Journal of Monetary Economics 57 (2010) 420–427

Contents lists available at ScienceDirect

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jme

Could making banks hold only liquid assets induce bank runs? ☆

James Peck [a,*], Karl Shell [b]

[a] Department of Economics, The Ohio State University, 440 Arps Hall, Columbus, OH 43210-1172, USA
[b] Department of Economics, Cornell University, USA

ARTICLE INFO

Article history:
Received 11 February 2009
Received in revised form
7 April 2010
Accepted 7 April 2010
Available online 10 April 2010

Keywords:
Bank runs
Bank stability
Deposit contracts
Glass–Steagall banking
Mechanism design
Portfolio restrictions
Sunspot equilibrium

ABSTRACT

Restrictions placed on bank portfolios are analyzed in a banking model designed to capture the role of checking accounts in facilitating transactions. Forcing banks to hold only liquid assets creates the incentive for liquidity-based runs. Even when a run does not occur, welfare is reduced as a result of overinvestment in the liquid asset.

1. Introduction

After the financial meltdown following the collapse of the building society Northern Rock and the investment bank Bear Stearns, research on bank runs (and, more generally, *financial instability*) has taken on new urgency. New regulatory oversight of banking, with the goal of making the sector less fragile, seems to be inevitable. Paul Volcker, former Fed Chairman and head of President Obama's Economic Recovery Advisory Board, remarked on March 6, 2009: "Maybe we ought to have a kind of two-tier financial system," harkening back to the divisions between commercial and investment banks mandated by the Glass–Steagall Act before it was repealed in 1999.[1] The subject of the present paper is whether legal restrictions imposed on bank portfolios are stabilizing or de-stabilizing.

Our analysis is built on Diamond and Dybvig (1983), but we introduce two innovations to this literature, capturing the role checking accounts and debit cards play in facilitating transactions. We assume that consumption opportunities are urgent and hence if checks or debit card transactions do not clear at par, the transaction is lost. We also envision that there is consumption ahead after the end of period 2; we capture this by positing utility of "left-over" bank balances. In order to allow for decisions about bank portfolio composition, it is assumed that there are two assets: one based on a liquid, lower-return technology and the other based on an illiquid, higher-return technology.[2] We reach the paradoxical finding that

☆ We dedicate this paper to the memory of Bruce Smith. We are grateful to Neil Wallace for criticism of an earlier draft. We thank Huberto Ennis and Todd Keister for stimulating discussions. We thank a referee and editor for help in clarifying and focusing our basic message. We are responsible for errors that persist.

* Corresponding author. Tel.: +1 614 292 0182; fax: +1 614 292 3906.
E-mail address: peck.33@osu.edu (J. Peck).

[1] Volcker's remarks were made at a conference at NYU Stern, and reported by Benjamin and Harper (2009a,b) on Bloomberg.com.

[2] The two-technology model has been studied by Haubrich and King (1990), Wallace (1996), and Cooper and Ross (1998).

0304-3932/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmoneco.2010.04.006

forcing banks to hold only liquid assets can create the incentive for liquidity-based runs: If portfolios are unrestricted, the optimal banking contract is immune to runs. If, however, banks are restricted from holding illiquid assets,[3] they *always* choose deposit contracts that are vulnerable to bank runs.[4] Hence Glass–Steagall-type regulations create instability in the face of panic-based runs.[5]

The unrestricted bank, or "unified system", avoids panic-induced runs, but it does entail some instability in the face of intrinsic shocks: The unrestricted bank depletes its cash when the realized fraction of impatient depositors is sufficiently large.[6] However, there is a trade-off between the risk of running out of cash and the benefit of higher asset returns (i.e., higher growth). The unified system optimally resolves the trade-off between liquidity and economic growth; in doing so, it maximizes social welfare. In the "separated system," restricted banks typically avoid the risk of running out of cash (i.e., non-run rationing of depositors). The explanation is that the restricted bank must perforce overinvest in the liquid asset to ensure that panic-based runs do not always occur. The over-abundance of the liquid asset allows the bank to avoid running out of cash even when the number of impatient depositors is highest. Therefore, avoiding the risk of running out of cash is the silver lining in an overly cautious growth-inhibiting policy, while the risk of bank runs is substituted in its place.

In fairness, it must be said that advocates of the return to Glass–Steagall banking are concerned about the moral hazards that induce bank executives to undertake exceedingly risky investments.[7] Evaluating whether or not portfolio restrictions mitigate moral hazard problems is beyond the scope of the present paper. If there are moral-hazard-reducing benefits from portfolio restrictions, they must be weighed against the cost of making the system vulnerable to panic-based runs and the slowing of economic growth.

In Diamond and Dybvig, banks provide insurance against the event that a consumer becomes impatient and must do her consumption "early" (if patient, she consumes "later"). We follow Diamond and Dybvig in building upon the stochastic nature of some urgent consumption opportunities, but we modify the model to more realistically capture the transactions role played by checking accounts. A depositor facing a consumption opportunity is often someone taking advantage of the convenience of writing a check or using a debit card to make an important purchase. The benefits of a demand deposit account would be severely limited if 100% payment were not made by the bank. The consumption opportunity would be lost or deferred. To capture the depositor's need for payments to clear at par, it is assumed that consumers increase their utility by a discrete amount if at least one unit of consumption is received during the period in which they have a consumption opportunity. Impatient consumers are those who find their best consumption opportunities in the first period, while patient consumers find their best consumption opportunities in the second period. In the model, banks not only provide insurance against impatience but they also facilitate the exercise of consumption opportunities. In another departure from Diamond and Dybvig and as a proxy for more complete intertemporal analysis, we assume that all consumers value "left-over" cash (beyond the demand for funds to finance these consumption opportunities) in the final period.[8] This assumption recognizes the obvious fact that even consumers who want immediate consumption also want to consume in the future. Utility is assumed to be a strictly concave function of the future consumption proxy.

The class of mechanisms is very broad and includes the *feasibility* of partial suspension of convertibility. The consumption good is perfectly divisible, although there is a discrete jump in utility at one unit of consumption. An immediate consequence of this assumption is that the equilibrium contract will not involve partial suspension of convertibility: The bank will always pay at par or not at all. We believe that allowing for consumption opportunities of varying amounts and varying utility jumps (to reflect the different items people want to purchase) is a difficult but worthwhile generalization, but it would not overturn the result about the separated system being vulnerable to bank runs.

Equilibrium bank runs *never* occur in the unified system, where the allocation maximizes social welfare, but in the separated system the optimal contract *always* has a run equilibrium. In the previous literature, bank-run equilibria were demonstrated either for particular parameter values or when contracts that suspend convertibility are prohibited.[9]

The left-over-consumption feature of the model provides the planner with substantial power for eliminating run equilibria in the unified system, because banks can use the assets in the illiquid technology to compensate consumers who do not withdraw in period 1. However, the existence of run equilibria in the separated system does not appear to depend on the left-over-consumption assumption. The crucial assumption is the two-technology framework. In the separated system, *when the liquidity shock is sufficiently close to its maximal value*, a patient consumer receives more from the bank by withdrawing in period 1 than by waiting for whatever liquid assets remain in period 2. Otherwise, reducing investment in

[3] Restricted banks are assumed to be unable to tap indirectly into assets in the illiquid technology. Restricting bank portfolios would not matter if the bank can costlessly enforce all contracts with its depositors (to return money) or institutions investing in the illiquid asset (subordinated debt contracts).

[4] Diamond and Rajan (2001) develop a model in which the possibility of a bank run affects bankers' bargaining power in renegotiating loan contracts with borrowers. If a run occurs, depositors capture the loans and renegotiate with borrowers directly. It is the *threat* of a run that disciplines bankers. In Diamond and Rajan, a run cannot occur in equilibrium.

[5] The bank in the separated system reflects portfolio restrictions that US commercial banks faced during the old Glass–Steagall Act, under which investment banks and commercial banks were separate by law.

[6] Champ et al. (1996) dub as "panics" large shocks that necessitate rationing. For us, a "panic-induced" run is a typical Diamond–Dybvig run.

[7] Liquidity risk is also a concern. See the report on financial reform by the Group of Thirty (2009), especially page 43.

[8] Our motivation is different from theirs, but Jacklin (1987) and Wallace (1996) also allow utility to depend in a general way on consumption in each period.

[9] A partial list of the literature following Diamond and Dybvig's seminal paper includes Cooper and Ross (1998), Wallace (1988, 1990, 1996), Green and Lin (2003), Peck and Shell (2003), Andolfatto et al. (2007), and Ennis and Keister (2003, 2008, 2009).

the liquid technology would provide higher returns, while satisfying all liquidity needs and providing the incentive for patient consumers to wait (when there is no run). Since a run is worse than the highest possible liquidity shock, running becomes a self-fulfilling prophecy. This logic applies whenever liquidity demand is uncertain, and does not require the maximal liquidity shock to be large.

2. The model

There are three periods and a continuum of consumers (the bank depositors) represented by the unit interval. In period 0, each consumer is endowed with y units of the consumption good. A fraction α of the consumers is *impatient*: each of these has a "consumption opportunity" in period 1, yielding incremental utility of $\overline{u}$ for 1 unit of consumption in period 1. If the consumption opportunity goes unfulfilled in period 1, these consumers face a diminished (or discounted) consumption opportunity in period 2, yielding incremental utility of $\beta\overline{u}$ for 1 unit of consumption in period 2, where the scalar β is less than unity. The remaining consumers are *patient*: each of these has a consumption opportunity, yielding incremental utility of $\overline{u}$ for 1 unit of consumption in period 2. Beyond these urgent consumption opportunities, both types of consumers derive utility from additional (left-over) consumption in period 2, and can costlessly store consumption across periods. Thus, impatient and patient consumers, respectively, have the reduced-form utility functions:

$$U_I(C_I^1,C_I^2)=\begin{cases}\overline{u}+u(C_I^1+C_I^2-1) & \text{if } C_I^1\geq 1,\\ \beta\overline{u}+u(C_I^1+C_I^2-1) & \text{if } C_I^1<1\end{cases}$$

and

$$U_P(C_P^1,C_P^2)=\overline{u}+u(C_P^1+C_P^2-1), \tag{1}$$

where C_I^t (resp., C_P^t) is the *total withdrawal* of an impatient (resp., patient) consumer from the bank in period t. Specification (1) is based on the assumption that the consumers will always be able to afford their consumption opportunities in period 2, and that $\overline{u}$ is high enough so that it is optimal to undertake available consumption opportunities. It is assumed that u is an increasing, smooth, and strictly concave function of terminal (or left-over) consumption, so we have $u'>0$ and $u''<0$.

Let f denote the probability density function for α, the fraction of the consumers who become impatient, which is assumed to be continuous and have support $[0,\overline{\alpha}]$, where $\overline{\alpha}<1$. In keeping with the assumption that consumers are *ex ante* identical, think about the following process: First, nature determines α according to f. Then, nature selects each particular consumer to be impatient with probability α and patient with probability $(1-\alpha)$. A consumer's type is her private information.

There are two constant-returns-to-scale technologies, an illiquid, higher-yield technology denoted as technology i, and a liquid, lower-yield technology denoted as technology ℓ. Investing 1 unit of period-0 consumption in technology i yields R_i units of consumption in period 2. Investing 1 unit of period-0 consumption in technology ℓ yields R_ℓ units of consumption if held until period 2, or 1 unit of consumption if harvested in period 1, $1<R_\ell<R_i$.

In period 0, the bank designs the demand-deposit contract, or banking mechanism. The bank seeks to maximize the *ex ante* expected utility of consumers. Focus on the post-deposit game, which starts after the mechanism is announced and deposits are in place.[10] An *optimal contract* solves the traditional planner's problem, which imposes the incentive compatibility condition: a patient consumer chooses period 2 *given that all other patient consumers choose period* 2. Thus, the planner computes the optimal contract under the assumption that a run does not take place. We then ask whether an optimal contract also admits a run equilibrium.[11]

The banking mechanism must respect the restrictions required by the timing of the post-deposit game: At the beginning of period 1, each consumer (now a depositor) learns her type and decides whether to arrive at the bank in period 1 or period 2. Consumers who choose period 1 are assumed to arrive in random order. Let z_j denote the position of consumer j in the queue. Because of the sequential service constraint, consumption must be allocated to consumers as they arrive to the head of the queue, as a function of the history of transactions up until that point. It is assumed that consumer j 's withdrawal can only be a function of her position, z_j, and that she has an opportunity to refuse to withdraw and return without prejudice in period 2. The bank cannot keep track of how many consumers have refused.[12] Let α_1 denote the measure of consumers who have actually *made a withdrawal* in period 1. In period 2, the bank chooses how to divide its remaining resources between those who have withdrawn in period 1 and those who have not.

[10] The literature assumes that a consumer could invest her endowment herself, instead of dealing with the bank. Our results do not depend on whether we allow a consumer to access technology ℓ privately, but we do require that unharvested "trees" cannot be traded. This is to rule out the case in which a patient depositor (claiming to be impatient) trades period-1 consumption withdrawn from the bank for unharvested trees. Jacklin (1987) has shown that such a market undermines the optimal contract, and his argument applies to our setting as well. See Haubrich (1988) for a more general analysis. Moreover, it may be reasonable simply to posit that only banks can provide the liquidity necessary to pay for urgent consumption opportunities.

[11] The solution to the planner's problem would no longer be optimal if it admits a run equilibrium and the planner takes into account a positive probability of a run taking place.

[12] Thus, z_j should really be interpreted as the measure of consumers who have already withdrawn from the bank in period 1 before consumer j has an opportunity to withdraw. The purpose of this restriction is to disallow the bank from telling a customer during a run that she may not withdraw in period 1 and that she also forfeits her claim to consumption in period 2.

A contract specifies the fraction of a consumer's endowment invested in technology ℓ, denoted by γ; her withdrawal in period 1 as a function of her arrival position, denoted by $c^1(z)$; and her withdrawal in period 2 from technology ℓ investments as a function of α_1 and whether the consumer made a withdrawal in period 1 or not, denoted respectively by $c_I^2(\alpha_1)$ and $c_P^2(\alpha_1)$.[13] That is, a consumer who receives $c_I^2(\alpha_1)$ from technology ℓ investments receives a total withdrawal in period 2 of $C_I^2(\alpha_1)=c_I^2(\alpha_1)+(1-\gamma)R_iy$. Similarly, a consumer who receives $c_P^2(\alpha_1)$ from technology ℓ investments receives a total withdrawal in period 2 of $C_P^2(\alpha_1)=c_P^2(\alpha_1)+(1-\gamma)R_iy$. It is assumed that parameters are such that nonnegativity constraints $C_I^2(\alpha_1)\geq 0$ and $C_P^2(\alpha_1)\geq 0$ never bind.

For the mechanism to be feasible, all remaining resources must be distributed in period 2. The resource constraint is given by

$$\alpha_1 c_I^2(\alpha_1)+(1-\alpha_1)c_P^2(\alpha_1)=\left[\gamma y-\int_0^{\alpha_1} c^1(z)\,dz\right]R_\ell. \tag{2}$$

Thus, the space of deposit contracts or mechanisms M is given by $M=\{\gamma,c^1(z),c_I^2(\alpha_1),c_P^2(\alpha_1)|$ Eq. (2) holds for all $\alpha_1\}$.

We analyze bank behavior in each of the two financial systems: (I) In the *separated financial system*, consumers place a fraction $(1-\gamma)$ of their wealth in technology *i*, whose return cannot be touched by the bank. In terms of resource constraint (2), this is equivalent to imposing the additional constraints: $c_P^2(\alpha_1)\geq 0$ and, more importantly, $c_I^2(\alpha_1)\geq 0$. Combined with incentive compatibility, these additional constraints give rise to overinvestment in technology ℓ and the possibility of bank runs. (II) In the *unified financial system*, the bank is able to invest in both technologies. This allows the bank more flexibility in smoothing consumption and preventing runs. For example, when $\overline{\alpha}$ consumers arrive in period 1 (the worst case scenario), the bank can liquidate all of its technology ℓ holdings, but differentially reward consumers from technology *i* in period 2. Consumers who arrive in period 1 might receive less than $(1-\gamma)R_iy$ in period 2, while consumers who wait might receive more than $(1-\gamma)R_iy$. In terms of resource constraint (2) this is equivalent to allowing $c_P^2(\alpha_1)$, or, more importantly, $c_I^2(\alpha_1)$, to be negative.

Definition 2.1. Consider either a unified financial system or a separated financial system, and a contract $m\in M$. Then the *post-deposit game* is said to have a *run equilibrium* if there is a Bayes–Nash equilibrium in which all consumers arrive in period 1 and a positive measure of patient consumers withdraw in period 1.

Given a contract, the solution concept in Definition 2.1 is Bayes–Nash equilibrium. However, it should be clear from the constructions given below that the equilibrium strategies are also consistent with sequential equilibrium. The definition of run equilibrium requires all patient consumers to choose period 1. It is not required that all withdraw. The bank might very well offer zero consumption after $\overline{\alpha}$ of the consumers have made withdrawals (and hence a run is known to be in progress). A positive measure of patient consumers is required to withdraw in order to rule out the degenerate case in which patient consumers arrive in period 1 with the intention of refusing all offers, since this is equivalent to waiting until period 2.

3. The unified system

In this section, the planner's problem is described, the solution to which yields an optimal contract for the unified system. The optimal contract attains the full-information optimal outcome in equilibrium. Assuming that a patient depositor will choose not to run when indifferent between running and not running, the optimal contract does not have a run equilibrium.[14]

Attention is focused on environments in which it is beneficial to provide for consumption opportunities whenever the resources are available. It is then desirable that the impatient consumers choose period 1 and the patient consumers choose period 2 for making their urgent withdrawals. Also, impatient consumers unable to withdraw in period 1 and patient consumers should take advantage of their consumption opportunities in period 2.[15] Since the amount of a withdrawal in period 1 greater than one unit would be stored for period 2, there is no reason for the bank to provide more than one unit in period 1; hence $y\gamma\leq\overline{\alpha}$ holds, and optimal contracts must satisfy

$$c^1(z)=1 \quad \text{for } z\leq\gamma y. \tag{3}$$

Given (3) and the fact that patient consumers wait until period 2, the *ex ante* welfare W is given by

$$\begin{aligned} W=&\int_0^{\gamma y}[\,\overline{u}+(1-\alpha)u((1-\gamma)yR_i+c_P^2(\alpha)-1)+\alpha u((1-\gamma)yR_i+c_I^2(\alpha))]f(\alpha)\,d\alpha \\ &+\int_{\gamma y}^{\overline{\alpha}}[(1-\alpha+\gamma y)\overline{u}+(\alpha-\gamma y)\beta\overline{u}+(1-\alpha)u((1-\gamma)yR_i+c_P^2(\alpha)-1)+(\alpha-\gamma y)u((1-\gamma)yR_i+c_P^2(\alpha)-1) \\ &+\gamma y u((1-\gamma)yR_i+c_I^2(\alpha))]f(\alpha)\,d\alpha. \end{aligned} \tag{4}$$

[13] Here, in an abuse of notation, the subscripts *I* and *P* refer to consumers *claiming* to be impatient and patient respectively.

[14] See Theorem 3.2. Even without this assumption, such runs could be avoided at negligible cost to the bank. Alternatively, with sunspots and the propensity to run, the probability of a bank run, at the optimal contract to the full pre-deposit game, is zero.

[15] These conditions will be met if $\beta\overline{u}$ is large, relative to the marginal utility u' of "left-over" consumption.

Maximand (4) captures the fact that impatient consumers who are rationed in period 1 cannot be prevented from receiving the (higher) consumption that the patient consumers receive in period 2. The only relevant incentive compatibility constraint is that a patient consumer must be better off waiting until period 2 than accepting one unit in period 1, given that the other patient consumers wait. Conditional on being patient and being offered $c^1=1$, the conditional density for α is denoted by f_P. Note that in general $f_P(\alpha)$ is different from $f(\alpha)$. For example, if nothing is learned from observing $c^1=1$, then $f_P(\alpha)$ can be calculated as

$$f_P(\alpha)=\frac{(1-\alpha)f(\alpha)}{\int_0^{\overline{\alpha}}(1-a)f(a)\,da}.$$

The incentive compatibility constraint is

$$\int_0^{\overline{\alpha}} u(c_P^2(\alpha)+(1-\gamma)yR_i-1)f_P(\alpha)\,d\alpha \geq \int_0^{\overline{\alpha}} u(c_I^2(\alpha)+(1-\gamma)yR_i)f_P(\alpha)\,d\alpha. \tag{5}$$

Resource constraint (2) can be simplified to yield

$$\alpha_1 c_I^2(\alpha_1)+(1-\alpha_1)c_P^2(\alpha_1)=(\gamma y-\alpha_1)R_\ell \quad \text{if } \alpha_1 \leq \gamma y,$$

$$\gamma y c_I^2(\alpha_1)+(1-\gamma y)c_P^2(\alpha_1)=0 \quad \text{if } \alpha_1 > \gamma y. \tag{6}$$

An optimal contract under the unified system is the solution to the following problem:

$$\max_{\gamma,c_I^2(\alpha_1),c_P^2(\alpha_1)} \quad W$$
$$\text{s.t.} \quad (5) \text{ and } (6). \tag{7}$$

The next theorem establishes that an optimal contract necessarily rations consumers in period 1 when the fraction of impatient consumers arriving in period 1 is sufficiently large, i.e., when α_1 is close to $\overline{\alpha}$, equal to $\overline{\alpha}$, or greater than $\overline{\alpha}$. The intuition for this result is that, if consumers were never rationed (no matter the realization of α) in period 1, then society through over-caution would be overinvesting in liquid technology ℓ.[16]

Theorem 3.1. *An optimal contract in the unified system satisfies $\gamma y<\overline{\alpha}$. The "first" γy impatient consumers to arrive are fully served by the bank in period 1. There is a positive probability that $\alpha>\gamma y$ holds, in which case $(\alpha-\gamma y)$ impatient consumers are rationed. Patient consumers do not withdraw in period 1, and there is full consumption smoothing, i.e.,*

$$c_I^2(\alpha_1)=c_P^2(\alpha_1)-1 \quad \text{for all } \alpha_1 \leq \gamma y. \tag{8}$$

The formal proof of Theorem 3.1 is available in the web appendix. It shows that an optimal contract offers complete consumption smoothing, according to (6) and (8). Since the patient consumers receive the same consumption, whether they arrive in period 1 or period 2, it is obviously incentive compatible. The simple form of the consumption opportunities ensures that an optimal contract has a simple, realistic solution in which there is full, but not partial, suspension of convertibility.

Theorem 3.2. *There exists an optimal contract for the unified system. For any optimal contract, the corresponding allocation is socially optimal, maximizing W subject only to the resource constraint (6). Assuming that a patient depositor will choose not to run when indifferent between running and not running, there is an optimal contract that does not have a run equilibrium.*

The formal proof of Theorem 3.2 can be found in the web appendix. It is shown that conditions (6) and (8) imply the incentive compatibility constraint (5). The optimal contract achieves the constrained first-best. There is no run equilibrium because all of the illiquid asset is available to ensure incentive compatibility and depositors value their end-of-second-period bank balances.

4. The separated system

In order to evaluate the impact of portfolio restrictions on banks, it is assumed that the bank cannot gain access to the funds invested in technology *i*. As in Section 3, attention is restricted to environments in which it is optimal to provide one unit of consumption to consumers arriving in period 1, whenever technology ℓ assets are available.[17] Since second-period withdrawals from the bank must come from technology ℓ investments that were not harvested in period 1, the separated system is quite different from the unified system. When α is high enough, some impatient consumers are rationed in the

[16] For an analogous example, building a bridge designed to survive a 100-year storm might make more economic sense than building a more expensive bridge to survive a 500-year storm.

[17] That is, $c^1(z)=1$ holds for $z \leq \overline{\alpha}$. Not wanting to hoard technology ℓ assets is a stronger assumption in the separated system than in the unified system. We will see that incentive compatibility binds in the separated system, and refusing to liquidate technology ℓ assets allows the bank to reduce its technology ℓ overinvestment while still satisfying incentive compatibility. However, if we were to rewrite the expressions in problem (10) to allow for the possibility of hoarding technology ℓ assets, Theorem 4.2 and the overinvestment component of Theorem 4.3 continue to hold. Details are available from the authors.

unified system, yet full consumption smoothing is optimal. When α is high in the separated system, it may be impossible to provide those arriving in period 2 with one unit of consumption from technology ℓ investments at the optimal γ, so full consumption smoothing might be impossible. Excessive investment in technology ℓ might be necessary in order to satisfy incentive compatibility, so an optimal contract may require $\gamma y > \overline{\alpha}$. Finally, an optimal contract might be subject to bank runs, which can only be avoided at significant welfare cost.

In the separated system, *ex ante* welfare, the incentive compatibility constraint, and the resource constraint are as given in expressions (4), (5), and (6). The restriction that the bank cannot gain access to investments of technology i is expressed simply as

$$c_I^2(\alpha_1) \geq 0 \quad \text{and} \quad c_P^2(\alpha_1) \geq 0 \quad \text{for all } \alpha_1. \tag{9}$$

Notice that constraints (6) and (9) imply $c_I^2(\alpha_1) = c_P^2(\alpha_1) = 0$ for $\alpha_1 > \gamma y$. If all of the technology ℓ investments are liquidated in period 1, then withdrawals from the bank must be zero in period 2. An optimal contract under the separated system is a solution to the following planner's problem:

$$\max_{\gamma, c_I^2(\alpha_1), c_P^2(\alpha_1)} \quad W$$
$$\text{s.t.} \quad (5),\ (6),\ \text{and } (9). \tag{10}$$

In the unified system, when α is sufficiently high, all of the technology ℓ investments are harvested and some consumers are rationed in period 1. In the separated system: when α is close to $\overline{\alpha}$, it is typically the case that not all technology ℓ investments are harvested and no one is rationed. The intuition is that more investment in technology ℓ is needed in order to satisfy incentive compatibility since technology i resources cannot serve this purpose. While some technology ℓ investments may remain for patient consumers in state $\overline{\alpha}$, the next lemma establishes that in state $\overline{\alpha}$ the consumers who choose period 1 withdraw more than those who choose period 2. Denote the solution to (10) by $m^* = \{\gamma^*, (c_I^2(\alpha_1))^*, (c_P^2(\alpha_1))^*\}$.

Lemma 4.1. *Any optimal contract in the separated system, which solves problem* (10), *satisfies* $(c_P^2(\overline{\alpha}))^* < 1$.

The proof of Lemma 4.1 can be found in the web appendix. It is the cornerstone of the argument that any optimal contract in the separated system has a run equilibrium. The intuition behind Lemma 4.1 is that too much would have been invested in technology ℓ if patient consumers arriving in period 2 were to receive 1 unit or more of consumption in state $\overline{\alpha}$. In this case, reducing γ would not lead to rationing, would not violate incentive compatibility or nonnegativity constraints, but would yield a higher return on investment and increase welfare. In fact, the only reason to save any consumption at all for period 2 in state $\overline{\alpha}$ is to satisfy the incentive compatibility constraint. Lemma 4.1 therefore implies that, when many other consumers withdraw in period 1, a patient consumer is better off accepting 1 unit of consumption in period 1 than waiting. Thus, for large liquidity shocks, withdrawing in period 1 yields a higher utility realization than waiting.[18]

Theorem 4.2. *Any optimal contract in the separated financial system always has a run equilibrium.*

The proof of Theorem 4.2 can be found in the web appendix. There is an intuitive explanation for why run equilibria always exist in the separated system. In an optimal contract for the separated financial system, $(c_P^2(\overline{\alpha}))^*$ must be less than 1, else too much is invested in technology ℓ. Therefore, in the event of a run, consumers arriving in period 2 receive less than one unit. Consumers arriving in period 1 are better off, since they either receive 1 unit of consumption if $z_j \leq \overline{\alpha}$, while they can refuse to withdraw and delay their arrival until period 2 if $z_j > \overline{\alpha}$. The proof is a bit more intricate, since we must rule out the possibility that a fraction greater than $\overline{\alpha}$ of the consumers withdraw in period 1, possibly leaving more than one unit of consumption per capita in period 2.

The next theorem provides a *weak sufficient condition* for the optimal liquid asset investment under the separated financial system to be greater than it is under the unified financial system and for the absence of non-run rationing of impatient consumers of the restricted bank.

Theorem 4.3 (*Overinvestment in the liquid asset*). *Assume* $\overline{\alpha} < 1/R_\ell$ *holds. An optimal contract for the separated financial system does not ration consumers in period* 1 *in the no-run equilibrium, and invests more in technology* ℓ *than any optimal contract for the unified financial system.*

5. Numerical example

A computed numerical example can be found in the web appendix. The following are calculated: For the unified system, the optimal proportion γ of investment in the liquid asset and the probability of running out of cash; for the restricted bank, the optimal proportion γ of investment in the liquid asset and the value of $c_P^2(\overline{\alpha})$. γ in the restricted bank is nearly double that in the unrestricted bank. Since in the restricted bank $c_P^2(\overline{\alpha}) < 1$, there will be a panic-based run to the optimal contract. If the probability of the run is small enough, then tolerating the run is optimal. The cut-off probability, below which runs are tolerated, is calculated in the web appendix.

[18] Incentive compatibility implies that for small liquidity shocks, waiting yields a higher utility realization, which explains why a patient consumer waits when she expects the other patient consumers to wait (and she does not observe α).

6. Conclusions

We have made two important innovations in the bank-runs model which are meant to capture the timing of consumption opportunities. The opportunities are urgent: If checks do not clear at par, the transactions are lost. The opportunities can arise at any time: There is typically a long future of potential shopping beyond periods 1 and 2. We capture this future by assigning utility to the bank balance "left over" after period 2.

It is not surprising that the unrestricted bank delivers more *ex ante* welfare than the restricted bank, because the restricted bank faces one more constraint than the unrestricted bank. What might be surprising at first blush is that the unrestricted bank is more stable (in fact, perfectly stable) in the face of (extrinsic) panic-based shocks than is the restricted bank. On the other hand, restricted banking leads society to overinvest in the liquid asset, so the risk of running out of cash due to an intrinsic shock (large number of impatient depositors) is lower. This reduction in the risk of running out of cash argues, however, *against* restricted banking because the unified system strikes the *optimal* balance between economic growth during "good times" and lack of liquidity when there is a large number of impatient depositors. Portfolio restrictions do prevent the risk of "running out" of bank cash due to intrinsic shocks, but they also create the undesirable risk of bank runs triggered by extrinsic shocks.

Our analysis in its present state does not prove that imposing Glass–Steagal restrictions would be a mistake, although it does suggest that one should be skeptical about the purported stability benefits. Before using the model to offer policy advice, moral hazard should be included.[19] Beyond that, it would be worthwhile to study the sensitivity of the results to: (1) consumption opportunities and utility functions that vary across depositors, (2) intrinsic uncertainty about the returns R_i and R_ℓ[20] and about the liquidation values from harvesting the assets in the first period, (3) extensions to multiple periods and overlapping generations, and (4) time inconsistency, introduced into the bank runs literature by Ennis and Keister (2008).

One might think of the separated system as containing "narrow" banks, and the unified system as containing "wide" banks, but that impression would be wrong. There are many versions of the narrow banking proposal, which has a long history dating back before Friedman (1959). Portfolio restrictions, sometimes a 100% reserve requirement, is part of the proposal, but a crucial element is the obligation for the bank to honor a pre-specified withdrawal option in all circumstances. See Wallace (1996). As Freixas and Rochet (2008, p. 223) explain, "the term *narrow banking* … refers to a set of regulatory constraints on banks' investment opportunities that will make them safe in any possible event." This amounts to a restriction that the space of contracts is limited to a menu of consumption bundles that is independent of the history of withdrawals, thereby ruling out suspension schemes.

Besides narrow banking, Freixas and Rochet (2008) discuss a second remedy to banking instability: suspension of convertibility. Allowing deposit contracts to specify withdrawals that are fully contingent on the history can sometimes yield optimal welfare and eliminate bank-run equilibria. This works for the unified system.[21] In the separated system, on the other hand, the banks' portfolios are restricted to technology ℓ assets, but even though suspension schemes are allowed, bank-run equilibria *always* exist at the optimal contract. This is in contrast to the widespread opinion that "narrow-banking" type portfolio restrictions induce stable banking.

In response to the current financial crisis, we urge that proposals to restrict the activities of financial firms, and claims that such restrictions promote stability, be scrutinized carefully.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jmoneco.2010.04.006.

References

Andolfatto, D., Nosal, E., Wallace, N., 2007. The role of independence in the Green–Lin Diamond–Dybvig model. Journal of Economic Theory 137, 709–715.

Benjamin, M., Harper, C., 2009a. Volcker Urges Dividing Investment, Commercial Banks (Update 1), March 6 ⟨Bloomberg.com⟩.

Benjamin, M., Harper, C., 2009b. Glass–Steagall's Specter Returns to Haunt Wall Street (Update 2), March 10 ⟨Bloomberg.com⟩.

Champ, B., Smith, B., Williamson, S., 1996. Currency elasticity and banking panics: theory and evidence. Canadian Journal of Economics 29, 828–864.

Cooper, R., Ross, T.W., 1998. Bank runs: liquidity costs and investment distortions. Journal of Monetary Economics 41, 27–38.

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

Diamond, D.W., Rajan, R.G., 2001. Liquidity risk, liquidity creation, and financial fragility: a theory of banking. Journal of Political Economy 109 (2), 287–327.

Ennis, H.M., Keister, T., 2003. Economic growth, liquidity, and bank runs. Journal of Economic Theory 109 (2), 220–245.

Ennis, H.M., Keister, T., 2008. Commitment and equilibrium bank runs. Federal Reserve Bank of New York Staff Report No. 274, revised August 2008.

Ennis, H.M., Keister, T., 2009. Run equilibria in the Green–Lin model of financial intermediation. Journal of Economic Theory 144 (5), 1996–2020.

[19] Our analysis also does not include deposit insurance or credit chains (or other systemic problems) as analyzed by Kiyotaki and Moore (1997).

[20] Yu Zhang, in preliminary work, has extended the analysis to the case in which R_i is stochastic, a case where there is intrinsic uncertainty in asset returns.

[21] However, in other environments without portfolio restrictions, the optimal contract does tolerate bank runs with positive probability. See Peck and Shell (2003).

Freixas, X., Rochet, J.-C., 2008. Microeconomics of Banking, second ed. The MIT Press, Cambridge, MA, London, England.
Friedman, M., 1959. A Program for Monetary Stability. Fordham University Press, New York.
Green, E.J., Lin, P., 2003. Implementing efficient allocations in a model of financial intermediation. Journal of Economic Theory 109 (1), 1–23.
Group of Thirty, 2009. Financial Reform: A Framework for Financial Stability, ISBN 1-56708-146-0.
Haubrich, J.G., 1988. Optimal financial structure in exchange economies. International Economic Review 29 (2), 217–235.
Haubrich, J.G., King, R.G., 1990. Banking and insurance. Journal of Monetary Economics 26, 361–386.
Jacklin, C.J., 1987. Demand deposits, trading restrictions and risk sharing. In: Edward, J.P., Wallace, N. (Eds.), Contractural Arrangements for Intertemporal Trade. University of Minnesota Press, Minneapolis.
Kiyotaki, N., Moore, J., 1997. Credit-cycles. Journal of Political Economy 105 (2), 211–248.
Peck, J., Shell, K., 2003. Equilibrium bank runs. Journal of Political Economy 111 (1), 103–123.
Wallace, N., 1988. Another attempt to explain an illiquid banking system: the Diamond and Dybvig model with sequential service taken seriously. Quarterly Review of the Federal Reserve Bank of Minneapolis 12 (4), 3–16.
Wallace, N., 1990. A banking model in which partial suspension is best. Quarterly Review of the Federal Reserve Bank of Minneapolis 14 (4), 11–23.
Wallace, N., 1996. Narrow banking meets the Diamond–Dybvig model. Quarterly Review of the Federal Reserve Bank of Minneapolis 20 (1), 3–13.