

MONEY AND BANKING AS INSURANCE*

Fan Liang

Loyola Marymount University

Florian Madison

Claremont McKenna College

Randall Wright

Zhejiang University and Wisconsin School of Business

10th October 2026

Abstract

This paper analyzes various arrangements for the provision of liquidity insurance in monetary economies. The model has agents facing random desires/needs to make payments, spurring demand for money or money-like assets. Holding such assets constitutes self insurance, but that is inefficient relative to a mutual arrangement that pools risk, like banking in standard (nonmonetary) models. However, mutual insurance raises incentive issues when agents cannot commit. Getting insurance from third parties interpreted as bankers mitigates these incentive problems but introduces others. Different from many banking papers, we study infinite-horizon environments, allowing us to incorporate currency, reputational considerations, and endogenous dynamics.

Keywords: Money, banking, insurance, liquidity

JEL Classification numbers: E42, E58, G20

*For input, we thank Chao Gu, Jamie McAndrews, Daniel Sanches, and Ziqi Qiao. Wright acknowledges support from the Ray Zemon Chair in Liquid Assets and the Ken Burdett Chair in Search Theory and Applications at the University of Wisconsin. The usual disclaimers apply. Emails: fan.liang@lmu.edu, florian.madison@claremontmckenna.edu, and randall.wright@wisc.edu.

The Precautionary-motive: To provide for contingencies requiring sudden expenditure and for unforeseen opportunities of advantageous purchases.
Keynes (1936).

1 Introduction

Liquid assets – money or money-like payment instruments – allow one to respond quickly and at favorable terms to random events. These events can be good, say, you finally find the vintage car you always wanted, or bad, say, your car breaks down. This paper views liquidity preference as insurance, in line with Keynes’ (1936) precautionary motive. We analyze various arrangements for the provision of liquidity insurance in monetary economies. Holding liquid assets constitutes self insurance, but that can be inefficient compared to a mutual risk-pooling arrangement, similar to Diamond and Dybvig (1983), although that is a finite-horizon model without money. In any case, mutual insurance raises incentive issues if agents cannot commit. That generates a role for bankers that mitigate some incentive problems but introduce others.

Our monetary model is based on Lagos and Wright (2005), a framework that has proved useful in many previous studies of liquidity (see Lagos et al. 2017 and Rocheteau and Nosal 2017 for surveys). Liquidity shocks are captured by random desires or needs to acquire something and, moreover, *to pay for it*. Agents can hold money or money-like assets to facilitate these transactions, but again that can be dominated by mutual insurance in a coalition, where those that realize they have more liquidity than they need transfer some to those that want more than they have. This is like, say, fire insurance that makes transfers ex post to those whose houses go up in smoke, given they pay their premiums ex ante, and it is worth emphasizing that these transfers are typically made in cash, not housing.

Again, our mutual insurance coalition is similar to the bank in Diamond-Dybvig models, even if those models do not generally have money. A monetary version is developed in Berentsen et al. (2007), but a key novelty here is that we focus on incentive

problems due to *limited commitment*: ex post agents may renege on transfers they would promise ex ante. Now, economic theory tells us how to implement cooperation without commitment using punishments. A paradigm following Kiyotaki and Moore (1997) punishes renegers by seizing assets, while one following Kehoe and Levine (1993) banishes them from the coalition, which can be interpreted in terms of reputation. But these devices only work well for some parameters – agents must be sufficiently patient, easy to monitor, etc.

For other parameters, full insurance through a mutual arrangement is not incentive compatible, so the coalition can implement only partial coverage, or maybe none at all, in which case some self insurance can still be desirable. Yet there is another possibility. Suppose there are agents not in the coalition, and of no value to the coalition in terms of gains from trade for the usual reasons such as mismatch between preferences and endowments. Could they be of value for strategic reasons? Yes, if they have attributes that make them relatively trustworthy – they may be patient, easy to monitor, etc.

Given these trustworthy agents, others can park their assets with them and delegate to them the task of doling out liquidity to those that want/need it ex post. This is related to the endogenous emergence of banking in Gu et al. (2013a) in a model that has neither money nor an insurance motive, but is similar to our approach in the way it handles incentives. Of course, there are many papers with some claim to banks emerging endogenously, but they do not generally incorporate money, reputational considerations or endogenous dynamics, features that take center stage here.¹

It should go without saying that it is desirable to develop models combining money and banking, which go together like love and marriage, gin and tonic, or peanut butter and jelly. Money and banking are both institutions meant to ameliorate frictions in the exchange process, and it is interesting to see how they interact for a variety of reasons

¹The banking literature is too vast to survey here; see Freixas and Rochet (2023) for that. We mention that a reason many banking papers do not have fiat money, reputational considerations, or endogenous dynamics is that serious analysis requires going beyond finite-horizon models, which are all too common in the banking literature.

(e.g., to better understand connections between bank regulation and monetary policy). It should also go without saying that we are not the first to attempt integrating banking and monetary theory, and there are a few papers using a related approach, but they do not generally deal with our incentive issues or derive our dynamic results.²

Indeed, the idea that money and banking go together goes back a long way (e.g., Gurley and Shaw 1960; Tobin 1963), even if the early discussions do not analyze liquidity the way we can using modern monetary economics. Still, in terms of substance, like some of this earlier work, we discuss differences between fiat and commodity money, which is not only relevant historically, but is relevant now with e-monies that can easily bear interest, including central bank digital currency. Similarly, we discuss fractional reserves, as in recent work by Sanches (2016), Chari and Phelan (2014) and Lee (2025), although they do not study our limited commitment issues.

Also, those papers do not investigate endogenous dynamics the way we do, with the exception of Lee (2025), but again he does not feature limited commitment. Delving into endogenous dynamics allows us to address the venerable notion that banks engender instability, as emphasized by Friedman (1959) and many others (see the survey by Gu et al. 2023 for more discussion and references). Papers based on Diamond-Dybvig only consider one kind of instability, the possibility of bank runs. Our interest goes beyond that: the possibility of generating fluctuations in bank activity and the resulting allocation as self-fulfilling prophecies.

To summarize, the intended contribution is to develop an integrated theory of money, banking, and insurance with limited commitment. Section 2 describes the environment. Section 3 discusses monetary equilibrium. Sections 4 and 5 discuss mutual insurance and banking. Section 6 investigates dynamics. Section 7 concludes.

²We already mentioned Berentsen et al. (2007), who put banks in Lagos-Wright and consider incentive problems for agents taking out loans, different from our concern with incentive problems for banks. He et al. (2008) use a similar model to study bank deposits as payment instruments that are less likely than cash to get lost or stolen. Related work includes Williamson (2012), Matsuoka and Watanabe (2019), Andolfatto et al. (2020), Bethune et al. (2022) and Gu et al. (2023). These papers do not deal with incentive issues or endogenous dynamics the way we do.

2 A Monetary Economy

Time is discrete and continues forever. Each period has two subperiods: first there is a decentralized market (DM) with frictions detailed below; then there is a frictionless centralized market (CM). A set of agents with mass 1 are called consumers, or sometimes buyers, as they may want to make a purchase in the DM, while a set with mass n are called producers, or sometimes sellers, as they can provide it; later a third type is introduced but we do not need them yet. DM trading opportunities arise randomly: with probability $\alpha \in (0, 1)$, which can depend on market tightness $1/n$, a consumer wants to acquire something, denoted q , and is in contact with its producer.³

The event that a consumer has a DM purchase opportunity is realized after the CM closes but before the DM opens. This generates liquidity mismatch – consumers with no DM purchase opportunity have liquid assets they do not need, while those with such an opportunity might like more than they have. In DM trade, the consumer gets $u(q)$ and the producer gets $-c(q)$. In the CM, all agents buy a numeraire good x , different from q , and supply labor h for utility $U(x) - h$, where $U', u', c' > 0$, $U'', u'' < 0 \leq c''$, $u(0) = c(0) = 0$ and $u'(0)/c'(0) = \infty$. The efficient q^* is defined by $u'(q^*) = c'(q^*)$. The common discount factor is $\beta = (1 + r)^{-1}$.

Goods x and q are perishable, but there is a storable object m that, as in standard asset-pricing theory (Lucas 1978), pays a dividend ρ in CM numeraire. When m is used as a DM payment instrument, it is called commodity money if $\rho > 0$ and fiat money if $\rho = 0$ (Wallace 1980). We can also say that m is a *money-like asset* when $\rho > 0$. In the interest of stationarity, when $\rho > 0$ we keep the supply M constant; when $\rho = 0$, the supply can vary and we still get stationarity if the CM price of M , denoted ϕ , falls at the rate M grows, but for simplicity we fix M even when $\rho = 0$.

³This CM-DM structure features an asynchronization of expenditures and receipts: sometimes you want something in the DM while your income accrues in the CM, so you must pay either with assets acquired in the past or debt settled in the future. Also, assets can be used as a medium of exchange, as in Kiyotaki-Wright models, or as collateral in support of deferred settlement, as in Kiyotaki-Moore models, and the results are usually the same (Lagos 2011), but not always (Madison 2024).

The key features of the DM are: consumers lack commitment, so they cannot simply acquire q with a promise to pay in the next CM; and they are anonymous, so we cannot support credit without commitment by taking away defaulters' future credit, as in Kehoe and Levine (1993), or taking away some of their assets, as in Kiyotaki and Moore (1997). Under these assumptions, M has an essential role as a DM payment instrument (Kocherlakota 1998). To be precise about language, in monetary theory, according to Wallace (2010), M is said to be *essential* when the set of incentive-feasible allocations is bigger or better with M than without it. One can similarly think about the essentiality of other institutions including banks (Araujo and Minetti 2011).

Any agent can bring m out of the CM, but if assets are scarce, as discussed below, their price ϕ will exceed the fundamental value ρ/r due to a liquidity premium. In this situation, consumers but not producers will choose $m > 0$, since the former are the ones in need of DM liquidity. For tractability, as explained below in Lemma 1, the framework is designed so all consumers bring the same $m = M$ out of the CM.

As regards modeling strategy, let us say more about liquidity shocks. Each period our consumers randomly have a demand for q , which generates a demand for m , since m is used to pay for q . Papers following Diamond and Dybvig (1983) also have shocks to the demand for, in our notation, q , but it is satisfied by taking q out of the bank and eating it, not taking out m and buying it. Similarly, in papers following Duffie et al. (2005), liquidity shocks are random changes in individual asset valuations, in our notation ρ , and gains from trade arise when low- ρ agents with assets meet high- ρ agents without, but again there is no derived demand for a payment instrument. For studying money and banking, the way we see it, our liquidity shocks seem better.⁴

Many extensions of the baseline model are available. For one, we can allow some

⁴It is worth recalling Hicks (1962): “We must remember the purposes for which liquid assets are held: to pay existing debts, in so far as they cannot be paid out of new receipts, and to meet necessary new expenditures in so far as they cannot be financed in the same way.” Our setup obviously captures “to meet necessary new expenditures” but it can also capture “to pay existing debts” by saying α is not the arrival of a chance to acquire q , but the arrival of a bill for a previous acquisition.

DM credit: you get q now and pay in the CM, with a debt limit that can be exogenous or endogenous (Gu et al. 2016). If credit is unconstrained and available in all DM transactions, there is no role for liquid assets; if it is constrained or available in only some DM transactions, there is a role for liquid assets, and the results are similar to what we get below. Given this is understood, we ignore DM credit in what follows.

We can also assume buyers in the DM can always produce something sellers want, so barter is an option in every trade, but if that has a high cost for buyers or low utility for sellers, there is still a role for money because it improves the terms of trade given everyone likes money that can be used to buy stuff, e.g., CM goods. This captures the idea that liquidity lets one respond to random events at favorable terms, as emphasized by Engineer and Shi (1998, 2001). Or, it may be that buyers can get something sellers really want, but it takes time, and meanwhile the opportunity can be lost and the deal can fall through. This captures the idea that liquidity lets one respond to events quickly, as in Silveira and Wright (2010) or Agarwal and Rempel (2026).⁵

Also note that while we often think of $c(q)$ as a production cost, it can be reinterpreted as an opportunity cost – the utility sellers give up by trading away q – in a pure exchange setting. Also, while we usually call q a good and those buying it consumers, we can alternatively say q is an input and those buying it are firms with production function $u(q)$, or we can say q is an asset and $u(q)$ is the return buyers get while $c(q)$ is the return sellers get from holding it. The point is that liquidity matters in many contexts and it is good to know the model can accommodate.

3 Self Insurance

Consider the consumers' problem (the producers' problem is similar, but easier since they have no need for liquidity, and so is omitted). Let their CM and DM value

⁵These generalizations are relevant because the two attributes of an asset – it can be traded quickly and at favorable terms – have long been associated with liquidity (see Lippman and McCall 1986 and references therein). Given this is understood, our baseline model abstracts from barter.

functions be $W(m)$ and $V(m)$. Then, a consumer in the CM solves

$$W(m) = \max_{x, h, \hat{m}} \{U(x) - h + \beta V(\hat{m})\} \text{ st } x = (\rho + \phi)m + h - \phi\hat{m} - T, \quad (1)$$

where $(\rho + \phi)m$ is the value of the asset taken into the CM, $\hat{m}$ is the amount taken out, h is labor income assuming 1 unit of h produces 1 unit of x , and T is a tax if positive or a transfer if negative, which provides a simple way to change the supply of M .

The following result is standard, and easy to prove when CM utility is linear in h , although Wong (2016) shows it holds for any $U(x, 1 - h)$ with $U_{11}U_{22} = U_{12}^2$.

Lemma 1. *$\hat{m}$ is independent of m ; $W(m)$ is linear in m with slope $\rho + \phi$.*

The first part of Lemma 1 says all consumers take the same $\hat{m}$ out of the CM, regardless of the m they bring in. The second part implies consumers are not risk averse wrt CM wealth, but they are wrt q . In particular, in the DM

$$V(\hat{m}) = \alpha[u(q) + W(\hat{m} - p)] + (1 - \alpha)W(\hat{m}), \quad (2)$$

where α is the probability of a trade, q the quantity acquired, and $p \leq \hat{m}$ the payment.

One can use various methods to determine the DM terms of trade (p, q) . We use Kalai's (1977) bargaining solution, which implies that for consumers to get q they must transfer to producers the value $v(q) = \sigma c(q) + (1 - \sigma)u(q)$, where $\sigma > 0$ is the buyers' bargaining power.⁶ Since $W'(m) = \rho + \phi$ is sellers' marginal value of money in the CM, q solves $v(q) = (\phi + \rho)p$. Recall that q^* solves $u'(q^*) = c'(q^*)$, and let $p^* = v(q^*)$ be the payment needed to get q^* . Then, as usual, we have:

$$p = \begin{cases} p^* & \text{if } p^* \leq \bar{p} \\ \bar{p} & \text{otherwise} \end{cases} \quad \text{and } q = \begin{cases} q^* & \text{if } p^* \leq \bar{p} \\ v^{-1}[\bar{p}(\phi + \rho)] & \text{otherwise} \end{cases} \quad (3)$$

⁶Kalai's solution specifies axioms similar to but not the same as Nash. It is used in many papers with liquidity considerations since Aruoba et al. (2007) because it has several advantages. In fact, Nash and Kalai give the same outcome if $p \leq \hat{m}$ is slack, but not if it binds, except in special cases. One such case is $\sigma = 1$ (buyers have all the bargaining power). When $\sigma = 1$ and $c(q) = q$, as in the examples below, the outcome is the same as Walrasian price taking (which might make sense if buyers and sellers meet in large groups instead of bilaterally, as in Rocheteau and Wright 2005). The point is that there is nothing crucial or special about the way we determine the terms of trade.

Notice we use $\bar{p}$ in the payment constraint here, even though we know $\bar{p} = \hat{m}$ in this case, because $\bar{p}$ always appears in the constraint but later $\bar{p} > \hat{m}$ is possible. Also note that market clearing means $\hat{m} = M$.

To derive the Euler equation for $\hat{m}$, take the FOC from (1), $\phi = \beta V'(\hat{m})$, and insert $V'(\hat{m})$ from (2) using $(\phi + \rho)p = v(q)$. This yields

$$\phi_{t-1} = \beta (\phi_t + \rho) [1 + \alpha \lambda(q_t)], \quad (4)$$

where we are careful with the timing to emphasize that (4) is a difference equation, and $\lambda(q)$ is the *liquidity premium*. One can check that $\lambda(q)$ is the Lagrange multiplier on $p \leq \bar{p}$ and satisfies:

$$\lambda(q) = \begin{cases} u'(q)/v'(q) - 1 & \text{if } q < q^* \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Therefore (4) holds whether or not the constraint binds.

Definition 1. *With self insurance, equilibrium is a path for (q_1, ϕ_1) satisfying difference equation (4) and bargaining condition (3) with $\bar{p} = \hat{m} = M$.*

In addition, equilibrium paths must be bounded and non-negative, but we do not include that explicitly in every definition. In fact, until Section 6 we focus on *stationary monetary equilibrium*, or SME, which means steady state with $\phi_1, q_1 > 0$. As regards notation, subscript 1 on (q_1, ϕ_1) indicates self insurance. Also, $\rho = 0$ ($\rho > 0$) implies there is always (never) a non-monetary equilibrium, $\phi = q = 0$, but obviously we ignore that to concentrate on monetary outcomes.

Proposition 1. *(Self insurance) When $\rho > 0$, SME exists, is unique, and $q_1 = q^*$ if $M \geq M_1 \equiv rv(q^*)/(1+r)\rho$ while $q_1 < q^*$ if $M < M_1$. When $\rho = 0$, SME exists iff $\sigma > \bar{\sigma}_1 \in (0, 1)$, is unique for generic parameters, and $q_1 < q^*$ for all M .*

No proof is given since this is standard (e.g., Gu and Wright 2016). Although this baseline result is not especially novel, our interpretation highlights the role of

m as insurance against “contingencies requiring sudden expenditure and unforeseen opportunities of advantageous purchases.” It also highlights differences between fiat and commodity money: q^* is attainable if $\rho > 0$ and $M \geq M_1$, while q^* is never attainable with $\rho = 0$.⁷ Further, if $\rho = 0$ SME pins down real balances $Z \equiv \phi M$, and hence q , independent of M – classical neutrality. In contrast, if $\rho > 0$ then $\partial q / \partial M = 0$ if $M \geq M_1$ but $\partial q / \partial M > 0$ if $M < M_1$ – classical neutrality does not hold for commodity money, i.e., changing its supply is not generally neutral.

4 Mutual Insurance

Self insurance entails a liquidity mismatch because consumers choose $\hat{m}$ in the CM before knowing whether they need it in the DM. Suppose now that they act as a coalition, where once shocks are realized those without DM trade opportunities transfer assets to those with DM opportunities. This is mutual insurance. While the coalition can in principle be small – two people will suffice if their shocks were perfectly negatively correlated – as a rule bigger is better for pooling risk, so we generally work with the grand coalition, which in this context means all consumers.

As a benchmark, suppose consumers can commit to each other (but still not to DM producers). Then they design a contract $\Gamma = (\pi, \tau)$, where $\pi \leq \hat{m}$ is the transfer from those without DM demand and $\tau = (1 - \alpha) \pi / \alpha$ is the transfer to those with DM demand. The CM value function still satisfies (1), but now

$$V(\hat{m}) = \alpha [u(q) + W(\hat{m} + \tau - p)] + (1 - \alpha) W(\hat{m} - \pi), \quad (6)$$

where $p \leq \bar{p} = \hat{m} + \tau$. The coalition chooses $\hat{m}$ and Γ to maximize

$$-\phi_{t-1} \hat{m}_t + \beta \alpha [u(q_t) + (\hat{m}_t + \tau_t - p_t)(\phi_t + \rho)] + \beta(1 - \alpha)(\phi_t + \rho)(\hat{m}_t - \pi_t), \quad (7)$$

⁷With Kalai bargaining and $\rho = 0$, $q = q^*$ in SME if M grows at rate $\beta - 1 < 0$, which is the Friedman Rule. Also, $q = q^*$ even away from the Friedman Rule if bargaining is replaced by the clever mechanism in Hu et al. (2009) and β is not small. As this is well known, we stick to constant M and bargaining. Also, Andolfatto (2013) gives a rigorous justification for not running the Friedman Rule or deflation more generally: the required fiscal policy may not be incentive feasible.

where again we are being careful with the timing. The first term is the cost of bringing $\hat{m}$ out of the CM; the other two are the surpluses of consumers with and without DM spending opportunities.

For now we can set $\pi = \hat{m}$ for the following reason: even if M/α is more than they want to spend in the DM, since W is linear consumers do not care who brings extra cash to the next CM as long as they all have the same probability α (although this will change below). Then $\bar{p} = M/\alpha$. Emulating Section 3, we get

$$\phi_{t-1} = \beta (\phi_t + \rho) [1 + \lambda(q_t)]. \quad (8)$$

Notice (8) is just like (4) except the α multiplying $\lambda(q)$ is replaced by 1.

Using subscript 2 on q_2 and ϕ_2 to indicate mutual insurance, we have this:

Definition 2. *With mutual insurance and commitment, equilibrium is a path for (q_2, ϕ_2) satisfying (8) and (3) with $\hat{m} = M$ and $\bar{p} = M/\alpha$.*

This definition does not have details of Γ_2 , but clearly $\pi_2 = M$ and $\tau_2 = (1 - \alpha) M/\alpha$. Also, here and in what follows it is understood that SME is a steady state of the system in the definition; we do not repeat this every time.

The next result follows directly from the above discussion:

Proposition 2. *(Mutual insurance, commitment) The results are qualitatively the same as Proposition 1 in terms of existence and uniqueness of SME, but the threshold for M to achieve $q_2 = q^*$ is $M_2 \equiv r\alpha v(q^*)/(1 + r)\rho < M_1$, where M_1 is from self insurance, and $q_2 \geq q_1$, where q_1 is from self insurance, with strict inequality if $q_1 < q^*$.*

Intuitively, since mutual insurance lets agents share liquidity, they get the same marginal value whether or not they trade in the DM. This is efficiency on the extensive margin. Mutual insurance also promotes efficiency on the intensive margin, since the money held by DM buyers M/α exceeds the per-buyer supply M .

Now suppose consumers cannot commit. Then those without DM demand might renege on the transfer π . Assume that deviating from Γ (reneging on π) is detected

with probability $\mu_2 \in [0, 1]$, and if a deviation is detected, the coalition seizes a fraction $\delta_2 \in [0, 1]$ of the deviant's assets and brands the agent a reneger to be excluded from the coalition, but not from the economy entirely, in the future.⁸

The key incentive condition, or IC, is

$$(\phi + \rho)(\hat{m} - \pi) + W(0) \geq \mu_2[(\phi + \rho)(1 - \delta_2)\hat{m} + W_1(0)] + (1 - \mu_2)[(\phi + \rho)\hat{m} + W(0)]. \quad (9)$$

Here W_1 is the continuation value for those caught deviating, which means self insurance: the LHS is the compliance payoff; the RHS is the deviation payoff, where with probability $1 - \mu_2$ a deviant gets away with it and keeps a clean reputation (avoids being branded as a reneger). We can rewrite (9) as

$$\pi \leq \frac{\Lambda + \mu_2 \delta_2 (\phi + \rho) \hat{m}}{\phi + \rho}, \quad (10)$$

where $\Lambda = \mu_2[W(0) - W_1(0)]$ is called the *reputation value*, what you expect to lose from a deviation that probabilistically gets you banned from the insurance coalition.

Higher Λ makes agents less likely to misbehave, or more trustworthy. This is endogenized below, but consider first exogenous Λ . The coalition maximizes (7) as before, but now subject to (10). Also, without commitment we cannot be sure $\pi = \hat{m}$ since bigger π makes deviations more tempting, similar to Peck and Setayesh (2021).

Definition 3. *With mutual insurance, no commitment, and exogenous Λ , equilibrium is the same as the commitment case except Γ is constrained by (10).*

With multipliers ξ_1 , ξ_2 and ξ_3 , the FOCs wrt $(\pi, \tau, \hat{m})$ are:

$$0 = -(1 - \alpha)(\phi + \rho) - \xi_1 + \xi_2(1 - \alpha) - \xi_3(\phi + \rho) \quad (11)$$

$$0 = -\xi_2 + [\lambda(q) + 1](\phi + \rho) \quad (12)$$

$$0 = -\phi_{t-1}/\beta + [\alpha\lambda(q_t) + 1](\phi_t + \rho) + \xi_1 + \xi_3(\phi_t + \rho)\mu_2\delta_2. \quad (13)$$

⁸These punishment devices, seizure and exclusion, are featured in Kiyotaki-Moore and Kehoe-Levine models, respectively. Various stories rationalize $\delta_2 < 1$, like letting renegers abscond with some assets. For $\mu_2 < 1$, since any individual is measure 0, withholding their contribution to the pool of funds to be doled out is not observable unless agents are monitored, like the IRS audits tax compliance, and $\mu_2 < 1$ says the monitoring technology is imperfect.

Proposition 3. *(Mutual insurance, no commitment, exogenous Λ) Existence and uniqueness of SME are the same as above, but now there are thresholds $\underline{\Lambda} \geq 0$ and $\bar{\Lambda} > \underline{\Lambda}$, defined in the proof, such that: (a) $\Lambda < \underline{\Lambda} \Rightarrow \Gamma_a$ with $\pi_a < M$, and compared to the commitment case $q_a < q_2$; (b) $\Lambda \in [\underline{\Lambda}, \bar{\Lambda}) \Rightarrow \Gamma_b$ with $\pi_b = M$ and $q_b < q_2$; (c) $\Lambda \geq \bar{\Lambda} \Rightarrow$ the same outcome as commitment $\Gamma_c = \Gamma_2$ and $q_c = q_2$. In all cases $\tau = (1 - \alpha)\pi/\alpha$.*

The proof is in Appendix A. We now know that for any Λ there is a unique SME – to be precise, as in Proposition 1, when $\rho = 0$ existence of SME requires $\sigma > \bar{\sigma}$ where $\bar{\sigma} > 0$, and when $\rho > 0$ it only requires $\sigma > 0$, but the key point is that whenever SME exists it is unique. Its nature depends on Λ relative to the thresholds $\bar{\Lambda}$ and $\underline{\Lambda}$ in Proposition 3. If $\Lambda \geq \bar{\Lambda}$, then IC is slack, so we get the commitment outcome, regime (c) in Proposition 3. Otherwise (10) binds, and we get two subcases. If $\Lambda \in [\underline{\Lambda}, \bar{\Lambda})$, then the value of money falls to satisfy IC (since M is fixed, this means the demand for money and hence ϕ fall) but we still get $\pi = \hat{m}$, regime (b). If $\Lambda < \underline{\Lambda}$, then $\pi < \hat{m}$, regime (a), where consumers use both mutual and self insurance.

We now endogenize Λ as in Alvarez and Jermann (2000) for nonmonetary and Gu et al. (2016) for monetary models. First, pick an arbitrary Λ and derive consumers' equilibrium payoff, say $W = f(\Lambda)$. Then compute the implied value of reputation as a function of W , say $\Lambda' = g(W)$. This yields $\Lambda' = P(\Lambda) \equiv g \circ f(\Lambda)$ where $g \circ f$ denotes a composite function. An endogenous reputation value is a fixed point $\Lambda = P(\Lambda)$.

The following is obvious from $\Lambda = \mu_2[W(0) - W_1(0)]$:

Lemma 2. *If $\mu_2 = 0$, there is exactly one endogenous Λ , which is $\Lambda = 0$.*

We now concentrate on $\mu_2 > 0$ and $\delta_2 > 0$, implying $P(0) > 0$. Also, $P(\Lambda)$ is constant for $\Lambda \geq \bar{\Lambda}$, so for big Λ we know $P(\Lambda) < \Lambda$. Moreover, Appendix A shows this:

Lemma 3. *If $\mu_2\delta_2 > 0$, then $P(\Lambda)$ is continuous.*

Now the following is immediate from the intermediate value theorem:

Proposition 4. (*Mutual insurance, no commitment, endogenous Λ*) There exists a solution to $\Lambda = P(\Lambda)$, and given Λ , the results are the same as Proposition 3. There can be multiple solutions to $\Lambda = P(\Lambda)$ and hence multiple SME.

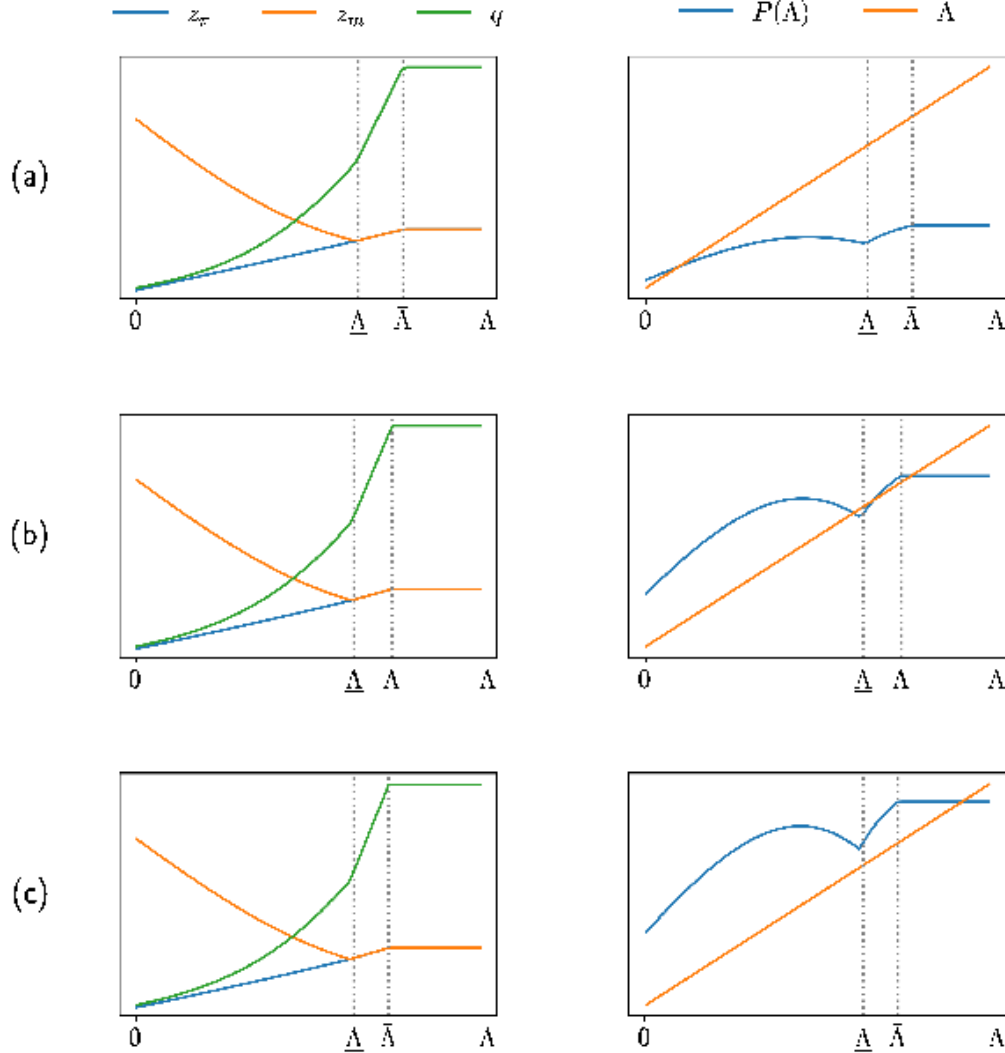


Fig. 1: Mutual insurance, exogenous Λ (left) and endogenous Λ (right).

The results are illustrated in Fig. 1.⁹ On the left, the panels show the real value of

⁹All examples in the paper use DM utility $u(q) = A [(q + b)^{1-a} - b^{1-a}] / (1 - a)$, similar to CRRA utility except $b > 0$ forces $u(0) = 0$ for all a , which can be handy. Also, $c(q) = q$ and $v(q) = q$, where the latter can be understood as bargaining with $\sigma = 1$, but that gives the same outcome as Walrasian price taking. Parameter values for all graphs are listed in Appendix C.

liquidity taken out of the CM, $z_m = M(\phi + \rho)$, the real transfer from those who do not need liquidity, $z_\pi = \pi(\phi + \rho)$, and the quantity q , all as functions of Λ . On the right, the panels show $P(\Lambda)$, and when it crosses the 45° line we have an endogenous Λ . From top to bottom, the graphs use increasing values of μ_2 .

Focusing first on the graphs on the left, as Λ increases from $\Lambda < \underline{\Lambda}$ to $\Lambda \in [\underline{\Lambda}, \bar{\Lambda})$ to $\Lambda \geq \bar{\Lambda}$ we pass through regimes (a), (b) and (c) in Proposition 3. Intuitively, low Λ makes z_π small and z_m big, as consumers top up mutual insurance with some self insurance because the former provides little risk sharing. This bolsters money demand and hence ϕ . As Λ increases, z_π rises and consumers rely less on self insurance, making z_m fall, but on net, liquidity is higher as indicated by q . When Λ increases to $\underline{\Lambda}$, consumers rely solely on mutual insurance, but IC still binds. As Λ passes $\underline{\Lambda}$ mutual insurance does even better, and now z_π and z_m move together (they are equal). When Λ hits $\bar{\Lambda}$, IC becomes slack, as if we had commitment.

Now looking at the graphs on the right, notice $P(\Lambda)$ can be non-monotone because there are two channels in play for $\Lambda \in (0, \underline{\Lambda})$,

$$P'(\Lambda) = \frac{\mu_2}{1 - \beta} \left[\underbrace{\beta(1 - \alpha)\lambda(q_a)}_{\text{insurance channel}} - \underbrace{\frac{\rho(M - m_1)}{\phi_a + \rho} \phi'_a(\Lambda)}_{\text{opportunity cost channel}} \right], \quad (14)$$

where subscript a denotes the allocation under Γ_a and m_1 is money held under self insurance with ϕ_a . Higher Λ relaxes (10) and improves risk sharing, raising $W(0)$, which is the positive insurance channel. The second channel runs through the opportunity cost of carrying real balances, $i \equiv \phi_a / [\beta(\phi_a + \rho)] - 1$, which is increasing in ϕ_a . Since $\phi'_a(\Lambda) < 0$, higher Λ lowers i , and by the envelope theorem this raises $W(0)$ and $W_1(0)$ in proportion to real balances in each case, $(\phi_a + \rho)M$ and $(\phi_a + \rho)m_1$. If $m_1 > M$, the opportunity cost channel is negative.

When P is non-monotone it is natural to expect multiple endogenous values Λ . Of course, non-monotonicity is not necessary for multiplicity – if $\rho = 0$ the second term in (14) vanishes and P is linearly increasing in $[0, \underline{\Lambda})$, but as shown in Fig. 1, P is

concave in $(\underline{\Lambda}, \bar{\Lambda})$ and hence can still have multiple fixed points – but the decomposition in (14) is useful because it explains the non-monotonicity. As the graphs show, high μ_2 and low μ_2 yield a unique fixed point, while intermediate μ_2 yields three, one in each regime (other examples have multiple fixed points in the same regime). Heuristically, multiplicity arises because a higher (lower) reputation makes the cooperative coalition better (worse) and that makes reputation more (less) valuable. Or, as Ernest Hemingway famously put it, “The way to make people trustworthy is to trust them.”

Fig. 2 shows the equilibrium correspondence, i.e., the values of z_m , z_π , q and the endogenous Λ as μ_2 varies, where μ_2 is a “deep” parameter (as usual one can always go “deeper” by endogenizing monitoring, but here it is exogenous). In the graphs on the left, the specification is such that we get uniqueness, while on the right it is such that we get multiplicity. As usual, when there are multiple equilibria, the comparative statics can flip – the middle curves on the right show that higher μ_2 can make Λ , z_m , z_π and q all fall.

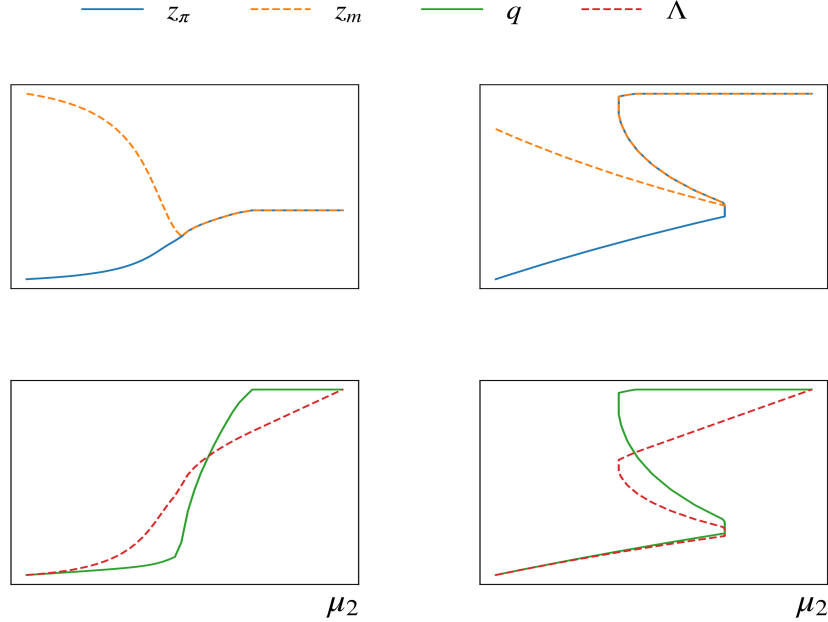


Fig. 2: Mutual insurance, endogenous Λ , uniqueness (left) or multiplicity (right).

In terms of the bigger picture, in case it is not obvious, our mutual insurance arrangement is similar to what gets called banking in the Diamond-Dybvig literature despite some differences.¹⁰ One important difference is that our coalition deals in money, not goods. Another difference from your typical banking paper is our emphasis on commitment frictions that can mean the coalition is not totally trustworthy. One implication is that to satisfy liquidity needs, consumers sometimes use both self and mutual insurance – i.e., both deposits and cash, as is common in reality but not in many models. Also, we do not consider bank runs, but emphasize the coexistence of equilibria with different levels of economic activity and welfare, which can be interpreted as fragility, and this does not require ad hoc restrictions on contracting, only IC constraints due to limited commitment.

That said, a perhaps delicate issue is that our approach is in the spirit of Nash equilibrium, where agents take Λ as given and optimize, leading to something like a best response $\Lambda' = P(\Lambda)$, and we look for fixed points. One might imagine a solution concept closer to the *core*, and when there are multiple fixed points one might think a coalition ought to agree to implement a good one – i.e., bad outcomes might get blocked by some coalitions, if not unilateral deviations (Corbae et al. 2003 proceed along these lines in monetary models). However, if one worries about this, it can be finessed as in Aliprantis et al. (2007a,b) and Gu et al. (2013b) in models of money and credit.

The trick is to have agents randomly relocated each period to interact with a different set of agents, which rules out the kind of blocking just discussed. Then it makes sense for them to take as given Λ and solve for Λ' , with no way to block bad outcomes,

¹⁰This is consistent with conventional wisdom, i.e., Wikipedia: “[In Diamond-Dybvig] the bank is, in effect, modeled as a *coalition* of depositors who pool their resources to provide liquidity insurance against individual consumption shocks,” and “the fundamental insight rests on the bank acting as a *mutual structure*” (emphasis added). The interpretation can depend on whether we think of the institution as insurance or banking but to a large extent that is superficial in the following sense: Appendix B shows how to recast the model in terms of a bank taking deposits and making loans, and the outcome is payoff equivalent. We can go the other way and make the arrangement look more like insurance by saying consumers pay a premium ex ante to be eligible for indemnities ex post, and again the outcome is payoff equivalent.

since those in their coalition next period are not in their coalition this period.¹¹ As important as these details may be, we do not want to get bogged down with them here. The time has come to speak of bankers, not just banks.

5 Bankers

If mutual insurance does not work well, must consumers rely on self insurance? Suppose there are other agents not in the coalition, say, because they have no need for DM trade, but they enjoy CM goods like anyone else. They are our *potential* bankers, whose key attribute is that they may be relatively trustworthy. It is useful to have more than one such agent for incentive reasons – the threat of taking away your banking privileges is more credible when you are not irreplaceable – but we will have only one *active* banker. As Huang (2019) shows in a related context, having more than one active banker has a benefit – they can each be smaller and hence less tempted to misbehave – but also a cost – they all must be monitored.

In the spirit of Diamond (1984), assume it is prohibitively costly to monitor more than one, so there is just one active banker, and the idea is for consumers to park their money with that agent and delegate the task of doling it out for DM use. A banker-coalition contract is $\Gamma \equiv (\pi, \tau, y)$, where y is a fee paid to the banker in CM numeraire, $\pi \leq \hat{m}$ is money given to the banker before the DM shocks, and $\tau \leq \eta(1 - \alpha)\pi/\alpha$ is money doled out after the shocks, where η represents a *reserve requirement* saying the cash doled out cannot exceed a fraction η of the cash taken in. To be clear, anyone going to the DM can use their own $\hat{m}$, whether they or the banker hold it before shocks are realized, plus they can get $\tau = \eta(1 - \alpha)\pi/\alpha$ additional money from the banker. So $\eta = 0$ represents narrow banking while $\eta > 0$ represents fractional reserves.¹²

¹¹We might interpret this as a limited communication friction. It is, by the way, reminiscent of some older banking papers where relocation shocks play a key role (Champ et al. 1996; Bencivenga and Smith 1991; Smith 2002; Bhattacharya et al. 2005).

¹²We thank Jamie McAndrews, who has expertise in these matters, for conversations concerning how to incorporate narrow and fractional reserve banking in the model.

Consumers' $W(m)$ still satisfies (1) but now

$$V(\hat{m}) = -y + \alpha[u(q) + W(\hat{m} + \tau - p)] + (1 - \alpha)W(\hat{m} - \pi),$$

which includes the fee y . The banker's CM value function is

$$J(\Pi) = \Pi + \beta J(\Pi), \quad (15)$$

where $\Pi = y + (\phi + \rho)[(1 - \alpha)\pi - \alpha\tau]$ is profit, including y plus the value of any real balances left over after doling as much as regulation permits. We also impose $\Pi \geq 0$, because it seems natural, although that does not matter for the key results.

The banker and the coalition bargain over Γ and $\hat{m}$. Let $S = S^c + S^b$ be the total surplus from the contract, where $S^b = \Pi$ and the surplus for coalition members is

$$S^c = -\phi_{t-1}\hat{m}_t/\beta - y_t + \alpha[u(q_t) + (\phi_t + \rho)(\hat{m}_t + \tau_t - p_t)] + (1 - \alpha)(\phi_t + \rho)(\hat{m}_t - \pi_t) - S_1^c,$$

where S_1^c is the surplus from self insurance. We assume Kalai bargaining between the consumer coalition and the banker, with θ the power of the former (not to be confused with σ , consumers' power when dealing with DM producers).

The bargaining outcome can be found by solving

$$\max_{\pi, \tau, y, \hat{m}} S^c \text{ st } S^c \leq \theta(S^c + S^b), \quad (16)$$

with additional constraints $\Pi \geq 0$, $\pi \leq \hat{m}$, and $\alpha\tau \leq \eta(1 - \alpha)\pi$. The solution is $\Gamma_3 \equiv (\pi_3, \tau_3, y_3)$ and $\hat{m}_3$, where the subscript indicates there are now three types – ex post DM buyers, ex post DM non-buyers, and the banker. It will turn out that $\tau_3 = \eta(1 - \alpha)\hat{m}_3/\alpha$, so the banker doles out as much as regulation allows, which one can describe as no excess reserves, and $y_3 = (1 - \theta)(S^b + S^c) - (\phi + \rho)(1 - \alpha)(1 - \eta)\hat{m}_3$.

Emulating the above procedure now delivers the Euler equation

$$\phi_{t-1} = \beta(\phi_t + \rho)\{1 + [\alpha + (1 - \alpha)\eta]\lambda(q_t)\}, \quad (17)$$

which looks like (8) except for η , and now $p \leq \bar{p} = [\alpha + \eta(1 - \alpha)]M/\alpha$ is the DM constraint. Thus, liquidity in the DM is lower than with mutual insurance if $\eta < 1$

and the same if $\eta = 1$. Importantly, our method is to assume mutual insurance is not viable, and compare banking to self insurance. When we say it is not viable, we mean the extreme case where there can be no mutual insurance, as guaranteed by setting $\Lambda = 0$ if it is exogenous, or setting $\mu_2 = 0$ if Λ is endogenous. Therefore, in what follows, mutual insurance is off the table.¹³

Definition 4. *With bankers and commitment, equilibrium is a path for (q_3, ϕ_3) satisfying (17) and (3) with $\hat{m}_3 = M$ and $\bar{p} = [\alpha + \eta(1 - \alpha)]M/\alpha$.*

To simplify the exposition, the next result focuses on $\rho > 0$ (sorting out $\rho = 0$ is left as a simple exercise). Then the above discussion implies:

Proposition 5. *(Banking with commitment) SME exists, is unique, and $q_3 = q^*$ iff $M \geq M_3$ where*

$$M_3 \equiv \frac{r v(q^*)\alpha}{\rho(1+r)[\alpha + (1-\alpha)\eta]}. \quad (18)$$

Notice $M_3 = M_1$ if $\eta = 0$, $M_3 \in (M_2, M_1)$ if $\eta \in (0, 1)$, and $M_3 = M_2$ if $\eta = 1$. Hence, an arrangement with a banker is a substitute for mutual insurance, but how good a substitute depends on η . In general, to increase q when it is below q^* we can either increase η or M . Fig.3 displays ϕ as a function of M , basically a standard money demand curve, for different η . Notice ϕ is truncated below by its fundamental value, ρ/r , which is 0 for fiat money.

Suppose bankers lack commitment – e.g., they can abscond with resources under their control to buy CM goods, but more generally bankers could misbehave in any number of ways in the spirit of “cash diversion” models (Biais et al. 2007; DeMarzo and Fishman 2006). Let $\mu_3 \in [0, 1]$ be the probability such opportunistic behavior is detected. If such behavior is detected, the coalition seizes the payment y plus a fraction

¹³This way of proceeding is parallel to the mechanism-design approach in monetary economics (Kocherlakota 1998; Wallace 2010), where it is easy to check in any standard model that if credit is viable it dominates money. So money cannot be essential when credit is viable, but can be essential when credit is not viable. Here, if mutual insurance is viable it dominates any arrangement with a banker, so the latter cannot be essential, unless mutual insurance is not viable.

δ_3 of the resources the banker is handling, plus it can revoke the banker's charter, which means no more banking.

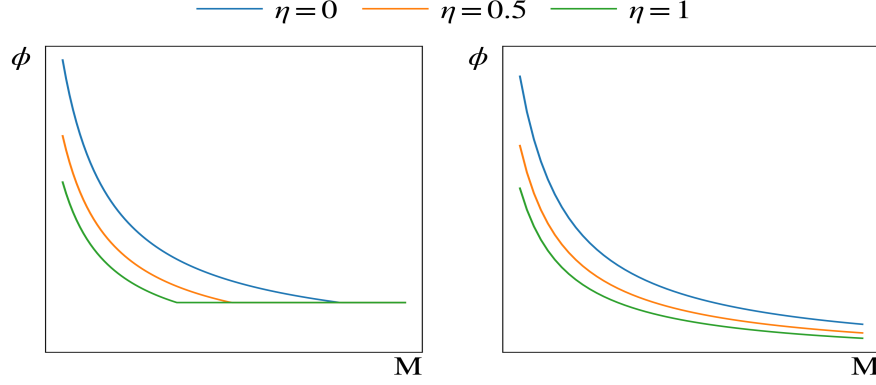


Fig. 3: Banking, ϕ vs M for commodity (left) or fiat (right) money.

IC for the banker means Γ must satisfy

$$\Pi_t + \beta J_{t+1} \geq \mu_3(1 - \delta_3)(1 - \alpha)\pi_t(\phi_t + \rho) + (1 - \mu_3)[(1 - \alpha)\pi_t(\phi_t + \rho) + y_t + \beta J_{t+1}], \quad (19)$$

where J_{t+1} is the continuation value for a banker in good standing and $\Pi = y + (\phi + \rho)[(1 - \alpha)\pi - \alpha\tau]$ is legitimate bank earnings.¹⁴ Rearranging, we get

$$\tau \leq \frac{\Phi + \mu_3[y + \delta_3(1 - \alpha)(\phi + \rho)\pi]}{(\phi + \rho)\alpha}, \quad (20)$$

where $\Phi \equiv \mu_3\beta J_{t+1}$ indicates the value of the banker's reputation, or *franchise value*, which is what gets put in jeopardy by misbehaving.

Analogous to Λ with mutual insurance, before endogenizing Φ , let us start with it exogenous. The contracting problem chooses $\Gamma = (\pi, \tau, y)$ and $\hat{m}$ to maximize S^c given $\pi \leq \hat{m}$, $\alpha\tau \leq \eta(1 - \alpha)\pi$, $S^c \leq \theta(S^c + S^b)$, $\Pi \geq 0$ and (20). If $\xi_1 - \xi_5$ are the multipliers,

¹⁴Note that β appears in (19) but not (9) because disenfranchising bankers affects their payoff starting next period, whereas excluding consumers from mutual insurance affects their payoff starting in the same period.

FOCs wrt (π, τ, y) and $\hat{m}$ are:

$$0 = -(1 - \alpha)(\phi + \rho) - \xi_1 + \xi_2(1 - \alpha)\eta + (1 - \alpha)(\phi + \rho)[\xi_3 + \xi_4 + \mu_3\delta_3\xi_5] \quad (21)$$

$$0 = (\phi + \rho)[\lambda(q_3) + 1] - \xi_2 + \xi_3(\phi + \rho)[(\theta - 1)[\lambda(q_3) + 1] - \theta] - (\phi + \rho)[\xi_4 + \xi_5] \quad (22)$$

$$0 = -1 + \xi_3 + \xi_4 + \mu_3\xi_5 \quad (23)$$

$$0 = [-\phi_{t-1}/\beta + \alpha u'(q_{3,t})\partial q_{3,t}/\partial \hat{m}_t + (1 - \alpha)(\phi_t + \rho)][1 + \xi_3(\theta - 1)] + \xi_1. \quad (24)$$

Definition 5. *With bankers, no commitment, and exogenous Φ , equilibrium is the same as the commitment case except Γ is constrained by (19).*

It is convenient to let Q be the equilibrium q with commitment and $\theta = 1$. The Appendix proves this:

Lemma 4. *There is a threshold $\bar{\Phi}$ such that for $\Phi \geq \bar{\Phi}$ IC for the banker is slack, so the outcome is the same as Proposition 5. For $\Phi < \bar{\Phi}$ IC binds, S^b exceeds the unconstrained bargaining solution $S^b = (1 - \theta)S^c/\theta$, and four types of contracts are possible depending on which of these bind, $\pi \leq M$ and $\Pi \geq 0$: Γ_A has $\pi < M$ and $\Pi > 0$; Γ_B has $\pi < M$ and $\Pi = 0$; Γ_C has $\pi = M$ and $\Pi = 0$; Γ_D has $\pi = M$ and $\Pi > 0$. In each case, $\tau = \eta(1 - \alpha)\pi/\alpha$ and $q < Q$.*

This says that if IC binds, then the banker needs more than the fraction $1 - \theta$ of the total surplus to stay honest – i.e., the banker gets a rent due to the ability to misbehave. Also, as mentioned above, $\tau = \eta(1 - \alpha)\pi/\alpha$ says the banker does out as much as possible. In all four cases in the Lemma where IC binds, $q < Q$ because $(\phi + \rho)\hat{m}$ carried out of the CM is lower than under commitment, where we recall that Q is the equilibrium q with commitment and $\theta = 1$.

The next step is to characterize the contract for a given Φ .

Proposition 6. *(Bankers, no commitment, exogenous Φ) There exist the thresholds $\underline{\mu} \geq 0$ and $\bar{\mu} > \underline{\mu}$, defined in the proof, such that Γ varies with Φ according to:*

- I. For $\mu_3 < \underline{\mu}$, there exist $\underline{\Phi} < \hat{\Phi} < \bar{\Phi}$ such that $\Phi < \underline{\Phi} \Rightarrow \Gamma_A$, $\Phi \in [\underline{\Phi}, \hat{\Phi}) \Rightarrow \Gamma_B$, $\Phi \in [\hat{\Phi}, \bar{\Phi}) \Rightarrow \Gamma_C$, and $\Phi \geq \bar{\Phi} \Rightarrow \Gamma_3$.
- II. For $\mu_3 \in [\underline{\mu}, \bar{\mu})$, there exist $\underline{\Phi} < \Phi^* < \tilde{\Phi} < \bar{\Phi}$ such that $\Phi < \underline{\Phi} \Rightarrow \Gamma_A$, $\Phi \in [\underline{\Phi}, \Phi^*) \Rightarrow \Gamma_B$, $\Phi \in [\Phi^*, \tilde{\Phi}) \Rightarrow \Gamma_D$, $\Phi \in [\tilde{\Phi}, \bar{\Phi}) \Rightarrow \Gamma_C$, and $\Phi \geq \bar{\Phi} \Rightarrow \Gamma_3$.
- III. For $\mu_3 \geq \bar{\mu}$, there exist $\tilde{\Phi} < \bar{\Phi}$ such that $\Phi < \tilde{\Phi} \Rightarrow \Gamma_D$, $\Phi \in [\tilde{\Phi}, \bar{\Phi}) \Rightarrow \Gamma_C$, and $\Phi \geq \bar{\Phi} \Rightarrow \Gamma_3$.

Proposition 6 maps the franchise value Φ to the set of contracts that can arise $\{\Gamma_A, \Gamma_B, \Gamma_C, \Gamma_D, \Gamma_3\}$ conditional on μ_3 . So we organize the presentation according to where μ_3 falls in relation to $\underline{\mu}$ and $\bar{\mu}$. Consider first a low $\mu_3 < \underline{\mu}$. Then only contracts Γ_A, Γ_B , and Γ_C can satisfy IC. For $\Phi < \underline{\Phi}$, we get Γ_A with $\pi < \hat{m}$ and $\Pi > 0$. As Φ increases, consumers first reduce the bank's profit to 0, meaning we get Γ_B , followed by an increase in π to $\pi = \hat{m}$, until eventually IC becomes slack and $\Gamma = \Gamma_3$ for $\Phi \geq \bar{\Phi}$.

Next consider an intermediate $\mu_3 \in [\underline{\mu}, \bar{\mu})$. The difference from $\mu_3 < \underline{\mu}$ is that now as Φ increases from $\underline{\Phi}$, instead of directly going to $\pi = \hat{m}$ with $\Pi = 0$, first the banker gets $\Pi > 0$ with $\pi = \hat{m}$ for $\Phi \in [\Phi^*, \tilde{\Phi})$. Then, once Φ increases to $\Phi \geq \tilde{\Phi}$, Π is reduced to 0, the contract Γ_C , until eventually $\Gamma = \Gamma_3$ for $\Phi > \bar{\Phi}$. Lastly, consider $\mu_3 \geq \bar{\mu}$. Now the only IC contracts are Γ_C and Γ_D . In this case, monitoring is sufficient for $\pi = \hat{m}$. While bankers need positive profit for a low Φ , as Φ increases, profit can be reduced to 0 at $\Phi \geq \tilde{\Phi}$ until eventually IC is slack at $\Phi > \bar{\Phi}$.

We now endogenize Φ . Similar to mutual insurance, we construct the mapping $\Phi' = H(\Phi)$, analogous to $\Lambda' = P(\Lambda)$, and look for a fixed point $\Phi = H(\Phi)$.

Proposition 7. *(Bankers, no commitment, endogenous Φ) For $\mu_3\delta_3 = 1$, we get SME with Γ_3 . For $\mu_3\delta_3 < 1$, we get at least one SME with Γ_A or Γ_D .*

Note that only contracts Γ_A and Γ_D are possible with Φ endogenous, since we need $\Pi > 0$ for a positive franchise value. Also, the proof of Proposition 6 implies that as

we increase Φ , the switch from Γ_B to Γ_D at Φ^* generates a discontinuity; the transition from Γ_A to Γ_B , by contrast, is continuous, which suffices for existence because $H(0) > 0$ under Γ_A while $H(\Phi) = 0 < \Phi$ for every $\Phi \in [\underline{\Phi}, \Phi^*)$ under Γ_B . A second discontinuity arises at $\Phi = \bar{\Phi}$, where the terms of trade shift from a regime in which bankers capture more than $1 - \theta$ of the total surplus to what they would get with Kalai bargaining and no IC. Multiple SME can coexist, as under mutual insurance, but the difference is that here the outcome is shaped by bargaining between the coalition and banker.

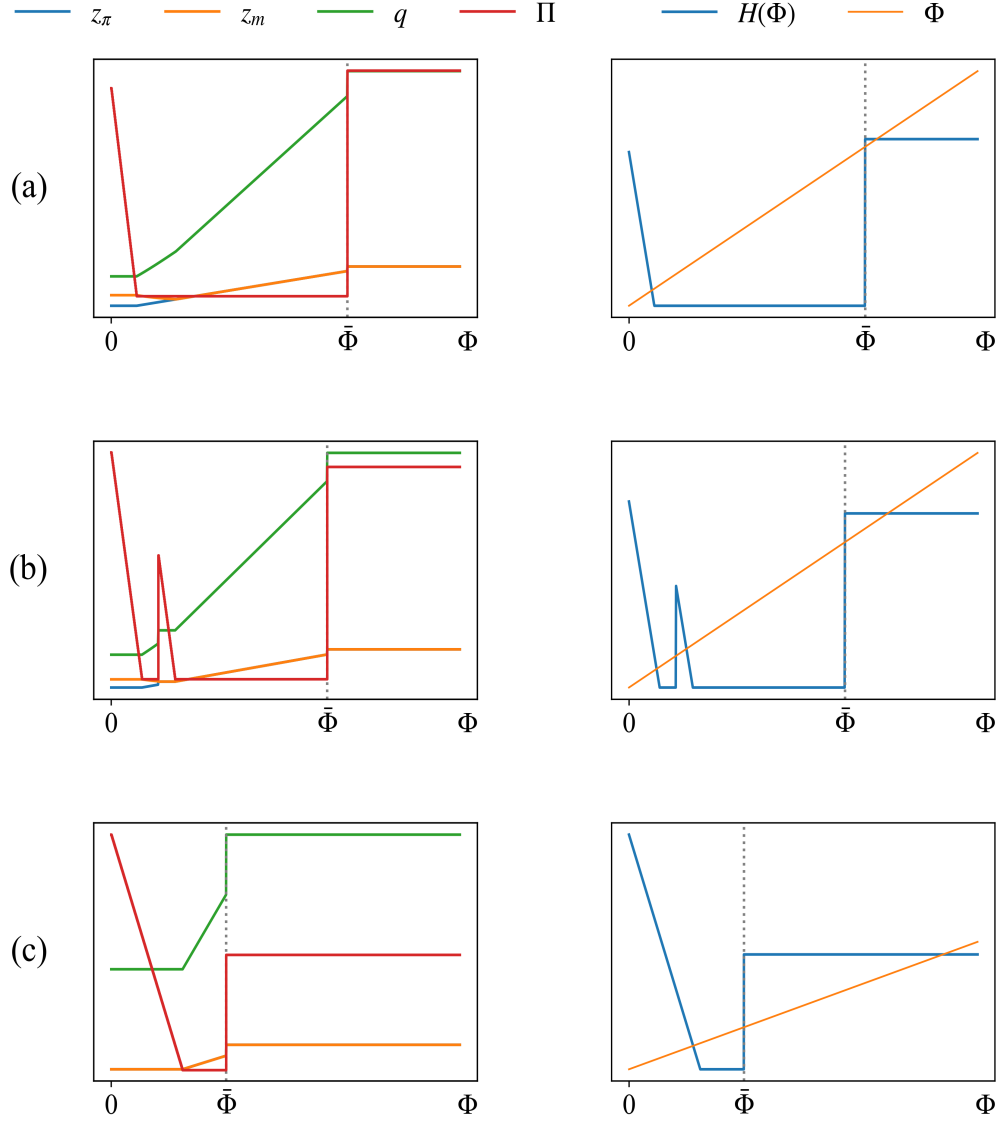


Fig. 4: Banking, exogenous Φ (left) and endogenous Φ (right).

Fig. 4 illustrates the results. As with mutual insurance, the left panel displays outcomes taking Φ as given while the right displays $\Phi' = H(\Phi)$. From panels (a) to (c), μ_3 increases. The left plot of panel (a) illustrates how contracts switch with Φ as described above for $\mu_3 < \underline{\mu}$, the left plot of panel (b) does the same for $\underline{\mu} \leq \mu_3 < \bar{\mu}$, and the left plot of panel (c) does it for $\mu_3 \geq \bar{\mu}$.

On the right, in panel (a), H crosses the 45° line at Γ_A and Γ_3 . In panel (b), there

are crossings at Γ_A , Γ_D , and Γ_3 . Lastly, in panel (c), there are fixed points at Γ_D and Γ_3 . This reflects the role of beliefs: if consumers believe the banker is trustworthy, the bank has a high franchise value, and the equilibrium contract is Γ_3 ; if consumers doubt the banker's trustworthiness, the franchise value is low and $\Gamma = \Gamma_D$ and/ or Γ_A . Note that higher bargaining power of the banker shifts up the flat part to the right of $\bar{\Phi}$, making coexistence of undistorted and distorted equilibria more likely. Lastly, note that $\partial \bar{\Phi} / \partial \mu_3 < 0$, indicating that better monitoring lowers the franchise value needed to keep the IC slack.

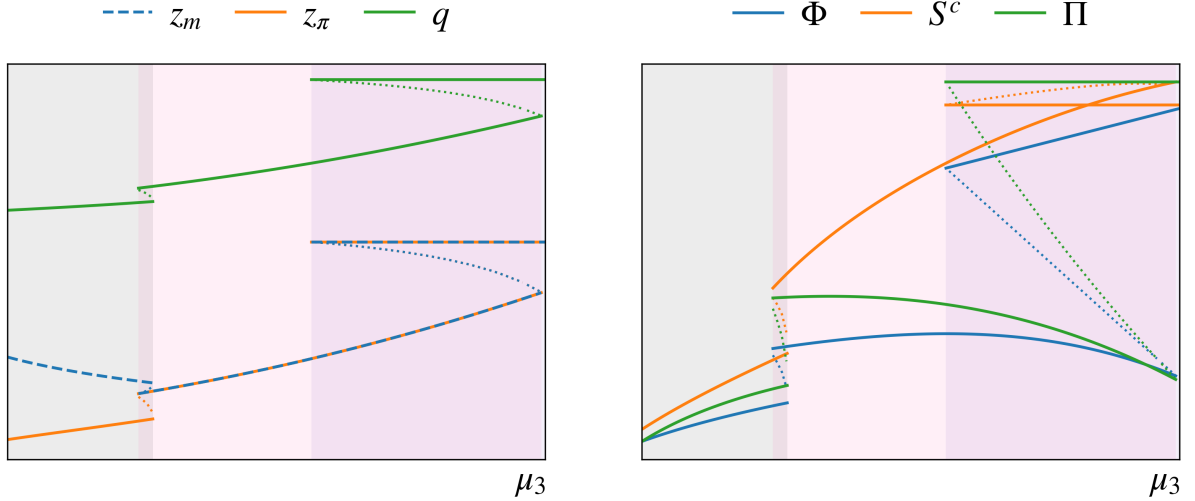


Fig. 5: Banking, endogenous Φ .

Fig. 5 plots z_m , z_π , q , S^c , Π and the endogenous Φ as μ_3 varies. Gray, pink, and purple shading indicate the regions in which the equilibrium contract is Γ_A , Γ_D , and Γ_3 , respectively, and dotted segments represent jumps. For low μ_3 , the contract is unique and given by Γ_A , with $\pi < m$. As μ_3 increases, Γ_D emerges with $\pi = m$ and coexists with Γ_A over an intermediate range. As μ_3 rises further, at first Γ_D is unique until eventually Γ_3 appears and coexists with Γ_D thereafter. Note that multiplicity persists even for very large μ_3 . The reason is that a deviator's outside option is zero

here but positive under mutual insurance, which is why banking delivers multiplicity even when monitoring is good, while mutual insurance delivers uniqueness. This is an example of the idea that banking – interpreted here as an arrangement with an explicit banker, as opposed to mutual insurance – might contribute to fragility or instability. We take this idea further in Section 6.

6 Dynamics

We now consider deterministic and stochastic dynamics – cycles and sunspots – where outcomes vary over time as self-fulfilling prophecies.¹⁵ We focus on the model with banking, since self insurance and mutual insurance are really special cases with $\eta = 0$ and $\eta = 1$, given $\theta = 1$. Defining $L(z) \equiv \lambda \circ v^{-1}(z)$, we can write the Euler equation in terms of real balances as

$$z_t = F(z_{t+1}) \equiv \beta z_{t+1} [L(z_{t+1}) (\alpha + \eta - \alpha\eta) + 1] + \rho M [1 + \eta/\alpha - \eta].$$

This is shown in the right panel of Fig. 6 for $\rho > 0$. As we already know, there is always a SME at $\phi = 0$ if $\rho = 0$, and never if $\rho > 0$, and in either case there is a unique SME with $\phi > 0$. In SME, $\bar{z}_3 = F(\bar{z}_3)$, which could be above or below z^* , where z^* is the value of z that gets q^* .

As is standard (see the Azariadis 1993 textbook), there is a two-cycle equilibrium when $F'(\bar{z}_3) < -1$, where $\bar{z}_3$ is the SME. In this equilibrium, z oscillates between $z_A < \bar{z}_3$ and $z_B > \bar{z}_3$. A two-state sunspot equilibrium is a stochastic version of a two-cycle with z fluctuating between $z_A, z_B \in \mathcal{N}(\bar{z}_3)$, where $\mathcal{N}(\bar{z}_3)$ is a neighborhood of $\bar{z}_3$, with $z_A < \bar{z}_3 < z_B$ and switching probabilities (ϵ_A, ϵ_B) .

For sunspot equilibrium, it is standard and useful to reduce the stochastic system

¹⁵Papers studying the dynamics of debt limits over time using similar methods include Gu et al. (2013b) and Bethune et al. (2018a,b). They show debt limits can vary as self-fulfilling prophecies over time. Carapella and Williamson (2015) do something similar but different: they show how debt limits can vary as self-fulfilling prophecies in the cross section, rather than the time series.

to

$$\begin{aligned} z_A &= \epsilon_A F(z_B) + (1 - \epsilon_A) F(z_A) \\ z_B &= \epsilon_B F(z_A) + (1 - \epsilon_B) F(z_B). \end{aligned}$$

This can be solved for

$$\epsilon_A = \frac{z_A - F(z_A)}{F(z_B) - F(z_A)} \text{ and } \epsilon_B = \frac{F(z_B) - z_B}{F(z_B) - F(z_A)}.$$

If $\epsilon_j \in (0, 1)$ for $z_A, z_B \in \mathcal{N}(\bar{z}_3)$, we have a sunspot equilibrium. It is easy to check that $\epsilon_j \in (0, 1)$ holds if $F'(\bar{z}_3) < -1$, the same as the condition for a two-cycle.

Proposition 8. *(Banking regulation and dynamics) Suppose we vary η , and adjust ρM so that the SME stays the same $\bar{q}$. Then $|F'(\bar{z})|$ is strictly increasing in η . Consequently, if two-cycles and sunspot equilibria near $\bar{z}$ exist for η_1 , they also exist for $\eta_2 > \eta_1$, but not vice versa.*

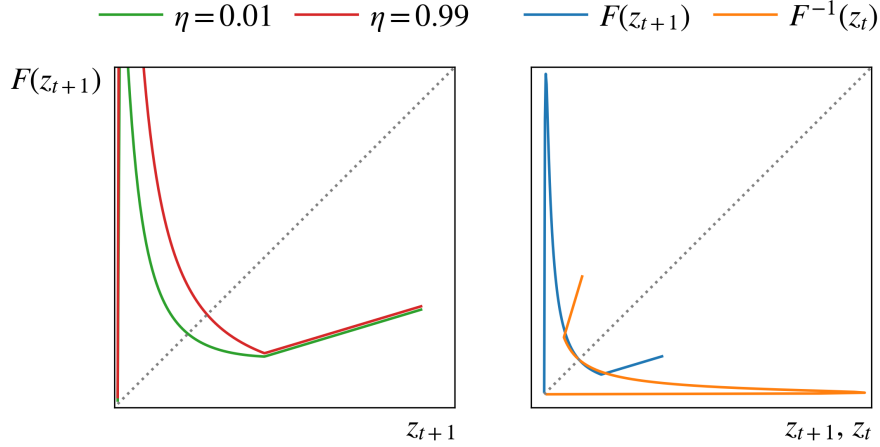


Fig. 6: Dynamic system $F(z)$ for different η (left); F and F^{-1} (right).

Reducing reserve requirements increases instability in the sense that $\partial F'(z)/\partial \eta < 0$, making it more likely to get $F'(\bar{z}_3) < -1$, and hence to get cycles and sunspots, when

η is higher. This is similar to the conclusion in Lee (2025) and in Gu et al. (2023) in related models. This is shown in the left panel of Fig. 6. To be clear, cyclic and sunspot equilibria can also arise under mutual and self insurance, but are more likely with banking when η is high.

Intuitively, complicated dynamics occur in many models with liquidity considerations because liquidity is to some extent self-referential. A lower η shifts $F(z)$ up, which increases $\bar{z}_3$ and that is good because it gets us closer to z^* and q^* , but also makes it more likely that $F'(\bar{z}_3) < -1$, which is bad if one is concerned with the possibility of endogenous volatility.¹⁶ The general idea is that some policies – here, higher η – can improve the liquidity of M , which makes it more likely that liquidity considerations engender volatility.

There is another way to construct a sunspot equilibrium, different from the method used in many related papers.¹⁷ Suppose agents believe that equilibrium prices ϕ will be correlated with an extrinsic i.i.d. variable ε following a distribution ψ with sufficiently small variance and $\mathbb{E}[\varepsilon] = 0$. The realization of ε occurs at the beginning of the DM and is not related to the fundamentals. Agents solve an expected utility maximization problem given that $\tau = (1 - \alpha)\eta\pi/\alpha$, $\pi = \hat{m}$, $y = (1/\theta - 1)S^c - (\phi + \rho)[(1 - \alpha)\pi - \alpha\tau]$, and ϕ follows a stochastic process:

$$\max_{\hat{m}_t} \mathbb{E}_{\varepsilon_t} \left\{ -\phi_{t-1}\hat{m}_t + \beta \left\{ \alpha G(\varepsilon_t) + (1 - \alpha)[\phi(\varepsilon_t) + \rho](\hat{m}_t - \pi_t) - y_t \right\} \right\} \quad (25)$$

with $G(\varepsilon_t) \equiv u \circ q(\varepsilon_t) + [\phi(\varepsilon_t) + \rho](\hat{m}_t + \tau_t - p_t)$. The FOC implies

$$-\phi_{t-1} + \beta \mathbb{E}_{\varepsilon_t} \left\{ [\phi(\varepsilon_t) + \rho][1 + (\alpha + \eta - \alpha\eta)\lambda \circ q(\varepsilon_t)] \right\} = 0, \quad (26)$$

where $q(\varepsilon) = v^{-1}\{[\phi(\varepsilon) + \rho](\hat{m} + \tau)\}$ and $p = \hat{m} + \tau$ if $[\phi(\varepsilon) + \rho](\hat{m} + \tau) < v(q^*)$; and $q(\varepsilon) = q^*$ and $p = v(q^*)/[\phi(\varepsilon) + \rho]$ otherwise. This says the Euler equation holds in

¹⁶In terms of welfare, it is possible that a cycle dominates steady state if we start at $z_B > \bar{z}$ and β is low. This is not new – see, e.g., Gu et al. (2013).

¹⁷This method goes back to Woodford (1986), with an early application by Spear et al. (1990). We learned a lot from the lecture notes in Spear (2026).

expectation over the sunspot distribution, but does not pin down the realization-by-realization pricing rule.

To construct such a rule, we start from the deterministic Euler equation,

$$-\phi_{t-1} + \beta(\phi_t + \rho)[1 + (\alpha + \eta - \alpha\eta)\lambda(q_t)] = 0, \quad (27)$$

and perturb it by setting the residual equal to the sunspot shock rather than zero:

$$-\phi_{t-1} + \beta(\phi_t + \rho)[1 + (\alpha + \eta - \alpha\eta)\lambda(q_t)] = \varepsilon_t. \quad (28)$$

Since $\mathbb{E}[\varepsilon] = 0$, taking expectations over (28) recovers (26), confirming that agents who face the stochastic pricing rule implied by (28) satisfy their first-order condition.

Again, (28) can be rewritten as $z_t = F(z_{t+1}, \varepsilon_{t+1})$. Let us focus on the case where F is monotonically decreasing in a neighborhood of steady state $\bar{z}$, with F^{-1} well-defined for a sufficiently small variance of the distribution ψ . By construction, the unperturbed Euler equation is recovered by taking the expectation over the sunspot distribution. This ensures that agents who optimize under the stochastic pricing rule satisfy their Euler equation in expectation, confirming that the construction yields a bona fide rational expectations equilibrium. The condition $F'(\bar{z}) < -1$ guarantees that the stochastic system admits a stationary invariant distribution supported on a neighborhood of $\bar{z}$. Thus, there is an invariant Markovian equilibrium in which the expectation that prices are random is self-fulfilling (see Spear 2026 for details).

Fig. 7 illustrates the resulting sunspot equilibrium using the perturbed system (28), where for simplicity ε is e or $-e$ with equal probability. The left panel shows persistent fluctuations in z_t around the steady state, driven by the extrinsic sunspot. The oscillatory pattern reflects the negative slope of F : a high value of z_t maps to a low z_{t+1} and vice versa, generating high-frequency reversals. The amplitude of the fluctuations exceeds the size of the underlying shock, as the condition $|F'(\bar{z})| > 1$ amplifies small expectational perturbations through the pricing equation.

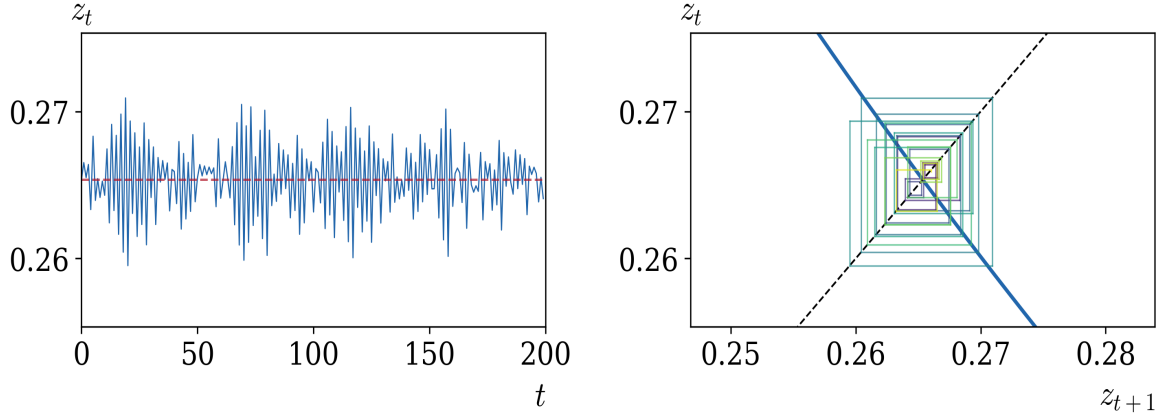


Fig. 7: Sunspot equilibrium, time series (left) and phaseplane (right).

Furthermore, the right panel visualizes this mechanism: the $F(z)$ curve crosses the 45-degree line with a slope steeper than -1 , and the stochastic trajectory spirals around $\bar{z}$ in a rectangular pattern characteristic of alternating overshooting. These results demonstrate that commodity money prices can exhibit excess volatility arising purely from self-fulfilling expectations, not from fundamental shocks to trading opportunities or dividends, highlighting an intrinsic source of price-level instability under commodity money regimes.

7 Conclusion

This paper studied different insurance arrangements in monetary economies. To self insure, agents can hold liquid assets, as in standard monetary models, like Lagos and Wright (2005). That can be dominated by mutual insurance in a coalition, similar to Diamond and Dybvig (1983), although our arrangement deals with money or money-like assets rather than goods. While others have put banking into similar monetary models, including Berentsen et al. (2007), a key difference here is the way we introduce a lack of commitment for the bank or banker. In particular, commitment frictions might impede the viability of mutual insurance, and that can give rise to a role for agents

outside the coalition that act as bankers. This mitigates some incentive problems, but introduces others when the banker has limited commitment, as in Gu et al. (2013a), but that paper has neither an insurance motive nor money.

So many of the components of our analysis appear elsewhere, but it is the combination of these features that gives rise to the most interesting outcomes. Moreover, we study endogenous dynamics (cyclic and sunspot equilibria), as many others do in different contexts. Also, as in Lee (2025), we ask how details of the banking arrangement, including a reserve regulation, contribute to endogenous volatility. Furthermore, our construction of sunspot equilibria is different from many other papers. While these methods go back to Woodford (1986), as far as we know, they have not previously been applied in this kind of model of money or banking.

Many extensions can be suggested, in addition to those discussed above. Some might involve generalizations or alternative formulations of the theory. It is known from various applications of the monetary framework that one can incorporate various specifications for search – e.g., directed rather than random matching – and price formation – e.g., posting rather than bargaining – and for some issues it can make a difference. Other ideas involve policy applications. One such extension could involve government deposit insurance, perhaps interpreted as the public sector insuring agents against the possibility of private liquidity insurance going wrong, possibly due to the malfeasance of bankers, not (just) panic among depositors. There is plenty of room for future work.

References

- A. Agarwal and M. Rempel (2026) Stockpiling liquidity to acquire innovation. Working Paper.
- C. Aliprantis, G. Camera and D. Puzzello (2007a) Anonymous markets and monetary trading. *JME* 54, 1905-28.
- C. Aliprantis, G. Camera and D. Puzzello (2007b) Contagion Equilibria in a Monetary Model *Econometrica* 75, 277-82.
- F. Alvarez and U. Jermann (2000) Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica* 68, 775-97.
- D. Andolfatto (2013) Incentive-feasible deflation. *JME* 60, 383-90.
- D. Andolfatto, A. Berentsen and F. Martin (2020) Money, banking, and financial markets. *RES* 87, 2049-86.
- L. Araujo and R. Minetti (2011) On the essentiality of banks. *IER* 52, 679-91.
- S. Aruoba, G. Rocheteau and C. Waller (2007) Bargaining and the value of money. *JME* 54, 2636-55.
- C. Azariadis (1993) *Intertemporal Macroeconomics*. Blackwell.
- V. Bencivenga and B. Smith (1991) Financial intermediation and endogenous growth. *RES* 58, 195-209.
- A. Berentsen, G. Camera and C. Waller (2007) Money, credit and banking. *JET* 135, 171-95.
- Z. Bethune, T. Hu and G. Rocheteau (2018a) Indeterminacy in credit economies. *JET* 175, 556-84.
- Z. Bethune, T. Hu and G. Rocheteau (2018b) Optimal credit cycles. *RED* 27, 231-45.
- Z. Bethune, G. Rocheteau, T. Wong and C. Zhang (2022) Lending relationships and optimal monetary policy. *RES* 89, 1833-72.
- J. Bhattacharya, J. Haslag and A. Martin (2005) Heterogeneity, redistribution and the Friedman rule. *IER* 46, 437-54.
- B. Biais, T. Mariotti, G. Plantin and J. Rochet (2007) Dynamic security design: Convergence to continuous time and asset pricing implications. *RES* 74, 345-90.
- F. Carapella and S. Williamson (2015) Credit markets, limited commitment, and government debt. *RES* 82, 963-90.

- B. Champ, B. Smith and S. Williamson (1996) Currency elasticity and banking panics: Theory and evidence. *Can. J. Econ* 29, 828-64.
- V. Chari and C. Phelan (2014) On the social usefulness of fractional reserve banking. *JME* 65, 1-13.
- D. Corbae, T. Temzelides and R. Wright (2003) Directed Matching and Monetary Exchange. *Econometrica* 71, 731-56.
- P. DeMarzo and M. Fishman (2006) Agency and optimal investment dynamics. *RFS* 20, 151-88.
- D. Diamond (1984) Financial intermediation and delegated monitoring. *RES* 51, 393-414.
- D. Diamond and P. Dybvig (1983) Bank runs, deposit insurance, and liquidity. *JPE* 91, 401-19.
- D. Duffie, N. Garleanu and L. Pedersen (2005) Over-the-counter markets. *Econometrica* 73, 1815-47.
- M. Engineer and S. Shi (1998) Asymmetry, imperfectly transferable utility, and the role of fiat money in improving terms of trade. *JME* 41, 153-83.
- M. Engineer and S. Shi (2001) Bargains, barter, and money. *RED* 4, 188-209.
- X. Freixas and J. Rochet (2023) *Microeconomics of Banking*, 3rd ed. MIT Press.
- M. Friedman (1959) *A Program for Monetary Stability*, Fordham University Press.
- C. Gu, F. Mattesini, C. Monnet and R. Wright (2013a) Banking: A new monetarist approach. *RES* 80, 636-62.
- C. Gu, F. Mattesini, C. Monnet and R. Wright (2013b) Endogenous credit cycles. *JPE* 121, 940-65.
- C. Gu, F. Mattesini and R. Wright (2016) Money and credit redux. *Econometrica* 84, 1-32.
- C. Gu, C. Monnet, E. Nosal and R. Wright (2023) Diamond-Dybvig and beyond: On the instability of banking. *EER* 154.
- C. Gu and R. Wright (2016) Monetary mechanisms. *JET* 163, 644-57.
- J. Gurley and E. Shaw (1960) *Money in a Theory of Finance*. Brookings.
- P. He, L. Huang and R. Wright (2008) Money, banking, and monetary policy. *JME* 55, 1013-24.

- J. Hicks (1962) Liquidity. *Economic Journal* 72, 787-802.
- T. Hu, J. Kennan and N. Wallace (2009) Coalition-proof trade and the Friedman rule in the Lagos-Wright model. *JPE* 117, 116-37.
- A. Huang (2019) On the number and size of banks: Efficiency and equilibrium. *mimeo*.
- E. Kalai (1977) Proportional solutions to bargaining situations: Interpersonal utility comparisons. *Econometrica* 45, 1623-30.
- T. Kehoe and D. Levine (1993) Debt-constrained asset markets. *RES* 60, 865-88.
- J. M. Keynes (1936) *The general theory of employment, interest, and money*. Macmillan.
- N. Kiyotaki and J. Moore (1997) Credit cycles. *JPE* 105, 211-48.
- N. Kocherlakota (1998) Money is memory. *JET* 81, 232-51.
- R. Lagos (2011) Asset prices, liquidity, and monetary policy in an exchange economy. *JMCB* 43, 521-52.
- R. Lagos, G. Rocheteau and R. Wright (2017) Liquidity: A new monetarist perspective. *JEL* 55, 371-440.
- R. Lagos and R. Wright (2005) A unified framework for monetary theory and policy analysis. *JPE* 113, 463-84.
- H. Lee (2025) On the instability of fractional reserve banking. *EER* 178.
- S. Lippman and J. McCall (1986) An operational measure of liquidity. *AER* 76, 43-55.
- Lucas, R. (1978) Asset prices in an exchange economy. *Econometrica* 46, 1429-45.
- F. Madison (2024) Asymmetric information in frictional markets for liquidity: Collateralized credit vs asset sale. *JEDC* 159.
- T. Matsuoka and M. Watanabe (2019) Banking crises and liquidity in a monetary economy. *JEDC* 108.
- J. Peck and A. Setayesh (2021) Bank runs and the optimality of limited banking. *mimeo*.
- G. Rocheteau and E. Nosal (2017) *Money, Payments, and Liquidity*. MIT Press.
- G. Rocheteau and R. Wright (2005) Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium. *Econometrica* 73, 175-202.
- D. Sanches (2016) On the welfare properties of fractional reserve banking. *IER* 57, 935-54.

- R. Silveira and R. Wright (2010) Search and the market for ideas. *JET* 145, 1550-73.
- B. Smith (2002) Monetary policy, banking crises, and the Friedman rule. *AER* 92, 128-34.
- S. Spear (2026) Overlapping generations models. *mimeo*.
- S. Spear, S. Srivastava and M. Woodford (1990) Indeterminacy of stationary equilibrium in stochastic overlapping generations models. *JET* 50, 265-84.
- J. Tobin (1963) Commercial banks as creators of ‘money.’ *Cowles Discussion Paper* 388.
- N. Wallace (1980) The overlapping generations model of fiat money. *Models of Monetary Economies*, ed. J. Kareken and N. Wallace. FRB Minneapolis.
- N. Wallace (2010) The mechanism-design approach to monetary theory. *Handbook of Monetary Economics*, vol. 3, ed. B. Friedman and M. Woodford. North-Holland.
- S. Williamson (2012) Liquidity, monetary policy, and the financial crisis: A new monetarist approach. *AER* 102, 2570-605
- M. Woodford (1986) Stationary sunspot equilibrium: The case of small shocks around a deterministic steady state. Online at <https://blogs.cuit.columbia.edu/mw2230/old-papers/>.
- T. Wong (2016) A tractable monetary model under general preferences. *RES* 83, 402-20.

Appendix A: Proofs

Proposition 3: We proceed in three steps: first characterize the allocations in each regime and establish existence and uniqueness of the equilibrium within it; then derive the threshold reputation values; and lastly verify their ranking.

Step 1: From the second FOC $\xi_2 = [\lambda(q) + 1](\phi + \rho) > 0$, so $\alpha\tau = (1 - \alpha)\pi$ holds in all regimes. Substituting ξ_2 into the first FOC gives

$$\xi_1 = (1 - \alpha)(\phi + \rho)\lambda(q) - \xi_3(\phi + \rho). \quad (29)$$

Case Γ_a ($\pi < \hat{m}$, IC binds; $\xi_1 = 0$, $\xi_3 > 0$). Setting $\xi_1 = 0$ in (29) yields $\xi_3 = (1 - \alpha)\lambda(q_a)$. Substituting into the third FOC with $\hat{m} = M$ gives the Euler equation

$$\frac{\phi_a}{\beta} = (\phi_a + \rho)[1 + (\alpha + (1 - \alpha)\mu_2\delta_2)\lambda(q_a)], \quad (30)$$

where $q_a = v^{-1}((\phi_a + \rho)[M + (1 - \alpha)\pi_a/\alpha])$ and π_a satisfies IC at equality,

$$(\phi_a + \rho)\pi_a = \Lambda + \mu_2\delta_2(\phi_a + \rho)M. \quad (31)$$

Since (31) is linear in $(\phi_a + \rho)\pi_a$, substituting it into $v(q_a) = (\phi_a + \rho)[M + (1 - \alpha)\pi_a/\alpha]$ eliminates π_a , $v(q_a) = z(\phi_a)$, where

$$z(\phi) \equiv (\phi + \rho)M \frac{\alpha + (1 - \alpha)\mu_2\delta_2}{\alpha} + \frac{(1 - \alpha)\Lambda}{\alpha}$$

is affine and strictly increasing, so (30) reduces to one equation in one unknown,

$$\frac{\phi}{\phi + \rho} = \beta \left[1 + (\alpha + (1 - \alpha)\mu_2\delta_2)\lambda(v^{-1}(z(\phi))) \right]. \quad (32)$$

If $\rho > 0$, the LHS of (32) is continuous and strictly increasing in ϕ from 0 to 1, while the RHS is continuous and non-increasing, since $z' > 0$ and λ is strictly decreasing on $(0, q^*)$ and zero thereafter, with $\text{RHS} \geq \beta > 0 = \text{LHS}$ at $\phi = 0$ and $\text{RHS} \rightarrow \beta < 1 = \lim \text{LHS}$ as $\phi \rightarrow \infty$. Hence ϕ_a exists and is unique, and q_a , π_a and ξ_3 follow uniquely. If $\rho = 0$, the LHS is identically 1, so (32) pins down $\lambda(q_a) = r/[\alpha + (1 - \alpha)\mu_2\delta_2]$; since λ is strictly decreasing where positive, q_a is unique, and it exists iff $[\alpha + (1 - \alpha)\mu_2\delta_2]\lambda(0) > r$, the analogue of the existence condition $\sigma > \bar{\sigma}_1$ in Proposition 1, which defines $\bar{\sigma}_2$, with $\bar{\sigma}_2 \leq \bar{\sigma}_1$ since $\alpha + (1 - \alpha)\mu_2\delta_2 \geq \alpha$; then ϕ_a is unique from the affine z .

Since $\alpha + (1 - \alpha)\mu_2\delta_2 < 1$, comparing (30) with the Euler equation under commitment shows that consumers share a lower marginal benefit for a given q relative to Γ_2 . Note also that $q_a < q_2$. From $\xi_3 = (1 - \alpha)\lambda(q_a) > 0$, which requires $\lambda(q_a) > 0$, we know $q_a < q^*$. Thus, if $q_2 = q^*$ this result is immediate. The proof of $q_a < q_2$ for $q_2 < q^*$ is by contradiction. Suppose $q_a \geq q_2$. In steady state, the Euler equations can be written

$$\frac{\phi_2}{\phi_2 + \rho} = \beta[1 + \lambda(q_2)], \quad \frac{\phi_a}{\phi_a + \rho} = \beta[1 + (\alpha + (1 - \alpha)\mu_2\delta_2)\lambda(q_a)].$$

Since $q_a < q^*$, $\lambda(q_a) > 0$, and $\alpha + (1 - \alpha)\mu_2\delta_2 < 1$, so $(\alpha + (1 - \alpha)\mu_2\delta_2)\lambda(q_a) < \lambda(q_a)$. Moreover, $q_a \geq q_2$ implies $\lambda(q_a) \leq \lambda(q_2)$, as λ is strictly decreasing on $(0, q^*)$. Chaining these two inequalities: $(\alpha + (1 - \alpha)\mu_2\delta_2)\lambda(q_a) < \lambda(q_2)$. Since $\phi/(\phi + \rho)$ is strictly increasing in ϕ , $\phi_a < \phi_2$.

Now compare the liquidity available to DM buyers. Under full commitment, $\pi = M$ and $\tau = (1 - \alpha)M/\alpha$, so $v(q_2) = (\phi_2 + \rho)M/\alpha$. In Case Γ_a , $\pi_a < M$, so $v(q_a) = (\phi_a + \rho)(M + (1 - \alpha)\pi_a/\alpha)$. Hence

$$\phi_a + \rho < \phi_2 + \rho \text{ and } M + \frac{(1 - \alpha)\pi_a}{\alpha} < \frac{M}{\alpha},$$

where the second inequality follows from $\pi_a < M$. Therefore $v(q_a) < v(q_2)$, which gives $q_a < q_2$, a contradiction.

Case Γ_b ($\pi = \hat{m} = M$, IC binds; $\xi_1 > 0$, $\xi_3 > 0$). IC pins down the equilibrium price,

$$(\phi_b + \rho)(1 - \mu_2\delta_2)M = \Lambda, \tag{33}$$

one linear equation in one unknown with solution $\phi_b = \Lambda/[(1 - \mu_2\delta_2)M] - \rho$, so existence and uniqueness are immediate, and ϕ_b is strictly increasing in Λ . With $\pi = M$ and $\tau = (1 - \alpha)M/\alpha$, $q_b = v^{-1}[(\phi_b + \rho)M/\alpha]$. Substituting (29) into the third FOC yields

$$\frac{\phi_b}{\beta} = (\phi_b + \rho)[1 + \lambda(q_b)] - \xi_3(\phi_b + \rho)(1 - \mu_2\delta_2), \tag{34}$$

where the subtracted quantity in the last term is positive, so consumers demand lower real balances compared to full commitment; ξ_3 from (34) and ξ_1 from (29) follow uniquely. Since $\Lambda < \bar{\Lambda}$ in this regime, (33) immediately gives $\phi_b < \phi_2$.

Case Γ_2 ($\xi_3 = 0$). IC is slack and the Euler equation reduces to $\phi_2/\beta = (\phi_2 + \rho)[1 + \lambda(q_2)]$, the commitment outcome; this is (32) with the coefficient $\alpha + (1 - \alpha)\mu_2\delta_2$

replaced by 1 and $z(\phi) = (\phi + \rho)M/\alpha$, so the same argument gives existence and uniqueness.

Step 2: Within Case Γ_a , Λ enters (32) only through z , with $\partial z/\partial \Lambda = (1 - \alpha)/\alpha > 0$, so ϕ_a is continuous and strictly decreasing in Λ . Hence, from (31), $\pi_a = \Lambda/(\phi_a + \rho) + \mu_2 \delta_2 M$ is continuous and strictly increasing in Λ , from $\mu_2 \delta_2 M < M$ at $\Lambda = 0$ to ∞ . Thus π_a reaches M at a unique $\underline{\Lambda} > 0$; setting $\pi_a = M$ in (31),

$$\underline{\Lambda} = (\underline{\phi} + \rho)(1 - \mu_2 \delta_2)M,$$

where $\underline{\phi}$ solves (30) with $q = v^{-1}((\underline{\phi} + \rho)M/\alpha)$. The requirement $\pi_a < M$ of Case Γ_a therefore holds iff $\Lambda < \underline{\Lambda}$.

Let $\bar{\Lambda}$ be the value at which IC becomes slack in Case Γ_b . Since ϕ_b and q_b are strictly increasing in Λ by (33), (34) implies $\xi_3(1 - \mu_2 \delta_2) = 1 + \lambda(q_b) - \phi_b/[\beta(\phi_b + \rho)]$ is strictly decreasing in Λ , reaching 0 exactly when (34) coincides with the commitment Euler equation, i.e., when $\phi_b = \phi_2$. Substituting into (33),

$$\bar{\Lambda} = (\phi_2 + \rho)(1 - \mu_2 \delta_2)M,$$

so $\xi_3 > 0$ iff $\Lambda < \bar{\Lambda}$, and IC is slack (Case Γ_2) iff $\Lambda \geq \bar{\Lambda}$. As for $\xi_1 = (\phi_b + \rho)[(1 - \alpha)\lambda(q_b) - \xi_3]$: substituting ξ_3 from (34), $\xi_1 > 0$ iff

$$\frac{\phi_b}{\beta(\phi_b + \rho)} > 1 + [\alpha + (1 - \alpha)\mu_2 \delta_2]\lambda(q_b),$$

where the RHS is (30) evaluated at $\pi = M$, whose unique solution is $\underline{\phi}$. Since the difference between the two sides is strictly increasing in ϕ , the inequality holds iff $\phi_b > \underline{\phi}$, i.e., by (33), iff $\Lambda > \underline{\Lambda}$. Hence $\xi_1 > 0$ on $(\underline{\Lambda}, \bar{\Lambda})$, $\xi_1 = 0$ at $\Lambda = \underline{\Lambda}$, and $\xi_1 < 0$ for $\Lambda < \underline{\Lambda}$, ruling out Case Γ_b there.

Step 3: Thresholds take the form $(\phi + \rho)(1 - \mu_2 \delta_2)M$ evaluated at $\underline{\phi}$ and ϕ_2 . Since $\alpha + (1 - \alpha)\mu_2 \delta_2 < 1$, the coefficient on $\lambda(q)$ in (30) is strictly smaller than in the commitment Euler equation, so the monotonicity argument in Step 1 implies $\underline{\phi} < \phi_2$, and therefore $\underline{\Lambda} < \bar{\Lambda}$. By Steps 1 and 2, the conditions defining Case Γ_a , Case Γ_b and Case Γ_2 hold precisely on $[0, \underline{\Lambda})$, $[\underline{\Lambda}, \bar{\Lambda})$ and $[\bar{\Lambda}, \infty)$, which partition $[0, \infty)$. Hence for each Λ exactly one contract type is consistent with equilibrium, and within it the allocation is unique by Step 1. SME exists and is unique for every $\Lambda \geq 0$ when $\rho > 0$; when $\rho = 0$, it exists iff $\sigma > \bar{\sigma}_2$ and is unique. ■

Lemma 3: Within each regime it is obvious that (ϕ, q, π) is continuous in Λ . It remains to verify continuity across regime boundaries.

At $\Lambda = \underline{\Lambda}$: in Case Γ_a , $\pi_a \rightarrow M$ and binding IC gives $\underline{\Lambda} = (\underline{\phi} + \rho)(1 - \mu_2\delta_2)M$. In Case Γ_b at the same $\underline{\Lambda}$, IC pins down $(\phi_b + \rho)(1 - \mu_2\delta_2)M = \underline{\Lambda}$ with $\phi_b = \underline{\phi}$. Since both cases have $\pi = M$ and the same ϕ , they yield the same $\underline{q} = v^{-1}((\underline{\phi} + \rho)M/\alpha)$ at the cutoff $\underline{\Lambda}$.

At $\Lambda = \bar{\Lambda}$: in Case Γ_b , IC $(\phi_b + \rho)(1 - \mu_2\delta_2)M = \bar{\Lambda} = (\phi_2 + \rho)(1 - \mu_2\delta_2)M$, so $\phi_b \rightarrow \phi_2$ and $q_b \rightarrow q_2$, matching the commitment case.

Since allocations are continuous in Λ , consumer surplus and hence $P(\Lambda)$ are continuous. ■

Lemma 4: The FOCs (21)-(24) imply this: If $\xi_5 > 0$ then $\xi_2 > 0$ and $\xi_3 = 0$; otherwise $\xi_1 > 0$, $\xi_2 > 0$, $\xi_3 > 0$ and $\xi_4 \geq 0$. So if IC binds $\xi_5 > 0$, and we must give bankers more than the share $1 - \theta$ of total surplus. And if IC is slack, $\xi_5 = 0$, then bargaining characterizes Γ_3 . In any case, $\xi_2 > 0$. As a result, since $\pi \leq \hat{m}$ and $\Pi \geq 0$ may or may not bind, there are four possible contracts labeled $\Gamma_i = \Gamma_A, \Gamma_B, \Gamma_C, \Gamma_D$ with Γ_A satisfying $\pi < \hat{m}$ and $\Pi > 0$, Γ_B satisfying $\pi < \hat{m}$ and $\Pi = 0$, Γ_C satisfying $\pi = \hat{m}$ and $\Pi = 0$, and Γ_D satisfying $\pi = \hat{m}$ and $\Pi > 0$.

Next we characterize each contract and the implied q . These depend on Φ and μ_3 . Define $\kappa \equiv \eta(1 - \mu_3) + \mu_3(1 - \delta_3)$. Since IC binds $\xi_5 > 0$, the Kalai bargaining constraint is slack, $\xi_3 = 0$, and θ drops out of the FOCs. From the FOC wrt y with $\xi_3 = 0$, $\xi_4 + \mu_3\xi_5 = 1$. The FOC wrt $\hat{m}$ with $\xi_3 = 0$ reduces to

$$0 = -\phi/\beta + (\phi + \rho)[1 + \alpha\lambda(q)] + \xi_1. \quad (35)$$

Since the reserve requirement always binds, $\xi_2 > 0$, substituting $\tau = \eta(1 - \alpha)\pi/\alpha$ and $\Pi = y + (\phi + \rho)(1 - \alpha)(1 - \eta)\pi$ into IC yields

$$(1 - \alpha)(\phi + \rho)\pi\kappa = \Phi + \mu_3\Pi. \quad (36)$$

Γ_A : ($\pi < M$, $\Pi > 0$; $\xi_1 = \xi_4 = 0$). FOCs yield $\lambda(q_A) = (1 - \delta_3)/\eta + 1/\mu_3 - 1$, and (35) with $\xi_1 = 0$ gives $\phi_A + \rho = \rho/[1 - \beta(1 + \alpha\lambda(q_A))]$, with π_A from $v(q_A) = (\phi_A + \rho)(M + \eta(1 - \alpha)\pi_A/\alpha)$. Now q_A, ϕ_A, π_A are independent of Φ . If $Q = q^*$, then $q_A < Q$. If $Q < q^*$, suppose by way of contradiction that $q \geq Q$.

Since λ is strictly decreasing on $(0, q^*)$, $\lambda(q) \leq \lambda(Q)$. Because $\alpha < \alpha + (1 - \alpha)\eta$, we know $\alpha \lambda(q) < [\alpha + (1 - \alpha)\eta] \lambda(Q)$. Comparing Euler equations:

$$\frac{\phi}{\phi + \rho} = \beta[1 + \alpha \lambda(q)] < \beta[1 + (\alpha + (1 - \alpha)\eta) \lambda(Q)] = \frac{\phi_{3,\theta=1}}{\phi_{3,\theta=1} + \rho}.$$

Since $\phi/(\phi + \rho)$ is strictly increasing in ϕ , $\phi < \phi_{3,\theta=1}$. Since $\pi < M$,

$$M + \frac{\eta(1 - \alpha) \pi}{\alpha} < M + \frac{\eta(1 - \alpha) M}{\alpha}.$$

Using these inequalities,

$$v(q) = (\phi + \rho) \left(M + \frac{\eta(1 - \alpha) \pi}{\alpha} \right) < (\phi_{3,\theta=1} + \rho) \frac{M[\alpha + (1 - \alpha)\eta]}{\alpha} = v(Q).$$

Since v is strictly increasing, $q < Q$, contradicting $q \geq Q$. Hence, $q_A < Q$.

$\Gamma_B : (\pi < M, \Pi = 0; \xi_1 = 0, \xi_4 > 0)$. Equation (35) with $\xi_1 = 0$ provides

$$\phi/\beta = (\phi + \rho)[1 + \alpha\lambda(q)]. \quad (37)$$

From (36) with $\Pi = 0$, $\pi = \Phi/[(1 - \alpha)(\phi + \rho)\kappa]$. We can also show $q_B < Q$ using the same proof by contradiction used for Γ_A .

$\Gamma_C : (\pi = M, \Pi = 0; \xi_1 > 0, \xi_4 > 0)$. From (36), $\phi + \rho = \Phi/[(1 - \alpha)M\kappa]$. Both ϕ and q increase with Φ . This holds until IC becomes slack at $\Phi = \bar{\Phi}$, where

$$\bar{\Phi} \equiv (1 - \alpha)(\phi_{3,\theta=1} + \rho) M \kappa,$$

with $\phi_{3,\theta=1}$ solving the commitment Euler equation with $\theta = 1$. Thus, $\phi_C < \phi_{3,\theta=1}$ and $q_C < Q$.

$\Gamma_D : (\pi = M, \Pi > 0; \xi_1 > 0, \xi_4 = 0)$. FOCs yield

$$\phi/\beta = (\phi + \rho)[1 + (\alpha + (1 - \alpha)\eta)\lambda(q)] - (1 - \alpha)(\phi + \rho)(1 - \eta - \delta_3 + \eta/\mu_3), \quad (38)$$

with $v(q) = (\phi + \rho)M[\alpha + (1 - \alpha)\eta]/\alpha$, determining ϕ_D and q_D independently of Φ . Because the RHS of (38) is smaller than that of the full commitment Euler equation, $\phi_D < \phi_{3,\theta=1}$ and $q_D < Q$. ■

Proposition 6: There are five steps. First, we identify the Φ -cutoffs for each case. Second, the parallel surplus lines. Third and fourth, the thresholds $\bar{\mu}$ and $\underline{\mu}$. Fifth, the three orderings.

Step 1: Φ -cutoffs for each case.

$\Gamma_A : (\pi < M, \Pi > 0)$. From (36), $\Pi_A = [(1 - \alpha)(\phi_A + \rho)\pi_A\kappa - \Phi]/\mu_3$, which is positive iff $\Phi < \underline{\Phi}$, where

$$\underline{\Phi} \equiv (1 - \alpha)(\phi_A + \rho)\pi_A\kappa,$$

is the cutoff Φ such that $\Pi = 0$ under Γ_A . Γ_A exists for $\Phi \in [0, \underline{\Phi})$ when $\pi_A \in (0, M)$.

$\Gamma_B : (\pi < M, \Pi = 0)$. From (36) with $\Pi = 0$ and $v(q) = (\phi + \rho)[M + \eta(1 - \alpha)\pi/\alpha]$, the binding IC can be written $\Phi = \Psi(q)$, where

$$\Psi(q) \equiv (\alpha\kappa/\eta)[v(q) - (\phi(q) + \rho)M],$$

with $\phi(q)$ determined by the self-insurance Euler equation (37). Since λ is strictly decreasing on $(0, q^*)$ and $\phi/(\phi + \rho)$ is strictly increasing in ϕ , (37) implies $\phi'(q) < 0$; combined with $v' > 0$, both terms give $\Psi'(q) > 0$. Hence q , and with it π , increases strictly and continuously with Φ until $\pi = M$ at $\Phi = \hat{\Phi}$, where

$$\hat{\Phi} \equiv (1 - \alpha)(\hat{\phi} + \rho)M\kappa,$$

with $\hat{\phi}$ solving (37) at $\pi = M$. Thus, $\hat{\Phi}$ is the cutoff such that $\pi = M$ under Γ_B . The switching from Γ_A to Γ_B at $\Phi = \underline{\Phi}$ is continuous: they share the same Euler equation, so π , ϕ , q are preserved and only Π drops to zero.

$\Gamma_C : (\pi = M, \Pi = 0)$. This regime holds until IC becomes slack at $\Phi = \bar{\Phi}$, where

$$\bar{\Phi} \equiv (1 - \alpha)(\phi_{3,\theta=1} + \rho)M\kappa.$$

$\Gamma_D : (\pi = M, \Pi > 0)$. From (36), $\Pi_D > 0$ iff $\Phi < \tilde{\Phi}$, where

$$\tilde{\Phi} \equiv (1 - \alpha)(\phi_D + \rho)M\kappa$$

is the cutoff Φ such that $\Pi = 0$ under Γ_D .

Step 2: Parallel surplus lines. With Γ_A and Γ_D , allocations (π, ϕ, q) are independent of Φ ; only Π adjusts, at rate $d\Pi/d\Phi = -1/\mu_3$ from (36). Since allocations are fixed, Π adjusts entirely through the fee, $dy/d\Phi = d\Pi/d\Phi = -1/\mu_3$, and since y enters S^c with a negative sign, $dS_A^c/d\Phi = dS_D^c/d\Phi = 1/\mu_3$. Hence, S_A^c , consumer surplus under Γ_A , is parallel to S_D^c , consumer surplus under Γ_D . In other words, $S_A^c - S_D^c$ is independent of Φ . This implies: (i) surpluses of Γ_A and Γ_D never cross, so their ranking is invariant to Φ ; and (ii) the switch from Γ_A to Γ_D cannot occur through a surplus crossover; it occurs only when Γ_A is about to become infeasible, $\Pi_A = 0$ at $\Phi = \underline{\Phi}$.

Step 3: Cutoff $\bar{\mu}$. By Step 2, the ranking of Γ_A and Γ_D is determined at $\Phi = 0$. Define $\Delta(\mu_3) \equiv S_A^c(\Phi = 0; \mu_3) - S_D^c(\Phi = 0; \mu_3)$, which is continuous in μ_3 . Higher μ_3 relaxes the IC constraint (19) in proportion to the funds intermediated, and thus raises S_D^c relative to S_A^c : Δ is strictly decreasing in μ_3 . Define $\bar{\mu}$ as the unique solution to $\Delta(\bar{\mu}) = 0$, with the convention $\bar{\mu} = 1$ if $\Delta > 0$ throughout. For $\mu_3 < \bar{\mu}$, $\Delta(\mu_3) > 0$, so Γ_A delivers strictly higher consumer surplus than Γ_D at all Φ where both exist. For $\mu_3 \geq \bar{\mu}$, $\Delta(\mu_3) \leq 0$, so Γ_D weakly dominates Γ_A , which never appears in equilibrium.

Step 4: Cutoff $\underline{\mu}$. When $\mu_3 < \bar{\mu}$, Γ_A exists for $\Phi \in [0, \underline{\Phi})$ and ceases to exist at $\underline{\Phi}$ where $\Pi_A \rightarrow 0$. At this point, Γ_B takes over as the continuation contract (the Γ_A -to- Γ_B switching is continuous: $S_B^c(\underline{\Phi}) = S_A^c(\underline{\Phi})$). By Steps 2–3, $S_B^c(\underline{\Phi}) - S_D^c(\underline{\Phi}) = \Delta(\mu_3) > 0$, so Γ_B initially beats Γ_D at $\underline{\Phi}$. However, within the Γ_B region ($\Phi > \underline{\Phi}$), π increases with Φ (Step 1), breaking the parallel-lines property: S_B^c is no longer parallel to S_D^c . Γ_D may therefore overtake Γ_B at some $\Phi^* \in (\underline{\Phi}, \hat{\Phi})$, where Φ^* is defined implicitly by $S_B^c(\Phi^*; \mu_3) = S_D^c(\Phi^*; \mu_3)$.

Define $\underline{\mu}$ as the supremum of μ_3 values for which Γ_B dominates Γ_D in $[\underline{\Phi}, \hat{\Phi}]$:

$$\underline{\mu} \equiv \sup\{\mu_3 < \bar{\mu} : S_B^c(\Phi; \mu_3) \geq S_D^c(\Phi; \mu_3) \ \forall \Phi \in [\underline{\Phi}, \hat{\Phi}]\},$$

with the convention $\underline{\mu} \equiv 0$ if the set is empty; by construction $\underline{\mu} \leq \bar{\mu}$. Since higher μ_3 raises S_D^c relative to S_B^c (as in Step 3, favoring full insurance $\pi = M$), the set is an interval: dominance of Γ_B holds for all $\mu_3 < \underline{\mu}$ and fails for all $\mu_3 \in [\underline{\mu}, \bar{\mu})$.

Step 5: The three orderings. To visualize switching from one contract type to another as Φ increases, we use the notation $\Gamma_i \xrightarrow{\Phi=X} \Gamma_j$ to indicate that the contract switches from Γ_i to Γ_j at the threshold $\Phi = X$. Reading left to right traces out the Γ as Φ increases from 0.

Ordering I ($\mu_3 < \underline{\mu}$): by Step 4, Γ_B dominates Γ_D in $[\underline{\Phi}, \hat{\Phi}]$. After Γ_A ceases to exist, the constraint $\Pi \geq 0$ binds, pinning down Γ_B , and π increases smoothly from π_A to M . As Φ increases, Γ switches as follows:

$$\Gamma_A \xrightarrow{\underline{\Phi}} \Gamma_B \xrightarrow{\hat{\Phi}} \Gamma_C \xrightarrow{\bar{\Phi}} \Gamma_3.$$

The order of cutoffs is as follows: $\underline{\Phi} < \hat{\Phi}$ because π increases strictly within Γ_B ($\Psi'(q) > 0$, Step 1); $\hat{\Phi} < \bar{\Phi}$ because the self-insurance Euler coefficient $\alpha < \alpha + (1 - \alpha)\eta$ implies $\hat{\phi} < \phi_{3,\theta=1}$.

Ordering II ($\underline{\mu} \leq \mu_3 < \bar{\mu}$): Γ_D overtakes Γ_B at Φ^* before π reaches M . After Γ_A ceases to exist at $\underline{\Phi}$, Γ_B takes over initially ($\Pi = 0$, $\pi < M$), but Γ_D provides higher consumer surplus once Φ exceeds Φ^* , at which point $\pi = M$ with positive bank profit resumes:

$$\Gamma_A \xrightarrow{\underline{\Phi}} \Gamma_B \xrightarrow{\Phi^*} \Gamma_D \xrightarrow{\tilde{\Phi}} \Gamma_C \xrightarrow{\bar{\Phi}} \Gamma_3.$$

The order of cutoffs is as follows: $\underline{\Phi} < \Phi^*$ because $S_B^c(\underline{\Phi}) > S_D^c(\underline{\Phi})$ (by the parallel-lines property and $S_A^c > S_D^c$), so Γ_B strictly dominates Γ_D at $\underline{\Phi}$ and the crossover Φ^* must occur strictly later. $\Phi^* < \tilde{\Phi}$ because the crossover can occur only where Γ_D is feasible, $\Pi_D(\Phi^*) \geq 0$, i.e., $\Phi^* \leq \tilde{\Phi}$; and if $\Phi^* = \tilde{\Phi}$, then Γ_D is infeasible for all $\Phi > \Phi^*$, so Γ_D never strictly dominates Γ_B , contradicting $\mu_3 \geq \underline{\mu}$. $\tilde{\Phi} < \bar{\Phi}$ because (38) gives $\phi_D < \phi_{3,\theta=1}$.

Ordering III ($\mu_3 \geq \bar{\mu}$): Γ_D weakly dominates Γ_A at every Φ (by Step 3). Γ_A never appears in equilibrium:

$$\Gamma_D \xrightarrow{\tilde{\Phi}} \Gamma_C \xrightarrow{\bar{\Phi}} \Gamma_3.$$

This completes the proof. ■

Proposition 7: Define the mapping $\Phi' = H(\Phi)$. A stationary equilibrium is a fixed point $\Phi = H(\Phi)$.

Step 1: Boundary values of H .

Orderings I and II. At $\Phi = 0$, the equilibrium contract is Γ_A with $\Pi_A(0) = (1 - \alpha)(\phi_A + \rho) \pi_A \kappa / \mu_3 > 0$. Hence, $H(0) > 0$. At $\Phi = \underline{\Phi}$, Γ_A ceases to exist because $\Pi_A(\underline{\Phi}) = 0$, and Γ_B takes over with $\Pi_B = 0$. Hence, $H(\underline{\Phi}) = 0 < \underline{\Phi}$.

Ordering III. At $\Phi = 0$, the equilibrium contract is Γ_D with $\Pi_D(0) = (1 - \alpha)(\phi_D + \rho) M \kappa / \mu_3 > 0$. Hence, $H(0) > 0$. At $\Phi = \tilde{\Phi}$, Γ_D ceases to exist because $\Pi_D(\tilde{\Phi}) = 0$,

and Γ_C takes over with $\Pi_C = 0$. Hence, $H(\tilde{\Phi}) = 0 < \tilde{\Phi}$.

Step 2: Continuity of H .

Within Γ_A . Bank profit $\Pi_A = [(1 - \alpha)(\phi_A + \rho) \pi_A \kappa - \Phi]/\mu_3$ is continuous and strictly decreasing in Φ , since ϕ_A, π_A are independent of Φ .

At the boundary $\underline{\Phi}$. The switching from Γ_A to Γ_B is continuous: both cases share the self-insurance Euler equation, so π, ϕ , and q are preserved at $\underline{\Phi}$. Since $\Pi_A(\underline{\Phi}) = 0 = \Pi_B(\underline{\Phi})$, the mapping H is continuous at $\underline{\Phi}$.

Within Γ_D . Bank profit $\Pi_D = [(1 - \alpha)(\phi_D + \rho) M \kappa - \Phi]/\mu_3$ is continuous, since ϕ_D is independent of Φ .

At the boundary $\tilde{\Phi}$. The switching from Γ_D to Γ_C is continuous: both have $\pi = M$, and at $\tilde{\Phi}$, Γ_D has $\Pi_D = 0$. The binding IC in Γ_C gives $\phi + \rho = \Phi/[(1 - \alpha)M\kappa]$, which at $\Phi = \tilde{\Phi}$ yields $\phi_C(\tilde{\Phi}) + \rho = \tilde{\Phi}/[(1 - \alpha)M\kappa] = \phi_D + \rho$. Thus, allocations are preserved, and H is continuous at $\tilde{\Phi}$.

Step 3: By the intermediate value theorem, there exists a SME with Γ_A under orderings I and II and Γ_D under ordering III. ■

Proposition 8: F is $F'(z) = \beta[(\alpha + (1 - \alpha)\eta)(L(z_+) + z_+L'(z_+)) + 1]$, where $L(z_+) + z_+L'(z_+)$ could be positive or negative. It is easy to see $\partial|F'(z)|/\partial\eta > 0$ for a given z , so F rises or falls faster with a higher η . Now to make a fair comparison between η_1 and $\eta_2 > \eta_1$, M is adjusted so $\bar{q}$ stays the same: $\bar{q} = v^{-1}(\bar{z}) = v^{-1}[(\phi + \rho)(1 + (1 - \alpha)\eta_1/\alpha)M_1] = v^{-1}[(\phi + \rho)(1 + (1 - \alpha)\eta_2/\alpha)M_2]$, where M_1 and M_2 are money supplies under η_1 and η_2 . If $F'(\bar{z}) < -1$ under η_1 , then under any $\eta_2 > \eta_1$ the slope is even more negative. By continuity of $F'(\bar{z})$ in η there exists a threshold η^* at which $F'(\bar{z}) = -1$. Choosing $\eta_1 < \eta^* < \eta_2$ show cycles occur with η_2 but not η_1 . ■

Appendix B: Mutual Insurance as Banking

Suppose instead of contracting over $\Gamma = (\pi, \tau)$, the coalition functions as a competitive bank taking deposits and making loans. After the DM shock non-demanders deposit $d \leq \hat{m}$ at a nominal net deposit rate i_d with the coalition, yielding $d(1 + i_d)$ in the next CM. In turn, DM demanders borrow ℓ at the net loan rate i_ℓ from the coalition and repay $\ell(1 + i_\ell)$ in the CM. Market clearing means $\alpha\ell = (1 - \alpha)d$. We assume here that repayment of loans and deposits in the CM can be perfectly enforced.

Consumers enter the CM with m , plus loan obligations ℓ and deposits d to solve:

$$W(m, \ell, d) = \max_{x, h, \hat{m}} \{U(x) - h + \beta V(\hat{m})\} \quad (39)$$

$$\text{st } x = h + (\phi + \rho)m - \phi\hat{m} - T + (\phi + \rho)[d(1 + i_d) - \ell(1 + i_\ell)]. \quad (40)$$

As usual the FOC for $\hat{m}$ is $\beta V'(\hat{m}) = \phi$. Also, $W_m = \phi + \rho$, $W_\ell = -(1 + i_\ell)(\phi + \rho)$ and $W_d = (1 + i_d)(\phi + \rho)$. In the DM,

$$V(\hat{m}) = \alpha \max_{\ell} \{u(q) + W(\hat{m} + \ell - p, \ell, 0)\} + (1 - \alpha) \max_{d \leq \hat{m}} \{W(\hat{m} - d, 0, d)\}, \quad (41)$$

with $p \leq \hat{m} + \ell$.

Differentiating wrt ℓ using $\partial q / \partial \ell = (\rho + \phi) / v'(q)$ and $v(q) = (\phi + \rho)p$, we get loan demand

$$1 + \lambda(q) = 1 + i_\ell, \quad (42)$$

with $\lambda(q)$ as in the baseline model. Deposit supply is $d = \hat{m}$ for $i_d > 0$. Combining (39), (40) and (41) using $d = \hat{m}$ and $\partial q / \partial \hat{m} = (\phi + \rho)(1 + \partial \ell / \partial \hat{m}) / v'(q)$, we get

$$\phi_{t-1} = \beta(\phi_t + \rho) \{ \alpha[1 + \lambda(q_t)] + (1 - \alpha)(1 + i_{d,t}) \}. \quad (43)$$

Zero profit implies $\alpha \ell(\phi + \rho)(1 + i_\ell) = (1 - \alpha)d(\phi + \rho)(1 + i_d)$, so $i_d = i_\ell$. Using $m = M$ reduces (43) combined with (42) to (8), the Euler equation in the baseline model. Hence mutual insurance can be made to look more like (a monetary version of) a standard bank. ■

Appendix C: Parameter Values

Fig 1: $M = 1$, $\alpha = 0.2$, $\rho = 0.01$, $\beta = 0.9$, $\delta_2 = 0.1$, $A = 1$, $a = 0.55$ and $b = 0.00001$.

For μ_2 , the panels from top to bottom use 0.23, 0.63 and 0.75.

Fig 2: Same as in Fig 1.

Fig 3: $A = 1$, $b = 0$, $a = 0.5$, $\alpha = 0.5$ and $\beta = 0.9$, with $\rho = 0.15$ for commodity money and $\rho = 0$ for fiat money.

Fig 4: $M = 1$, $\alpha = 0.1$, $\rho = 0.001$, $\beta = 0.9$, $\delta_3 = 0.96$, $\eta = 0.5$, $\theta = 0.8$, $A = 1$, $a = 0.2$ and $b = 0.01$. For μ_3 , the panels from top to bottom use 0.67, 0.7 and 0.85.

Fig 5: Same as Fig 4 but $\theta = 0.9$.

Fig 6: $M = 1$, $\alpha = 0.5$, $\beta = 0.3$, $\rho = 0.01$, $\theta = 1$, $A = 0.15$, $a = 2.4$ and $b = 0.018$.

Fig 7: Same as Fig 6 plus $e = 0.001$ and $\eta = 0.99$.