

Bank Portfolio Restrictions and Equilibrium 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 and
New York University, Department of Economics, 269 Mercer Street, New York, NY 10003 USA; e-mail: ks22@cornell.edu; telephone (212) 998-8927; fax (212) 995-4186

December 3, 1999

Abstract and Headnote

We put “runs” back in the bank runs literature. A unified bank, one that invests in both liquid and illiquid assets, can easily avoid runs but it still faces a small probability of non-run rationing of depositors. In a separated financial system, the bank only holds relatively liquid assets; it is subject to runs with small probability, but because of its overinvestment in the liquid asset it is immune to non-run rationing of depositors. Surprisingly, legal restrictions on the bank’s portfolio are either ineffective or they increase the fragility of the bank.

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

Keywords: Bank runs, deposit contracts, payments mechanism, mechanism design, sunspot equilibrium.

Correspondent: 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.

Proposed Running Head: Equilibrium Bank Runs

* We thank the Thorne Fund at Cornell University for research support and Patrick Bolton, John Duggan, Peter Howitt, Bruce Smith and seminar participants at IUPUI, NYU and Princeton for helpful discussions and comments. We are grateful to Neil Wallace for extremely useful criticism of an earlier draft.

1 Introduction

The theoretical literature on bank runs is based on the early work of Bryant (1980) and the now classic model of Diamond and Dybvig (1983) — henceforth DD. While DD claim that with simple deposit contracts bank runs are possible, they show that bank run equilibria can be avoided with more sophisticated mechanisms that allow for the suspension of convertibility. When the fraction of impatient depositors is random, DD find that government-provided deposit insurance can eliminate runs and achieve the full-information first best. In two important papers, Wallace (1988, 1990) argues that the sequential service constraint should apply to the government as well as the bank, in which case the full-information first-best allocation is not achievable.¹ The constrained-efficient contract entails partial suspension of convertibility. Green and Lin (1996) extend Wallace’s framework to include more general distributions. They show that the constrained-efficient outcome can always be implemented by a mechanism that employs partial suspension of convertibility and excludes bank runs.² Cooper and Ross (1998) do find equilibrium bank runs, but their banks are restricted to simple deposit contracts that do not allow for suspension of convertibility.

Our goal is to put “runs” back in the bank runs literature. In particular, we investigate the possibility of equilibrium runs on banks that can write sophisticated contracts in which the current withdrawal depends on the history of withdrawals; if a bank is restricted to only very simple deposit contracts, these simple contracts are not optimal in a broader sense. Bank runs are historical facts. If bank runs were impossible, then much of banking policy would be directed toward a non-issue. In our model, equilibrium bank runs are possible and government policies that restrict banks to holding only (lower-yield) liquid assets might

¹See also McCulloch and Yu (1998).

²Other authors analyze market based solutions in which it is assumed that patient depositors always wait before withdrawing their funds. See, e.g. Jacklin (1987) and Diamond (1997). In this literature, runs are assumed away.

actually increase the likelihood of bank runs.³

We build on the basic DD model and its successors, but there are significant differences:

(1) In DD, banks provide insurance against the event that a consumer becomes impatient and must do her consumption “early” (if patient, she consumes “later”). We follow DD in building our model upon the stochastic nature⁴ of some urgent consumption opportunities or urgent needs, but we modify the model to capture the transactions and payments roles played by checking accounts. For us, a depositor facing a consumption opportunity is someone who requires immediate liquidity to make an important purchase or someone taking advantage of the convenience of writing a check. The benefits of a demand deposit account would be severely limited if 100% payment were not made by the bank: either the cash transaction could not be completed or the check would bounce. We attempt to capture the payment role of banks by assuming that when the consumer finds a “consumption opportunity” it is of the nature of an indivisible good. We assume that the impatient consumers find their consumption opportunities in the first period, while the patient consumers find their consumption opportunities in the second period. In another departure from DD and as a proxy for more complete intertemporal analysis, we assume that all consumers value “left-over” consumption goods (beyond the demand for funds to finance these indivisible consumption opportunities) in the final period.⁵ Utility is assumed to be a strictly concave function of “left-over” consumption.

(2) We allow for intrinsic uncertainty: in particular, we allow the proportion of impatient

³Diamond 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. However, it is the *threat* of a run that disciplines bankers, and a run can not occur in equilibrium.

⁴For a general-equilibrium analysis of this type of intrinsic uncertainty, see Peck (1996).

⁵Our motivation is somewhat different, but our utility specification is a special case (except for the indivisibilities) of those in Jacklin (1987) and Wallace (1996).

consumers to be random.⁶ We introduce a publicly observed extrinsic (or sunspots) random variable on which depositors can coordinate their actions. It seems to us that it is the combination of intrinsic uncertainty *and* extrinsic (or strategic, market) uncertainty that is essential to financial intermediation — especially to the potential fragilities in financial intermediation.

(3) 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.⁷ We introduce an *ad hoc* (possibly zero) cost to the bank of serving as a consolidated or unified financial institution that holds both types of investments. We also allow for the possibility that the government might restrict banks to holding only the liquid asset in their portfolios.

(4) We follow Wallace and others in treating the bank's deposit contract as a mechanism, but we place some restrictions on this mechanism. While the bank must satisfy sequential service, we do not allow the bank to punish customers who were denied service in period 1 or who refused service in period 1. We also assume the bank is not allowed to offer pure lotteries in its deposit contract.

We analyze two financial systems. In the first, the unified financial system, the bank provides payment services and invests in both types of assets. In the second, the separated⁸ financial system, the bank provides payment services but only invests in the liquid asset. The separated financial system might emerge because of regulations which restrict bank

⁶See also DD, Green and Lin (1996), and Wallace (1988, 1990).

⁷This two-technology approach was used by Wallace (1996). The restrictions on the bank in Wallace's formal models are different from ours. See also Cooper and Ross (1998).

⁸Our separated financial system yields a bank that is close to what advocates of narrow banking recommend, although Friedman would perhaps restrict banks from holding government bonds. See the introduction to Wallace (1996). The differences are largely of interpretation. We are thinking that the separated bank does not hold the most illiquid assets in its portfolio, while the narrow bank is restricted to holding only the most liquid assets. Of course, with only two asset types, these ideas are the same. Nonetheless, advocates of narrow banking should be startled by our results.

investment to relatively liquid assets or because the real (nonregulatory) costs of financial integration lead to a separated financial system.

The unified bank can easily eliminate runs, but its constrained-efficient contract does allow for non-run rationing⁹ of impatient customers when the proportion of impatient customers is large. For the separated bank, the constrained efficient contract allows for runs with sufficiently small probability. Consumers are rational depositors in such banks. Compared to a unified bank with no extra transactions costs, the separated financial system heavily overinvests in the more liquid asset. For this reason, non-run rationing in the separated bank is impossible or extremely unlikely.

The social policy implications of this analysis might be surprising — at least to some of the early proponents of narrow banking. If the government restricts banks to holding only the more liquid assets, then either this restriction will have no effect because the real costs of unification are high or this restriction will make the financial system more fragile by introducing a positive probability of bank runs.

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 an urgent need for 1 unit of consumption in period 1. The remaining consumers are *patient*: each of these has an urgent need for 1 unit of consumption in period 2.¹⁰ Beyond these urgent

⁹This type of rationing is socially desirable since otherwise there would be overinvestment in preparation for the worst case outcome. Champ, Williamson, and Smith (1996) argue that large real shocks that necessitate rationing should be called panics. Thus, our unified financial system can avoid extrinsic panics of the self-fulfilling DD variety, but opens up the possibility of intrinsic panics and rationing.

¹⁰A slightly more realistic version would break the strict dichotomy between patient and impatient consumers. In each of several periods, each consumer has a probability of an urgent opportunity. This probability well might depend on the consumer's past history, but would not necessarily be zero even if an urgent

“consumption opportunities,” both types of consumers derive utility from additional consumption in period 2, and can costlessly store consumption across periods. Thus, impatient and patient consumers, respectively, have the reduced form utility functions:

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

and

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

where C_i^t is the *total withdrawal* of a type i consumer from the bank in period t . I stands for impatient and P stands for patient. The positive scalar \bar{u} is the utility from the (indivisible) consumption opportunity. Specification (1) is based on the assumption that the patient consumer will always be able to afford her consumption opportunity, 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 $[0, \bar{\alpha}]$. In keeping with our assumption that consumers are identical, *ex ante*, we have the following process in mind.¹¹ First, nature determines α according to f . Then, nature selects each particular consumer to be impatient with probability α and patient with probability $(1-\alpha)$. Conditional on being patient, the density for α , denoted as f_P , can be calculated as

consumption occurred in the past.

¹¹The continuum model is convenient, but there are technical issues regarding the law of large numbers which we ignore. Our main results still hold with a finite number of consumers, although the expressions and calculations become more complicated.

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

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. We assume that $1 < R_B < R_A$ holds.

In period 0, the bank designs a demand-deposit contract, which we call the banking mechanism. We assume that the bank seeks to maximize the *ex-ante* expected utility of consumers. The banking mechanism must respect the following timing considerations. In period 0, consumers must be willing to make the required deposit¹². 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. We further assume that consumer j 's withdrawal can only be a function of her position, z_j ¹³, and that she has an opportunity to refuse to withdraw and return without prejudice in

¹²A consumer could invest her endowment herself, instead of dealing with the bank. It does not matter whether or not 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. Ruling out this asset market is merely to posit that only banks can provide the liquidity necessary to pay for urgent consumption opportunities.

¹³From a mechanism design standpoint, it might seem strange not to allow consumers to send messages

period 2. The bank cannot keep track of how many consumers have refused.¹⁴ Let α_1 denote the measure of consumers who have 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.¹⁵

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 $c_P^2(\alpha_1)$.¹⁶ 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 a mechanism to be feasible, all remaining resources must be distributed in period 2. Letting λ denote a fixed transactions cost, to be motivated below, we then have

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

to the bank. This is almost without loss of generality, since consumers will send whatever message gives them the most consumption. However, we are ruling out the bank offering a lottery to learn a consumer's type, and punishing consumers in period 1 after a patient consumer has arrived. Besides being costly to implement, these lotteries and punishments hardly correspond to accepted conservative banking practices.

¹⁴Thus, 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 they may not withdraw in period 1 and that they also forfeit their claims to consumption in period 2.

¹⁵In principle, what a consumer receives in period 2 could depend on how much she received in period 1, and not just on whether she made a withdrawal. This distinction is not important here, since the optimal contract always provides 1 unit of consumption in period 1 until the bank runs out.

¹⁶Here, in a slight abuse of notation, the subscripts I and P refer to consumers *claiming* to be impatient and patient respectively.

Thus, the space of 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 have two financial systems in mind. First, 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 the overinvestment in technology B and the possibility of bank runs. Offsetting the inefficiencies caused by separation, the transactions cost might be less in the separated system. Hence we normalize transaction costs to be zero in the separated system; if λ is positive, it is the transactions cost incurred by the unified system beyond those incurred by the separated system.

Second, under the *unified financial system*, the bank is able to invest in both technologies. This allows the bank a great deal of flexibility to smooth consumption and prevent 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 arrived in period 1 might receive less than $(1 - \gamma)R_A y$ in period 2, while consumers who waited 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 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 allow for the possibility that the unified system might incur transactions costs beyond those incurred in

a separated system. Thus, λ could represent the cost of writing and enforcing subordinated debt contracts, or the cost associated with the bank's moral hazard problem with respect to the issuers of subordinated debt contracts. Under the interpretation in which the bank directly holds technology A assets, λ could represent the cost associated with combining or linking a depositor's several accounts, or the cost associated with the bank's increased moral hazard problem with respect to its depositors.¹⁷

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

Our definition of run equilibrium requires all patient consumers to choose period 1. They are not all required to withdraw, since the bank might very well offer zero consumption after $\bar{\alpha}$ of the consumers have made withdrawals (and a hence 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. Following the literature, we focus for the moment on the post-deposit game. In Section 5, with the help of extrinsic uncertainty we analyze equilibrium runs in the full pre-deposit game.

3 The Unified System

In this section, we describe the planner's problem, the solution to which yields the optimal contract for the unified system. The optimal contract very nearly, but not quite, attains

¹⁷Plausible parameters yield an optimal contract in the separated system that pays a real interest rate whose expectation is near zero and whose variance is small. In the unified system, the range of second-period consumption is much greater, as is the temptation for the bank to cheat. See Calomiris and Kahn (1991) for an explicit analysis of moral hazard issues and embezzlement in banking.

the full-information optimal outcome in equilibrium. Although we do not rule out a run equilibrium at the optimal contract, we will argue that this possibility is not quantitatively interesting. On the other hand, we demonstrate below that the optimal contract entails a positive probability of (non-run) rationing of consumers in period 1.

We restrict attention to environments in which it is beneficial to provide for consumption opportunities whenever the resources are available. It is then clearly desirable that the impatient consumers choose period 1 and the patient consumers choose period 2 for making their urgent withdrawals. 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 $y\gamma \leq \bar{\alpha}$. Thus we can restrict the search to contracts in which we have

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

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

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

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 incentive compatibility constraint to worry about 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. Thus, we have

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

Resource constraint (2) can be simplified to yield

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

The optimal contract under the unified system is the solution to the following planner's problem:

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

The next theorem establishes that the optimal contract necessarily rations consumers in period 1 when the number of impatient consumers is sufficiently large, i.e. the realization of α is close to its maximum. 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 the liquid technology, B .

Theorem 3.1: In the optimal contract for the unified system, consumers in period 1 are rationed when the realization α is sufficiently close to its maximum possible value, $\bar{\alpha}$. That is, for optimality in the unified system, we have $\gamma y < \bar{\alpha}$.

Proof: Suppose that, at the optimal contract, $\gamma y = \bar{\alpha}$ holds. We will treat γ as a parameter, consider the problem of choosing the allocation that maximizes W subject to (5) and (6), and show that $\partial W/\partial \gamma$, evaluated at $\gamma = \bar{\alpha}/y$, is negative. From the envelope theorem, we will conclude that the optimal γ is strictly less than $\bar{\alpha}/y$.

Given $\gamma = \bar{\alpha}/y$, the functions $\{c_I^2(\alpha), c_P^2(\alpha)\}$ that maximize W subject only to resource constraint (6) entail full consumption smoothing. That is, we have

$$c_I^2(\alpha) = c_P^2(\alpha) - 1 \quad \text{for all } \alpha. \quad (8)$$

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

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

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. \square

The proof of Theorem (3.1) suggests that we can approximate the optimal contract by one that 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. It would be fully optimal as well, if not for the fact that some impatient consumers are rationed in period 1 when the realization of α is close to $\bar{\alpha}$. A planner who could observe consumers' types would give these rationed consumers $c_I^2(\alpha)$, but they instead receive $c_P^2(\alpha)$, which is one unit too much. Thus, for those few states in which rationing occurs, a few impatient consumers would have marginal utility below that of the other consumers. A slight increase in $c_I^2(\alpha)$ and reduction in $c_P^2(\alpha)$, in states where rationing occurs, would yield higher welfare. However, we show below in example (3.2) that the dropoff in welfare is negligible in practice.¹⁸

¹⁸Calculating the exact optimal contract is difficult, because the small adjustments described here cause

In the next example, we specify the following basic parameters: the utility function $u(c)$, the utility from satisfying the urgent opportunity \bar{u} , the interest factor on the illiquid asset R_A , and $f(\alpha)$. We then compute (or at least approximate) the function γ invested in the liquid asset and welfare W as functions of the interest factor R_B on the liquid technology and the transaction cost of financial unification λ .

Example 3.2: Consider the following parameter specification:

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

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

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 *ad hoc* cost of integration λ , and the interest factor R_B on the liquid technology. We do not compute directly the solution to planning problem (7), but we compute instead tight upper and lower bounds on the optimal solution. As a lower bound, we compute the values of γ and W for the optimal contract involving full consumption-smoothing, (6) and (8), which is feasible and incentive compatible. As an upper bound, we compute the optimal contract when the planner has full information about each consumer's type. The only distinction between the two contracts is that the "full consumption smoothing" contract is unable to distinguish between patient consumers and impatient consumers who were unable to receive consumption

incentive compatibility to be (barely) violated. Further changes in the allocation and possibly γ are required to reestablish incentive compatibility. We conjecture that the optimal contract will actually have a run equilibrium, since incentive compatibility when the patient wait is probably less restrictive than when there is a run. We do not explore these issues here, because the unified system is nearly full-information efficient (except for λ), and runs can be avoided at negligible resource cost.

in period 1. Since the probability of being an impatient consumer who is rationed is less than 0.002, and the imbalance in marginal utilities is small even in these circumstances, the two contracts yield nearly identical values for γ and W .

Our results are summarized in Table 1.¹⁹

	“Cons. Smoothing” lower bound		“Full Information” upper bound	
	γ	W	γ	W
$R_B = 1.05$ $\lambda = 0$	0.04747180271	0.8891732	0.04734903564	0.8893609
$R_B = 1.05$ $\lambda = 0.01$	0.04746679331	0.7886871	0.04734341373	0.7890072
$R_B = 1.05$ $\lambda = 0.02$	0.04746176340	0.6881982	0.04733776756	0.6885445
$R_B = 1.08$ $\lambda = 0$	0.04895748003	0.9590241	0.04890419892	0.9590912
$R_B = 1.08$ $\lambda = 0.01$	0.04895534931	0.8587239	0.04890179176	0.8587233
$R_B = 1.08$ $\lambda = 0.02$	0.04895320845	0.7582438	0.04889937484	0.7583353

Table 1: **The Liquidity Proportion γ and Welfare W in the Unified System**

Table 1 indicates that the difference in welfare between the “full-consumption-smoothing” lower bound and the “full-information” upper bound is several orders of magnitude less important than the effects of modest changes in R_B or λ . The tightness of these welfare bounds will allow us, in section 4, to compare the optimal contract for the unified system with the optimal contract for the separated system.

The relatively small change in λ from 0.01 to 0.02 has a large impact on welfare relative to its impact on γ . This makes sense, because λ represents a fixed resource cost, which directly reduces welfare by approximately (0.01) multiplied by marginal utility evaluated

¹⁹Computations were made using Maple version 5 running on Windows 95. The code is available from the authors for the purposes of replicating the results.

at consumption of 10, which is $(0.01) \times 100/10 = 0.1$. The impact on γ is a second-order effect. Second-period consumption is reduced when λ increases, so marginal utility is slightly higher. Thus, the tradeoff between the higher yields in technology A versus the higher risk of rationing impatient consumers in period 1 tilts slightly towards more investment in technology A and lower γ . Notice also that 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. The optimal γ is less than 0.05, which follows from Theorem (3.1) and the parameter specification, $\bar{\alpha}/y = 0.5/10 = 0.05$. The fact that γ is close to .05 indicates that the probability of rationing is small at the optimal contract. The intuition is that, at the margin, the loss of \bar{u} multiplied by the probability of rationing is traded off against the difference in yields, $R_A - R_B$. Whenever the consumption opportunity is important relative to $R_A - R_B$ rationing will be unlikely.

4 The Separated System

Here we assume that the bank cannot gain access to the funds invested in technology A . One motivation for this restriction is that the bank chooses not to incur transactions costs, such as the costs associated with writing subordinated debt contracts with other financial institutions. Another type of transactions cost lies in committing not to cheat depositors, i.e., preventing false claims by the bank of a large number of early withdrawals that would necessitate small second-period withdrawals. (As Example 4.4 below demonstrates, the real return on deposits is likely to be fairly predictable in the separated system, so the cost of committing not to cheat might be smaller.) Under this interpretation, the optimal system depends on the transactions cost λ .

Another motivation for the analysis of the separated system is the presence of legal restrictions on the activities of banks. For example, we can interpret the Glass-Steagall Act as

preventing banks from directly investing in technology A , and circumventing these restrictions might be costly. Under this interpretation, removing these statutory and regulatory restrictions might result in the emergence of the unified system. A comparison of the separated system to the unified system with $\lambda = 0$ might indicate how costly the restrictions are, and how banking institutions might evolve if we eliminated the restrictions.

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. However, 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 possible. 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, so full consumption smoothing may be impossible. “Excessive” investment in technology B may be necessary in order to satisfy incentive compatibility, so the optimal contract may require $\gamma y > \bar{\alpha}$. Finally, the optimal contract may be subject to bank runs, as long as the runs are sufficiently infrequent.

In the separated system, ex-ante welfare, the incentive compatibility constraint, and the resource constraint are as given in expressions (4), (5), and (6), where we set $\lambda = 0$. The restriction that the bank cannot gain access to investments of technology A is expressed simply as follows:

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

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

following planner's problem:

$$\begin{aligned} \max_{\gamma, c_T^2, c_P^2} W & \quad (11) \\ \text{subject to (5), (6), and (10).} \end{aligned}$$

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.

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

Proof: Suppose instead that $(c_P^2(\bar{\alpha}))^* \geq 1$ holds at the optimal contract, m^* , and let γ^* be the optimal value of γ . Since resources remain in period 2, it follows that $\bar{\alpha} \leq \gamma^* y$. Treating γ as a parameter, let $W(\gamma)$ denote welfare, as a function of γ , which solves:

$$\begin{aligned} \max_{c_T^2, c_P^2} W & \quad (12) \\ \text{subject to (5), (6), and (10).} \end{aligned}$$

The optimal consumption allocation in m^* must solve (12) for $\gamma = \gamma^*$. Given γ^* , and for $\alpha \leq \bar{\alpha}$, the solution to (12) must offer full consumption-smoothing. That is, consumption in state α is uniquely determined by (6) and

$$(c_P^2(\alpha))^* = 1 + (c_T^2(\alpha))^* \quad (13)$$

This is because, given γ^* , choosing consumption according to (6) and (13) maximizes welfare subject only to the resource constraint. However, inequality (5) is satisfied because patient consumers are indifferent between arriving in period 1 and period 2, and the inequalities in (10) are satisfied because $(c_P^2(\alpha))^* \geq 1$, so we know that $(c_I^2(\alpha))^* \geq 0$ for all α . Therefore, given γ^* , choosing consumption according to (6) and (13) is optimal for the more tightly constrained problem (12). It also follows that the Lagrangean multipliers on constraints (5) and (10) are zero. Using (6) and (13) and applying the envelope theorem, one can show:

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

However, it follows that reducing γ improves welfare, contradicting the fact that γ^* is part of the optimal contract, m^* . \square

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 the optimal contract in the separated system always admits a run equilibrium.

Theorem 4.2: The optimal banking contract in the separated financial system has a run equilibrium.

Proof: We know that the optimal contract satisfies $c^1(z) = 1$ for all $z \leq \bar{\alpha}$. Since we are considering the possibility of 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

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

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 optimal contract. 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.²⁰ 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 $c_I^2(z) = 0$ for $z > \bar{\alpha}$, because giving period-2 consumption to those who withdraw in period 1 only increases the incentive to run. Thus, for $z > \bar{\alpha}$, second period consumption is given by

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

Differentiating with respect to z in (15) yields

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

which is negative for $z < z^*$, since inequality (14) does not hold. Thus $c_P^2(z)$ is decreasing in z for $z < z^*$.

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 $c_P^2(z)$ is continuous and decreasing in z for $z < z^*$. Therefore, it is a best response for consumer j to accept $c^1(z_j)$. □

²⁰Remember, the optimal contract is the one that provides the highest welfare in the equilibrium in which the patient consumers wait until period 2.

There is an intuitive explanation for why run equilibria always exist in the separated system. In the optimal banking contract for the separated financial system, c_P^2 must be less than 1, or 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 sufficient condition for the optimal liquid asset investment under the separated financial system to be greater than it is under the unified financial system.

Theorem 4.3 (Overinvestment in the Liquid Asset): If $\bar{\alpha} < 1/R_B$ holds, then the optimal banking contract for the separated financial system does not ration consumers in period 1, and invests more in technology B than the optimal banking contract for the unified financial system.

Proof: From Theorem (3.1), the optimal contract in the unified system satisfies $\gamma < \bar{\alpha}/y$. In the separated system, the optimal contract 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, $\gamma^* \geq \bar{\alpha}/y$ holds, so consumers are not rationed in period 1. Second, γ^* exceeds the optimal fraction invested in technology B for the unified system. \square

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, while in the separated system, the fraction of resources invested in technology B is more than 0.09.

Next we compute the optimal contract in the separate system.

Example 4.4: We use the same parameter values specified in Example (3.2). We also specify: $\lambda = 0$ and $R_B = 1.08$.

It can be shown that the optimal solution to (12) satisfies $c_I^2(\alpha_1) = 0$ for all α_1 . The intuition is that giving consumption to impatient consumers in period 2 hurts incentive compatibility and forces more overinvestment in technology B , outweighing any consumption-smoothing advantages.²¹ Since Theorem (4.3) applies, it follows from (6) that we have

$$c_P^2(\alpha) = \frac{(10\gamma - \alpha)1.08}{1 - \alpha}. \quad (16)$$

Finding the optimal contract 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: $\gamma^* = 0.094445331$ and $W^* = 0.8687644$.

Now compare the optimal contract in the unified system to the optimal contract in the separated system, for parameters given by (9) and $R_B = 1.08$. When we have $\lambda = 0$ in both

²¹This result is derived by first obtaining a lower bound for the multiplier on the incentive compatibility constraint, which allows us to obtain a lower bound for the multiplier on each state- α non-negativity constraint. Since these multipliers are positive, the constraints must bind, so the impatient receive no consumption from the bank in period 2. (Details are available from the authors.) If the parameters of the example are changed, it is possible that some of the non-negativity constraints do not bind.

cases, welfare is clearly higher in the unified system, even using the pessimistic lower bound from Table 1. This is because problem (12) involves maximizing over a strictly smaller set than problem (7), due to the additional non-negativity constraints. Technology B investment in the separated system is nearly double that in the unified system. If $\lambda = 0$ in the separated system, but the transactions cost of achieving the unified system requires $\lambda = 0.01$ or higher, then welfare is higher in the separated system, even using the optimistic upper bound for welfare in the unified system. Again, technology B investment in the separate system is nearly double that in the unified system.

Another distinction between the two systems is that the real interest rate that consumers receive *from technology B* investments is much more volatile in the unified system. In the unified system, impatient consumers nearly all receive 1 unit in period 1, but invest only about 0.489 units, for a real interest rate of over 100%. Patient consumers receive a real interest rate of 8% when $\alpha = 0$, but they receive a real interest rate of *negative* 100% when $\alpha = 0.5$. Of course, the overall return also reflects the fact that the patient consumer receives approximately 1 unit more than the impatient consumer from technology A investments. In the separated system, the impatient consumer receives a real interest rate of approximately 5.88%. The patient consumer receives a real interest rate ranging from 1.6%, when $\alpha = 0.5$, to 8% when $\alpha = 0$. □

Several conclusions can be drawn from our example, depending on how the *ad hoc* transactions cost parameter λ is to be interpreted. First, suppose that transactions costs are not significantly different in the two systems, so that $\lambda = 0$ holds for each system. Then efficiency can be enhanced by eliminating legal restrictions on banks' portfolios, such as the restrictions imposed by the Glass-Steagall Act. Moreover, our analysis suggests that in this case eliminating regulatory and other government restrictions would cause a major shift from liquid, lower-yield investments to illiquid, higher-yield investments. This shift would

necessitate occasional scarcities of liquidity (non-run rationing of period 1 consumption in the model). Perhaps surprisingly, allowing banks to freely invest in all technologies would reduce the likelihood of speculative bank runs, thereby making the system less fragile.²²

Second, suppose that “real” (i.e. nongovernment) transactions costs are significantly greater in the unified system than in the separated system.²³ Then the separated system is the optimal financial system. Furthermore, the optimal system is fragile, in the sense that sufficiently infrequent bank runs are preferable to investing so much in technology B that run equilibria can be eliminated.

5 Sunspots and Susceptibility to Runs

Postlewaite and Vives (1987) argue that, strictly speaking, run equilibria in DD are not equilibria at all, because consumers would not agree to the original contract if they knew that a run would take place. DD suggest that a run could take place in equilibrium with positive probability, triggered by some extrinsic random variable “sunspots,” as long as the probability of the run is sufficiently small. Here we formalize this notion, and calculate what “sufficiently small” is for an example.²⁴

At the beginning of period 1, each consumer learns her type and observes a sunspot variable, s , distributed uniformly on $[0,1]$.²⁵ Sunspots do not affect preferences, the likelihood of being impatient, endowments, or technology. Now the period in which a consumer arrives can depend on the realization of the sunspot variable s as well as the realization of her

²²A major caveat is that our analysis does not include deposit insurance or explicit moral hazard issues, nor does it include credit chains as analyzed by Kiyotaki and Moore (1997).

²³Setting $\lambda = 0.01$ corresponds roughly to losing 2% of the resources invested in technology B . The reader may decide whether this is a large or a small number.

²⁴Cooper and Ross (1998), without performing calculations, also model runs being triggered by sunspots.

²⁵The uniformity assumption is without loss of generality.

type. The space of mechanisms is unchanged; the solution concept is perfect Bayesian equilibrium. We have a run equilibrium if, for some set of realizations of s occurring with positive probability, all consumers arrive at the bank in period 1 and a positive measure of the patient consumers withdraw in period 1.

To fix ideas, suppose that the economy is susceptible to runs, in the following sense. Whenever we have $s < s_0$, then all consumers choose to arrive at the bank in period 1, and accept consumption offers of at least one unit, whenever such a run is consistent with equilibrium. When we have $s \geq s_0$, the equilibrium is selected in which all patient consumers wait for the second period. Of course, incentive compatibility must be satisfied. We will calculate the cutoff value of s_0 below which the optimal contract in the separated system has a run equilibrium.²⁶ We work in the context of our example, defined by the parameters in (9), $\lambda = 0$, and $R_B = 1.08$.

There are two approaches to eliminating runs within the separated financial system. One approach, which turns out to be the best here, is to continue to provide 1 unit of consumption in period 1 when available, and increase γ to the point at which $c_P^2(0.5) = 1$. From 16, we can calculate the optimal γ to avoid runs: $\gamma^{**} = 0.0962962963$. It is now feasible and welfare-maximizing to offer full consumption smoothing, as in (13).²⁷ Welfare in this contract that eliminates runs, W^{norun} , can now be calculated: $W^{norun} = 0.86506$. The other approach to eliminating runs is to ration consumers in period 1, thereby retaining enough consumption to guarantee that $c_P^2(0.5) = 1$. This approach is too costly here. However, if we change the

²⁶Since run – if they exist – equilibria could be eliminated in the unified system at negligible cost, even if such run equilibria exist, we focus here on the separated system. For this section, suppose that either the real transactions cost of the unified system is high, or that legal restrictions on banks' portfolio choices prohibit the unified system.

²⁷With full consumption smoothing, patient consumers are indifferent as to which period to choose. To eliminate run equilibria, we must give consumers slightly more consumption in period 2 in some states, making it a dominant strategy for patient consumers to wait. Therefore, we can eliminate runs with contracts arbitrarily close to the one considered here.

density function, f , so that high realizations of α are sufficiently unlikely, then the best way to avoid runs is to make the impatient consumers suffer in the extremely unlikely event of large α .²⁸

We now determine the optimal contract that has a run equilibrium, given that a run occurs with probability s_0 and that the equilibrium in which patient consumers wait occurs with probability $(1 - s_0)$. Patient consumers must be induced to wait when $s \geq s_0$, and there is no attempt to provide incentives when $s < s_0$, so the incentive compatibility constraint is unchanged. It can be shown that the incentive compatibility constraint binds, so investment in technology B is unchanged, $\gamma = \gamma^* = 0.094445331$. In other words, the probability of a run has no effect whatsoever on the optimal contract, as long as the probability is small enough so that the optimal contract tolerates runs. When we have $s < s_0$, there is no reason to punish consumers for running. It turns out that everyone will receive 1 unit of consumption in period 1, until the bank runs out. The gain from providing for the consumption opportunities of a significant minority of impatient consumers outweighs the foregone yield of R_B . Thus, the probability of receiving 1 unit in period 1 is γy , which equals 0.94445331, so the proportion of consumers taking advantage of their consumption opportunity in state α is $(1 - \alpha + \alpha\gamma y)$. Conditional on a run taking place, welfare is calculated to be: $W^{run} = 0.1502908$. Overall welfare is a weighted average of W^{run} and the welfare when a run does not take place, $W^* = 0.8687644$. Thus, the cutoff value of s_0 solves

$$s_0 W^{run} + (1 - s_0) W^* = W^{norun}. \quad (17)$$

From (17), we calculate the cutoff value of s_0 to be 0.0051557921. If the susceptibility to runs is less than 0.0051557921, then it is better to tolerate the unlikely event of a run than to

²⁸Even when the density is such that large realizations of α are extremely unlikely, it is still quite possible for the optimal contract, without fear of runs, not to ration consumers in period 1, an assumption we have maintained throughout. It might be worthwhile to invest enough in technology B so that patient consumers receive positive consumption in period 2, so there is no rationing, while not worthwhile to invest enough for them to receive 1 unit.

increase technology B investment to prevent runs.²⁹ In other words, the cost of eliminating this fragility from the financial system is too high, in terms of efficiency when the system is working smoothly. 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.

6 Concluding Remarks

We believe that the indivisibility in our model that is associated with the transactions and payments for urgent “consumption opportunities” is well-motivated, and it facilitates calculation of the optimal contract, but we conjecture that the indivisibility is not crucial for our basic results.³⁰ For the discussion below, we drop the indivisibility assumption—supposing instead that the impatient consumers have smooth utility functions over period-1 *and* period-2 consumption, and the patient consumers have smooth utility functions over period-2 consumption.

Under the unified system, all of the technology B investments will have to be liquidated in state $\bar{\alpha}$, or else higher welfare can be achieved by investing more in technology A , without affecting the consumption of the impatient. In the special case of deterministic α , technology B investments will be used exclusively for impatient consumers and will be fully liquidated. When α is random, rather than rationing the impatient consumers, the optimal contract will likely have period-1 consumption tapering off as more consumers arrive, exhibiting the “partial suspension of convertibility” obtained by Wallace (1988, 1990) and Green and Lin (1996). A reasonable conjecture would be that run equilibria are impossible under the

²⁹For the same parameters, the probability of a run on the unified system must be less than 6.89×10^{-5} for it to be an equilibrium to the pre-deposit game if indeed an equilibrium run is possible at all.

³⁰Although the indivisibility does leave open the possibility of runs in the unified bank but only with tiny probability.

optimal contract.

Under the separated system, the typical case will be overinvestment in technology B . For example, if α is deterministic, then the unified system harvests all of the technology B investments in period 1, leaving nothing for the patient consumers.³¹ To satisfy incentive compatibility, the optimal contract in the separated system will invest more in technology B and/or reduce consumption in period 1. When α is random, but not too volatile, we anticipate the same conclusion.

At the risk of being foolish, we conjecture that run equilibria are possible at the optimal contract in the separated system. Our basic intuition remains: that investment in technology B must satisfy incentive compatibility for each patient consumer, based on the other patient consumers waiting. A patient consumer is indifferent between choosing to arrive at the bank during period 1 and choosing to arrive during period 2, weighing all possible values of α . State $\bar{\alpha}$ is likely to be one of the states in which a patient consumer is better off choosing period 1, because this is the state in which the most investments are harvested in period 1 and the fewest remain for period 2. If so, then there is a run equilibrium.

What are the differences between our model and the closely related narrow banking model of Wallace (1996)? Wallace's technologies are slightly different from ours, allowing some liquidation value of technology A in period 1, and having technology B 's return between period 0 and 1 equal the return between period 1 and 2. These differences are probably not crucial. Wallace's utility functions do not exhibit any indivisibilities (such as our urgent consumption opportunities), but they are otherwise more general than ours. We have argued that the indivisibility is probably not crucial for most of our basic results. The extra structure we impose on utility functions allows us to maintain tractability while generalizing Wallace

³¹As in our setup, the patient consumers are compensated by a bigger share of the technology A investments in the unified system. Notice that it is difficult even to talk about the separated system without the impatient caring about future consumption. If the impatient were finished with consumption in period 1, what would happen to their investments in technology A ?

(1996) in other directions. In particular, we allow α to be stochastic and only impose the condition that the distribution be continuous, while Wallace requires α to be deterministic. We believe that the introduction of stochastic aggregate fundamentals, in our case the random α , dramatically changes the problem.

Wallace's working definition of narrow banking is that all of the bank's obligations to depositors must be met for *every* possible pattern of withdrawals. 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. On the other hand, Wallace's banks are permitted to invest in both technologies. Wallace shows that, without subordinated debt contracts, any allocation achievable with narrow banking is also achievable with autarky, so that narrow banking eliminates the role for banks. On the other hand, our notion of the separated financial system is closer to the description of narrow banking in the introduction to Wallace (1996), where banks are (only) restricted to holding liquid (short-term) assets. We allow deposit contracts to specify withdrawals that are fully contingent on the history, but restrict the banks' portfolios to technology B assets. Because our portfolio restrictions are different from Wallace's restrictions on the banking contract, banks play a very important role in our separated financial system.

In our model, as in Wallace (1996), the banking restrictions under consideration cannot improve welfare and can cause harm. If the only transactions cost preventing the unified system is due to government restrictions on banks' portfolios, then removing the restrictions will improve welfare. On the other hand, if there are other significant transactions costs which inhibit the necessary linkages implicit in the unified system, then removing the restrictions will not hurt, but it will not lead to change in the financial system. We find that the optimal contract in the separated financial system exhibits overinvestment in the liquid technology, and is subject to run equilibria. Considering that many have argued that Glass-Steagall type restrictions on bank portfolios are needed in order to ensure stability, our result may

come as a surprise. Not only are bank runs possible in the separated financial system, they are far more likely with portfolio restrictions than without the restrictions.³²

The sole intrinsic uncertainty in our formal model comes from the urgent transaction opportunities of consumers. We expect quite similar results in the case where the uncertainty is about investment returns, R_A and R_B , rather than about the aggregate importance parameter α .

References

Allen, Franklin and Gale, Douglas. "Financial Contagion." Mimeo., C.V. Starr Center at NYU and the Wharton Financial Institution Center at the University of Pennsylvania, October 1998.

Bryant, John. "A Model of Reserves, Bank Runs, and Deposit Insurance." *Journal of Banking and Finance*, 43 (1980): 749-61.

Calomiris, Charles W. and Kahn, Charles M. "The Role of Demandable Debt in Structuring Optimal Banking Arrangements." *American Economic Review*, 81(3) (1991): 497-513.

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

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

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

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

³²On the other hand, the unified system is subject to non-run rationing in period 1 if the realization of α is sufficiently large. Convincing people to choose the no-run equilibrium is a good thing, but totally avoiding non-run rationing in the unified system reduces welfare.

Diamond, Douglas W. and R.G. Rajan, "Liquidity Risk, liquidity creation, and financial fragility: A theory of banking," University of Chicago, revised July 6, 1998.

Freixas, Xavier and Rochet, Jean-Charles. *Microeconomics of Banking*, Cambridge, MA: MIT Press, 1998.

Green, Edward J. and Lin, Ping. "Implementing Efficient Allocations in a Model of Financial Intermediation." Mimeo., Federal Reserve Bank of Minneapolis and Southern Methodist University, October 1996.

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

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

Lagunoff, Roger and Schreft, Stacey. "A Model of Financial Fragility." Mimeo., Georgetown University and Federal Reserve Bank of Kansas, September 1998.

McCulloch, J. Huston and Yu, Tin-Teh. "Government Deposit Insurance and the Diamond-Dybvig Model." *The Geneva Papers on Risk and Insurance*, 23 (1998): 139-149.

Peck, James. "Demand Uncertainty, Incomplete Markets, and the Optimality of Rationing." *Journal of Economic Theory*, 70(2) (1996): 342-63.

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

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.

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.

Wallace, Neil. "Narrow Banking Meets the Diamond-Dybvig Model." *Quarterly Review*

of the Federal Reserve Bank of Minneapolis, 20(1) (1996): 3-13.