow bank invests all of its resources in the illiquid assets. The shadow bank subsequently sells these assets in the secondary market if it experiences a liquidity need at date 1 or holds onto them if it does not. Specifically, in this strategy the shadow bank acquires $\frac{e}{1+\lambda}$ units of the illiquid asset x at date 0. If $\theta = 0$, it sells x and puts $\tilde{I}_1 = p\frac{e}{1+\lambda}$ in the long-term investment at date 1. If $\theta = 1$, it holds x to maturity and invests $\tilde{I}_2 = R\frac{e}{1+\lambda}$ at date 2.

The value of this strategy, $\mathbb{E}[V(\tilde{I}_1 + \theta\tilde{I}_2)] = \mathbb{E}[V(\frac{pe}{1+\lambda}(1-\theta+\theta\frac{R}{p}))]$, decreases in the shadow banks' cost mark-up λ and increases in the private resale price p. The planner wants to increase p toward p^{fb} but is constrained by the banks' option of shadow banking. Larger λ makes this option less attractive, which allows the planner to increase p more without triggering an exodus of assets to the shadow sector.[25] If λ is large enough, i.e., $\bar{\lambda}$ or larger, the planner can lift p all the way up to p^{fb}.

Given Proposition 3, it is easy to see how the constrained optimum depends on the quality of the banks' outside option, measured by the cost λ. If the outside option is sufficiently unattractive, i.e., $\lambda > \bar{\lambda}$, the participation constraint does not bind and the first best is attainable. Note that, as long as $\lambda > \bar{\lambda}$, it does not matter how high λ is. If the outside option is attractive enough to bind, however, the constrained optimum becomes worse the better the outside option is: lower λ implies lower p^* and lower final investment made at date 1, I_1^*. This means the amount of liquidity available at date 1 is smaller, and the maturity mismatch is worse. Banks are provided with less insurance against the liquidity shock θ, and the illiquid asset is priced in the secondary market with a larger discount relative to its face value of R.

In the extreme case of $\lambda = 0$, the constrained-optimal price is $p^* = 1$, which coincides with the LF competitive equilibrium price, and the optimal allocation of final investment (I_1^*, I_2^*), given in (22), coincides with the competitive equilibrium allocation given in Theorem 1.

4.4 A pecuniary externality

With $\lambda = 0$, as we just saw, the ex ante participation constraint is so tight that the constrained-optimal price is $p^* = 1$, the constrained-optimal allocation coincides with the unregulated competitive equilibrium allocation, and thus the LF competitive equilibrium is constrained-efficient.

With $\lambda > 0$, however, there is a discrepancy between the competitive equilibrium and the constrained-optimal allocation. This market failure is due to what is often called a pecuniary externality. In our model, as in Farhi et al. (2009), the pecuniary externality is not due to incomplete markets, as in Geanakoplos and Polemarchakis (1986), but rather due to the fact that the resale price p enters the incentive compatibility constraint (7). If we extend the market structure from simple trade in the secondary market for the illiquid asset to the general one in which banks trade state-contingent claims—with one another or with a centralized counter-party—the pecuniary externality is still present and the competitive outcome is still inefficient. The reason is that in any competitive equilibrium, a firm—including a central counter-party—takes p as given, while the planner internalizes the impact of the primary allocation on the resale

[25] Equivalently, looking at (24), we can say that the planner wants to decrease $x_1(1)$ toward $x_1^{\mathrm{fb}}(1)$ but is constrained by the possibility of banks obtaining $\tilde{x}_1(1) = \frac{e}{1+\lambda}$ as shadow banks. By making $\tilde{x}_1(1)$ smaller, higher λ decreases the value of this outside option and thus relaxes the constraint faced by the planner.

price. Formally, it is easy to extend the analysis in Farhi et al. (2009) to show that even the general Prescott-Townsend competitive equilibrium is inefficient in the present environment.

The strength of the pecuniary externality can be quantified by the size of the wedge between the constrained-optimal price p^* and the competitive equilibrium price, which is 1. Clearly, (20) implies that this wedge is strictly increasing in λ for all $\lambda < \bar{\lambda}$, and for $\lambda \geq \bar{\lambda}$ it remains constant, at its maximum level. The ex ante participation constraint, therefore, reduces the strength of the pecuniary externality. The tighter this constraint, i.e., the lower the cost λ, the weaker the externality.

5 Comparative statics with shadow banking

In this section, we derive comparative statics for the optimal allocation. Our focus is on how the possibility of shadow banking affects the way in which the optimal allocation responds to changes in external conditions.

First, we consider changes to the return on illiquid assets, R. We show that with shadow banking, the asset resale price p must be suppressed below its first-best level when R is high but not so when R is low. Second, we consider an increase in the fraction of impatient banks, π. We show that shadow banking forces p to drop in response to an increase in π.[26]

Our exercise here is complementary to the exercise of extending the model to allow for aggregate uncertainty in R and π resolved at date 1, which has been previously studied in the literature.[27] Our analysis shows how the optimal allocation depends on the current external conditions at date 0, rather than how it should be adapted to future contingencies. Both exercises are informative for optimal liquidity regulation of banks.

5.1 Rate of return R

We start out by describing how the first-best allocation responds to changes in R.

Lemma 6 *Under Assumption 1, the first-best price p^{fb} is strictly increasing in R. Correspondingly, the liquid asset investment s_0^{fb} is increasing, the illiquid asset investment x_0^{fb} is decreasing, and both types' final investment, I_1^{fb} and I_2^{fb}, are increasing in R.*

The intuition for the above result is as follows. First-best optimality requires that ex post, i.e., at date 1, the marginal value of the liquid asset and the illiquid asset be the same. Under Assumption 1, higher R reduces the marginal value of the illiquid asset (because V' drops faster than R increases). Thus, the marginal value of the liquid asset must also be reduced, which

[26]For these results, it is not important that shadow banking remains an off-equilibrium threat in our model. In the extended model in Appendix B, shadow banks are on-equilibrium and analogous results hold.

[27]Notably in Allen and Gale (1998) and Allen and Gale (2004a), respectively.

requires that the quantity of the liquid asset be increased, i.e., $s_1^{\text{fb}}(0)$ goes up. By the ex ante resource constraint, the quantity of the illiquid asset must decrease, i.e., $x_1^{\text{fb}}(1)$ goes down. Together, these two changes imply that the retrade-market-clearing price $p^{\text{fb}} = s_1^{\text{fb}}(0)/x_1^{\text{fb}}(1)$ is increasing in R. With $I_1^{\text{fb}} = s_1^{\text{fb}}(0)$, the impatient types benefit from higher R. Also, because $x_1^{\text{fb}}(1)$ decreases slower than R increases, $I_2^{\text{fb}} = Rx_1^{\text{fb}}(1)$ increases. Hence both ex post types benefit from a higher rate of return on illiquid assets in the first-best economy.

It is worth emphasizing that in the first best, provision of insurance to the impatient banks implies that investment in the illiquid asset must go down as the return on this investment, R, goes up. Next, we show that the banks' option to become shadow banks constrains the provision of insurance in this economy when R is sufficiently high.

Proposition 4 *Under Assumption 1, there exists a unique threshold $\hat{R} > 1$ such that the shadow-banking constraint binds if and only if $R \geq \hat{R}$. For all $R < \hat{R}$, the first-best allocation is implementable and, consequently, the constrained-optimal illiquid asset price $p^* = p^{\text{fb}}$ is strictly increasing in R. For all $R > \hat{R}$, the constrained-optimal illiquid asset price $p^* = \frac{\lambda}{\pi} + 1 < p^{\text{fb}}$ does not increase in R, date-0 investment in the liquid and illiquid asset does not change with R, I_1^* does not change with R, and I_2^* increases proportionately to R.*

The intuition for this result follows from two observations. First, the shadow banking constraint is slack at the first-best allocation when R is close to 1. Clearly, when $R = 1$, the illiquid asset offers no excess return over the liquid asset, and the first-best retrade price is $p^{\text{fb}} = 1$. Due to the cost $\lambda > 0$, investing all-illiquid as a shadow bank is strictly dominated. Second, as we increase R, the value of shadow banking grows faster than does the value of remaining a bank. In particular, the market value of a shadow bank's assets at date 1, $\frac{p^{\text{fb}}e}{1+\lambda}$, grows faster than the market value of assets that banks receive at the first-best allocation, $s_1^{\text{fb}}(0) = p^{\text{fb}}x_1^{\text{fb}}(1) = \frac{p^{\text{fb}}e}{\pi p^{\text{fb}}+1-\pi}$.

These two observations explain the result in Proposition 4. If R is small, i.e., close to 1, the participation constraint is slack and the constrained optimal allocation coincides with the first-best allocation. As we increase R, the first best continues to be feasible, but the slack in the participation constraint declines. At $R = \hat{R}$, the participation constraint holds as an equality. In particular, the value of a shadow bank's assets at date 1 reaches the value of a bank's assets, $\frac{p^{\text{fb}}e}{1+\lambda} = \frac{p^{\text{fb}}e}{\pi p^{\text{fb}}+1-\pi}$, which implies $p^* = p^{\text{fb}} = \frac{\lambda}{\pi} + 1$. As R increases above $\hat{R}$, the first-best retrade price p^{fb} continues to increase but can no longer be implemented because any further increase in p would make a shadow bank's assets more valuable than the assets of a bank. The highest implementable price, $p = \frac{\lambda}{\pi} + 1$, is the constrained-optimal price for all $R \geq \hat{R}$.

In sum, the first-best optimal asset price is feasible when the rate of return on illiquid assets is low, but the threat of shadow banking implies that the price of illiquid assets must be suppressed below its first-best level when the return on illiquid assets is high. This price suppression is a macroprudential feature of the optimal allocation in our model: the asset price is capped when its return is high. It is worth emphasizing here that this macroprudential feature is due to the threat of shadow banking. Absent the possibility of shadow banking, the first-best price p^{fb} would be implementable at all R, i.e., the price would be always increasing in the return R. In a market implementation of the optimal allocation, which we discuss in Section 7.1, the need to suppress illiquid asset prices when their return is high will shape the optimal liquidity regulation policy.

Note also that the binding participation constraint restricts the amount of insurance that can be achieved among the banks. While the final investment of the patient banks, $I_2 = Rx_1^{\text{fb}}(1)$, is increasing in R in the region above $\hat{R}$, the final investment of impatient banks, $I_1 = s_1^{\text{fb}}(0)$, is not. The threat of shadow banking implies that only the patient banks benefit from gains in the rate of return on illiquid assets in the constrained optimum.

5.2 Expected liquidity demand π

Now we study how the optimal allocation depends on the level of aggregate demand for liquidity. We model high liquidity demand as a state in which a higher fraction of banks become impatient at date 1. Formally, let $\pi \in \{\underline{\pi}, \bar{\pi}\}$ with $0 < \underline{\pi} < \bar{\pi} < 1$. We will refer to $\underline{\pi}$ as the baseline state and $\bar{\pi}$ as the high liquidity demand state.

For comparative statics, we have three cases. First, it is possible that the bank's ex ante participation constraint does not bind in either state $\underline{\pi}$ or $\bar{\pi}$. Second, it may bind in both. Third, it may bind in just one.

5.2.1 Slack participation constraint

We start by examining the first case, in which the first-best allocation is implementable in both $\underline{\pi}$ and $\bar{\pi}$. The next proposition shows how the first-best allocation depends on π.

Proposition 5 *At the first-best allocation, s_0^{fb} is strictly increasing, and x_0^{fb} is strictly decreasing in π. Both I_1^{fb} and I_2^{fb} are strictly decreasing in π. Further, if the payoff function V exhibits constant relative risk aversion, the first-best illiquid asset price p^{fb} is constant in π.*

Although the initial investment in the liquid asset is larger in the high liquidity demand state $\bar{\pi}$, the ex post allocation of the liquid asset to each impatient bank is smaller. With more impatient banks, the economy is poorer in the state $\bar{\pi}$, as high anticipated demand for liquidity implies a smaller initial investment in the illiquid, high-yield asset at date 0. This decreases the average aggregate return earned on the initial resources e and lowers the amounts of assets allocated to both patient and impatient banks at date 1. As a result, both types attain lower final investment levels $I_1^{\text{fb}} = s_1^{\text{fb}}(0)$ and $I_2^{\text{fb}} = Rx_1^{\text{fb}}(1)$.

How the first-best equilibrium retrade price of the illiquid asset, p^{fb}, depends on π is ambiguous in general. Recall that p is set to equalize the values of assets allocated to the two ex post types, thus making the allocation incentive compatible. With asset allocations reduced for both types when π increases, the impact of higher π on p depends on the relative size of these reductions. In the special case in which the value function of the long-term investment, $V(I)$, exhibits constant relative risk aversion (CRRA), these reductions are proportional to each other. As a result, p^{fb} is independent of π in the CRRA case.

The negative wealth effect of high π also impacts the optimal allocation when the ex ante participation constraint is binding. In that case, as we show next, illiquid asset prices unambiguously fall when high liquidity demand is anticipated.

5.2.2 Binding participation constraint

Proposition 6 *In all states in which the participation constraint binds, s_0^* is strictly increasing and x_0^* is strictly decreasing in π. While I_1^* is strictly decreasing in π, I_2^* remains constant. The constrained-optimal illiquid-asset price, $p^* = \frac{\lambda}{\pi} + 1$, is strictly decreasing in π.*

Recall that incentive compatibility requires that the value of assets allocated to both ex post bank types be the same: $s_1^*(0) = p^* x_1^*(1)$. When the participation constraint binds, additionally, this common value is the same as the value of a shadow bank's assets under the Jacklin deviation of investing all-illiquid, i.e., $s_1^*(0) = p^* x_1^*(1) = p^* \frac{e}{1+\lambda}$. When π increases, all three of these values decrease due to the negative wealth effect. A simple but important observation here is that the value of a shadow bank's assets can only be decreased through a drop in p^*, because the quantity of the (illiquid) assets a shadow bank holds, $\frac{e}{1+\lambda}$, is independent of π and cannot be controlled by the planner. This drop in p^* simultaneously decreases the value of a patient bank's assets, as the optimal mechanism allocates illiquid assets, $x_1^*(1)$, to patient banks. To match this decline in the value of illiquid assets, the quantity of liquid assets allocated to impatient banks, $s_1^*(0)$, must be decreased proportionately to the drop in p^*.

Note that the ex-post patient banks are not hurt by an increase in π. Their allocation of assets, $x_1^*(1) = \frac{e}{1+\lambda}$, is independent of π, and so is their final investment $I_2^* = R x_1^*(1)$. The whole negative wealth effect of higher π is absorbed by the impatient banks through the drop in $I_1^* = s_1^*(0)$. From the ex ante perspective, all banks are negatively affected by an increase in π, as now they have a smaller chance of ending up patient.

Further, note that the amount invested in illiquid assets, x_0^*, and the retrade price, p^*, comove in π. Investment is low in a high liquidity demand state, when the fire-sale price of illiquid assets is low as well. Conversely, with low anticipated liquidity demand, both investment and the price of illiquid assets are high. This feature of our Diamond-Dybvig model with shadow banking is different from macroeconomic models with financial frictions. As discussed in Kiyotaki and Moore (2012) and Shi (2015), in response to a liquidity shock that tightens an asset resellability or a collateral constraint, these models predict an investment drop and a counterfactual asset price boom.

In sum, we have shown that without the threat of shadow banking, illiquid asset prices remain approximately constant in high liquidity demand states. With shadow banking, however, illiquid asset prices must drop when high demand for liquidity is anticipated. Such price dislocations, or fire sales, are often viewed as inefficient. In our model, however, they are constrained-efficient. In particular, fire sales are necessary to reduce the value of shadow banking when high demand for liquidity is anticipated. In Section 7.2, we discuss how optimal liquidity regulation policy can be organized to implement these fire-sale-like drops in asset prices.

5.2.3 The mixed case

Finally, we can have a mixed case, where the participation constraint binds in only one state. In particular, it follows from Proposition 5 and equation (24) that if the participation constraint binds in state $\underline{\pi}$, then it also binds in state $\bar{\pi} > \underline{\pi}$. Therefore, the only mixed case possible is the one in which the participation constraint is slack in the baseline state $\underline{\pi}$ and binds in the high liquidity demand state $\bar{\pi}$. It is easy to verify that the comparative statics in this case are a mixture of the results from the two previous cases. In the high liquidity demand state $\bar{\pi}$, the asset allocations to both types, and thus their final investment I_1 and I_2, decline, as in Proposition 5, while the price of the illiquid asset unambiguously declines, as in Proposition 6.

6 Optimal liquidity regulation

In this section, we discuss implementation of optimal allocations as competitive equilibria with government regulation. If $\lambda = 0$, we have $p^* = 1$ and the constrained optimum coincides with the LF competitive equilibrium defined in Section 3, so no intervention is needed. If $\lambda > 0$, however, $p^* > 1$ and thus, due to the pecuniary externality, the equilibrium allocation is inefficient. Next, we show how this market failure can be corrected with payment of interest on reserves (a subsidy to the liquid bank asset) and a proportional tax on the illiquid asset.

6.1 Competitive asset market equilibrium with IOR and a liquidity tax

Suppose the government pays interest on reserves (IOR) at the rate i and imposes a proportional (i.e., linear) liquidity tax τ on investment in the illiquid asset. The illiquid-asset tax is imposed at date 0, which means banks need to spend $1 + \tau$ dollars in order to originate one dollar worth of the illiquid asset x. With IOR paid to the banks at the rate i, in order to have one dollar of liquidity at date 1, banks need to invest only $(1 + i)^{-1}$ dollars in liquid reserves at date 0.

The government announces the policy rates (i, τ) before banks decide whether to remain a bank (i.e., subject to these policy instruments) or become a shadow bank (not subject to government regulation but having to use the inferior illiquid asset origination technology with the additional proportional cost λ). The government is committed to the preannounced rates. In equilibrium, all banks will at least weakly prefer to remain as banks rather than become shadow banks, and we assume that in the case of indifference they remain as banks. After this participation decision is made, banks invest and pay taxes at date 0, trade in the private market at date 1, and make their long-term investments I_t at dates $t = 1, 2$.

In the notation of Section 2, thus, the liquidity regulation/tax system we consider here is

$$\mathcal{T}(s_0, x_0) = -\frac{i}{1+i} s_0 + \tau x_0, \tag{25}$$

with which the budget constraint (1) reads

$$\frac{s_0}{1+i} + (1+\tau) x_0 \leq e. \tag{26}$$

Competitive equilibrium is defined as in Section 3: Banks maximize (4) subject to the budget constraint (26) at date 0, and budget constraints (2) and (3) at date 1. The government faces the budget constraint $\mathcal{T}(s_0, x_0) \geq 0$.

Next, we study banks' optimal investment choices subject to $\mathcal{T}(s_0, x_0)$ given in (25). We characterize the solution to the banks' problem in the relevant range for the illiquid asset retrade price p.

Lemma 7 *In the asset market economy with $\mathcal{T}$ given in (25), suppose $p \leq R$. Then, the investment portfolio (s_0, x_0) solving a bank's ex ante maximization problem has the bang-bang property: $s_0 = 0$ and $x_0 = \frac{e}{1+\tau}$ if $p > (1+\tau)(1+i)$; $s_0 = (1+i)e$ and $x_0 = 0$ if $p < (1+\tau)(1+i)$; and any (s_0, x_0) such that $\frac{s_0}{1+i} + (1+\tau)x_0 = e$ if $p = (1+\tau)(1+i)$. The ex ante expected value attained by a bank is*

$$\mathbb{E}\left[V\left(\max\left\{1+i, \frac{p}{1+\tau}\right\} e \left(1 - \theta + \theta \frac{R}{p}\right)\right)\right]. \tag{27}$$

The bank's behavior with IOR i and tax rate τ has the same structure as in the laissez-faire equilibrium of Theorem 1. With access to the secondary market for the illiquid asset at date 1, which is open before the long-term investment I_1 has to be made, only the total value $s_0 + px_0$ of a bank's portfolio of liquid and illiquid assets matters to the bank at date 0. Indeed, given a portfolio (s_0, x_0), an impatient bank will sell the illiquid asset and put $I_1 = s_0 + px_0$ in the long-term investment at date 1. A patient bank will spend all its cash in the retrade market increasing its holdings of illiquid assets, as the return on this investment is $R/p > 1$. At date 2, the patient bank invests $I_2 = (s_0 + px_0)R/p$. In either case, the level of I_t the bank can afford is strictly monotone in $s_0 + px_0$, and whether or not the bank can earn the extra return R/p is out of its control.

The banks' problem of allocating resources e between the two assets at date 0, therefore, boils down to maximizing the date-1 value of its asset holdings, $s_0 + px_0$, subject to the budget constraint (26). In this maximization problem, both the objective and the constraint are linear in s_0 and x_0. The solution therefore has the bang-bang structure described in the above lemma.

As an immediate implication, note that banks will not choose an interior portfolio of cash and the illiquid asset unless the after-tax rates of return on these two investments are the same, i.e., $1+i = \frac{p}{1+\tau}$. The constrained-optimal allocation (associated with the retrade price p^*) requires a positive initial investment in both assets. Thus, for this allocation to be implementable, the after-tax returns on the two assets must be equal. We move on to this discussion next.

6.2 Optimal IOR and liquidity tax

Theorem 2 *The optimal retrade price p^* and the associated optimal allocation A^* are an equilibrium in the economy with an IOR rate i and an illiquid-asset tax rate τ if and only if*

$$1+i = \frac{s_1^*(0)}{e}, \tag{28}$$

$$1+\tau = \frac{e}{x_1^*(1)}, \tag{29}$$

$$(1+i)(1+\tau) = p^*. \tag{30}$$

To make the optimal allocation and the optimal price p^* consistent with bank optimization, the policy rates (i, τ) decrease the return on the illiquid-asset investment and increase the return on the liquid-asset investment as of date 0. Indeed, without any taxes or subsidies and with the resale price $p^* > 1$, a bank's value would be maximized by investing all resources e in the illiquid asset at date 0 and holding to maturity if patient, or selling at date 1 if impatient. This strategy would give the bank $x_1(1) = e > x_1^*(1)$ units of the illiquid asset at date 1 if patient, or $s_1(0) = p^*e = \frac{s_1^*(0)}{x_1^*(1)}e > s_1^*(0)$ units of the liquid asset if impatient, which is more than what the optimal allocation assigns, state-by-state.[28] Policy rates (i, τ) in (28) and (29) decrease the value of this investment strategy. With these rates, a deviating bank can acquire at most $\frac{e}{1+\tau}$ units of the illiquid asset at date 0. As (29) shows, this gives it $x_1(1) = \frac{e}{1+\tau} = x_1^*(1)$ units of the illiquid asset at date 1. If impatient, the bank can sell its illiquid portfolio and obtain $s_1(0) = p^*\frac{e}{1+\tau} = (1+i)(1+\tau)\frac{e}{1+\tau} = s_1^*(0)$ units of the liquid asset, where the last equality uses (28). Thus, the all-illiquid investment strategy no longer delivers more assets than $(s_1^*(0), x_1^*(1))$, which makes the optimal allocation and the optimal price p^* consistent with equilibrium.

[28] That $x_1^*(1) < e$ follows from $p^* > 1$ and (21).

More generally, the rates i and τ in (28) and (29) change the structure of the return on initial resources e that a bank can earn regardless of its investment strategy. In the laissez-faire equilibrium of Theorem 1, with $p = 1$, any allocation of resources e to liquid and illiquid investment yields the same gross return of 1 between date 0 and date 1. Impatient banks must cash out at date 1, earning the total gross return of 1. Patient banks are able to earn the additional return of R between dates 1 and 2, which gives them the total return of R. In the optimal equilibrium of Theorem 2, banks are also indifferent with respect to the initial asset allocation. Indeed, $(1+\tau)(1+i) = p^*$ implies that any budget-feasible asset allocation produces the same return of $1 + i = \frac{p^*}{1+\tau}$ between date 0 and date 1. Impatient banks must cash out at date 1, earning the total return of $1 + i$. Patient banks earn an additional return of $\frac{R}{p^*}$ between dates 1 and 2, which gives them the total return of $(1 + i)\frac{R}{p^*} = \frac{R}{p^*}\frac{p^*}{1+\tau} = \frac{R}{1+\tau}$. As we see, a positive IOR rate i increases the total return earned by the impatient banks, from 1 to $1+i$, while a positive illiquid-asset tax τ reduces the total return earned by the patient banks, from R to $R/(1 + \tau)$. By doing so, i and τ improve the ex ante liquidity insurance provided to all banks in equilibrium. In fact, they implement the maximal level of liquidity insurance consistent with incentives and the threat of shadow banking.[29]

The exact magnitude of optimal rates (i, τ) is pinned down by the optimal allocation that is being implemented. As we saw in Section 4.2, the whole allocation is summarized by the optimal illiquid asset retrade price p^*. We can therefore obtain exact expressions for i and τ as a function of p^*. Substituting the values from (21) into (28) and (29), we get

$$i = \frac{(1-\pi)(p^*-1)}{\pi(p^*-1)+1}, \tag{31}$$

$$\tau = \pi(p^*-1). \tag{32}$$

These expressions confirm the optimal policy rates are indeed strictly positive, unless $p^* = 1$, which we recall is the case only if $\lambda = 0$. Further, we see that both i and τ increase in p^*, i.e., larger distortions are needed to implement a larger wedge $p^* - 1$.

Finally, we can use (19) and (20) to express the optimal rates as

$$i = \frac{1-\pi}{\pi}\frac{\min\{\lambda,\bar{\lambda}\}}{1+\min\{\lambda,\bar{\lambda}\}}, \tag{33}$$

$$\tau = \min\{\lambda,\bar{\lambda}\}. \tag{34}$$

As we see, optimal intervention is stronger when the threat of shadow banking is lower, i.e., when the cost of shadow banking λ is higher. Optimal policy rates become flat (no longer increasing) in λ for all $\lambda > \bar{\lambda}$. This, of course, is because the first-best optimum is attained with any $\lambda \geq \bar{\lambda}$, and the optimal price $p^* = p^{\text{fb}}$ no longer increases in λ, as shown in (20).

[29]Note also that positive policy rates i and τ flatten the after-tax yield curve.

6.3 Robustness to regulatory arbitrage

The policy rates (i, τ) given in (33) and (34) are the unique proportional tax/subsidy rates implementing the constrained-optimum as a competitive equilibrium subject to the threat of shadow banking. In our model, the ex ante choice to become a bank or a shadow bank is discrete, i.e., institutions cannot be both (or commingle banking and shadow banking activities). In this section, we point out that under the proportional tax system (i, τ), this restriction is innocuous, i.e., the equilibrium is robust to commingling of banking and shadow banking.

To see this, note first that if $\lambda > \bar{\lambda}$, then (34) implies $\tau = \bar{\lambda} < \lambda$. This means that a) shadow banks prefer to be banks as banks' cost of acquiring illiquid assets is lower and they can also earn IOR on liquid assets, and b) banks have no incentive to try to earn IOR on their holdings of liquid assets while moving their illiquid assets to shadow banking (perhaps by setting up an off-balance-sheet Structured Investment Vehicle), as the cost λ is higher than the tax τ. If $\lambda \leq \bar{\lambda}$, (34) implies that $\tau = \lambda$. Shadow banks cannot earn IOR, but they attain the same ex ante value as banks. This is because their optimal investment strategy is the Jacklin all-illiquid deviation, and under (i, τ) the banks are indifferent between all-illiquid, all-liquid, or an interior portfolio. Thus, banks have no incentive to move illiquid assets to the shadow banking sector because $\tau = \lambda$. Shadow banks could not benefit from having access to IOR because in order to collect it they would need to reduce their holdings of illiquid assets, which would decrease the "Jacklin premium" they can earn on the all-illiquid portfolio.

In addition to showing that equilibrium is robust to this kind of regulatory arbitrage, the above observations also make it clear that IOR is the only reason why banks are willing to hold liquidity (i.e., positive s_0) in equilibrium. With $\tau = \lambda$ and $i = 0$, banks would be sufficiently deterred from becoming shadow banks. But as banks, they would strictly prefer to hold only the illiquid asset if its resale price is p^*. This is because in the absence of i the return on the liquid asset is 1 and the return on the illiquid asset is $p^* \frac{1}{1+\tau} > 1$. The subsidy $i > 0$ to the liquid asset is just large enough to make banks indifferent to holding liquidity, which allows for positive liquid investment ex ante and supports the optimal price $p^* > 1$ as an equilibrium outcome.

Moreover, the robustness of the equilibrium to regulatory arbitrage depends crucially on the linearity of the subsidy-tax system (i, τ). This point can be easily seen in the following two examples.

First, consider a regulatory system in which there is no tax on illiquid assets, but instead the IOR paid to banks is funded with a lump-sum tax T levied on banks. With these instruments, a bank's budget constraint at date 0 is $(1 + i)^{-1} s_0 + x_0 \leq e - T$. It is easy to verify that under the assumption of no commingling of banking and shadow banking this system can implement the constrained-optimal allocation with the IOR rate $i = p^* - 1 > 0$ and the lump-sum tax $T = \pi e \frac{p^*-1}{\pi p^* + 1 - \pi} > 0$. In this system, however, the IOR rate is high (it covers the whole wedge

$p^* - 1 > 0$), which gives banks an incentive to engage in regulatory arbitrage in the following way. By setting up as shadow banks, they can avoid the tax T. By depositing their resources with another bank (one that remained as a bank), they can earn the high IOR. The value of this strategy dominates the equilibrium value because shadow banks, without paying T, still enjoy the same benefits as banks. This equilibrium, therefore, is not robust to regulatory arbitrage, i.e., it does depend on the assumption of no commingling.

Similarly, consider a system in which there is no IOR but instead there is a lump-sum subsidy S to banks funded by a proportional tax τ on illiquid bank assets. A bank's budget constraint is $s_0 + (1+\tau)x_0 = e + S$. It is easy to verify that this system can implement the constrained-optimal allocation with $\tau = p^* - 1$ and $S = (1-\pi)e\frac{p^*-1}{\pi p^* + 1 - \pi} > 0$ if no commingling is allowed. Banks, however, have an incentive to engage in regulatory arbitrage by setting up as banks, in order to collect S, while at the same time holding their illiquid assets in the shadow sector, which effectively lets them pay the cost λ on the investment in illiquid assets instead of paying the tax τ. Outside of the cases with λ higher than $\bar{\lambda}/\pi$, this cost is lower than the tax τ, which makes this strategy preferable to the on-equilibrium strategy. Therefore this equilibrium, too, is not robust to regulatory arbitrage.

7 Comparative statics for optimal policy rates

In this section, we discuss comparative statics for the optimal IOR rate, i, and the illiquid-asset tax rate τ. In particular, we focus on the impact of shadow banking on the response of these rates to changes in R and π.

Section 5 already presents comparative statics for the constrained-optimal allocation and the illiquid asset retrade price, p^*, in a direct mechanism, where the planner controls asset allocations as a function of the banks' reports of their type θ. Expressing these comparative statics in terms of the optimal policy rates helps us think about how optimal liquidity regulation policy should respond to changes in business conditions.

Following Section 5, we first discuss how the optimal policy rates (i, τ) depend on the economy-wide level of return on the illiquid bank assets, R. Then, we discuss how optimal policy rates depend on the aggregate level of demand for liquidity measured by the fraction of impatient banks, π.

7.1 Sensitivity of policy rates to the return R

Proposition 4 and equations (31) and (32) immediately imply the following corollary.

Corollary 1 *Let $\hat{R} > 1$ be the threshold defined in Proposition 4. The optimal policy rates i and τ are strictly increasing in R for all $R < \hat{R}$ but are constant, at $i = \frac{1-\pi}{\pi}\frac{\lambda}{1+\lambda}$ and $\tau = \lambda$, for all $R \geq \hat{R}$.*

Recall that in equilibrium all banks earn the same return between dates 0 and 1, $1 + i = \frac{p^*}{1+\tau}$, and patient banks additionally earn the return of $\frac{R}{p^*}$ between dates 1 and 2, which gives them a total return of $\frac{p^*}{1+\tau}\frac{R}{p^*} = \frac{R}{1+\tau}$. Higher policy rates provide better insurance against the liquidity shock, as they increase the impatient banks' return $1+i$ and decrease the patient banks' return $\frac{R}{1+\tau}$.

For all $R < \hat{R}$, the shadow-banking constraint is slack, which means that the first-period return that can be earned in the shadow sector, $\frac{p^*}{1+\lambda}$, is strictly less than the banks' first-period return $1 + i = \frac{p^*}{1+\tau}$.[30] Clearly, this means $\tau < \lambda$. As we increase R, the optimal retrade price $p^* = p^{\text{fb}}$ increases, and so do the optimal policy rates i and τ that implement it. The first-period return attainable in shadow banking, $\frac{p^*}{1+\lambda}$, increases faster than the banks' first-period return $\frac{p^*}{1+\tau}$, because τ increases and λ is constant.

At $R = \hat{R}$, banks are indifferent to joining the shadow sector because the first-period return in shadow banking becomes equal to the return earned by banks. Clearly, $\tau = \lambda$ when $R = \hat{R}$. As we increase R above $\hat{R}$, the first-best allocation calls for further increases in p, i, and τ, but these cannot be implemented. If τ were to exceed λ, shadow banking would yield a higher first-period return than banking. To keep banks from leaving the regulated sector, τ must remain capped at λ for all $R > \hat{R}$. With $p^* = \frac{\lambda}{\pi} + 1$ constant in R, the IOR rate i also remains constant in order to preserve the equality of banks' return on the liquid and illiquid assets, $1 + i = \frac{p^*}{1+\lambda}$, which implies $i = \frac{1-\pi}{\pi}\frac{\lambda}{1+\lambda}$.

In sum, the threshold $\hat{R}$ defines two regimes for the optimal liquidity policy: a "recession" regime, when R is low; and an "expansion" regime, when R is high. In recessions, the optimal policy rates are low and sensitive to R. In expansions, policy rates are high and completely insensitive to changes in R. Since the return on bank assets is low in recessions, this prediction of the model suggests that the IOR should be cut and any distortions suppressing banks' investment in illiquid assets (provision of credit to businesses) should be loosened in recessions. The exact scope of this policy loosening should be fine-tuned to the state of the economy in recessions. By contrast, in economic booms, the liquidity policy should be tightened, with high IOR and high tax-like distortions on banks' credit expansion. Once policy rate ceilings are reached, policy should not respond to the exact state of the economy in booms.

By holding i and τ constant for all $R \geq \hat{R}$, the optimal policy has a macroprudential effect of keeping asset prices from increasing while their underlying productivity, R, increases. As discussed in Section 5.1, this macroprudential feature is necessitated by the threat of shadow banking. Absent the possibility of shadow banking, the optimal policy rates would be always increasing in R implementing the always-increasing first-best asset price p^{fb}.

[30] In the second period, all patient banks and shadow banks earn the same return, so only the first-period return matters for a bank's decision on whether or not to join the shadow banking sector.

7.2 Sensitivity of policy rates to expected liquidity demand π

As in Section 5.2, let $\underline{\pi}$ denote the baseline state and $\bar{\pi} > \underline{\pi}$ the high liquidity demand state.

If the banks' ex ante participation constraint does not bind in either state $\underline{\pi}$ or $\bar{\pi}$, comparative statics for the optimal policy rates are a simple corollary of Proposition 5 and Theorem 2.

Corollary 2 *Suppose the ex ante participation constraint does not bind in either state $\underline{\pi}$ or $\bar{\pi}$. Then, the optimal IOR rate and the illiquid-asset tax rate satisfy $i(\underline{\pi}) > i(\bar{\pi})$ and $\tau(\underline{\pi}) < \tau(\bar{\pi})$.*

It may sound intuitive that when high liquidity demand is anticipated, the IOR should be increased in order to encourage banks to hold more liquidity. The above result shows that this intuition would not be correct in our model. By Proposition 5, investment in the liquid asset at date 0, s_0, does increase with π. In the implementation, however, the liquid-assets subsidy, i.e., the IOR rate i, decreases. That decrease is necessary due to the negative wealth effect of high anticipated liquidity demand. As the return on investment in the liquid asset, $1 + i$, decreases, the first-period return on the illiquid asset, $\frac{p^{\text{fb}}}{1+\tau}$, must also decline in order to keep banks indifferent. By Proposition 5, the price p^{fb} is approximately constant (exactly constant in the CRRA case). Therefore, the illiquid-asset tax τ must be increased.

Proposition 5 also shows that, with the ex ante participation constraint slack, the negative wealth effect of the high liquidity demand is shared by both types. This can be easily seen in the above corollary if we recall that the total return earned by the patient banks between dates 0 and 2 is $\frac{p}{1+\tau}\frac{R}{p} = \frac{R}{1+\tau}$. Clearly, with $\tau(\bar{\pi}) > \tau(\underline{\pi})$, this return is lower in the high liquidity demand state $\bar{\pi}$.

We now move on to the case in which the banks' ex ante participation constraint binds in both states $\underline{\pi}$ or $\bar{\pi}$. Comparative statics for the optimal policy rates are now a corollary of Proposition 6 and Theorem 2.

Corollary 3 *Suppose the participation constraint binds in both states $\underline{\pi}$ and $\bar{\pi}$. Then, the optimal IOR rate and the illiquid-asset tax rate satisfy $i(\underline{\pi}) > i(\bar{\pi})$ and $\tau(\underline{\pi}) = \tau(\bar{\pi})$.*

Recall that the first-period rate of return earned by banks is $1 + i = \frac{p^*}{1+\tau}$ and by shadow banks is $\frac{p^*}{1+\lambda}$. With the participation constraint binding, these rates of return match each other in both states $\underline{\pi}$ and $\bar{\pi}$. Immediately, this shows that $\tau(\underline{\pi}) = \tau(\bar{\pi}) = \lambda$. From Proposition 6 we also know that the price $p^* = \frac{\lambda}{\pi} + 1$ is strictly decreasing in π. This means that the first-period return in the economy, $1 + i = \frac{p^*}{1+\lambda}$, is low in the high liquidity demand state $\bar{\pi}$, reflecting the negative wealth effect of high π. Further, we can explicitly calculate $i = \frac{1-\pi}{\pi}\frac{\lambda}{1+\lambda}$.

Note also that the second-period return in the economy, $\frac{R}{p^*}$, is higher in the high liquidity demand state $\bar{\pi}$. The total rate of return that patient banks earn over two periods, i.e., between dates 0 and 2, $\frac{p^*}{1+\lambda}\frac{R}{p^*} = \frac{R}{1+\lambda}$, does not decline with π but, rather, remains constant. This way, consistent with Proposition 6, the patient banks share no part of the negative wealth effect of high liquidity demand when the ex ante participation constraint binds.

Finally, we can have the mixed case, where the participation constraint binds only in the high liquidity demand state $\bar{\pi}$. It is easy to verify that the comparative statics for the optimal policy rates are qualitatively the same as in Corollary 2, i.e., the IOR decreases and the illiquid-asset tax rate increases in the high liquidity demand state $\bar{\pi}$. Recall from Section 5.2.3, however, that in the mixed case the price of the illiquid asset unambiguously declines when high liquidity demand is anticipated.

In sum, optimal policy responds to high anticipated liquidity demand as follows. The IOR rate i must decline. The response of the illiquid-asset tax τ and the equilibrium asset price p^* depends on whether the shadow banking constraint binds. If the shadow banking constraint is slack, τ is increased but the equilibrium price of illiquid assets is approximately constant (exactly constant in the CRRA case). If the shadow banking constraint binds, by contrast, the tax on origination of illiquid assets remains constant but the price of illiquid assets responds with a fire-sale-like drop.

8 Optimal regulation via a minimum liquidity requirement

The previous two sections are focused on the IOR/tax implementation of the optimal allocation. In this section, we briefly discuss the corresponding implementation via a quantity restriction.

In a model without the possibility of shadow banking, i.e, without the ex ante participation constraint, Farhi et al. (2009) show how the first-best optimum p^{fb} can be implemented with a minimum liquidity requirement of the following form

$$\frac{s_0}{e} \geq \iota, \tag{35}$$

where ι is a policy parameter.[31] Thus, the quantity regulation of Farhi et al. (2009) takes $\mathcal{T}$ to be identically zero in the budget constraint (1) but instead imposes (35) as a constraint in a bank's maximization problem at date 0. Farhi et al. (2009) show that with

$$\iota = \frac{s_0^{\text{fb}}}{e}$$

the market equilibrium allocation coincides with the first-best optimal allocation associated with the first-best retrade price p^{fb}.

This result translates directly into our model, with the possibility of shadow banking restricting the implementable level of ι. In particular, it is not hard to check that the minimum liquidity requirement (35) implements the constrained-optimal allocation associated with p^* if

[31] Kucinskas (2015) studies a liquidity requirement of this form in a model with mutual funds.

the liquidity floor parameter is set as follows:

$$\iota = \frac{s_0^*}{e} = \min\left\{\frac{s_0^{\text{fb}}}{e}, \frac{\pi+\lambda}{1+\lambda}\right\}. \tag{36}$$

When the participation constraint does not bind, $\iota = \frac{s_0^{\text{fb}}}{e}$ as in Farhi et al. (2009). When it binds, however, ι cannot be set as high as $\frac{s_0^{\text{fb}}}{e}$. The possibility of shadow banking in this case necessitates that the liquidity requirement be loosened.

In particular, when $R \leq \hat{R}$, we have $s_0^* = s_0^{\text{fb}}$. By Proposition 5, s_0^{fb} is strictly increasing in R. Thus, the optimal minimum liquidity requirement is low when R is low, and it increases, i.e., becomes tighter, when the return on illiquid assets grows. When R reaches $\hat{R}$, the minimum liquidity requirement is pinned down by the binding shadow banking constraint. The optimal liquidity floor reaches the level $\frac{\pi+\lambda}{1+\lambda}$ and becomes insensitive to any further increases in R. Also, since both s_0^{fb} and $\frac{\pi+\lambda}{1+\lambda}$ are strictly increasing in π, the minimum liquidity requirement is tightened when high aggregate demand for liquidity is anticipated.

9 Conclusion

In this paper, we extend the pecuniary-externality-based theory of optimal liquidity regulation of banks by allowing for the possibility of shadow banking. We model shadow banking as an arbitrage-seeking activity aimed at avoiding regulation. With the additional constraint introduced by shadow banking, we obtain a tractable mechanism design problem in which the possibility of free riding by shadow banks restricts the set of implementable asset prices. We view our analysis as making three contributions.

First, we show that the possibility of shadow banking requires the implementation of a macroprudential cap on illiquid asset prices when returns on illiquid assets are high but not so when the returns are low. Also, in states of high liquidity demand, fire sales are optimal with shadow banking but not so without shadow banking. With shadow banking, fire sales generate positive comovement between asset prices and investment. These effects are driven by the need to keep the value of shadow banking below the value of formal banking. As such, they are novel to the literature on optimal liquidity regulation of banks.

Second, we discuss optimal policy through the lens of a simple implementation with a flat-rate tax on illiquid assets combined with a flat-rate subsidy to liquid assets similar to the payment of IOR. In 2008, the need to support market liquidity was the justification given for accelerating Congress's authorization for the Federal Reserve to pay IOR to depository institutions.[32] Our analysis provides a normative rationale for IOR consistent with this justification. Indeed, the payment of IOR is in our model a part of the policy implementing the optimal level

[32] See The President's Working Group on Financial Markets (2008).

of liquidity in equilibrium.

Third, the point that the presence of outside options can limit the strength of the pecuniary externality is likely to apply more broadly to economies with pecuniary externalities driven by access to private retrade markets, beyond the simple Diamond-Dybvig maturity-mismatch environment we study here.

There are several interesting ways in which our analysis can be extended. For example, shadow banking can be included in the model as an on-equilibrium phenomenon. In Appendix B, we introduce heterogeneity among banks in the cost of shadow banking λ. The constrained-optimal allocation, then, assigns those banks with low λ to the shadow sector. We verify that our results characterizing optimal liquidity regulations continue to hold in this extension of the model. In particular, we show there that optimal policy must implement a macroprudential asset price suppression as the return on illiquid assets increases and a fire-sale-like asset price discount when anticipated liquidity demand increases.

Further, our model can be extended to endogenize the cost of shadow banking λ. One possible way could be to introduce a friction in the secondary market for illiquid assets that puts shadow banks at a disadvantage relative to banks.[33] Another could be to allow the planner to spend resources ex ante to increase λ.[34] We conjecture that our results are robust to these extensions. Optimal policy would still be implemented with IOR and a tax on illiquid assets with tighter policy needed when the shadow banking constraint becomes tighter. In particular, the planner's incentive to spend resources on increasing λ would be absent in recessions, when R is low, while this margin would become active in expansions, when R is high.

Appendix A: proofs

Proof of Theorem 1

If $p < 1$, then banks' optimal choices are $(s_0, x_0) = (e, 0)$, $(n(0), I_1(0), I_2(0)) = (0, e, 0)$, and $(n(1), I_1(1), I_2(1)) = \left(\frac{e}{p}, 0, R\frac{e}{p}\right)$. Thus, $\mathbb{E}[n] = (1-\pi)\frac{e}{p} > 0$, i.e., the market-clearing condition (6) is violated, and so there is no equilibrium with $p < 1$. If $p > 1$, then banks' optimal choices are $(s_0, x_0) = (0, e)$, $(n(0), I_1(0), I_2(0)) = (-e, pe, 0)$, and $(n(1), I_1(1), I_2(1)) = (-e, pe, 0)$ for $p > R$ or $(n(1), I_1(1), I_2(1)) = (0, 0, Re)$ for $1 < p \leq R$. In either case, the market-clearing condition (6) is violated: with $p > R$, we have $\mathbb{E}[n] = -e < 0$ (both types want to sell); with $1 < p \leq R$, we have $\mathbb{E}[n] = -\pi e < 0$ (the impatient type wants to sell). Thus, there is no equilibrium with $p > 1$.

With $p = 1$, banks are indifferent among all pairs (s_0, x_0) such that $s_0 + x_0 = e$. Given any such pair (s_0, x_0) and $p = 1$, the optimal choices at date 1 are $(n(0), I_1(0), I_2(0)) = (-x_0, e, 0)$

[33] E.g., Diamond (1997) considers limited market access; Plantin (2014) considers adverse selection.

[34] E.g., as in Koeppl et al. (2014).

and $(n(1), I_1(1), I_2(1)) = (s_0, 0, Re)$. The market-clearing condition $-\pi x_0 + (1-\pi)s_0 = 0$ is satisfied if and only if $(s_0, x_0) = (\pi e, (1-\pi)e)$, which gives us a unique equilibrium. QED

Proof of Lemma 1

If A is incentive-feasible with a retrade price p, then redefining $s_1(\theta)$ to be $\hat{s}_1(\theta) = s_1(\theta) - pn(\theta)$, $x_1(\theta)$ to be $\hat{x}_1(\theta) = x_1(\theta) + n(\theta)$, and $\hat{n}(\theta) = 0$ delivers a new allocation that is resource feasible and incentive compatible with the same price p but has zero trade. These two allocations yield the same final investment plan $(I_1(\theta), I_2(\theta))$, hence they achieve the same level of welfare for banks. Because p is the same at both allocations, the value of becoming a shadow bank is not increased. Hence, the ex ante participation constraint is preserved at the new allocation with no trade. QED

Proof of Proposition 1

Necessity:

(i) By contradiction, if $s_1(0)+px_1(0) < s_1(1)+px_1(1)$, then type 0 would misreport to achieve a higher utility in the retrade market. Similarly, if $s_1(0) + px_1(0) > s_1(1) + px_1(1)$, then type 1 would misreport.

(ii) With $x_1(0) > 0$, the impatient type 0 could achieve a higher value than $V(s_1(0)+0Rx_1(0))$ (the value under no-trade) by selling $x_1(0)$ in the retrade market.

(iii) Because type 1 is indifferent between I_1 and I_2, he would choose to invest when it is relatively cheaper. If he chooses $s_1(1) > 0$, it must be the case that I_1 is weakly cheaper, that is, $\frac{R}{p} \leq 1$. Analogously, $x_1(1) > 0$ implies $\frac{R}{p} \geq 1$.

Sufficiency: The equal present value condition (i) implies truth-telling. The impatient type 0 will not want to retrade because of (ii). Likewise, the patient type 1 will not want to retrade at p because of (iii). QED

Proof of Lemma 2

(i) By contradiction, suppose one of the three RF constraints is slack at allocation A. There exists some $\delta > 1$ such that the allocation A^δ defined by $s_1^\delta(\theta) = s_1(\theta)\delta$, $x_1^\delta(\theta) = x_1(\theta)\delta$, $s_0^\delta = (\pi s_1(0) + (1-\pi)s_1(1))\delta$, and $x_0^\delta = (\pi x_1(0) + (1-\pi)x_1(1))\delta$ is still resource feasible. It follows from Proposition 1 that the new allocation A^δ remains incentive compatible under retrade price p. Like A, allocation A^δ satisfies the ex ante participation constraint because the retrade price p is the same, i.e., the deviator's value remains unchanged. Because A^δ achieves higher social welfare, A cannot be optimal.

(ii) We show that an incentive-feasible allocation A in which $s_1(1) > 0$ is dominated by another incentive-feasible allocation A' in which $s_1'(1) = 0$ and $p \leq R$. Note first that since A is IC, by part (iii) in Proposition 1, $s_1(1) > 0$ implies $p \geq R$. We construct A' separately for the case $p > R$ and for the case $p = R$.

(a) If $p > R$, then, by (iii) in Proposition 1, $x_1(1) = 0$ at A. By (ii) in Proposition 1, $x_1(0) = 0$, so $x_0 = 0$ at A. By (i) in Proposition 1, $s_1(0) = s_1(1) = e$ at A. The ex ante value delivered at allocation A is thus $V(e)$. This value is less than $\pi V(e) + (1-\pi)V(Re)$, which is delivered by allocation A' in which $p' = 1$ and

$$\begin{aligned} s_1'(1) &= x_1'(0) = 0, \\ s_1'(0) &= x_1'(1) = e. \end{aligned}$$

Allocation A' satisfies the ex ante participation constraint because if a shadow bank chooses $\tilde{s}_0$ and $\tilde{x}_0 = \frac{e-\tilde{s}_0}{1+\lambda}$, his total income at date 1 is $\tilde{s}_0 + p'\tilde{x}_0 = \frac{e+\lambda\tilde{s}_0}{1+\lambda} \leq e$, and so he receives (weakly) less date-1 income, and thus also the final value, than what he can get as a bank at allocation A'.

(b) If $p = R$, then, by (i) and (ii) in Proposition 1, $s_1(0) = s_1(1) + Rx_1(1)$. Construct an allocation A' by $p' = R$ (i.e., the same as at A), $s_1'(1) = 0$, and

$$\begin{aligned} s_1'(0) &= s_1(0) + \epsilon s_1(1), \\ x_1'(1) &= x_1(1) + \frac{\epsilon+1}{R} s_1(1), \end{aligned}$$

where $\epsilon \equiv \frac{(1-\pi)(R-1)}{\pi R + 1 - \pi}$. It follows from $s_1(0) = s_1(1) + Rx_1(1)$ that $s_1'(0) = s_1'(1) + Rx_1'(1)$, so A' remains incentive compatible. The resource constraint is satisfied by A' because the definition of ϵ implies

$$\pi s_1(0) + (1-\pi)(s_1(1) + x_1(1)) = \pi s_1'(0) + (1-\pi)(s_1'(1) + x_1'(1)).$$

Allocation A' also satisfies the ex ante participation constraint, as the deviator's value remains unchanged under $p' = R = p$.

(iii) By (iii) in Proposition 1, $x_1(1) > 0$ implies $p \leq R$.

QED

Proof of Lemma 3

It follows from $s_1(0) = px_1(1)$ and the resource constraint that $s_1(0) = \frac{pe}{\pi p + 1 - \pi}$ and $x_1(1) = \frac{e}{\pi p + 1 - \pi}$. Therefore, type 0's value is $V(s_1(0)) = V\left(\frac{pe}{\pi p + 1 - \pi}\right)$, and type 1's value is $V(x_1(1)R) =$

$V\left(\frac{pe}{\pi p+1-\pi}\frac{R}{p}\right)$. QED

Proof of Lemma 4

Fix a portfolio $(\tilde{s}_0, \tilde{x}_0)$ chosen by a shadow bank at date 0. At date 1, an impatient shadow bank will sell $\tilde{x}_0$ and invest $\tilde{I}_1 = \tilde{s}_0 + p\tilde{x}_0$. Because $p \leq R$, a patient shadow bank will hold onto $\tilde{x}_0$ and buy $\frac{\tilde{s}_0}{p}$ additional units of the illiquid asset, which gives it the final investment $\tilde{I}_2 = (\frac{\tilde{s}_0}{p} + \tilde{x}_0)R = (\tilde{s}_0 + p\tilde{x}_0)\frac{R}{p}$ at date 2. In both cases, thus, the final investment attained by the shadow bank depends on $(\tilde{s}_0, \tilde{x}_0)$ only though $\tilde{s}_0 + p\tilde{x}_0$, the value of the portfolio at date 1. The shadow bank's portfolio choice at date 0, thus, is equivalent to maximizing this value subject to the budget constraint (5). This is a linear problem with the bang-bang solution $\tilde{s}_0 + p\tilde{x}_0 = e \max\left\{1, \frac{p}{1+\lambda}\right\}$. In sum, the value attained by the impatient shadow bank is $V(\tilde{I}_1) = V\left(e \max\left\{1, \frac{p}{1+\lambda}\right\}\right)$, and the value attained by the patient shadow bank is $V(\tilde{I}_2) = V\left(e \max\left\{1, \frac{p}{1+\lambda}\right\}\frac{R}{p}\right)$. QED

Proof of Lemma 5

Given Lemma 3 and Lemma 4, the private investment constraint can be written as

$$f\left(\frac{p}{\pi p + 1 - \pi}\right) \geq f\left(\max\left\{1, \frac{p}{1+\lambda}\right\}\right),$$

where f is a strictly increasing function defined as

$$f(t) = \mathbb{E}\left[V\left(te\left(1 - \theta + \theta\frac{R}{p}\right)\right)\right]. \tag{37}$$

Applying f^{-1} to the above inequality, we have

$$\frac{p}{\pi p + 1 - \pi} \geq 1 \quad \text{and} \quad \frac{p}{\pi p + 1 - \pi} \geq \frac{p}{1+\lambda},$$

which gives us (16). QED

Proof of Proposition 2

The derivative of the objective function in (14) is $\frac{\pi(1-\pi)}{(\pi p+1-\pi)^2}G(p)$, where

$$G(p) \equiv V'\left(\frac{pe}{\pi p + 1 - \pi}\right)e - V'\left(\frac{Re}{\pi p + 1 - \pi}\right)Re.$$

At $p = 1$, it follows from Assumption 1 that $G(1) = V'(e)\,e - V'(Re)\,Re > 0$. At $p = R$, $G(R) = V'\left(\frac{Re}{\pi R+1-\pi}\right)(1 - R)e < 0$. Intermediate Value Theorem states that there exists $p^{\text{fb}} \in (1, R)$ such that $G(p^{\text{fb}})$ (and thus also the derivative of the objective function) is zero.

Because $V'(\cdot)$ is decreasing, $\frac{pe}{\pi p+1-\pi}$ is increasing in p, and $\frac{Re}{\pi p+1-\pi}$ is decreasing in p, we know that $G(p)$ is decreasing in p. Therefore p^{fb} is unique. QED

Proof of Proposition 3

The derivative of the social welfare function is positive at all $p \in [1, p^{\text{fb}})$ but is negative at $p > p^{\text{fb}}$. If $\lambda \geq \bar{\lambda}$, because p^{fb} is feasible, the constrained-optimal price $p^* = p^{\text{fb}}$. If $\lambda < \bar{\lambda}$, because the social welfare function increases at all $p \in [1, 1 + \frac{\lambda}{\pi}]$, the constrained-optimal price $p^* = 1 + \frac{\lambda}{\pi}$. Thus, in (16) the constraint $p \geq 1$ never binds and the constraint $p \leq 1 + \frac{\lambda}{\pi}$ binds if and only if $\lambda < \bar{\lambda}$. QED

Proof of Lemma 6

For each R, the first-best allocation satisfies

$$V'(s_1^{\text{fb}}(0)) = V'(Rx_1^{\text{fb}}(1))R. \tag{38}$$

We will show that $x_1^{\text{fb}}(1)$ is strictly decreasing in R. By contradiction, suppose $R < R^1$ and $x_1^{\text{fb}}(1)(R) \leq x_1^{\text{fb}}(1)(R^1)$. The resource constraint (13) implies $s_1^{\text{fb}}(0)(R) \geq s_1^{\text{fb}}(0)(R^1)$. Hence,

$$\begin{aligned} V'(Rx_1^{\text{fb}}(1)(R))R &= V'(s_1^{\text{fb}}(0)(R)) \\ &\leq V'(s_1^{\text{fb}}(0)(R^1)) \\ &= V'(R^1 x_1^{\text{fb}}(1)(R^1))R^1 \\ &\leq V'(R^1 x_1^{\text{fb}}(1)(R))R^1, \end{aligned}$$

but Assumption 1 implies that, for a fixed x, $V'(Rx)R$ is a strictly decreasing function of R, which gives us a contradiction. Hence $x_1^{\text{fb}}(1)$ is strictly decreasing in R. The resource constraint (13) thus implies that $s_1^{\text{fb}}(0)$ is strictly increasing in R. These imply that $p^{\text{fb}} = \frac{s_1^{\text{fb}}(0)}{x_1^{\text{fb}}(1)}$ is strictly increasing in R. $I_1^{\text{fb}} = s_1^{\text{fb}}(0)$ is strictly increasing. Because V is strictly concave, $V'(Rx_1^{\text{fb}}(1))R = V'(s_1^{\text{fb}}(0))$ is strictly decreasing in R, which implies that $I_2^{\text{fb}} = Rx_1^{\text{fb}}(1)$ is strictly increasing in R. QED

Proof of Proposition 4

By (19), the threshold $\bar{\lambda}$ below which the participation constraint binds is strictly increasing in p^{fb}. Lemma 6 thus implies that $\bar{\lambda}$ is strictly increasing in R. Let us denote this relation by $\bar{\lambda}(R)$. Given a fixed value of the cost parameter $\lambda > 0$, it is easy to verify that the banks' ex ante participation constraint will bind for R sufficiently large, where $\bar{\lambda}(R) > \lambda$, and not bind for R sufficiently close to 1, where $\bar{\lambda}(R)$ goes to zero and thus is smaller than $\lambda > 0$. By continuity and strict monotonicity of $\bar{\lambda}(R)$, there is a unique threshold $\hat{R} > 1$ such that $\bar{\lambda}(\hat{R}) = \lambda$.

The solution to the mechanism design problem in Proposition 3 and the characterization of the optimal policy rates in Theorem 2 apply here for each R. In particular, the optimal retrade price is

$$p^*(R) = \min\left\{p^{\text{fb}}(R), \frac{\lambda}{\pi} + 1\right\}.$$

Lemma 6 immediately implies that the constrained-optimal price $p^*(R)$ increases with R for all $R < \hat{R}$ and is constant, equal to $\frac{\lambda}{\pi} + 1$, at all $R \geq \hat{R}$. The properties of initial asset allocation and final investment follow from (21)-(22).

Proof of Lemma 7

The ex post budget constraints (2) and (3) can be collapsed to a single constraint

$$I_1 + \frac{p}{R} I_2 \leq s_0 + p x_0. \tag{39}$$

If $\theta = 0$ is realized, the bank will choose $I_1 = s_0 + p x_0$ and $I_2 = 0$. Because $p \leq R$, with $\theta = 1$ the bank will choose $I_1 = 0$ and $I_2 = (s_0 + p x_0) \frac{R}{p}$. Substituting these values into the bank's objective function gives us

$$V\left((s_0 + p x_0)\left(1 - \theta + \theta \frac{R}{p}\right)\right) \tag{40}$$

as the value that type θ bank achieves ex post. The ex ante problem of a bank can be thus written as maximization of $f(\frac{s_0 + p x_0}{e})$ subject to the budget constraint (26), where f is the strictly increasing function defined in (37). This problem is equivalent to maximization of $s_0 + p x_0$ subject to the budget constraint (26). This is a linear problem with the bang-bang solution structure given in the statement of the lemma. Substituting this solution into (40), we get (27). QED

Proof of Theorem 2

Assume i and τ are as in (28) and (29). Then (12) implies

$$(1 + i)(1 + \tau) = \frac{s_1^*(0)}{x_1^*(1)} = p^*.$$

By Lemma 7, it is individually optimal for each bank to invest ex ante in any portfolio such that $\frac{s_0}{1+i} + (1+\tau)x_0 = e$. In particular, the portfolio (s_0^*, x_0^*) does satisfy this condition because $s_0^* = \pi s_1^*(0)$ and $x_0^* = (1 - \pi) x_1^*(1)$ and so

$$\frac{s_0^*}{1 + i} + (1 + \tau) x_0^* = \frac{\pi s_1^*(0)}{\frac{s_1^*(0)}{e}} + \frac{e}{x_1^*(1)}(1 - \pi) x_1^*(1) = e.$$

Market clearing and government budget balance follow from the fact that the optimal allocation A^* is an incentive-feasible allocation with price p^*. In particular, the market-clearing condition at date 1, $\pi p^* x_0^* = (1-\pi)s_0^*$, follows from the fact that A^* and p^* satisfy (12). Thus, we have an equilibrium.

Conversely, suppose p^* and the allocation A^* are an equilibrium in the economy with IOR and the tax on illiquid assets, under some rates (i,τ). Because A^* is interior, Lemma 7 implies that (i,τ) must satisfy

$$(1+i)(1+\tau) = p^*. \tag{41}$$

Because A^* satisfies the banks' budget constraint at date 0, we have $\frac{s_0^*}{1+i} + (1+\tau)x_0^* = e$. Multiplying this by $(1+i)$ and using (41), we get

$$s_0^* + p^* x_0^* = (1+i)e. \tag{42}$$

Because A^* satisfies the banks' budget constraints at date 1, we have $s_0^* + p^* x_0^* = s_1^*(0)$ and $s_0^* + p^* x_0^* = p^* x_1^*(1)$. Using (42), the first of these conditions implies (28). Using (41) and (42), the second one implies (29). QED

Proof of Proposition 5

Equation (38), which holds for each π, implies that $x_1^{\text{fb}}(1)$ and $s_1^{\text{fb}}(0)$ are comonotone, i.e., change in the same direction as π changes. Suppose $x_1^{\text{fb}}(1)$ and $s_1^{\text{fb}}(0)$ are weakly increasing in π, i.e., $x_1^{\text{fb}}(1)(\underline{\pi}) \le x_1^{\text{fb}}(1)(\bar{\pi})$ and $s_1^{\text{fb}}(0)(\underline{\pi}) \le s_1^{\text{fb}}(0)(\bar{\pi})$. This leads to the following contradiction

$$\begin{aligned} e &= \underline{\pi} s_1^{\text{fb}}(0)(\underline{\pi}) + (1-\underline{\pi})x_1^{\text{fb}}(1)(\underline{\pi}) \\ &< \bar{\pi} s_1^{\text{fb}}(0)(\underline{\pi}) + (1-\bar{\pi})x_1^{\text{fb}}(1)(\underline{\pi}) \\ &\le \bar{\pi} s_1^{\text{fb}}(0)(\bar{\pi}) + (1-\bar{\pi})x_1^{\text{fb}}(1)(\bar{\pi}) \\ &= e, \end{aligned}$$

where the first inequality follows from $1 < p^{\text{fb}}(\underline{\pi}) = \frac{s_1^{\text{fb}}(0)(\underline{\pi})}{x_1^{\text{fb}}(1)(\underline{\pi})}$ and $\underline{\pi} < \bar{\pi}$. The RF conditions in Definition 1 imply $(1-\pi)x_1^{\text{fb}}(1) = x_0^{\text{fb}}$. Since both $(1-\pi)$ and $x_1^{\text{fb}}(1)$ are strictly decreasing in π, so is x_0^{fb}. By the RF condition at date 0, s_0^{fb} is strictly increasing in π.

If $V'(I) = I^{-\gamma}$, then (38) reads $s_1^{\text{fb}}(0)^{-\gamma} = x_1^{\text{fb}}(1)^{-\gamma} R^{1-\gamma}$, which implies that $p^{\text{fb}} = \frac{s_1^{\text{fb}}(0)}{x_1^{\text{fb}}(1)} = R^{\frac{\gamma-1}{\gamma}}$ is independent of π. QED

Proof of Proposition 6

In all states with the participation constraint binding, equation (24) implies that $x_1^*(1)$ is constant, equal to $\frac{e}{1+\lambda}$. Thus, $x_0^* = (1-\pi)x_1^*(1) = (1-\pi)\frac{e}{1+\lambda}$ is strictly decreasing in π, and

$s_0^* = e - x_0^*$ is strictly increasing in π. Equation (23) implies that $s_1^*(0) = \left(1 + \frac{1-\pi}{\pi}\frac{\lambda}{1+\lambda}\right)e$, which is strictly decreasing in π. Finally, (20) implies that $p^* = \frac{\lambda}{\pi} + 1$, which is strictly decreasing in π. QED

Appendix B: active shadow banking

In this appendix, we consider an extension of our model in which banks are heterogeneous in their valuation of access to the government safety net. This extension is a natural one given the large degree of heterogeneity observed among banks, measured for example by size.[35] As in the rest of the paper, we model this valuation as the cost of shadow banking λ.

By allowing λ to be small for some institutions, we obtain active shadow banking in equilibrium of the optimal mechanism. In particular, banks with the lowest cost λ, i.e., lowest relative valuation of access to the safety net, become shadow banks in equilibrium. Despite the added complexity, the extended model produces qualitatively the same policy prescriptions as our baseline model: asset prices must be suppressed when the return is high but not so when it is low, fire-sale-like price discounts are optimal in high expected liquidity demand states, and optimal liquidity regulation can be achieved by IOR and a flat-rate tax on illiquid assets.

For brevity, we omit proofs of the results stated in this appendix and demonstrate comparative statics using computed examples. Details of our analysis and proofs are available upon request.

Let us denote the cumulative distribution function for λ by F. Assume $F(\lambda) > 0$ for all $\lambda > 0$, i.e., the measure of banks with λ close to zero is positive. The mechanism design problem is very similar to that in Section 4 except that now it is efficient to have some banks become shadow banks. Indeed, the bank with the lowest cost of shadow banking, $\lambda = 0$, has the option of becoming a shadow bank and acquiring $e/(1+\lambda) = e$ units of the illiquid asset. To convince this bank to not become a shadow bank, the mechanism would have to make this bank at least indifferent, i.e., keep the initial investment of $x_0 = e$ feasible for all formal banks. This implies $\tau = 0$ and $p = 1$, as in the LF equilibrium. We will show that allowing for exit of some banks to shadow banking relaxes the ex ante participation constraint of the remaining banks, which allows for better insurance against the liquidity risk among them. The optimal mechanism delivers higher welfare than the LF equilibrium.

Further, it is easy to show that if a mechanism calls for a bank with cost λ' to become a shadow bank, then it is efficient to have all banks with $\lambda < \lambda'$ become shadow banks. An optimal mechanism will therefore select a threshold λ^* such that all banks with $\lambda \leq \lambda^*$ become shadow banks and banks with a larger cost of shadow banking remain as banks. Also, it is without loss of generality to assume that shadow banks invest all resources in the illiquid asset.

[35] See McCord and Prescott (2014) for evidence from the U.S.

This is because the marginal shadow bank, λ^*, is indifferent to remaining as a formal bank, all formal banks are in equilibrium indifferent to investing in either asset, and all shadow banks have $\lambda \leq \lambda^*$.

As we saw in Section 4, the ex ante expected payoff each formal or shadow bank attains in equilibrium is a strictly increasing function of the value of its asset in the retrade market at date 1. For a formal bank, the ex ante expected payoff is

$$\mathbb{E}\left[V\left((s_0 + px_0)\left(1 - \theta + \theta\frac{R}{p}\right)\right)\right].$$

For a shadow bank with cost λ, this payoff is

$$\mathbb{E}\left[V\left(p\frac{e}{1+\lambda}\left(1 - \theta + \theta\frac{R}{p}\right)\right)\right].$$

The marginal shadow bank is determined by the equality between these two expected payoffs, or, equivalently, between the values of the assets in the retrade market at date 1. Therefore, the marginal shadow bank, i.e., the bank with the cost λ^*, is determined by the following condition:

$$s_0 + px_0 = p\frac{e}{1+\lambda^*}. \tag{43}$$

We can now write the planner's objective function in this mechanism design problem as follows:

$$\begin{aligned} W(s_0, x_0, p, \lambda^*) \equiv & \int_0^{\lambda^*} \mathbb{E}\left[V\left(\frac{pe}{1+\lambda}\left(1 - \theta + \theta\frac{R}{p}\right)\right)\right] dF(\lambda) \\ & + (1 - F(\lambda^*))\,\mathbb{E}\left[V\left((s_0 + px_0)\left(1 - \theta + \theta\frac{R}{p}\right)\right)\right]. \end{aligned}$$

Note here that all formal banks achieve the same ex ante expected payoff, but shadow banks' payoffs are dispersed due to their heterogeneity in λ. The marginal shadow bank attains the same expected payoff as formal banks, which is the lowest payoff among all shadow banks.

The planning problem is to find (s_0, x_0, p, λ^*) that maximize $W(s_0, x_0, p, \lambda^*)$ subject to the RF constraint $s_0 + x_0 = e$, the ex ante participation condition (43), and the following market clearing condition at date 1:

$$p\pi\left(\int_0^{\lambda^*} \frac{e}{1+\lambda} dF(\lambda) + (1 - F(\lambda^*))\,x_0\right) = (1-\pi)\,(1 - F(\lambda^*))\,s_0. \tag{44}$$

In (44), total liquid-asset investment at date 0 is $(1 - F(\lambda^*))\,s_0$, as only formal banks invest in the liquid asset. Total investment in the illiquid asset is $\int_0^{\lambda^*} \frac{e}{1+\lambda} dF(\lambda) + (1 - F(\lambda^*))\,x_0$, with both shadow banks and formal banks holding illiquid assets. In the retrade market at date 1, impatient holders of illiquid assets want to sell $\pi\left(\int_0^{\lambda^*}\frac{e}{1+\lambda}dF(\lambda)+(1-F(\lambda^*))\,x_0\right)$ units of the illiquid asset, while patient banks supply $(1-\pi)\,(1-F(\lambda^*))\,s_0$ units of the liquid asset. In equilibrium, values of assets supplied on both sides must be equal for the retrade market to clear.

Using the constraints (43), (44), and the RF constraint $s_0+x_0=e$, we can express the objective function W as a function of a single variable. We can eliminate s_0, x_0, and p and write the objective W as a function of the threshold λ^* alone. Further and more intuitively, we can write the objective W as a function of the fraction of shadow banks $F(\lambda^*)$. We will treat $F(\lambda^*)$ as the policy choice variable of the planner, with the rest of the allocation being determined by the constraints of the planning problem. Under parametric assumptions on V and F, it is possible to show that $W(F(\lambda^*))$ is single-peaked.

In the reminder of this appendix, we explore the properties of the solution to this reduced mechanism design problem using parameterized examples.

Numerical results

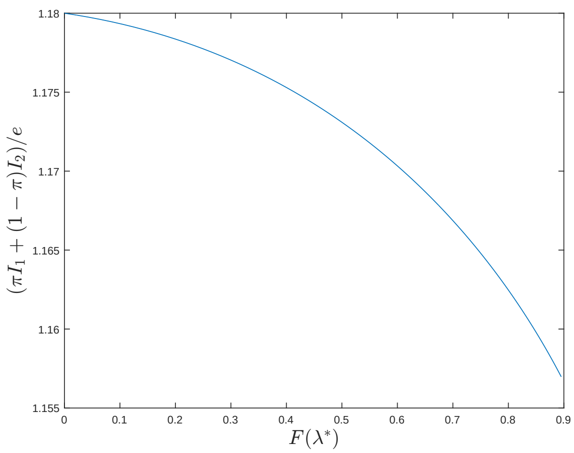
Let $1+\lambda$ follow a Pareto distribution, which means F has a density function $f(\lambda)=\alpha(1+\lambda)^{-(1+\alpha)}$, where $\alpha>0$ is a parameter. Let the banks' payoff from the final investment project I be a power function, $V(I)=\frac{I^{1-\sigma}}{1-\sigma}$, where $\sigma>1$, in line with Assumption 1.

Figure 1 illustrates the optimal choice of the fraction of shadow banks under the following parametrization: $\pi=0.1$, $R=1.2$, $\sigma=10$, and $\alpha=35$. This value of the shape parameter α implies that the median λ is about 2%.

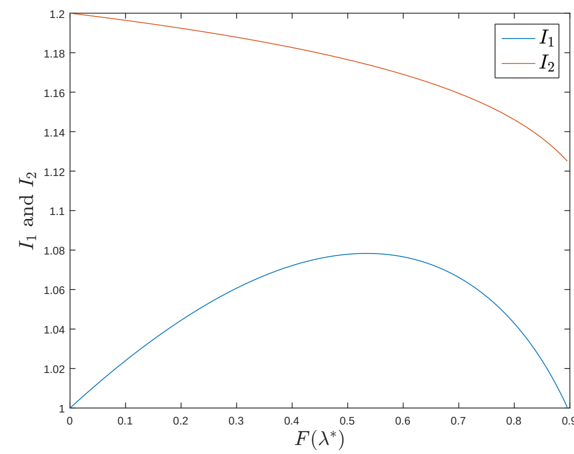
The first observation about this figure confirms that if the planner wants to keep shadow banking off equilibrium, then the LF allocation from Theorem 1 is the best that can be done. If the fraction of shadow banks is zero, as we see, the optimal retrade price is $p=1$, the aggregate return on resources e is $\pi 1+(1-\pi)R=1.18$, the liquid asset allocation is $s_0=\pi=0.1$, and final investment is $I_2=Re=1.2$ for the patient banks and $I_1=e=1$ for the impatient ones. Panel (f) of the figure shows that this outcome is suboptimal. It plots the ex ante expected social welfare W as a function of $F(\lambda^*)$. This function is single peaked, with a unique maximum achieved at the point where the fraction of shadow banks is 0.39.

The intuition for how the optimal fraction of shadow banks is determined is as follows. The benefit of allowing some banks to exit the regulated banking sector and become shadow banks is that the remaining banks can be induced to invest more in the liquid asset, which improves liquidity risk insurance. The cost is that shadow banks will free ride on market liquidity in the retrade market, which impedes insurance.

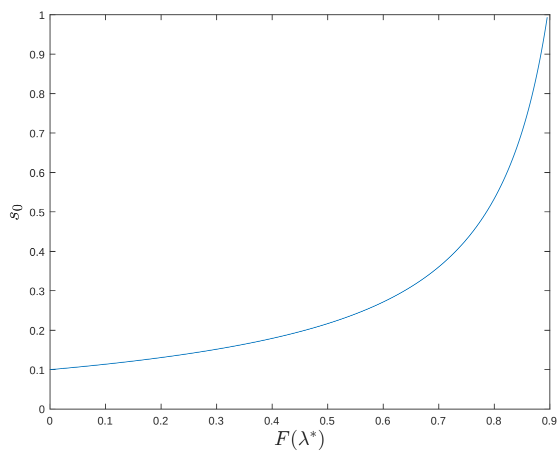
When the size of the shadow sector is small, allowing an additional bank to exit lets the planner support a higher illiquid asset retrade price p (see Figure 1 panel e), which, as in the baseline model, decreases the average return (panel a) but improves insurance (panel b). The



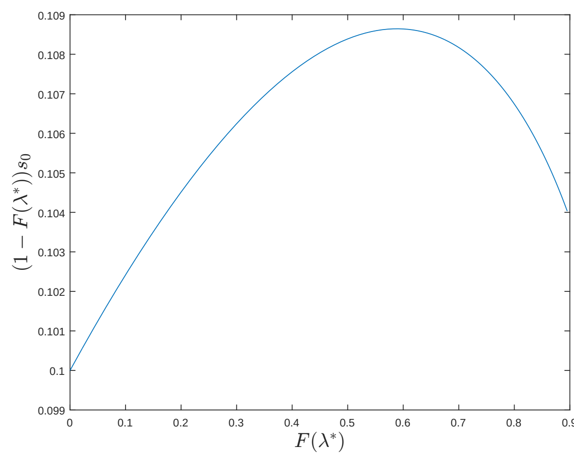
(a) Average total return in the formal sector



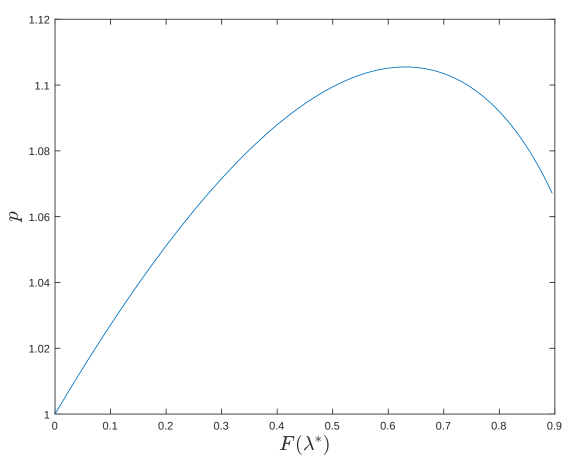
(b) Patient and impatient banks' final investment



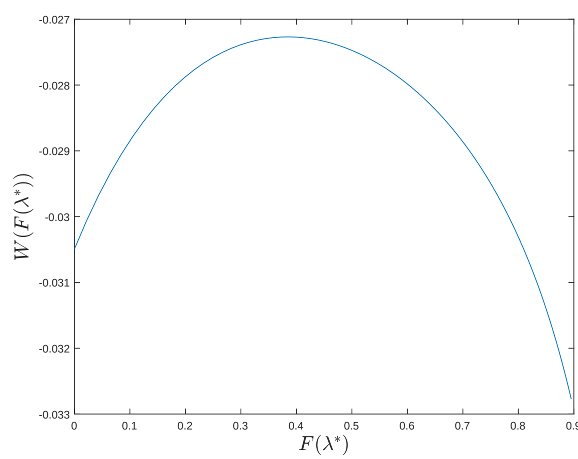
(c) Liquid investment by an individual bank



(d) Aggregate liquidity



(e) Illiquid asset retrade price



(f) Welfare

Figure 1: Trade-offs associated with the choice of the fraction of shadow banks, $F(\lambda^*)$.

price p goes up because the remaining formal banks are made to invest more in the liquid asset at date 0 (panel c), which increases the quantity of cash in the retrade market at date 1 (panel d). The economic trade-off between insurance and the average return is the same as in the baseline model in this paper.

When the shadow banking sector reaches a certain mass, p reaches a peak and begins to decline, and the trade-off changes. Additional liquid investment being forced on the remaining banks does not contribute much to the aggregate liquidity in the retrade market at date 1, as there are fewer and fewer banks (panels c and d). At the same time, as there are more and more shadow banks (investing only in the illiquid asset), the supply of illiquid assets in the retrade market increases. For this reason, the price p reaches a peak and begins to fall as the size of the shadow sector becomes large (panel e). When the shadow sector is large, thus, there is no trade-off between average return in the economy (panel a) and insurance achievable among the banks (panel b), as both decrease with additional shadow banks. Intuitively, this happens because the negative price impact from shadow banks' free riding on market liquidity can no longer be offset by making formal banks invest more in the liquid asset.

Welfare is maximized (panel f) before p reaches its peak (panel e).[36] Thus, even with active shadow banking, the optimal size of the shadow sector is small enough that the optimal retrade price p, and thus the rest of the allocation, continues to be determined by the trade-off between the average return and insurance, as in the baseline model. The fundamental economic intuition of our model, therefore, continues to apply in this extension with active shadow banking. With this trade-off unchanged, it is easy to verify that the optimum with active shadow sector can be implemented via interest on reserves i and a proportional tax on illiquid assets τ, as in Section 6.

Sensitivity to R and π

We now compute comparative statics in the model with active shadow banking. We use the same parametrization as in Figure 1.

First, we discuss comparative statics with respect to R. Figure 2 shows two regions, where the model's solution responds to changes in R quite differently. The optimal IOR rate i, the tax rate τ, and the illiquid asset price p are strictly increasing in R at low levels of R but are fairly flat in R at high levels of R. This is consistent with the analytical comparative statics results obtained for the baseline model in Section 7.1. The kink in the policy rates present in the baseline model at the point where the shadow banking constraint begins to bind is smoothed out here by the adjustment of the size of the shadow banking sector. The optimal size of the shadow banking sector increases in R, consistent with the intuition that the banks' participation

[36] This feature of the optimum is not specific to this numerical example. Clearly, there is no point pushing the size of the shadow sector up beyond the peak of p as beyond this peak both the average return and the degree to which banks can be insured decline.

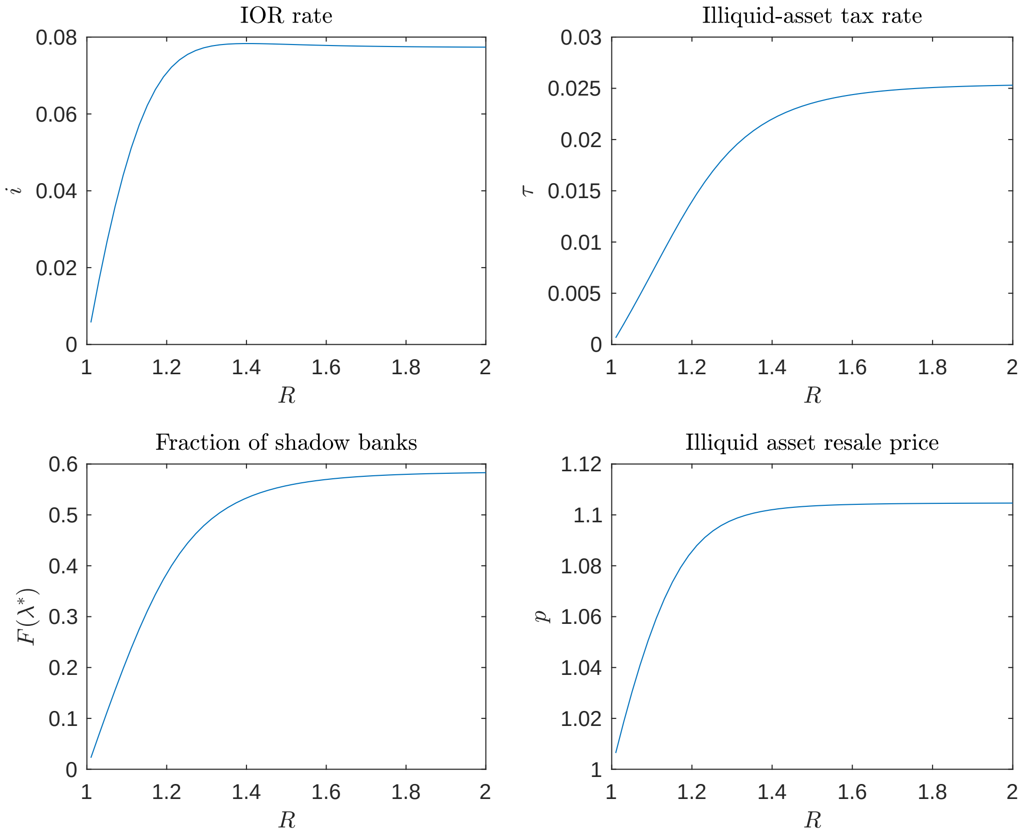


Figure 2: Comparative statics with respect to R with active shadow banking.

constraint is more likely binding when R is higher, which is also present in Section 7.1. The flat portion of the optimal price schedule represents the macroprudential suppression of asset prices, which is needed when asset returns are high, as in the baseline model.

Second, we discuss comparative statics with respect to the fraction of impatient banks, π. The parametrization is the same as above except that we hold the yield on illiquid assets constant at $R = 1.44$. The main results obtained for the baseline model in Section 5.2 and Section 7.2 continue to hold in the extended model in the empirically relevant range for π. In our parametrization, this range is $0 \leq \pi \leq 0.2$, meaning that a liquidity shock, where a bank needs to sell assets prematurely, occurs less frequently than once every five years on average. In this range of π, as shown in Figure 3, the price of the illiquid asset is strictly decreasing in π, consistent with the fire-sale result of Section 5.2. Also, the optimal IOR rate is decreasing in π and the illiquid-asset tax rate τ is increasing, as in Section 7.2 in the baseline model.

In the region of high π, the optimal tax rate τ is decreasing in π. This contrasts the baseline model, where τ is either increasing or flat in π. Two things are important in understanding this difference. First, τ is tied to the fraction of shadow banks here because $\tau = \lambda^*$, while no such tie exists in the baseline model because the fraction of shadow banks is always zero. Second,

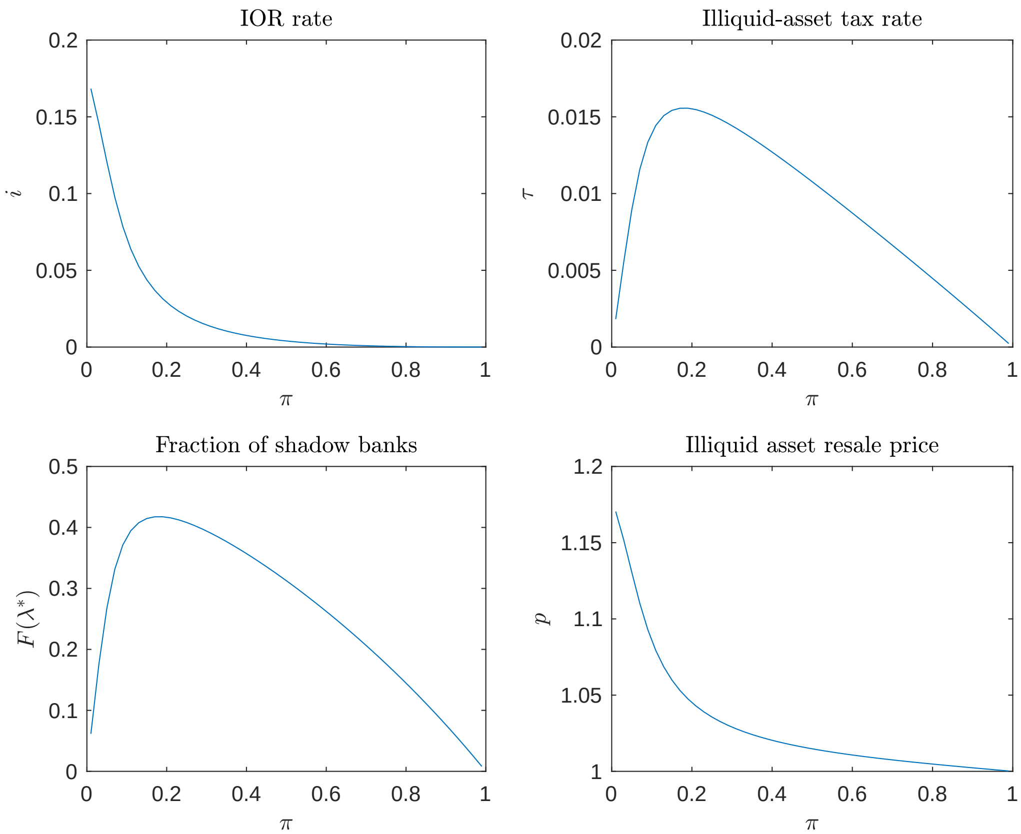


Figure 3: Comparative statics with respect to π with active shadow banking.

$F(\lambda^*)$ approaches zero as π goes to one. When all banks are impatient, there is no investment in illiquid assets. Shadow banks vanish, as they invest only in illiquid assets (bottom-left panel of Figure 3). As π approaches one, thus, λ^* goes to zero. However, these effects kick in only outside of the empirically relevant range of π.

References

Adrian, T. and A. B. Ashcraft (2012). Shadow banking: A review of the literature. *FRBNY Staff Report* (580).

Allen, F. (1985). Repeated principal-agent relationships with lending and borrowing. *Economics Letters 17*, 27–31.

Allen, F., E. Carletti, and D. Gale (2009). Interbank market liquidity and central bank intervention. *Journal of Monetary Economics 56*(5), 639 – 652. Carnegie-Rochester Conference Series on Public Policy: Distress in Credit Markets: Theory, Empirics, and Policy. November 14-15, 2008.

Allen, F. and D. Gale (1994). Limited market participation and volatility of asset prices. *The American Economic Review 84*(4), 933–955.

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

Allen, F. and D. Gale (2004a). Financial fragility, liquidity, and asset prices. *Journal of the European Economic Association 2*(6), 1015–1048.

Allen, F. and D. Gale (2004b). Financial intermediaries and markets. *Econometrica 72*(4), 1023–1061.

Bengui, J., J. Bianchi, and L. Coulibaly (2015). Shadow banking and public liquidity. *Working Paper*.

Bernanke, B. (2012). Fostering financial stability. Speech at the 2012 Federal Reserve Bank of Atlanta Financial Markets Conference, Stone Mountain, Georgia.

Bhattacharia, S. and D. Gale (1987). Preference shocks, liquidity, and central bank policy. In W. Barnett and K. Singleton (Eds.), *New Approaches to Monetary Economics*, pp. 69–88. Cambridge University Press.

Bianchi, J. (2011). Overborrowing and systemic externalities in the business cycle. *American Economic Review 101*(7), 3400–3426.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives 23*(1), 77–100.

Cole, H. L. and N. R. Kocherlakota (2001). Efficient allocations with hidden income and hidden storage. *The Review of Economic Studies 68*, 523–542.

Di Tella, S. (2014). Optimal financial regulation and the concentration of aggregate risk. Stanford GSB Working Paper.

Diamond, D. W. (1997). Liquidity, banks, and markets. *Journal of Political Economy 105*(5), 928–956.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*(3), 401–419.

Farhi, E., M. Golosov, and A. Tsyvinski (2009). A theory of liquidity and regulation of financial intermediation. *The Review of Economic Studies 76*(3), 973–992.

Farhi, E. and J. Tirole (2012). Collective moral hazard, maturity mismatch, and systemic bailouts. *The American Economic Review 102*(1), pp. 60–93.

Farhi, E. and J. Tirole (2017). Shadow banking and the four pillars of traditional financial intermediation. *NBER Working Paper No. 23930*.

Financial Stability Board (2012). Strengthening oversight and regulation of shadow banking.

Financial Stability Forum (2009). Reducing procyclicality arising from the bank capital framework. *Working Paper*.

Freixas, X., A. Martin, and D. Skeie (2011). Bank liquidity, interbank markets, and monetary policy. *Review of Financial Studies 24*(8), 2656–2692.

Gale, D. and T. Yorulmazer (2013). Liquidity hoarding. *Theoretical Economics 8*(2), 291–324.

Geanakoplos, J. and H. Polemarchakis (1986). Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete. In W. Heller, R. Starr, and D. Starrett (Eds.), *Essays in Honor of Kenneth Arrow*, Volume 3, pp. 65–95. Cambridge University Press.

Geanakoplos, J. and K. Walsh (2017). Inefficient liquidity provision. *Economic Theory* (forthcoming).

Golosov, M. and A. Tsyvinski (2007). Optimal taxation with endogenous insurance markets. *Quarterly Journal of Economics 122*(2), 487–534.

Gorton, G. and A. Metrick (2010). Regulating the shadow banking system [with comments and discussion]. *Brookings Papers on Economic Activity Fall*.

Greenwood, R. and D. Scharfstein (2013). The growth of finance. *Journal of Economic Perspectives 27*(2), 3–28.

Holmstrom, B. and J. Tirole (1998). Private and public supply of liquidity. *Journal of Political Economy 106*(1), 1–40.

House, C. L. and Y. Masatlioglu (2015). Managing markets for toxic assets. *Journal of Monetary Economics 70*, 84 – 99.

Huang, J. (2014). Banking and shadow banking. *Princeton University Working Paper*.

Jacklin, C. J. (1987). Demand deposits, trading restrictions, and risk sharing. In E. C. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*, pp. 26–47. Minneapolis, MN: University of Minnesota Press.

Kehoe, T. J. and D. K. Levine (1993). Debt-constrained asset markets. *The Review of Economic Studies 60*(4), 865–888.

Kiyotaki, N. and J. Moore (2012). Liquidity, business cycles, and monetary policy. *NBER Working Paper No. 17934*.

Koeppl, T., C. Monnet, and E. Quintin (2014). Efficient contract enforcement. *Economic Theory 55*(1), 161–183.

Kucinskas, S. (2015). Liquidity creation without banks. *Working Paper VU Amsterdam*.

Lorenzoni, G. (2008). Inefficient credit booms. *The Review of Economic Studies 75*(3), 809–833.

McCord, R. and E. S. Prescott (2014). The financial crisis, the collapse of bank entry, and changes in the size distribution of banks. *FRBR Economic Quarterly 100*, 23–50.

Moreira, A. and A. Savov (2014). The macroeconomics of shadow banking. *NBER Working Paper* (20355).

Ordonez, G. (2015). Sustainable shadow banking. *University of Pennsylvania Working Paper*.

Plantin, G. (2014). Shadow banking and bank capital regulation. *Review of Financial Studies*, 146–175.

Pozsar, Z., T. Adrian, A. Ashcraft, and H. Boesky (2012). Shadow banking. *Federal Reserve Bank of New York Staff Reports* (458), 1–35.

Shi, S. (2015). Liquidity, assets and business cycles. *Journal of Monetary Economics 70*(Supplement C), 116 – 132.

The President's Working Group on Financial Markets (2008, October). Statement HP-1177. http://www.treasury.gov/press-center/press-releases/Pages/hp1177.aspx.

Tirole, J. (2012). Overcoming adverse selection: How public intervention can restore market functioning. *American Economic Review 102*(1), 29–59.

Ueda, K. and B. Weder di Mauro (2013). Quantifying structural subsidy values for systemically important financial institutions. *Journal of Banking and Finance 37*(10), 3830–3842.