Could Making Banks Hold Only Liquid Assets Induce Bank Runs?*

by

James Peck
Department of Economics, The Ohio State University, 440 Arps Hall, Columbus, OH 43210-1172. email: peck.33@osu.edu; telephone: (614) 292-0182; fax: (614) 292-3906.

Karl Shell
Department of Economics, 402 Uris Hall, Cornell University, Ithaca, NY 14853-7601 USA. email: ks22@cornell.edu; telephone: (607) 255-5277; fax: (607) 255-8838

September 15, 2009

Abstract and Headnote

We consider the effects of restrictions placed on bank portfolios. The analysis is based on the Diamond-Dybvig model modified to capture the role of checking accounts in facilitating transactions. Forcing banks to hold only liquid assets can paradoxically create the incentive for liquidity-based runs on the bank. Even when a run does not occur, social welfare is reduced as a result of overinvestment in the liquid asset.

Journal of Economic Literature Classification Numbers: D82, G21, E42.

Keywords: bank runs, bank stability, deposit contracts, Glass-Steagall banking, mechanism design, portfolio restrictions, sunspot equilibrium

Corresponding Author: James Peck

---

* 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 for help in clarifying and focusing our basic message. We are responsible for errors that persist.

1 Introduction

We revisit the question of whether or not legal restrictions on bank portfolios contribute to the stability of banks. With the recent financial meltdown that followed the collapse of the building society Northern Rock and the investment bank Bear Stearns, the study of bank runs (and, more generally, *financial instability*) has taken on new urgency. New regulatory oversight of investment banking seems inevitable, with the goal of making the sector less fragile. On March 6, 2009, Paul Volcker, former Fed Chairman and head of President Obama's Economic Recovery Advisory Board, remarked, "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back to the divisions between commercial and investment banks mandated by the Glass-Steagall Act before it was repealed in 1999.[1]

Our model is built on the classic model of Diamond and Dybvig (1983). We assume 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 forcing banks to hold only liquid assets can create the incentive for liquidity-based runs on the bank. That is, the equilibrium probability of a bank run is zero if bank portfolios are unrestricted; however, when banks are restricted from holding illiquid assets,[3] in equilibrium they always choose deposit contracts that are vulnerable to bank runs.[4] Hence Glass-Steagall-type regulations

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

[2] This two-technology approach was used by Wallace (1996).

[3] Restricted banks are assumed to be unable to tap indirectly into assets in the illiquid technology, with subordinated debt contracts or other arrangements designed to overcome portfolio restrictions.

[4] Diamond and Rajan (1998) 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.

create instability in the face of panic based runs.[5]

Our unrestricted "unified system" avoids panic-induced runs but does entail some instability in the face of intrinsic shocks. In particular, the unrestricted bank depletes its cash when the realized fraction of impatient depositors is sufficiently large.[6] However, there is a tradeoff between the risk of running out of cash and the benefit of higher asset returns (i.e., higher growth). We show for the unified system that the resulting allocation optimally resolves this tradeoff, and maximizes social welfare. In our "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 over-invest in the liquid asset to ensure that panic based runs do not always occur, and this 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 of an overly cautious growth inhibiting policy, while an undesirable 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 our paper. Even 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.

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").

[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, Smith, and Williamson (1996) dub as "panics" large shocks that necessitate rationing. For us, a "panic-induced" run is a typical Diamond-Dybvig run.

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

We follow Diamond and Dybvig in building our model 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 this need for the bank to make payments at par, we posit 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 our 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 structure of consumers' utility functions, based on consumption opportunities and future consumption, departs from the previous literature and has implications for the existence of bank runs, as we explain below.

In our model, the class of mechanisms is very broad and includes the *feasibility* of partial suspension of convertibility. Equilibrium bank runs never occur in the unified system, where the allocation maximizes social welfare, but in the separated system, the optimal contract

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

always has a run equilibrium. In the previous literature, bank run equilibria were demonstrated either for certain parameter values or when contracts that suspended convertibility were prohibited.[9]

The consumption good is perfectly divisible in our model, although there is a discrete jump in utility at one unit of consumption. An immediate consequence of this assumption is that in our model 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 our result about the separated system being vulnerable to bank runs. The left-over consumption feature provides the planner with tremendous power to eliminate 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 leftover consumption. The crucial assumption is the two-technology framework. In the separated system, banks invest enough in the liquid technology so that patient consumers are indifferent between withdrawing and waiting to consume whatever liquid assets remain in period 2, given the expectation that other patient consumers wait. Since this indifference balances the preference to wait when few others withdraw and the preference to withdraw when many others withdraw, running becomes a self-fulfilling prophecy.

[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, Nosal, and Wallace (2007), and Ennis and Keister (2003, 2008, 2009).

2 The Model

There are three periods and a continuum of consumers (the potential 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:

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

and

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

where C_i^t (for "cash") is the *total withdrawal* of a type i consumer from the bank in period t. I stands for impatient and P stands for patient. Specification (1) is based on the assumption that the consumers will always be able to afford their consumption opportunities in period 2, and that $\bar{u}$ is high enough so that it is optimal to undertake available consumption opportunities. We assume 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 $[0, \bar{\alpha}]$, where $\bar{\alpha} < 1$. In keeping with our assumption that consumers are identical, *ex ante*, we have the following process in mind. 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, A, and a liquid, lower-yield technology, B. Investing 1 unit of period-0 consumption in technology A yields R_A units of consumption in period 2. Investing 1 unit of period-0 consumption in technology B yields R_B units of consumption if held until period 2, or 1 unit of consumption if harvested in period 1, $1 < R_B < R_A$.

In period 0, the bank designs the demand-deposit contract, or banking mechanism. We assume that the bank seeks to maximize the *ex-ante* expected utility of consumers. We focus on the post-deposit game, which starts after the mechanism is announced and deposits are in place.[10] A *no-run optimal contract* (NROC) 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.* We can then ask whether a NROC also admits a run equilibrium.

The banking mechanism must respect the restrictions required by the timing of the post-deposit game, described as follows. 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.

[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 B 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.

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. We further assume 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.[11] 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.

A contract specifies the fraction of a consumer's endowment invested in technology B, 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 B 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 $c_P^2(\alpha_1)$.[12] That is, a consumer who receives $c_I^2(\alpha_1)$ from technology B investments receives a total withdrawal in period 2 of $C_I^2(\alpha_1) = c_I^2(\alpha_1) + (1-\gamma)R_A y$. Similarly, a consumer who receives $c_P^2(\alpha_1)$ from technology B investments receives a total withdrawal in period 2 of $C_P^2(\alpha_1) = c_P^2(\alpha_1) + (1-\gamma)R_A y$. We assume 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

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

[11] 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.

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

Thus, the space of deposit contracts or mechanisms M is given by

$$M = \{\ \gamma, c^1(z), c_I^2(\alpha_1), c_P^2(\alpha_1) \mid \text{Equation (2) holds for all } \alpha_1\ \}.$$

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 A, 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 B 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 $\bar{\alpha}$ consumers arrive in period 1 (the worst case scenario), the bank can liquidate all of its technology B holdings, but differentially reward consumers from technology A in period 2. Consumers who arrive in period 1 might receive less than $(1-\gamma)R_A y$ in period 2, while consumers who wait might receive more than $(1-\gamma)R_A y$. 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.

The unified system can be interpreted in several ways. The most straightforward and natural interpretation is that the bank is allowed to hold technology A assets (stocks or mutual funds) as part of its portfolio. Another interpretation is that the bank can write subordinated debt contracts with firms investing in technology A, whereby the bank receives period-2 consumption in the event that sufficiently many consumers arrive in period 1.

We next define "run equilibrium" in the post-deposit game.

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 consistent with sequential equilibrium as well. Our 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 $\bar{\alpha}$ of the consumers have made withdrawals (and hence a run is known to be in progress). We require a positive measure of patient consumers to withdraw, 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, we describe the planner's problem, the solution to which yields a NROC for the unified system. The NROC we construct 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 NROC does not have a run equilibrium.[13]

We restrict attention to 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.[14] 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 we have

[13] See Theorem (3.2). Even without this assumption, when we introduce sunspots and the propensity to run, it will follow that the probability of a bank run, at the optimal contract to the full pre-deposit game, is zero. Alternatively, such runs can be avoided at negligible cost to the bank.

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

$y\gamma \leq \bar{\alpha}$, and we can restrict our search to contracts in which we have

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

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

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

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, the conditional density for α is denoted by f_P. Note: in general, $f_P(\alpha)$ is different from $f(\alpha)$. For example, if nothing is learned from observing c^1, then $f_P(\alpha)$ can be calculated as

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

Thus, we have

$$\begin{aligned} \int_0^{\bar{\alpha}} u(c_P^2(\alpha) + (1-\gamma)yR_A - 1)f_P(\alpha)d\alpha \geq \int_0^{\gamma y} u(c_I^2(\alpha) + (1-\gamma)yR_A)f_P(\alpha)d\alpha \\ + \int_{\gamma y}^{\bar{\alpha}} (\gamma y/\alpha)u(c_I^2(\alpha) + (1-\gamma)yR_A) + (1-\gamma y/\alpha)u(c_P^2(\alpha) + (1-\gamma)yR_A - 1)f_P(\alpha)d\alpha. \end{aligned} \tag{5}$$

Resource constraint (2) can be simplified to yield

$$\begin{aligned} \alpha_1 c_I^2(\alpha_1) + (1-\alpha_1)c_P^2(\alpha_1) &= (\gamma y - \alpha_1)R_B \quad \text{if } \alpha_1 \leq \gamma y \\ \gamma y c_I^2(\alpha_1) + (1-\gamma y)c_P^2(\alpha_1) &= 0 \qquad\qquad\quad\; \text{if } \alpha_1 > \gamma y. \end{aligned} \tag{6}$$

An NROC under the unified system is the solution to the following problem:

$$\begin{aligned} &\max_{\gamma, c_I^2(\alpha_1), c_P^2(\alpha_1)} W \\ &\text{subject to (5) and (6).} \end{aligned} \tag{7}$$

The next theorem establishes that an NROC 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 $\bar{\alpha}$, equal to $\bar{\alpha}$, or greater than $\bar{\alpha}$. 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 over-investing in liquid technology B.[15]

Theorem 3.1: An NROC in the unified system satisfies $\gamma y < \bar{\alpha}$. 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 we have full consumption smoothing, i.e.,

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

Proof: Given γ, the functions $\{c_I^2(\alpha_1), c_P^2(\alpha_1)\}$ that maximize W subject only to resource constraint (6) entail full consumption smoothing (8). However, the allocation defined by (6) and (8) also satisfies the incentive compatibility constraint (5), and therefore solves the more tightly constrained problem (7). Plugging (6) and (8) into the expression for W, and differentiating with respect to γ, we have

[15] 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.

$$\left(\frac{\partial W}{\partial \gamma}\right)_{\gamma=\bar{\alpha}/y} = \int_0^{\bar{\alpha}} y(R_B - R_A)u'[c_P^2(\alpha) - 1 + (y - \bar{\alpha})R_A]f(\alpha)d\alpha < 0.$$

Clearly, any contract for which we have $\gamma y > \bar{\alpha}$ is inferior to the one characterized by (6) and (8), since the former provides fewer resources available in period 2, with no compensating advantage in terms of consumption smoothing in period 2 or provision of consumption in period 1. □

The proof of Theorem (3.1) shows that a NROC 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 consumption opportunities ensures that a NROC has a simple, realistic solution in which there is full, but not partial, suspension of convertibility.

Theorem 3.2: There exists a NROC for the unified system. For any NROC, 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 a NROC that does not have a run equilibrium.

Proof: First, note that a solution to (7) must exist. Given γ, a sufficient condition for $\{c_I^2(\alpha_1), c_P^2(\alpha_1)\}$ to solve problem (7) is for (6) and (8) to hold. The specification of consumption for $\alpha_1 > \gamma y$ does not affect the objective or the incentive compatibility constraint. Plugging (6) and (8) into the expression for W, problem (7) can be transformed into an equivalent unconstrained problem of choosing γ to maximize W, which must have a solution satisfying $\gamma y < \bar{\alpha}$. Because (6) and (8) imply (5), it follows that the NROC is socially optimal.

Construct the contract, $(\gamma, c_I^2(\alpha_1), c_P^2(\alpha_1))$, as follows. Let γ be as in the solution to (7).

Consumptions are determined by the resource equation, (6), and

$$c_I^2(\alpha_1) = c_P^2(\alpha_1) - 1 \qquad\qquad \text{for all } \alpha_1,$$

It follows that patient consumers are indifferent between withdrawing in period 1 and waiting. Under the assumption that a patient consumer will choose not to run when indifferent between running and not running, the NROC does not have a run equilibrium. □

In the next example, we specify the following basic parameters: the utility function $u(\cdot)$ for left-over consumption, the utility $\bar{u}$ from satisfying the urgent opportunity, the interest factor on the illiquid asset R_A, the discount factor β for delayed consumption opportunities, and the density function $f(\alpha)$ for the proportion who are impatient.

Example 3.3:

$$y = 10,\ u(c) = 100\log(c) - 249,\ \bar{u} = 20,\ R_A = 1.1,\ \beta = 0.7,$$

$$\text{uniform distribution with } \bar{\alpha} = 0.5\text{: } f(\alpha) = \begin{cases} 2 \text{ for } \alpha \in [0, 0.5] \\ 0 \text{ otherwise.} \end{cases} \tag{9}$$

For the unified financial system, we compute γ, the proportion of wealth invested in technology B and the *ex-ante* welfare of consumers W for different values of the interest factor R_B on the liquid technology. For $R_B = 1.05$, we have $\gamma = 0.04544$ and $W = 0.8942$; for $R_B = 1.08$, we have $\gamma = 0.04807$ and $W = 0.9599$.[16] An increase in R_B increases welfare, due to the higher yield on technology B investments, and increases γ, because reducing the probability of rationing consumers in period 1 is now less costly, because the gap between the yields in technologies A and B has been reduced.

[16] Computations were made using Maple version 5. The code is available from the authors for the purposes of replicating the results.

4 The Separated System

In order to evaluate the impact of portfolio restrictions on banks, we assume that the bank cannot gain access to the funds invested in technology A. As in Section 3, we restrict attention to environments in which it is optimal to provide one unit of consumption to consumers arriving in period 1, whenever technology B assets are available.[17] Since second-period withdrawals from the bank must come from technology B 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 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 B investments at the optimal γ, so full consumption smoothing might be impossible. Excessive investment in technology B might be necessary in order to satisfy incentive compatibility, so a NROC may require $\gamma y > \bar{\alpha}$. Finally, a NROC 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 A is expressed simply as follows:

$$c_I^2(\alpha_1) \geq 0 \text{ and } c_P^2(\alpha_1) \geq 0 \text{ for all } \alpha_1. \tag{10}$$

Notice that constraints (6) and (10) imply $c_I^2(\alpha_1) = c_P^2(\alpha_1) = 0$ for $\alpha_1 > \gamma y$. If all of the technology B investments are liquidated in period 1, then withdrawals from the bank must be zero in period 2. An NROC under the separated system is a solution to the following

[17] That is, $c^1(z) = 1$ holds for $z \leq \bar{\alpha}$. Not wanting to hoard technology B 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 B assets allows the bank to reduce its technology B overinvestment while still satisfying incentive compatibility. However, if we were to rewrite the expressions in problem (11) to allow for the possibility of hoarding technology B assets, Theorem (4.2) and the overinvestment component of Theorem (4.3) continue to hold. Details are available from the authors.

planner's problem:

$$\max_{\gamma, c_I^2(\alpha_1), c_P^2(\alpha_1)} W \tag{11}$$
$$\text{subject to (5), (6), and (10).}$$

In the unified system, all of the technology B investments are harvested and some consumers are rationed in period 1 when α is sufficiently high. In the separated system when α is close to $\bar{\alpha}$, it is typically the case that not all technology B investments are harvested and no one is rationed. The intuition is that more investment in technology B is needed in order to satisfy incentive compatibility without using technology A resources. While some technology B investments may remain for patient consumers in state $\bar{\alpha}$, the next lemma establishes that in state $\bar{\alpha}$ the consumers who choose period 1 withdraw more than those who choose period 2. Denote the solution to (11) by $m^* = \{\gamma^*, (c_I^2(\alpha_1))^*, (c_P^2(\alpha_1))^*\}$.

Lemma 4.1: Any NROC in the separated system, which solves problem (11), satisfies $(c_P^2(\bar{\alpha}))^* < 1$.

Proof: Suppose instead that $(c_P^2(\bar{\alpha}))^* > 1$ holds. Since resources remain in period 2, it follows that $\bar{\alpha} \leq \gamma^* y$ holds. Therefore, it is possible to increase welfare by reducing γ and achieving full consumption-smoothing, $(c_P^2(\alpha_1))^* = 1 + (c_I^2(\alpha_1))^*$ for all α_1. It is easy to see that incentive compatibility and nonnegativity are satisfied, contradicting the fact that m^* solves (11).

Now suppose that $(c_P^2(\bar{\alpha}))^* = 1$ holds. We must have $(c_I^2(\bar{\alpha}))^* = 0$, or else (just as in the previous case) we can increase welfare by reducing γ, while maintaining full consumption-smoothing. It follows that we must have

$$(c_P^2(\alpha_1))^* - (c_I^2(\alpha_1))^* \geq 1 \tag{12}$$

for almost all α_1. Otherwise, for a positive-measure set of realizations of α_1, $c_P^2(\alpha_1)$ can be increased and $c_I^2(\alpha_1)$ can be reduced to satisfy the resource and nonnegativity constraints, which increases welfare and relaxes the incentive compatibility constraint. If inequality (12) is strict for a positive-measure set of realizations of α_1, then welfare can be feasibly increased by choosing $(c_P^2(\alpha_1))^*$ and $(c_I^2(\alpha_1))^*$ to satisfy full consumption-smoothing (where (12) holds as an equality) and the resource constraint, (6). Therefore, we have for all α_1,

$$(c_P^2(\alpha_1))^* - (c_I^2(\alpha_1))^* = 1. \tag{13}$$

Treating γ as a parameter, and solving the equations (6) and (13) for consumptions, we can define welfare as a function of γ, $W(\gamma)$. Because m^* solves (11), $W(\gamma)$ must be maximized at $\gamma = \gamma^*$. Applying the envelope theorem, one can show:

$$W'(\gamma^*) = y(R_B - R_A) \int_0^{\bar{\alpha}} u'[(1-\gamma^*)yR_A + (\gamma^* y - \alpha)R_B + \alpha - 1]f(\alpha)d\alpha < 0.$$

It follows that reducing γ improves welfare, contradicting the fact that γ^* is part of the NROC, m^*. □

The intuition behind Lemma (4.1) is that too much is invested in technology B if patient consumers arriving in period 2 receive at least 1 unit of consumption in state $\bar{\alpha}$. Reducing γ does not lead to rationing, and yields a higher return on investment. The only reason to save any consumption at all for period 2 in state $\bar{\alpha}$ is to satisfy non-negativity and incentive compatibility constraints. Since these constraints are not binding if $(c_P^2(\bar{\alpha}))^* \geq 1$ holds, too much has been invested in technology B. Applying Lemma (4.1), we next show that a NROC in the separated system always admits a run equilibrium.

Theorem 4.2: Any NROC in the separated financial system has a run equilibrium.

Proof: We know that a NROC satisfies $c^1(z) = 1$ for all $z \leq \bar{\alpha}$. Since we are considering the possibility of bank runs here, a patient consumer's decision to arrive in period 1 must also take into account $c^1(z)$ for all $z > \bar{\alpha}$. Let z^* be the smallest value of z, greater than or equal to $\bar{\alpha}$, such that the following inequality holds

$$c^1(z) \leq \frac{[\gamma y - \bar{\alpha} - \int_{\bar{\alpha}}^{z} c^1(a)da]R_B}{1-z}. \tag{14}$$

If there is no value of z satisfying (14), define z^* to equal 1. From Lemma (4.1), we know that $(c_P^2(\bar{\alpha}))^* < 1$ holds at the NROC. We must also have $(c_I^2(\bar{\alpha}))^* = 0$, or else higher welfare can be achieved by transferring consumption from those who arrived in period 1 to those who did not, while continuing to satisfy the constraints.[18] It follows that, setting $z = \bar{\alpha}$, the left side of (14) is equal to unity, and the right side of (14) is equal to $(c_P^2(\bar{\alpha}))^*$. Since inequality (14) is not satisfied, we have $z^* > \bar{\alpha}$.

We claim that there is a run equilibrium, in which all consumers arrive in period 1. Those for whom $z_j < z^*$ holds accept $c^1(z_j)$, and those for whom $z_j \geq z^*$ holds refuse $c^1(z_j)$ and do not withdraw in period 1.

Without loss of generality, we can assume $(c_I^2(\alpha_1))^* = 0$ for $\alpha_1 > \bar{\alpha}$, because giving period-2 consumption to those who withdraw in period 1 only increases the incentive to run. Thus, for $\alpha_1 > \bar{\alpha}$, second period consumption is given by

$$(c_P^2(\alpha_1))^* = \frac{[\gamma y - \bar{\alpha} - \int_{\bar{\alpha}}^{\alpha_1} c^1(a)da]R_B}{1-\alpha_1}. \tag{15}$$

Differentiating with respect to α_1 in (15) yields

$$\frac{\partial (c_P^2(\alpha_1))^*}{\partial \alpha_1} = \frac{(c_P^2(\alpha_1))^* - c^1(\alpha_1)R_B}{1-\alpha_1},$$

which is negative for $\alpha_1 < z^*$, since we have $c^1(\alpha_1)R_B > c^1(\alpha_1) > (c_P^2(\alpha_1))^*$, with the second inequality due to inequality (14) not holding. Thus $(c_P^2(\alpha_1))^*$ is decreasing in α_1 for $\alpha_1 < z^*$.

[18] Remember, the NROC is the one that provides the highest welfare, based on the equilibrium in which the patient consumers wait until period 2.

Given the acceptance/refusal behavior specified above, everyone will refuse $c^1(z^*)$, so we have $\alpha_1 = z^*$ with probability 1. By (14) and (15), it is a best response for consumer j to refuse if $z_j = z^*$, and $z_j > z^*$ is irrelevant. If $z_j < z^*$ holds, (14) and (15) imply $c^1(z_j) > (c_P^2(z_j))^* > (c_P^2(z^*))^*$, where the last inequality follows from the fact that $(c_P^2(\alpha_1))^*$ is continuous and decreasing in α_1 for $\alpha_1 < z^*$. Therefore, it is a best response for consumer j to accept $c^1(z_j)$. □

There is an intuitive explanation for why run equilibria always exist in the separated system. In an NROC for the separated financial system, $(c_P^2(\bar{\alpha}))^*$ must be less than 1, else too much is invested in technology B. 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 receive 1 unit of consumption if $z_j \leq \bar{\alpha}$, and they can refuse to withdraw and delay their arrival until period 2 otherwise. The proof is a bit more intricate, since we must rule out the possibility that a fraction greater than $\bar{\alpha}$ of the consumers withdraw in period 1, possibly leaving more than one unit of consumption per capita in period 2. Another perspective is that the desire to economize on technology B investments causes the incentive compatibility constraint to bind, so that a patient consumer is indifferent between period 1 arrival and period 2 arrival, assuming other patient consumers wait. If instead the other patient consumers arrive in period 1, those who wait are worse off, so that incentive compatibility is no longer satisfied.

The following 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 absense of non-run rationing of impatient consumers of the restricted bank.

Theorem 4.3 (Overinvestment in the Liquid Asset): If $\bar{\alpha} < 1/R_B$ holds, then any NROC for the separated financial system does not ration consumers in period 1 in the no-run equilibrium, and invests more in technology B than any NROC for the unified financial system.

Proof: From Theorem (3.1), an NROC in the unified system satisfies $\gamma < \bar{\alpha}/y$. In the separated system, an NROC must invest at least enough in technology B to provide 1 unit of consumption in period 2 when everyone is patient. That is, we must have $(c_P^2(0))^* \geq 1$, or else all patient consumers will choose period 1. Thus, the optimal fraction invested in technology B for the separated system, γ^*, satisfies $\gamma^* \geq 1/(R_B y)$. Since $R_B\bar{\alpha} < 1$ holds, we draw the following conclusions. First, $\gamma^* > \bar{\alpha}/y$ holds, so consumers are not rationed in period 1 unless there is a run. Second, γ^* exceeds the optimal fraction invested in technology B for the unified system. □

This overinvestment in the liquid asset, B, is likely to be substantial, as long as the maximum fraction of impatient consumers is relatively small. For example, if we have $\bar{\alpha} = 0.5$, $y = 10$, and $R_B = 1.08$, then in the unified system, the fraction of resources invested in technology B is less than 0.05, as a consequence of Theorem (3.1). In the separated system, the fraction of resources invested in technology B is more than 0.09, as consequence of the fact that $(c_P^2(0))^* \geq 1$ must hold, which implies $\gamma^* \geq 1/(R_B y)$. Notice that this lower bound on overinvestment (at least 80% more than under the unified system) applies to all utility functions consistent with our maintained assumptions.

Next we compute the NROC in the separated system, for the parameters of Example (3.3).

Example 4.4: Consider the parameter values specified in Example (3.3). We also specify $R_B = 1.08$.

It can be shown that the NROC satisfies $(c_I^2(\alpha_1))^* = 0$ for all α_1. Since Theorem (4.3) applies, it follows from (6) that we have

$$(c_P^2(\alpha_1))^* = \frac{(10\gamma - \alpha_1)1.08}{1-\alpha_1}. \tag{16}$$

Finding the NROC now reduces to finding the value of γ that maximizes welfare subject to the incentive compatibility constraint. The optimal γ will cause the incentive compatibility constraint to hold with equality, yielding: $\gamma^* = 0.09445$ and $W^* = 0.8688$. Comparing the NROC in the unified system to the NROC in the separated system, we see that technology B investment in the separated system is nearly double that in the unified system.

The run equilibrium to our NROC–like the run equilibrium in Diamond and Dybvig (1983)–is not really an equilibrium to the pre-deposit game, because consumers would not deposit their funds if they knew that a run would take place. See Postlewaite and Vives (1987). Diamond and Dybvig suggest that a run could take place in equilibrium with positive probability, triggered by "sunspots," as long as the probability of the run is sufficiently small. Cooper and Ross (1998) and Peck and Shell (2003) formalize this notion. Consider the context of our example, defined by the parameters in (9), and $R_B = 1.08$. From (16), we can calculate the optimal γ to avoid runs, $\gamma^{**} = 0.09630$. By comparing welfare in the optimal contract that avoids runs and welfare in the optimal contract that tolerates runs, one can show that the cutoff propensity to run is 0.005521. If the propensity to run is less than 0.5521%, then it is better to tolerate the unlikely event of a run than to increase technology B investment to prevent runs.[19]

5 Conclusions and Concluding Remarks

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

[19] See the papers by Lagunoff and Schreft (1998) and Allen and Gale (1998) for an analysis of financial crises based on local interactions. Our notion of fragility is distinct from theirs, yet not altogether unrelated.

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 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 *against* restricted banking, however, because the unified system strikes the *optimal* balance between growth during good times and lack of liquidity when there is a large number of impatient depositors. Thus, portfolio restrictions prevent the worthwhile risk of planned "running out" of bank cash due to intrinsic shocks, but create a new and undesirable risk, of bank runs triggered by extrinsic shocks.

Our model in its present state does not prove that imposing Glass-Steagal restrictions would be a mistake, although it suggests that one should be skeptical about the purported stability benefits. Before using the model to offer policy advice, moral hazard should be included.[20] Beyond that, it would be worthwhile to study the sensitivity of our results to: (1) consumption opportunities and utility functions that vary across depositors, (2) intrinsic uncertainty about the returns R_A and R_B 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 with-

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

drawal option in all circumstances. See Wallace (1996). As Freixas and Rochet (2008, page 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 is the case for our unified system.[21] In our separated system, on the other hand, we restrict the banks' portfolios to technology B assets and allow for suspension schemes, yet we find that bank run equilibria always exist at the NROC. This is in contrast to the widespread opinion that "narrow-banking" type portfolio restrictions provide stability.

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.

References

[1] Allen, Franklin and Gale, Douglas. "Financial Contagion." *Journal of Political Economy,* 108(1) (2000): 1-33.

[2] Andolfatto, David; Nosal, Ed; and Wallace, Neil. "The Role of Independence in the Green-Lin Diamond-Dybvig Model," *Journal of Economic Theory,* 137 (2007): 709-715.

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

[3] Benjamin, Matthew and Christine Harper. "Volcker Urges Dividing Investment, Commercial Banks (Update 1)," Bloomberg.com, March 6, 2009.

[4] Benjamin, Matthew and Christine Harper. "Glass-Steagall's Specter Returns to Haunt Wall Street (Update 2)," Bloomberg.com, March 10, 2009.

[5] Champ, Bruce; Smith, Bruce; and Williamson, Stephen. "Currency Elasticity and Banking Panics: Theory and Evidence," *Canadian Journal of Economics*, 29 (1996): 828-864.

[6] Cooper, Russell and Ross, Thomas W. "Bank Runs: Liquidity Costs and Investment Decisions." *Journal of Monetary Economics* 41 (1998): 27-38.

[7] Diamond, Douglas W. "Liquidity, Banks and Markets." *Journal of Political Economy,* 105(5) (1997): 928-56.

[8] Diamond, Douglas W. and Dybvig, Philip H. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy,* 91 (1983): 401-19.

[9] Diamond, Douglas W. and Rajan, R. G. "Liquidity Risk, liquidity creation, and financial fragility: A theory of banking," *Journal of Political Economy,* 109(2) (2001): 287-327.

[10] Ennis, Huberto M. and Keister, Todd. "Economic Growth, Liquidity, and Bank Runs," *Journal of Economic Theory,* 109(2) (2003): 220-45.

[11] Ennis, Huberto M. and Keister, Todd. "Commitment and Equilibrium Bank Runs," Federal Reserve Bank of New York Staff Report No. 274, revised August 2008.

[12] Ennis, Huberto M. and Keister, Todd. "Run Equilibria in the Green-Lin Model of Financial Intermediation," *Journal of Economic Theory*, 144(5) (2009): 1996-2020.

[13] Freixas, Xavier and Jean-Charles Rochet, *Microeconomics of Banking*, Second Edition. Cambridge, Massachusetts and London, England: The MIT Press, 2008.

[14] Friedman, Milton, *A Program for Monetary Stability*, New York: Fordham University Press, 1959.

[15] Green, Edward J. and Lin, Ping. "Implementing Efficient Allocations in a Model of Financial Intermediation." *Journal of Economic Theory,* 109(1) (2003): 1-23.

[16] Group of Thirty. "Financial Reform: A Framework for Financial Stability," ISBN 1-56708-146-0 (2009).

[17] Haubrich, Joseph G. "Optimal Financial Structure in Exchange Economies," *International Economic Review*, 29(2) (1988), 217-235.

[18] Jacklin, Charles J."Demand Deposits, Trading Restrictions, and Risk Sharing." In *Contractural Arrangements for Intertemporal Trade*, edited by Edward J. Prescott and Neil Wallace. Minneapolis: University of Minnesota Press, 1987.

[19] Kiyotaki, Nobuhiro and Moore, John. "Credit-Cycles." *Journal of Political Economy,* 105(2) (1997): 211-48.

[20] Lagunoff, Roger and Schreft, Stacey. "A Model of Financial Fragility." *Journal of Economic Theory,* 99(1-2) (2001): 220-64.

[21] Peck, James and Shell, Karl. "Equilibrium Bank Runs." *Journal of Political Economy,* 111(1) (2003): 103-23.

[22] Postlewaite, Andrew and Vives, Xavier. "Bank Runs as an Equilibrium Phenomenon." *Journal of Political Economy,* 95(3) (1987): 485-491.

[23] Wallace, Neil. "Another Attempt to Explain an Illiquid Banking System: The Diamond-Dybvig Model with Sequential Service Taken Seriously." *Quarterly Review of the Federal Reserve Bank of Minneapolis,* 12(4) (1988): 3-16.

[24] Wallace, Neil. "A Banking Model in Which Partial Suspension is Best." *Quarterly Review of the Federal Reserve Bank of Minneapolis,* 14(4) (1990): 11-23.

[25] Wallace, Neil. "Narrow Banking Meets the Diamond-Dybvig Model." *Quarterly Review of the Federal Reserve Bank of Minneapolis,* 20(1) (1996): 3-13.