# Bank panics in scale economies*

David Andolfatto
*Federal Reserve Bank of St. Louis*
*Simon Fraser University*

Ed Nosal
*Federal Reserve Bank of Chicago*

September 2017

**Abstract**

Sequential service plays an important role in the explanation for bank panics in Diamond and Dybvig (1983). We question the empirical relevance of sequential service and ask whether panics are possible absent this constraint. We replace sequential service in the Peck and Shell (2003) model with a fixed cost of investment and demonstrate that panics are possible in a large region of the parameter space. A scale economy is not sufficient in itself to generate multiple equilibria in the withdrawal game. Aggregate liquidity risk, private information over liquidity preference, the communications protocol, and the scale economy all interact to produce the result.

*We thank conference participants at the 2nd Annual Missouri Macro Workshop, the 2017 Summer Workshop on Money, Banking, Payments, and Finance, at the Bank of Canada, as well as seminar participants at the Federal Reserve Banks of Chicago and St. Louis, the National University of Singapore, Arizona State University and Simon Fraser University. We owe a special thanks to Todd Keister whose comments on a prior draft led to a number of substantial improvements. The views expressed here are our own and should not be attributed to the Federal Reserve Banks of Chicago and St. Louis, or the Federal Reserve System. JEL Codes: G01, G21, G28

# 1 Introduction

At its core, banking is centered on the business of liquidity transformation — the use of demandable debt to finance non-marketable, but otherwise safe, higher-return investments. The success of this business model depends on many factors, not the least of which is widespread confidence in the ability of banks to meet their short-term obligations. The use of demandable debt to fund illiquid assets, however, allegedly injects a sense of fragility to the banking system. In particular, even if bank assets are fundamentally sound, a loss of depositor confidence for any reason whatsoever can trigger a wave of redemptions that ultimately erodes the capacity of banks to achieve their promised returns. In this way, a purely psychological fear of mass redemptions can manifest itself as a self-fulfilling prophecy, a so-called *bank panic*.

The banking model developed by Diamond and Dybvig (1983) is the first attempt to formalize the notion of bank panics in the context of a model where liquidity transformation is necessary for efficient risk-sharing. They emphasize two prerequisite conditions for the existence of bank panics. First, information relating to individual liquidity preference must be private information, since misrepresentation of liquidity preference—a defining characteristic of panic—is otherwise not possible.[1] Second, redemption requests by depositors must be executed on a first-come, first-served basis without full knowledge of the period's aggregate withdrawal demand since uncertainty over redemption payouts is needed to produce a panic. While Diamond and Dybvig (1983) do not formally establish how private information and *sequential service* result in panics under optimal contractual arrangements, their basic intuition was ultimately verified by Peck and Shell (2003).

Although sequential service is necessary for the existence of bank panics in the Peck and Shell (2003) model, Andolfatto, Nosal and Sultanum (2016) show how their fragility result vanishes under a broader class of mechanisms. Furthermore, Green and Lin (2003) demonstrate that sequential service need not induce fragility if depositors can roughly calculate their queue position before making their withdrawal decision.[2] On the basis of these latter results, it appears that sequential service is not sufficient to induce fragility. But is

[1] Moreover, as Bryant (1980) points out, private information along this dimension provides a natural way to motivate the use of demandable debt.

[2] In particular, depositors communicating their withdrawal request later in the period know that they are likely to be near the end of the sequential service queue.

it necessary? This is an important question because, in our view at least, it seems likely that sequential service can potentially be bypassed if it poses a sufficiently important threat to bank sector stability. The next paragraph explains our reasoning.

While sequential service is usually modeled as a constraint, it is perhaps better thought of as a mechanism design issue that is tailored to accommodate some underlying friction in the environment. Sequential service evidently amounts to the notion that at least some important investment or expenditure opportunities have an ephemeral quality about them, making an immediate "on demand" funding facility highly desirable, or even necessary. The empirical relevance of this property depends in part on what one takes to be the appropriate length of a period in the Diamond-Dybvig (1983) model. If it is the length of an agricultural growing season, then a farmer needing funds early in the season may well face disaster if he had to wait until the end of the season to access them. On the other hand, if the length of period is in the order of days, the cost of replacing demand debt with notice debt—or instituting a short banking holiday—may present little more than a minor inconvenience.[3] Moreover, the benefit of doing so may be significant since it grants a bank time to better assess and deal with the true state of liquidity demand.

Whether the properties of an environment give rise to sequential service is an empirical matter. However, we do not believe that the possibility panics vanish if the sequential service constraint is eliminated. To be clear, it seems apparent enough that *some* restriction on communications is needed, if for no other reason than to provide a rationale for banking. After all, if individuals could costlessly remain in contact with each other at all times, liquidity transformation could take place in a financial market after liquidity preferences are realized (Jacklin, 1987). There must be something to prevent financial markets from displacing banks as liquidity providers.[4] It is by no means obvious, however, that the relevant property must manifest itself as a sequential service constraint—at least, in the way this constraint is usually modeled (Wallace, 1988).

To investigate if sequential service is necessary for the existence of bank

[3]An example of transactional notice debt in the United States is the negotiable order of withdrawal (NOW) account. NOW accounts technically required seven days notice, but they were used as everyday checking accounts.

[4]See also Andolfatto, Berentsen and Martin (2017) and the references cited therein.

panics, we reconsider the Peck and Shell (2003) model but without a sequential service constraint. This implies that the bank can wait to make a full assessment of the period's liquidity demand before releasing funds. Absent any other modification to the environment, it is well-known that the efficient allocation is panic-free. Our innovation is to introduce a fixed cost to investment activity, so that a scale economy is present. In particular, we assume that a relatively high risk-adjusted rate of return on investment is attainable only if the investment is funded beyond some minimum threshold, which we take as a parameter, possibly varying across different asset classes. If funding for investment falls below this minimum threshold, risk-adjusted returns will be (modestly) negative. In Peck and Shell (2003) and its antecedents, this minimum threshold is taken to be zero, so that the investment technology is always linear. Our main result shows how private information, aggregate liquidity risk, communication protocols, and increasing returns to scale all conspire to render banks fragile, even in the absence of sequential service.

The efficient bank contract in our model induces a coordination game with multiple equilibria, just as in Peck and Shell (2003). The difference in their approach relative to ours is in what motivates *patient* depositors—those with no pressing liquidity needs—to misrepresent themselves as being *impatient*—those with pressing liquidity needs. In Peck and Shell (2003) sequential service implies that it optimal deposit contract provides higher returns to early arriving depositors that request withdrawals and lower returns—which will be negative when redemption demand is large—for those who arrive later. As a result, if redemptions turn out to be very heavy, there will not be much left over for those patient depositors leaving their funds in the bank. The form of the deposit contract suggests that the expectation of a heavy redemption event can become a self-fulfilling prophecy. Intuitively, a patient depositor may prefer to masquerade as an impatient depositor in hopes of being early in the service queue and receiving a high(er) payout. In our model, if the probability of a heavy redemption state is remote, then the amount of available for investment will almost surely exceed the minimum threshold needed to achieve the high risk-adjusted returns. If, however, redemptions turn out to be very heavy, resources will fall short of the minimum threshold, resulting in depositor losses on any funds left in the bank. Here, the expectation of a heavy redemption event can become a self-fulfilling prophecy because the minimum threshold funding will is not expected to be met, which implies that the investment will generate negative returns. In our environment the scale economy assumption replaces sequential service as the mechanism that generates the instability.

The paper is organized as follows. Section 2 describes the economic environment. In Section 3, we characterize the set of efficient incentive-compatible allocations for economies subject to private information and potential scale economies in investment opportunities. We establish the existence of panic equilibria in Section 4. Section 5 considers an application of our model to the question of narrow and shadow banking. The extension provides some support for the notion that low real rates of return on safe asset classes can promote financial instability through a reach-for-yield behavior. Section 6 compares our model environment to others in the literature and highlights some policy insights implied by our model. Finally, we provide a discussion of pertinent issues in Section 7.

## 2 The model

Our model is a variant of Peck and Shell (2003). The economy has three dates, $t = 0, 1, 2$, and a finite number $N \geq 3$ of *ex ante* (date 0) identical individuals. Individuals are subject to a preference shock at date $t = 1$ that determines their type: *impatient* or *patient*. Let $0 < \pi < 1$ denote the probability that an individual is impatient. Let $\pi_n$ denote the probability that $0 \leq n \leq N$ individuals are impatient. We assume that individual types are *i.i.d.* so that $\pi_n = \binom{N}{n}\pi^n(1-\pi)^{N-n}$. The distribution of types has full support, $0 < \pi_n < 1$ for all $n$.

Impatient individuals want to consume at date 1 only. Patient individuals are willing to defer consumption to date 2; technically, they are indifferent between consuming at dates 1 and 2. Let $c_t$ represent the consumption of an individual at date $t$. Date 0 preferences are given by

$$U(c_1, c_2) = \pi u(c_1) + (1-\pi)u(c_1 + c_2), \tag{1}$$

where $u(c) = c^{1-\sigma}/(1-\sigma)$ and $\sigma > 1$. Note that the state in which all $N$ agents (strictly) prefer to consume at date 1 occurs with positive probability, $\pi_N > 0$.

Each individual is endowed with $y$ units of date 1 output. There exists a technology that transforms $k$ units of date 1 output into $F_\kappa(k)$ units of date

2 output according to

$$F_\kappa(k) = \begin{cases} rk & \text{if} \quad k < \kappa \\ Rk & \text{if} \quad k \geq \kappa \end{cases}, \tag{2}$$

where $0 < r < 1 < R$ and $0 \leq \kappa < Ny$. The high rate of return $R$ is available only if the level of investment exceeds a minimum scale requirement of $\kappa$.[5] When the minimum scale $\kappa$ is not met, the rate of return reflects the cost of intermediated storage, indexed by the parameter $1 - r$. Technology (2) is a generalization of the standard specification used in the literature, which assumes $\kappa = 0$, in which case $F_0(k) = Rk$ for all $k > 0$.

There are two benefits to cooperation in this economy. First, there are the usual gains associated with sharing risk. Second, and absent from the standard model, scale economies are more easily attained when resources are pooled.

In what follows, we refer to a risk-sharing arrangement that pools resources and exploits scale economies as a *bank.* Individuals who deposit resources with the bank are referred to as *depositors.* A bank can be viewed as a resource-allocation mechanism that pools the resources of the $N$ depositors before they learn their types. In exchange for deposits, the bank issues state and time-contingent deposit liabilities. What makes our environment a bank problem, instead of a standard insurance problem, is that the bank cannot verify depositor types. Specifically, we assume that the depositor's type—his liquidity preference—is revealed only to the depositor. Because liquidity preference is private information, the optimal risk-sharing arrangement will include options to withdraw funds on demand. It is in this sense that the optimal contract resembles conventional demand deposit liabilities (Bryant, 1980).

The timing of events and communication protocols are as follows. At date 0, depositors decide whether to participate in a risk-sharing arrangement or not. Participation entails depositing their $y$ endowment of claims to date 1 goods and agreeing to the terms of the contract governing the returns on future redemptions. After the participation decision, depositors disperse to remote locations, as in Townsend (1987). Although depositors remain

[5] One could easily generalize the analysis to permit multiple threshold levels and associated rates of return. For most of our analysis, we stick to one threshold level for the sake of demonstrating how our mechanism works in the simplest manner possible.

largely incommunicado with each other, they are able to communicate with the bank at date 1 and date 2, subject to some restrictions. In particular, we follow Peck and Shell (2003) and assume that it is prohibitively costly for depositors to contact the bank more than once after their initial deposit at date 0.[6] Hence, depositors can contact the bank either at date 1 or at date 2, but not at both dates.[7] Finally, in contrast to Peck and Shell (2003), we assume that date $t = 1, 2$ withdrawal requests are processed only after all date $t = 1, 2$ requests have been made, i.e., we do not impose a sequential service constraint.

Our communications protocol implies that the date 1 consumption payments specified in the bank contract need only be conditioned on the number of depositors $m$ who contact the bank at date 1, where $m \in \{0, 1, ..., N\}$. In particular, if $m$ depositors contact the bank at date 1, then each depositor receives $c_1(m)$ units of date 1 consumption. Depositors who contact the bank at date 2 each receive $c_2(m) = F_\kappa[Ny - mc_1(m)]/(N - m)$ units of date 2 consumption. Hence, the bank offers depositors a contract in the form of a promised allocation $(\mathbf{c}_1, \mathbf{c}_2)$, where $\mathbf{c}_1 = [c_1(1), \ldots, c_1(N)]$ and $\mathbf{c}_2 = [c_2(0), c_2(1), \ldots, c_2(N-1)]$.

By construction, the allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is feasible. Because liquidity preferences are private information, depositors may misrepresent themselves to the bank. To ensure that $(\mathbf{c}_1, \mathbf{c}_2)$ promotes efficient resource allocation, the allocation should be structured in a manner that gives depositors an incentive to represent their preferences truthfully. We now describe the strategic interaction among agents as a *withdrawal game* that depositors play.

Suppose that all $N$ individuals deposit their claims at the bank at date 0.[8] In between dates 0 and 1, depositors learn their types and a withdrawal

[6] A less restrictive communications protocol would allow depositors to contact the bank at higher frequency and, in particular, even in periods when they do not want to withdraw funds. Peck and Shell (2003) examine this less restrictive protocol in their appendix and show that their fragility result is unaffected. The communcations protocol plays a critical role in our set up and we postpone discussion of this and related matters to Section X.

[7] This protocol is adopted in part for its descriptive realism, which is to say, depositors typically visit their bank only when they want to make a withdrawal.

[8] Cooper and Corbae (2002) study a deposit game with increasing returns to intermediation and examine if this game has multiple equilibrium. As we are interested in the withdrawal game, everything we have to say presumes that all $N$ participate in the bank. If we invoke a standard equilibrium refinement, we can show (below) that the unique equilibrium to our deposit game has all $N$ agents depositing their endowment at that bank at

game is played at dates 1 and 2. The withdrawal game is simple: each depositor $j \in \{1, 2, ..., N\}$ simultaneously chooses an action $t_j \in \{1, 2\}$, where $t_j$ denotes the date depositor $j$ contacts the bank. Depositor $j$ knows only his own type when he chooses $t_j$. In particular, depositor $j$ does not know the number of impatient depositors $n$ in the economy. A *strategy profile* $\boldsymbol{t} \equiv \{t_1, t_2, ..., t_N\}$ implies an $m \in \{0, 1, ..., N\}$, the number of depositors that contact the bank at date 1. Since efficiency dictates that impatient depositors consume at date 1 and patient depositors consume at date 2, a *truth-telling strategy* is a strategy profile that has impatient depositors contacting the bank at date 1 and patient depositors contacting the bank at date 2. A truth-telling strategy implies that $m = n$.

A strategy profile $\mathbf{t}$ and its associated $m$ constitutes a Bayes-Nash *equilibrium* of the withdrawal game with allocation $(\mathbf{c}_1, \mathbf{c}_2)$ if $t_j \in \mathbf{t}$ is a best response for depositor $j$ against $\mathbf{t}_{-j} \equiv \{t_1, ..., t_{j-1,} t_{j+1}, ..., t_N\}$ for all $j \in \{0, 1, ..., N\}$. An allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is said to be *incentive compatible* (IC) if the truth-telling strategy is an equilibrium for the withdrawal game.

Since $c_1(m) > 0$, it is always a strictly dominant strategy for impatient depositors to visit the bank at date 1. A patient depositor tells the truth by contacting the bank at date 2; he has an incentive to do so—assuming that all other patient depositors visit at date 2—iff

$$\sum_{n=0}^{N-1} \Pi^n u\left[c_2(n)\right] \geq \sum_{n=0}^{N-1} \Pi^n u\left[c_1(n+1)\right], \tag{3}$$

where $\Pi^n$ is the conditional probability that there are $n$ impatient individuals given there is at least one patient individual and

$$\Pi^n = \frac{\binom{N-1}{n}(1-\pi)^{N-n-1}\pi^n}{\sum_{n=0}^{N-1}\binom{N}{n}(1-\pi)^{N-n-1}\pi^n}.$$

If a feasible allocation $(\mathbf{c}_1, \mathbf{c}_2)$ satisfies (3), then there exists an equilibrium where all depositors play the truth-telling strategy.

If an allocation $(\mathbf{c}_1, \mathbf{c}_2)$ satisfies (3), there *may* exist other equilibrium outcomes in the withdrawal game. In particular, there may exist an equilibrium in which depositors play a *panic strategy.* A panic strategy is a strategy profile that has all depositors contacting the bank at date 1, i.e., $t_j = 1$ for all $j$ and, as a result, $m = N$ for any $n \leq N$.

date 0.

## 3 Efficient incentive-compatible allocations

In this section we characterize the properties of efficient incentive-compatible allocations. We begin with the standard case where the return to investment is invariant to its scale, $\kappa = 0$. We then study the case in which the return to investment is subject to a scale economy, $\kappa > 0$.

### 3.1 Linear technology

We first characterize the unconstrained efficient allocation. The unconstrained allocation *assumes* that impatient depositors contact the bank at date 1 and patient depositors contact the bank at date 2. This implies that $n$ impatient depositors contact the bank at date 1. The unconstrained efficient allocation is given by an allocation $(\mathbf{c}_1, \mathbf{c}_2) \equiv \{c_1(n), c_2(n)\}_{n=0}^{N}$ that maximizes the expected utility of the representative, *ex ante* identical depositor,[9]

$$\max_{\{c_1(n)\}} \sum_{n=0}^{N} \pi_n \{nu\left[c_1(n)\right] + (N-n)u\left[c_2(n)\right]\} \tag{4}$$

subject to the resource constraints

$$\begin{aligned} nc_1(n) &= Ny - k(n) && (5) \\ Rk(n) &= (N-n)c_2(n) && (6) \end{aligned}$$

which, when combined, yields

$$nc_1(n) + \frac{(N-n)c_2(n)}{R} = Ny, \tag{7}$$

for all $n \in \{0, 1, ..., N\}$. Let $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ denote the solution to the problem above. It is easy to show that a unique solution exists and satisfies

$$u'[c_1^*(n)] = Ru'[c_2^*(n)] \;\forall n < N, \tag{8}$$

[9] Green and Lin (2000) provide a characterization of the efficient allocation when there is no sequential service and the investment technology is linear. We present this allocation here because it is relevant for the efficient allocation for the scale economy.

with $c_2^*(N) = 0$ and the resource constraint (7). Given our CES preference specification, the solution is available in closed-form,

$$c_1^*(n) = \frac{Ny}{n + (N-n)R^{1/\sigma - 1}} \tag{9}$$

$$c_2^*(n) = R^{1/\sigma} c_1^*(n), \tag{10}$$

for all $n < N$ with $c_1^*(N) = y$ and $c_2^*(N) = 0$. Note that for all $n < N$ depositors engage in risk-sharing since $y < c_1^*(n) < c_2^*(n) < Ry$ . Moreover, because $\sigma > 1$ and $R > 1$ imply $R^{1/\sigma - 1} < 1$, it follows that both $c_1^*(n)$ and $c_2^*(n)$ are decreasing in $n$.

**Property 1** $c_2^*(n) > c_1^*(n) > c_1^*(n+1)$ for all $n \in \{0, 1, ...N-1\}$.

One implication of Property 1 is that the short and long-term rates of return on deposits, $c_1^*(n)/y$ and $c_2^*(n)/y$, respectively, are both decreasing in the level of date 1 redemption activity, $n$. Wallace (1988) interprets $c_1^*(n) > c_1^*(n+1)$ as a partial suspension scheme which, by construction, is efficient here.

Using (6) and (9), the efficient investment schedule, $k^*(n)$, is

$$k^*(n) = \left[ \frac{(N-n)R^{1/\sigma - 1}}{n + (N-n)R^{1/\sigma - 1}} \right] Ny. \tag{11}$$

Notice that $k^*(n)$ is decreasing $n$. A higher value of $n$ means that aggregate demand for early withdrawals increases. To accomodate this higer aggregate demand, funding for the longer-term capital investment will be scaled back. Note that high realizations for $n$ can be interpreted as recessionary events or investment collapses associated with large numbers of depositors making early withdrawals. These events, however, are driven by economic fundamentals—this source of return uncertainty of deposits has nothing to with bank fragility. A bank could mitigate the economic impact of these "fundamental runs" by expanding its depositor base, $N$. Our full support assumption, however, implies that the probability that all depositors desire early withdrawal, $\pi_N$, remains strictly positive even as $N \to \infty$.

There are two important results associated with allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$. First, it follows immediately from Property 1 that it $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ is incentive-compatible.

In particular, since $c^*(n) > c^*(n+1)$ for all $n < N$, allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ satisfies the incentive compatibility condition (3).

Second, the truth-telling equilibrium that implements $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ in the withdrawal game is *unique*. To see this, first note that it is a dominant strategy for impatient depositors to contact the bank at date 1 since $c_1(m) > 0$ for all $m \in \{1, 2, ..., N\}$. It is also a dominant strategy for the patient depositor to contact the bank at date 2 for any conjecture $m$, since $c_2^*(m) > c_1^*(m) > c_1^*(m+1)$, i.e., a patient depositor always receives a higher payoff by postponing his withdrawal to the later date. Since it is a dominant strategy for a patient individual to contact the bank at date 2, the allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ can be uniquely implemented as an equilibrium in dominant strategies. We summarize the linear technology case with the following proposition,

**Proposition 1** *[Green and Lin, 2000]. The unconstrained efficient allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ is uniquely implementable as a Bayes-Nash equilibrium of the withdrawal game when depositor types are private information and the investment technology is linear.*

Proposition 1 implies that private information is not in itself an obstacle to implementing the unconstrained efficient allocation uniquely. Notice too that the Peck and Shell (2003) communications protocol does not hinder the implementation of the unconstrained efficient allocation under private information and linear returns. Proposition 1 implies that bank panics do not exist when investments are subject to constant returns to scale.

### 3.2 Scale economies

Let $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ and $\hat{\mathbf{k}}$ denote the unconstrained efficient allocation and the associated unconstrained efficient level capital investment, respectively, in a scale economy with minimum scale $\kappa > 0$. Note that when $n = N$, achieving scale is irrelevant since all depositors are impatient. Therefore, we always have that $\hat{k}(N) = k^*(N) = 0$ and $\hat{c}_1(N) = c_1^*(N)$. Suppose that the minimum scale $\kappa$ is such that $k^*(j+1) < \kappa < k^*(j)$. In this situation, if there are at most $j$ impatient depositors—or, equivalently at least $N - j$ patient depositors—then available funding under allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ will exceed $\kappa$ and, as a result, the return on investment will be $R$. Therefore, if $k^*(j+1) < \kappa < k^*(j)$, we have

**Property 2** $\{\hat{c}_1(n), \hat{c}_2(n)\} = \{c_1^*(n), c_2^*(n)\}$ for all $n \in \{0, 1, ..., j, N\}$.

That is, the unconstrained efficient allocation in the scale economy corresponds to the unconstrained efficient allocation in the linear economy for all states $j$, where $k^*(j) \geq \kappa$.

Qualitatively speaking, the remaining allocations $\{\hat{c}_1(n), \hat{c}_2(n)\}$, $n = j + 1, ..., N-1$, will take one of two forms. Since $k^*(n) < \kappa$ for $n = j+1, ..., N-1$, the capital funding associated with these allocations will be characterized by either $\hat{k}(n) = \kappa$ or $\hat{k}(n) < \kappa$. That is, the consumption allocations will be designed so that capital is either exactly equal to minimum scale $\kappa$ or falls short of it. Since the investment return associated with the former is $R > 1$ and $r < 1$ for the latter, if $\hat{k}(n) = \kappa$, then we have

$$\hat{c}_1(n) = \frac{Ny - \kappa}{n} \hat{c}_2(n) = \frac{R\kappa}{N - n} \text{ and } \hat{k}\,(n) = \kappa; \tag{12}$$

if $\hat{k}(n) < \kappa$, then we have

$$\hat{c}_1(n) = \frac{Ny}{n + (N-n)r^{1/\sigma - 1}} \tag{13}$$

$$\hat{c}_2(n) = r^{1/\sigma} \hat{c}_1(n), \tag{14}$$

and

$$\hat{k}(n) = \left[ \frac{(N-n)r^{1/\sigma - 1}}{n + (N-n)r^{1/\sigma - 1}} \right] Ny. \tag{15}$$

Notice that (13)-(15) replicates (9)-(11), respectively, but where $R$ is replaced by $r$. Since $\sigma > 1 > r$, (13)-(14) imply that $\hat{c}_2(n) < \hat{c}_1(n) < y$, i.e., impatient depositors receive *higher* payments than patient depositors and both payments are *less* than their initial endowment deposit, $y$. Intuitively, there is a trade-off between the two choices of capital: Capital allocation $\hat{k}(n) = \kappa$ maximizes the aggregate, 2 period consumption but has poor risk-sharing features; capital allocation $\hat{k}(n) < \kappa$ provides for efficient risk sharing but at low levels of consumption.

We now fully characterized the unconstrained efficient allocation for $\kappa = 2y$ and for model parameters imply that $k^*(N-2) < \kappa = 2y < k^*(N-3)$.[10]

[10]In Appendix X we show that, qualitatively speaking, this parameterization is without loss of generality. We adopt this parameterization because it is "easy" to characterize the unconstrained efficient allocation.

This parameterization implies that if $n \leq N-3$—i.e., there are at most $N-3$ impatient depositors—then the unconstrained efficient allocation for $n$ is given by $(c_1^*(n), c_2^*(n))$ since $k^*(n) > \kappa = 2y$. If, however, $n \in \{N-2, N-1\}$, then the unconstrained efficient allocation will be given by either (12) or (13)-(14) since $k^*(n) < \kappa = 2y$.

Let's first examine the case where there are $n = N-2$ impatient depositors. If the "high-return, high-level investment" option, $\hat{k}^H(N-2) = \kappa = 2y$, is chosen, then (12) implies that $\hat{c}_1^H(N-2) = y$ and $\hat{c}_2^H(N-2) = Ry$. If instead the "low-return, low-level investment" option, $\hat{k}^L(N-2) < \kappa$, is chosen, then (13)-(14) implies $\hat{c}_2^L(N-2) < \hat{c}_1^L(N-2) < y$. Clearly, the high-return, high-level investment option dominates the low-return, low-investment option since $\hat{c}_1^H(N-2) > \hat{c}_1^L(N-2)$ and $\hat{c}_2^H(N-2) > \hat{c}_1^L(N-2)$. Therefore, we have

**Property 3** When $n = N-2$, the unconstrained efficient allocation in the scale economy is given by the high-return, high-level investment option, where $\hat{k}^H(N-2) = 2y$, $\hat{c}_1^H(N-2) = y$ and $\hat{c}_2^H(N-2) = Ry$.

Now let's examine the case where there are $n = N-1$ impatient depositors—or, equivalently, one patient depositor. If the high-return, high-level investment, $\hat{k}^H(N-1) = \kappa = 2y$, is chosen (12) implies that

$$\hat{c}_1^H(N-1) = \frac{N-2}{N-1}y \tag{16}$$

$$\hat{c}_2^H(N-1) = 2Ry. \tag{17}$$

Since $\hat{c}_1^H(N-1) < y < R < \hat{c}_2^H(N-2)$, notice that this investment option comes at the cost of rather inefficient risk-sharing. In particular, notice that the payoff to the impatient depositor is less than his autarkic payoff, $y$. If, instead the low-return, low-level investment option, $\hat{k}^L(N-1) < \kappa$ is chosen, then (13)-(14) imply that

$$\hat{c}_1^L(N-1) = \frac{N}{N-1+r^{1/\sigma-1}}y, \tag{18}$$

$$\hat{c}_2^L(N-1) = r^{1/\sigma}c_1^L(N-1). \tag{19}$$

Inspecting conditions (18) and (19), leads us to

**Lemma 1** *For $r$ arbitrarily close to (but less than) unity, $\hat{c}_1^L(N-1) \approx \hat{c}_2^L(N-1) \approx y = \hat{c}_1(N)$, where $\hat{c}_2^L(N-1) < \hat{c}_1^L(N-1) < y$.*

Lemma 1 tells us that if $r$ is close to unity, then the payouts to patient and impatient depositors are about equal to $y$. Let's assume that $r < 1$ is arbitrarily close to unity. Then, by Lemma 1, the expected utility payoff associated with the low-return, low-level investment option is approximately equal to $u(y)$. From (16)-(17), the expected utility associated with the the high-return, high-level investment option is

$$\left(\frac{N-1}{N}\right) u\left(\frac{N-2}{N-1}y\right) + \left(\frac{1}{N}\right) u\left(2Ry\right).$$

Since this investment option has poorer risk-sharing properties than the low-return, low-level investment option, we would expect the benefit of the former option to diminish with depositors' appetite for risk. Indeed, we can demonstrate that for preferences with $\sigma \geq 2$, the expected utility associated with the low-return, low-level investment option exceeds that of the high-return, high-level investment option, i.e.,[11]

$$\left(\frac{N-1}{N}\right) u\left(\frac{N-2}{N-1}y\right) + \left(\frac{1}{N}\right) u\left(2Ry\right) < u\left(y\right).$$

Therefore, we have the following,

**Property 4** When $n = N-1$, $r < 1$ sufficiently close to unity and $\sigma \geq 2$, the unconstrained efficient allocation in the scale economy is given by the low-return, low-level investment option, where $\hat{c}_1^L\left(N-1\right)$ and $\hat{c}_2^L(N-1)$ are determined by (18) and (19), respectively.

Property 4 implies that when $n = N-1$, the bank "breaks the buck" in the sense that for every unit that individuals deposit at the bank, they receive less than a unit payoff at date 1, as well as date 2. The empirical relevance of this observation is discussed in Section 5, below.

Properties 1-4 fully characterize the unconstrained efficient allocation in a scale economy where $k^*(N-2) < \kappa = 2y < k^*(N-3)$, $r < 1$ is sufficiently

[11] See Appendix 1 for the proof.

close to unity and $\sigma \geq 2$. In particular, the unconstrained efficient allocation, $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$, is given by

$$\{\{c_1^*(n), c_2^*(n)\}_{n=0}^{N-3}, \hat{c}_1^H(N-2), \hat{c}_2^H(N-2), \\ \hat{c}_1^L(N-1), \hat{c}_2^L(N-1), c_1^*(N), c_2^*(N)\}$$

We now show that the unconstrained efficient allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ is incentive-compatible. Impatient depositors do not have an incentive to misrepresent themselves, so they always contact the bank at date 1. Regarding patient depositors, in states all states $n \leq N-2$, we have $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ (from Properties 1, 2 and 3); and in state $n = N-1$, we have $\hat{c}_2(N-1) < \hat{c}_1(N) \approx y$ (from Property 4). Assuming that all other patient depositors contact the bank at date 2, a patient depositor will contact the bank at date 2 if the unconstrained efficient allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ satisfies (3) or, equivalently, if

$$\sum_{n=0}^{N-2} \Pi^n u\left[\hat{c}_2(n)\right] - \sum_{n=0}^{N-2} \Pi^n u\left[\hat{c}_1(n+1)\right] \geq \Pi^{N-1} u\left[\hat{c}_1(N-1)\right] - \Pi^{N-1} u\left[\hat{c}_2(N)\right]. \tag{20}$$

Since $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ for all $n \leq N-2$, the left side is strictly greater than zero. When $r < 1$ is arbitrarily close to unity, the right side is positive but arbitrarily close to zero. Hence, (20) is satisfied with a strict inequality.[12] Therefore, we have the following result,

**Proposition 2** *The unconstrained efficient allocation* $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ *can be implemented as a truth-telling equilibrium of the depositor game in the scale economy characterized by* $\kappa = 2y$ *and* $\sigma \geq 2$ *with* $r < 1$ *arbitrarily close to 1.*

A few remarks regarding the implementation result are in order before we proceed to bank panic equilibria.

1. Allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ respects the assumed communication protocol that depositors can contact the bank at date 1 *or* date 2 but not both. In

[12] Since $\hat{c}_2(n) > \hat{c}_1(n) > \hat{c}_1(n+1)$ when $n \leq N-2$, $r < 1$ need *not* be arbitrarily close to unity to have allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ satisfy incentive-compatibility. The condition that $r < 1$ is arbitrarily close to unity simply guarantees that the incentive-compatibility condition (3) will hold with strict inequality. We discuss this in more detail in Remark 4, below.

contrast to allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$, this protocol imposes a binding restriction on the allocation in one state of the world, $n = N - 1$. To see this, suppose that depositors could contact the bank at both dates 1 *and* 2. Then in state $n = N - 1$, the efficient bank allocation would provide a payment equal to $y$ to *all* depositors at date 1. Notice that this payment dominates the payments from allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ since $y > \hat{c}_1^L(N-1) > \hat{c}_2^L(N-1)$. But assuming that depositors can contact the bank at both dates 1 *and* 2 in our environment would essentially be equivalent to assuming the existence of *no* banking communication restrictions at all. In standard Diamond-Dybvig banking models, a sequential service (communications) restriction is always imposed since markets would otherwise eliminate the need for banks. We interpret our communication restriction as a weaker form of the more common sequential service restriction.

2. It is straightforward to show that the expected utility of allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ exceeds autarky. In autarky all agents consume their endowment at date 1 and receive an expected utility equal to $u(y)$, assuming $y < \kappa$. Allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ delivers depositors an expected utility that strictly exceeds $u(y)$ in states $n \in \{0, \ldots, N-2\}$, an expected utility that equals $u(y)$ in state $n = N$ and an expected utility that is arbitrarily less than $u(y)$ in state $n = N - 1$.

3. We assumed that $\kappa = 2y$. The qualitative properties of the unconstrained efficient allocation remain valid for more general cases. See Appendix X for a discussion.

4. Lemma 1 and Proposition 2 assumed that $r < 1$ is arbitrarily close to 1. In this case we are able to show that the low-return low-level investment option is strictly preferred to the high-return high-level investment option and that allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ is strictly preferred to autarky. Clearly, $r$ can be made much less than unity without upsetting either of these results.

Proposition 2 tells us that allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ can be implemented as a truth-telling equilibrium. We now examine if the truth-telling equilibrium is the *unique* equilibrium for the depositor game.

# 4 Panic equilibria

We now investigate if deposit contact $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ generates outcomes other than the truth-telling equilibrium. In particular, we are interested if there exists a panic equilibrium—an equilibrium where all $N$ individuals contact the bank at date 1 regardless of their true liquidity needs, i.e., $m = N$ for all $n \in \{0, 1, \ldots, N-1\}$. Our main result is stated in the following proposition,

**Proposition 3** *The unconstrained efficient allocation* $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ *admits a panic equilibrium.*

To check the validity of Proposition 3, we propose an equilibrium strategy profile where all $N$ depositors contact the bank at date 1 and ask whether a patient depositor has an incentive to play the strategy and contact the bank at date 1. If a patient depositor plays the proposed equilibrium strategy profile and contacts the bank at date 1, he receives a consumption payoff equal to $\hat{c}_1(N) = y$. If, instead, he deviates from the proposed equilibrium strategy profile and contacts the bank at date 2, then he receives a consumption payoff equal to $\hat{c}_2^L(N-1) < y$, since all other $N-1$ depositors contact the bank at date 1. Clearly, a patient depositor does not have an incentive to deviate from the proposed equilibrium strategy profile and, as a result, a panic equilibrium exists.[13]

When the unconstrained efficient allocation is characterized by the low-return, low-level investment option being implemented state $n = N-1$, a bank panic equilibrium always exists. The low-return, level investment option will, in fact, be implemented in this state whenever the return associated with the investment, $r$, is not "too negative." In contrast, the existence of bank run equilibria in standard banking environments, which feature linear investment technologies and sequential service, depend on the specification of the sequential service constraint and, possibly, the correlation of depositor types.

[13]Notice that if $r = 1$, then a patient depositor would be indifferent between misrepresenting himself or not. Thus, a panic equilibrium remains possible even if $r = 1$, though it seems unlikely to survive any reasonable equilibrium refinement.

### 4.1 Equilibrium panics and eliminating panics

Proposition 4 says that if allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ is in place, then there exists a panic equilibrium. The important question to address is if a bank would ever offer allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$, assuming that all $N$ agents deposit their endowments at date 0. Following Peck and Shell (2003) we show that one can construct a sunspot equilibrium that supports a panic equilibrium. A sunspot is an extrinsic event observed after individuals make their bank deposits but before they learn their type that occurs with probability $\theta$, where $0 \leq \theta \leq 1$. The sunspot equilibrium is characterized by the probability $\theta$ and the allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$. The equilibrium strategy profile for the sunspot equilibrium is: when the sunspot is *not* observed, an event that occurs with probability $1-\theta$, depositors play the truthtelling equilibrium strategies described in Proposition 2; and when the sunspot is observed, an event that occurs with probability $\theta$, all depositors play the panic equilibrium strategies described in Proposition 3. We now verify that a sunspot equilibrium exists for some values of $\theta$.

Let $V(\kappa, R, \theta)$ denote the expected utility associated with allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ when a sunspot occurs with probability $\theta$ assuming that depositors play the above described sunspot equilibrium strategies, i.e.,

$$V(2y, R, \theta) \equiv (1-\theta)E\left[U(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)\right] + \theta u(y).$$

Clearly, $V(2y, R, \theta)$ is strictly decreasing and continuous in $\theta$ for all $\theta \in (0, 1]$, with $V(2y, R, \theta = 0) = u(y)$. In contrast to Peck and Shell (2003), a banking arrangement will always emerge as an equilibrium since in our environment $V(2y, R, \theta) > u(y)$ for all $\theta > 0$.[14]

The risk-sharing arrangement that prevails will depend on the magnitude of $\theta$, the probability of a sunspot. In particular, just as in Peck and Shell (2003), for $\theta$ sufficiently large, the bank may want to eliminate a panic equilibrium by offering an allocation that is *panic proof*. The most obvious

[14] In our environment, if a sunspot is observed, then each depositor receives a payoff of $u(y)$; if a sunspot is not observed, then the expected payoff to the representative depositor—who does not yet know his type—is $E\left[U(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)\right] > u(y)$. Hence, as long as $\theta > 0$, a banking arrangement will always emerge in equilibrium. In Peck and Shell (2003), if $\theta$ is sufficiently high agents will prefer autarky (no-banking) since in their in environment, which assumes a sequential service constraint, the expected utility associated with playing the panic equilibrium is for their constrained efficient allocation strictly less than $u(y)$.

way to render the risk-sharing arrangement panic proof is for the bank to invest in at least $\kappa$ units of capital in all states $m < N$. (In state $m = N$, all depositors contact the bank at date 1 so the bank will return all of its $Ny$ potential investment assets to the depositors.) When $\kappa = 2y$, the best panic-proof allocation, $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, is given by $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ for all $m \neq N-1$ and for $m = N-1$,

$$[\bar{c}_1(N-1), \bar{c}_2(N-1)] = [y(N-2)/(N-1), 2Ry].$$

allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ is incentive compatible and panic proof. It is incentive compatible because $\bar{c}_2(m) > \bar{c}_1(m+1)$ for all $m \leq N-1$. Because allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ is incentive compatible, it is also panic proof. Specifically, a patient individual has no incentive to contact the bank at date 1 even if he thinks that the remaining $N-1$ depositors contact the bank at date 1 since $\bar{c}_2(N-1) = 2Ry > y = \bar{c}_1(N)$. Let $Q(\kappa, R)$ denote the expected utility associated with the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, i.e.,

$$Q(2y, R) \equiv EU[(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)].$$

Suppose that $Q(2y, R) > u(y)$. Then for all $N < \infty$ there exists a $\theta_0 \in (0, 1)$ such that

$$V(2y, R, \theta_0) = Q(2y, R) \tag{21}$$

since, (i) $V(2y, R, \theta)$ is strictly continuously decreasing in $\theta$, (ii) $V(2y, R, \theta = 0) > Q(2y, R)$ and (iii) $V(2y, R, \theta = 1) = u(y)$. Hence for any $\theta < \theta_0$, depositors strictly prefer allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, which leaves them exposed to a bank panic. If, however, $\theta > \theta_0$, then depositors will prefer the panic-free allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ to the panic-prone allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$. In this case—assuming that $Q(2y, R) > u(y)$—the bank will offer the no-panic allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$ to depositors, i.e., there will not be a panic in equilibrium.

Interestingly, it need not be the case that $Q(2y, R) > u(y)$. To see this, notice that for $n \leq N-2$

$$\left(\frac{n}{N}\right) u[\bar{c}_1(n)] + \left(\frac{N-n}{N}\right) u[\bar{c}_2(n)] > u(y)$$

and that for $\sigma \geq 2$,

$$\left(\frac{N-1}{N}\right) u\left(\frac{N-2}{N-1} y\right) + \left(\frac{1}{N}\right) u(2Ry) < u(y).$$

Hence, $\pi_{N-1}$ is sufficiently large, then it is possible that $Q(2y, R) < u(y)$, which implies that autarky is preferred to the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$. If $Q(2y, R) < u(y)$, then the equilibrium outcome for the economy will be for the bank to offer allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to depositors—since $V(2y, R, \theta) > u(y) > Q(2y, R)$—which depositors will accept by depositing their endowment at date 0. Notice that allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ carries with it the risk of bank panics occurring with probability $\theta$. We summarize the results in the section in the following proposition,

**Proposition 4** *If either $Q(2y, R) < u(y)$ or $Q(2y, R) > u(y)$ and $\theta < \theta_0$, then depositors prefer the panic-prone allocation $(\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2)$ to the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$; otherwise, depositors prefer the panic-proof allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$.*

We close this section by noting that the expected cost—in the form of poorer risk-sharing—associated with eliminating panics is decreasing in $N$ since the probability $\pi_{N-1}$ is strictly decreasing in the size of the depositor base $N$. It follows then that larger banks or a more interconnected banking system can be made more stable at a lower expected cost. Thus, the constrained-efficient risk-sharing arrangement may entail a panic-proof allocation if the susceptibility of a society to coordination failure, as indexed here by $\theta$, is sufficiently high, or if the banking system has a sufficiently large depositor base or is otherwise sufficiently well diversified as indexed by the parameter $N$.

# 5 Reach for yield

Assume there are two fundamentally risk-free investments in the economy parameterized by $\{(\kappa_1, R_1), (\kappa_2, R_2)\}$, where $(\kappa_1, \kappa_2) = (0, 2y)$, $(R_1, R_2) = (\delta, R)$ and $1 \leq \delta < R$. The reader is invited to think of $(\kappa_1, R_1)$ as a money market fund or bank invested in safe government bonds and $(\kappa_2, R_2)$ as representing a fundamentally safe class of business investments. Alternatively, one might think of $(\kappa_1, R_1)$ as being funded by narrow banks and $(\kappa_2, R_2)$ by shadow banks.

Let $(\mathbf{c}_1^{1*}, \mathbf{c}_2^{1*})$ denote the unconstrained efficient allocation associated with the money fund investment $(\kappa_1, R_1) = (0, \delta)$. Since $\kappa_1 = 0$, the money fund is panic free. Let $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$ denote the unconstrained efficient allocation associated with the bank funding business investments, where $(\kappa_2, R_2) = (2y, R)$. Assume that the probability of a sunspot $\theta$ is characterized by $0 < \theta < \theta_0$, where $\theta_0$ is defined in (21). Since $0 < \theta < \theta_0$, bank allocation $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$ is exposed to panic attacks and is preferred to a panic-free allocation.

Let $V(0, \delta, 0)$ represent the expected utility associated with allocation $(\mathbf{c}_1^{1*}, \mathbf{c}_2^{1*})$ and let $V(2y, R, \theta)$ denote the expected utility associated with allocation $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$. Clearly, $V(0, \delta, 0)$ is strictly increasing in $\delta$ with $V(0, 1, 0) = u(y)$. Given the properties of $V(0, \delta, 0)$ and $V(2y, R, \theta)$, it is evident that for a given $\theta \in (0, \theta_0)$, there exists a $\delta_0 \in (1, R)$ such that

$$V(0, \delta_0, 0) = V(2y, R, \theta). \tag{22}$$

Hence, there exists a risk-return trade off, where the risk is extrinsic here. Because our model assumes that investors have identical preferences, only one of the two funds will emerge in equilibrium; the one which generates the highest expected utility for depositors will be observed.[15]

Our two investment economy can help us put some structure of the concept of “reach for yield,” a term that is used extensively in the popular press. To begin, consider an environment in which $\delta_0 < \delta < R$. In this case, depositors prefer the narrow banking arrangement $(\mathbf{c}_1^{1*}, \mathbf{c}_2^{1*})$ because $V(0, \delta, 0) > V(2y, R, \theta)$. Now suppose that the rate of return on investment funded in the narrow banking sector declines to $\delta'$, where $1 < \delta' < \delta_0$. The cause of this decline in the “safe real interest rate” is immaterial here.[16] The decline in the safe real interest rate will induce a portfolio reallocation. In particular, since $V(2y, R, \theta) > V(0, \delta', 0)$, depositors are motivated to move their resources out of the narrow bank and into the shadow bank that is offering the allocation $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$. But the new risk-sharing arrangement $(\hat{\mathbf{c}}_1^2, \hat{\mathbf{c}}_2^2)$ is subject to panics. Hence, there is a sense in which our model supports the notion of low real interest rates motivating a reach-for-yield behavior that

[15] Modeling preference heterogeneity with respect to risk-tolerance would produce a model in which risk-sharing arrangements with different risk-return characteristics could coexist. We believe that much of the following intuition would survive such a generalization.

[16] Perhaps it is induced by central bank policy as an attempt to bolster the economy in the face of an imminent recession. Alternatively, imagine a wave of pessimism in a part of the globe resulting in a flight to the safety of U.S. treasury debt, leading to a decline in real yields.

renders the financial system less stable and more prone to panic attacks; see, for example, Stein (2013).

# 6 Discussion

## 6.1 Relation to the literature

Central to the concept of bank panic (or *run*) in the seminal work of Diamond and Dybvig (1983) is the notion of widespread misrepresentation of liquidity needs on the part of depositors who have no urgent use for funds. If these depositors could somehow be persuaded to remain calm and confident that others are not prone to panic, then everything would work out just fine. If communication channels are imperfect, however, depositors can at best only guess at how others are likely to behave. This, in turn, leads to the prospect of a coordination game possessing multiple equilibria, with each outcome indexed by an arbitrary initial belief over how others are likely to represent themselves. The theoretical challenge put forth by Diamond and Dybvig (1983) is to identify what conditions, if any, could plausibly give rise to a coordination game in which individually-rational depositors inadvertently coordinate on a socially suboptimal outcome.

While Diamond and Dybvig (1983) stress the importance of private information and sequential service. In the first part of their paper, they demonstrate how the efficient allocation is achieved with a "simple" contract and how this same contract induces a coordination game in which a bank panic exists. Stability is within grasp, however, if banks can credibly promise to suspend convertibility in the event reserves are exhausted.[17] In the second part of their paper, they introduce aggregate liquidity risk so that the suspension of convertibility is no longer a costless mechanism for preventing panics.[18] They do not, however, formally demonstrate the existence of bank

[17] In the absence of aggregate liquidity risk, a credible threat to suspend is never exercised in equilibrium, so that the policy eliminates panics costlessly. Mention Ennis and Keister and their model of limited commitment.

[18] In particular, because a bank is no longer able to tell whether a period of heavy redemption activity is being driven by fundamental factors or psychology, suspending convertibility runs the risk of leaving some depositors without the funds they urgently need.

panics. Instead, they demonstrate how deposit insurance is sufficient to ensure bank stability.[19]

Peck and Shell (2003) were the first to establish the existence of equilibrium bank panics in a broad class of optimal mechanisms. Both private information and sequential service play a critical role in establishing their result. In their model, depositors wishing to make a withdrawal communicate their request to the bank. These requests are assumed to arrive sequentially and in random order within the period under consideration. The sequential service constraint is formalized in the manner suggested by Wallace (1988). In particular, while the withdrawal limit for each depositor in a given queue position can be conditioned on the history of withdrawals to that point in (intra-period) time, it cannot be conditioned on the number of redemption requests made for the remainder of the period, since these are unknown to the bank *ex ante*. But in a similar mechanism studied by Green and Lin (2003), the same sequential service constraint does not render banking system unstable. The two papers differ along along a few dimensions, only one of which is important.

First, Green and Lin (2003) assume that depositors can communicate with the bank in both the early and late periods. The reasonableness of this assumption depends largely on what one views as the cost of communications. Suppose that communications entail the depositor physically visiting the bank. Then we would expect depositors to visit the bank only in the event they wanted to make a withdrawal. Indeed, this is the protocol adopted by Peck and Shell (2003) in the body of their paper. In Green and Lin (2003), truth-telling patient depositors "visit" the bank in the early period for the sole purpose of communicating their liquidity preference. Such information could, in principle, be useful to the bank if it helps ascertain the period's aggregate liquidity demand. However, Peck and Shell (2003) demonstrate in their appendix that their conclusion continues to hold under the Green and Lin (2003) communication protocol. We wish to stress this point here because our fragility result fails to hold under the Green and Lin (2003) protocol.

If the Peck and Shell (2003) fragility result is robust to the Green and Lin (2003) communications protocol, then what explains their different con-

[19] Wallace (1988) points out that their deposit insurance scheme ignores sequential service. Ennis and Keister (deposit insurance paper) do it correctly.

clusions? Ennis and Keister (2009) attribute the discrepancy to what the two mechanisms assume about the information available to depositors when they play the withdrawal game. In the Green and Lin (2003), depositors are assumed to know their position in the sequential service queue *before they arrive*, whereas in Peck and Shell (2003), they do not. This additional information concerning queue position permits the application of a powerful backward induction argument to rule out multiplicity that can be applied generally if depositor types are *i.i.d.* (Andolfatto, Nosal and Wallace, 2007).[20] The intuition is as follows. Suppose a patient depositor knows he will be the last in line if the claims to be impatient. Then any resources he might acquire at that time could be more profitably held in the bank where he earns interest (recall, the rate of return on his funds is independent of the scale of the bank's investments generating that return). Understanding this, the second-to-last patient depositor similarly has an incentive to tell the truth, and so on.

There is the question of how to interpret the possibility of knowing one's queue position prior to making a withdrawal. Green and Lin (2003) invoke some notion of "clock time" occuring within each period. Depositors realize their liquidity preference shocks at random times throughout the period. Those realizing their liquidity needs near the end of the period know that they are likely to be close to the end of the queue. Alternatively, one could imagine depositors realizing their liquidity needs simultaneously, with their deposit contract assigning each depositor to a queue position *ex ante*.[21] This latter arrangement resembles the tranching of debt into senior and junior components. The ability to do can evidently prevent panics. Nosal and Wallace (2009), however, point out that incentive constraints are relaxed–hence, risk-sharing improved–when depositors do not know their queue position. As not assigning queue positions *ex ante* evidently opens the door to bank panics, an interesting trade-off between efficiency and stability seems apparent. Andolfatto, Nosal and Sultanum (2016), however, demonstrate how this trade-off vanishes when allocations are permitted to condition payoffs on all available information.[22] Unfortunately, their mechanism relies on the Green

[20] Ennis and Keister (2009) show how the argument breaks down for correlated types.

[21] In the context of a movie theater, this would be like selling seats with queue positions exiting the builiding in case of fire.

[22] Green and Lin (2000) and all subsequent papers limit their analysis to direct mechanisms. Andolfatto, Nosal and Sultanum (2016) consider an indirect mechanism (which permits a richer message space relative to a direct mechanism). They demonstrate that

and Lin (2003) communications protocol–it is infeasible under the Peck and Shell (2003) protocol that we assume here.

Thus, while the Peck and Shell (2003) communication protocol is not critical for their results, it is for ours. The choice of their protocol was determined in part by its descriptive realism. As Peck and Shell (2003, p. 105) remark, "It is hard to imagine people visiting their bank for the purpose of telling them that they are not interested in making any transactions at the present time." Having said this, it is nevertheless of some interest to ask why this is not the case–especially if the practice helps make a bank less fragile. The question becomes more pertinent when one recognizes that physical visits are important only to the extent that this the most economical way of contacting a bank. This may have been true some generations ago, but is less true in more recent times. On the other hand, even with the advent of the telephone and email, perhaps it is simply too costly to communicate one's liquidity needs at high enough frequency.

## 6.2 Recent financial stability regulations

Our theory suggests that bank panics are very likely to exist in any environment where there is: (i) a private information over liquidity needs, and (ii) a desire to finance fundamentally safe investments beyond a minimum scale with demandable or short-term debt. Organizations that fund their working capital using bank credit lines or commercial paper seem particularly vulnerable. If funding in this form is suddenly pulled in sufficient volume, these organizations could see the value of their long-term operations significantly decline. This, in turn, can reinforce a bleak outlook on the backing of their remaining debt. Something along these lines seems to have occurred on September 16, 2008, when the Reserve Primary Fund broke the buck. News of this event triggered a large wave of redemptions in the money market sector, especially from funds invested in commercial paper. The wave of redemptions ceased only after the U.S. government announced it would insure deposits in money market funds, essentially rendering them panic-free.[23]

Even though at that time mutual funds allowed their depositors to with-

the optimal indirect mechanism does not reveal depositor queue position, but nevertheless manages to implement the constrained-efficient allocation uniquely.

[23] See Kacperczyk and Schnabl (2010).

draw their investments on demand with impunity at a fixed par exchange rate, our model suggests that if these funds were priced using a net-asset-valuation (NAV) method, there might still have been a panic. In our model, promised rates of return are made contingent on market conditions, i.e., aggregate redemption demand, and this can be interpreted as a form of NAV pricing of liabilities. Our efficient risk-sharing arrangement is permitted to break the buck in very heavy redemption states. But note that our flexible NAV-like pricing structure does not eliminate panics.

On July 23, 2014, the Securities Exchange Commission announced money market reforms that included the requirement of a floating NAV for institutional money market funds, as well as the use of liquidity fees and redemption gates to be administered in periods of stress or heavy redemption.[24] While our model suggests that NAV pricing of demandable liabilities by itself is not sufficient to prevent panics, the use of liquidity fees and redemption gates is consistent with eliminating panics in our model economy. For example, the difference in consumption levels between panic-prone and panic-free allocations, $\hat{c}_1(N-1) - \bar{c}_1(N-1) > 0$ described in Section 4.1, can be interpreted as a liquidity fee that depositors pay to obtain funds when redemption activity is judged by the directors of a market fund to be unusually high. This liquidity free prevents the bank-panic equilibrium.

Other post-financial crisis regulations also take aim at reducing banks' reliance on short-term borrowing. For example, recent Basil *liquidity ratio* regulations are designed to incentivize banks to borrow longer term. The liquidity ratio requires that banks be able to withstand a significant liquidity outflow for a period of 30 days. A bank is better able to survive such a liquidity event if it lends short, which means that it will receive cash during the liquidity event, and borrows long, which means there is a high probability that is will not have to pay off loans during the liquidity event. In the context of our model, this regulation can be interpreted as providing the bank an incentive to have at least $\kappa$ units of its loans in the form of long-term 2-period debt. This long term debt pays off at date 2. This implies that the bank will always has at least $\kappa$ invested in the high return project. An implication of this long-term borrowing is that economy will be panic free. There is, however, an additional cost is associated with long-term borrowing. In the event that all $N$ depositors withdraw early for fundamental reasons,

[24] See: https:// www.sec.gov/ News/ PressRelease/ Detail/ PressRelease/ 1370542347679

which is an event that occurs with low probability when $N$ is large, the bank will only be able to distribute $(N-2)y$ units date 1 consumption. This implies that $2y$ units will be wasted. If, however, the probability of a sunspot, $\theta$, is sufficiently high, then a regulation that requires banks to undertake some long-term borrowing can generate higher welfare than the allocation associated with only short-term borrowing that is subject to a panic.[25]

### 6.3 Bank panics and bank size

Based on an examination of the historical record of financial crises, Calomiris and Gorton (1991) conclude that institutional factors, such as branch bank laws and bank cooperation arrangements, play an important role in determining both the frequency and magnitude of bank panics. Williamson (1989) studies the institutional differences between the banking systems in Canada and the United States during the U.S. National Banking era. The upshot of the historical evidence is that banking systems with fewer but larger banks appear more resilient to panics. Note that "largeness" here should be taken to include the propensity for banks to engage in branch banking activities and to cooperate with other banks in an interbank market, possibly in conjugation with a central bank. In our model, the parameter indexing the size of a bank or banking system is $N$. Our model predicts that larger banks or a more interconnected banking system are better because larger and more diversified banks are less likely to experience extremely high redemption states, thereby lowering the expected cost associated with extra haircuts in high redemptions states needed to discourage panic behavior.

[25] The bank may want to invest in long-term debt, even though it is costly, for commitment reasons. For example, if the bank invests in short term debt and promises to keep at least $\kappa$ invested for all $m \leq N-1$, as in allocation $(\bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2)$, then if $N-1$ depositors contact the bank at date 1 and the bank lacks commitment, it may have an incentive to to let the investment fall below $\kappa$. A bank will have such an incentive if its objective is to maximize the welfare of the entire community, conditional of the number of contacts it receives at date 1.

# 7 Conclusions

To be added

# 8 References


1. Allen, Franklin and Douglas Gale (1998). "Optimal Financial Crises," *The Journal of Finance*, 53(4): 1245-1284.

2. Andolfatto, David, Nosal, Ed and Neil Wallace (2007). "The Role of Independence in the Green-Lin Diamond-Dybvig Model," *Journal of Economic Theory*, 137(1): 709–715.

3. Andolfatto, David, Nosal, Ed and Bruno Sultanum (2016). "Preventing Bank Runs," Forthcoming in *Theoretical Economics*.

4. Andolfatto, David, Berentsen, Aleksander and Fernando Martin (2017). "Money, Banking, and Financial Markets," Manuscript: Federal Reserve Bank of St. Louis.

5. Bryant, John (1980). "A Model of Reserves, Bank Runs, and Deposit Insurance," *Journal of Banking Finance*, 43: 749–761.

6. Calomiris, Charles and Gary Gorton (1991). "The Origins of Banking Panics: Models, Facts, and Bank Regulation," in *Financial Markets and Financial Crises*, edited by Glenn Hubbard, University of Chicago Press, 109–174.

7. Capiro, Gerard, Jr. and Daniela Klingebiel (1997). "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?" in M. Bruno and B. Pleskovic, eds., *Annual World Bank Conference on Development Economics*. Washington DC: World Bank, 79–104.

8. Cavalcanti et al.

9. Cooper, Russell and Ross, Tom (1998). "Bank Runs: Liquidity Costs and Investment Distortions," *Journal of Monetary Economics* 41, 27–38.

10. Cooper, Russell and Dean Corbae (2002). "Financial Collapse: A Lesson from the Great Depression," *Journal of Economic Theory*, 107: 159–190.

11. Diamond, Douglas and Philip Dybvig (1983). "Deposit Insurance and Liquidity," *Journal of Political Economy*, 91(June): 401–419.

12. Ennis, Huberto M. and Todd Keister (2006). "Bank Runs and Investment Decisions Revisited," *Journal of Monetary Economics*, 52(2): 217–232.

13. Ennis and Keister (2007). Commitment and Equilibrium Bank Runs. I can only find working paper version.

14. Ennis, Huberto M. and Todd Keister (2009). "Run Equilibria in the Green-Lin Model of Financial Intermediation," *Journal of Economic Theory*, (September): 1,996-2,020.

15. Ennis, Huberto M. and Todd Keister (2010). "On the Fundamental Reasons for Bank Fragility," FRB Richmond *Economic Quarterly* 96, First Quarter: 33-58.

16. Gorton, Gary (1988). "Banking Panics and Business Cycles," *Oxford Economic Papers*, 40: 751-781.

17. Green, Edward and Ping Lin (2000). "Diamond and Dybvig's Classic Theory of Financial Intermediation: What's Missing?" Federal Reserve Bank of Minneapolis *Quarterly Review*, 24 (Winter): 2–13.

18. Green, Edward and Ping Lin (2003). "Implementing Efficient Allocations in a Model of Financial Intermediation," *Journal of Economic Theory*, 109(1); 1–23.

19. Hughes, Jospeh P. and Loretta J. Mester (2013). "Who Said Large Banks Don't Experience Scale Economies? *Journal of Financial Intermediation,* 22(4): 559–585.

20. Jacklin, Charles J. (1987). "Demand Deposits, Trading Restrictions, and Risk Sharing," In *Contractual Arrangements for Intertemporal Trade*, edited by Edward C. Prescott and Neil Wallace. Minneapolis: University of Minnesota Press, 26-27.

21. Kacperczyk, Marcin and Philipp Schnabl (2010). "When Safe Proved Risky: Commercial Paper during the Financial Crisis 2007–2009," *Journal of Economic Perspectives*, 24(1): 29–50.

22. Keiste, Todd and Vijay Narasiman (2016). "Expectations vs. Fundamentals-based Bank Runs: When Should Bailouts be Permitted?" *Review of Economic Dynamics*, 21: 89-104.

23. Nosal, Ed and Neil Wallace (2009). "Information Revelation in the Diamond-Dybvig Model," Federal Reserve Bank of Chicago Policy Discussion Paper Series (December).

24. Peck, James and Karl Shell (2003). "Equilibrium Bank Runs," *Journal of Political Economy*, 111(1): 103–123.

25. Stein, Jeremy (2013). "Yield-Oriented Investors and the Monetary Transmission Mechanism," remarks presented at "Banking, Liquidity and Monetary Policy," a Symposium Sponsored by the Center for Financial Studies in Honor of Raghuram Rajan.

26. Townsend, Robert M. (1987). "Economic Organization with Limited Communication," *The American Economic Review*, 77(5): 954–971.

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

28. Williamson, Stephen (1989). "Bank Failures, Financial Restrictions, and Aggregate Fluctuations: Canada and the United States, 1870–1913," Federal Reserve Bank of Minneapolis *Quarterly Review*, Summer: 20–40.