

Andolfatto, David; Nosal, Ed

Working Paper

Shadow bank runs

Working Paper, No. 2020-14

Provided in Cooperation with:

Federal Reserve Bank of Atlanta

Suggested Citation: Andolfatto, David; Nosal, Ed (2020) : Shadow bank runs, Working Paper, No. 2020-14, Federal Reserve Bank of Atlanta, Atlanta, GA, <https://doi.org/10.29338/wp2020-14>

This Version is available at:

<https://hdl.handle.net/10419/244295>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Shadow Bank Runs

David Andolfatto and Ed Nosal

Working Paper 2020-14

August 2020

Abstract: Short-term debt is commonly used to fund illiquid assets. A conventional view asserts that such arrangements are run-prone in part because redemptions must be processed on a first-come, first-served basis. This sequential service protocol, however, appears absent in the wholesale banking sector—and yet, shadow banks appear vulnerable to runs. We explain how banking arrangements that fund fixed-cost operations using short-term debt can be run-prone even in the absence of sequential service. Interventions designed to eliminate run risk may or may not improve depositor welfare. We describe how optimal policies vary under different conditions and compare these to recent policy interventions by the Securities and Exchange Commission and the Federal Reserve. We conclude that the conventional view concerning the societal benefits of liquidity transformation and its recommendations for prudential policy extend far beyond their application to depository institutions.

JEL classification: G01, G21, G28

Key words: shadow banks, bank runs, short-term debt

<https://doi.org/10.29338/wp2020-14>

The authors thank conference participants at the second annual Missouri Macro Workshop; the 2017 Summer Workshop on Money, Banking, Payments, and Finance, at the Bank of Canada; the 2017 Canadian Macro Study Group in Ottawa; the 36th anniversary of the Diamond-Dybvig model held at Washington University in 2019; as well as seminar participants at the Federal Reserve Banks of Atlanta, Chicago, Cleveland and St. Louis, the National University of Singapore, Arizona State University, the University of Hawaii, Indiana University, Simon Fraser University, and the London School of Economics. The authors owe special thanks to Todd Keister, whose comments on a prior draft led to a number of substantial improvements. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of St. Louis, or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to David Andolfatto, Federal Reserve Bank of St. Louis, Federal Reserve Bank Plaza, St. Louis, MO 63102, david.andolfatto@stls.frb.org, or Ed Nosal, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309, ed.nosal@atl.frb.org.

Federal Reserve Bank of Atlanta working papers, including revised versions, are available on the Atlanta Fed's website at www.frbatlanta.org. Click "Publications" and then "Working Papers." To receive e-mail notifications about new papers, use frbatlanta.org/forms/subscribe.

SHADOW BANK RUNS*

David Andolfatto[†]

Federal Reserve Bank of St. Louis

Ed Nosal

Federal Reserve Bank of Atlanta

June 8, 2020

Abstract

Short-term debt is commonly used to fund illiquid assets. A conventional view asserts that such arrangements are run-prone in part because redemptions must be processed on a first-come, first-served basis. This sequential service protocol, however, appears absent in the wholesale banking sector—and yet, shadow banks appear vulnerable to runs. We explain how banking arrangements that fund fixed-cost operations using short-term debt can be run-prone even in the absence of sequential service. Interventions designed to eliminate run risk may or may not improve depositor welfare. We describe how optimal policies vary under different conditions and compare these to recent policy

*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, the 2017 Canadian Macro Study Group in Ottawa, the 36th Anniversary of the Diamond-Dybvig model held at Washington University in 2019, as well as seminar participants at the Federal Reserve Banks of Atlanta, Chicago, Cleveland and St. Louis, the National University of Singapore, Arizona State University, University of Hawaii, Indiana University, Simon Fraser University and the London School of Economics. 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 Atlanta and St. Louis, or the Federal Reserve System. JEL Codes: G01, G21, G28

[†]Federal Reserve Bank Plaza, St. Louis, MO 63102. Telephone: 314.444.4714. Email: david.andolfatto@stls.frb.org

interventions by the Security and Exchange Commission and the Federal Reserve. We conclude that the conventional view concerning the societal benefits of liquidity transformation and its recommendations for prudential policy extend far beyond their application to depository institutions.

1 Introduction

The use of short-term debt to fund illiquid assets is a common practice in financial markets. The stability of these “banking” arrangements often rests uncomfortably on the state of creditor confidence. When confidence wanes, the structure collapses. But if this is the case, then what accounts for their widespread use? And what, if anything, is the role of policy?

A traditional view contends that liquidity mismatch serves certain private interests at the expense of the broader community. Proponents of the Chicago Plan, a group of prominent economists who evidently shared this sentiment, went so far as to recommend the abolition of fractional reserve banking (Fisher 1936). The legislation that emerged at the time was, for better or worse, considerably less drastic. In particular, the Banking Act of 1935 continued to allow fractional reserve banking, but only with the support of federal deposit insurance. Several decades later, Diamond and Dybvig (1983) provided the theoretical justification for exactly this type of solution.

While deposit insurance provides the support needed to comfort small depositors, suppliers of short-term financing outside the commercial banking sector do not have similar assurances. The 2007-2008 financial crisis revealed the fragility of financial intermediaries such as money mutual funds (MMFs) and investment banks belonging to the so-called shadow banking sector. These institutions experienced runs by their lenders similar to those in the traditional retail banking sector in the pre-deposit insurance era (Bernanke 2009, Gorton 2010 and Gorton and Metrick 2010).

It is tempting to conclude that the rationale supplied by Diamond and Dybvig (1983) applies equally to the banking arrangements we observe in the wholesale sector. But the application is not so obvious. In particular, the fragility of the banking arrangements in Diamond and Dybvig (1983) relies critically on the assumption that depositor withdrawal requests must be processed on a first-come, first-served basis. This so-called sequential service constraint is notably absent in wholesale settings. And absent sequential service, the use of uninsured short-term financing is *not* unstable in the Diamond and Dybvig (1983) model.

The purpose of our paper is to demonstrate that sequential service is not necessary to render banking arrangements unstable in the Diamond and Dybvig (1983) model when the investments banks finance are subject to fixed

costs of production. Fixed costs imply increasing returns to scale at low levels of production.¹ Investments with this property can offer creditors attractive rates of return when production operates at scale. But if production is scaled back for any reason at all—including a sudden lack of funding—then the net rate of return on the underlying investment declines as unit production costs rise. The use of short-term debt to finance investments in operations with this property is potentially unstable. In particular, if confidence vanishes and investors call their loans, production collapses, unit costs rise, and the net return on the underlying investment declines—thereby justifying the initial lack of confidence. This mechanism is absent in Diamond and Dybvig (1983) because returns there are assumed to be linear.²

If our theory is correct, then Diamond and Dybvig’s (1983) view concerning the societal benefits of liquidity transformation and their recommendations for prudential policy extend far beyond their application to depository institutions. And, not surprisingly, legislators and regulators have enacted several money market reforms since the 2007-08 financial crisis. On July 23, 2014, for example, the Securities Exchange Commission announced the requirement of a floating net asset value (NAV) pricing 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 activity.³ In an earlier version of this paper (Andolfatto and Nosal 2018) we warned that NAV pricing would not in itself render money funds stable, though we expressed a more favorable view of liquidity fees and redemption gates.⁴ The strains exhibited by prime MMFs in March 2020, however, suggest we were too opti-

¹Our analysis could also be applied to the case in which increasing returns are directly a property of the business of banking itself, which seems consistent with available evidence (e.g., Mester 2008, Wheelock and Wilson 2017 and Corbae and D’Erasmo 2018).

²Ennis and Keister (2010) provide a useful survey of the literature spawned by Diamond and Dybvig (1983).

³A liquidity fee is a payment that the investor incurs to withdraw funds; a gate limits the amount of funds an investor can withdraw. See <https://www.sec.gov/News/PressRelease/Detail/PressRelease/1370542347679>. These reforms were motivated largely by an event 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. See Kacperczyk and Schnabl (2010).

⁴Our view on NAV pricing was made in contrast to Cochrane (2014, pg. 198), who expressed a more optimistic view of the stabilizing effects of NAV pricing.

mistic on this score.⁵ Calm was restored to the market only after the Federal Reserve implemented its emergency lending facilities for commercial paper and money funds on March 17-18, 2020; see Sengupta and Xue (2020). These developments suggest that for regulatory purposes, it may be necessary to treat shadow banks as *de facto* depository institutions.

While our approach is complementary to explanations of bank instability that rely on sequential service, it is of some interest to highlight their different implications along an important dimension. In particular, consider the risk-free, linear-return asset modeled by Diamond and Dybvig (1983) and, indeed, employed throughout the literature (e.g., Peck and Shell, 2003). Our model nests this standard specification when fixed costs are absent. Such an asset could reasonably be interpreted as a portfolio of illiquid U.S. Treasury securities. If so, then the Diamond and Dybvig (1983) model predicts that uninsured narrow banks and government money mutual funds are run-prone, owing to the sequential service friction. Our setup, in contrast, suggests that such intermediaries are run-proof. This is because unlike, say, the investments financed through commercial paper, the promised rate of return on U.S. Treasury securities is not sensitive to the scale of government operations.

The paper is organized as follows. Section 2 reviews an economic environment without sequential service and fixed costs that will serve as a benchmark. We demonstrate that the optimal risk-sharing arrangement in the basic environment is run-proof. In Section 3 we introduce a fixed cost into the basic environment and derive the conditions under which risk-sharing arrangements are run-prone. We then characterize the optimal risk-sharing arrangement as a function of the fixed cost and report the conditions under which an optimal risk-sharing arrangement is run-proof or run-prone. Section 4 asks if central bank liquidity provision can eliminate shadow bank runs. We summarize and conclude in Section 5.

⁵To be fair, we advocated for a rules-based policy, whereas the legislation permits fund managers to exercise discretion.

2 Shadow banks with linear returns

2.1 The environment

Our model is based on the Green and Lin (2000, 2003) version of the Diamond and Dybvig (1983) model. We begin by describing their model without sequential service and without fixed costs.

There are three dates, $t = 0, 1, 2$ and a finite number $N \geq 3$ of *ex ante* identical individuals. Individuals have preferences defined over consumption at dates 1 and 2, denoted c_1 and c_2 , respectively. Individuals receive a preference shock between $t = 0$ and $t = 1$ that determines their type: *impatient* or *patient*. An impatient individual only values consumption in date 1; let $Au(c_1)$ denote the utility payoff from consuming c_1 for an impatient individual. A patient individual is indifferent between consuming at dates 1 and 2; let $u(c_1 + c_2)$ denote the utility payoff from consuming $c_1 + c_2$ for a patient individual. Assume $u(c) = c^{1-\sigma}/(1-\sigma)$ with $\sigma > 1$. *Ex ante* preferences are given by

$$E_0 [\pi Au(c_1) + (1 - \pi)u(c_1 + c_2)], \quad (1)$$

where E_0 represents an expectations operator at $t = 0$ and $1 \leq A \leq R$ is a parameter.⁶

Each individual is endowed with y units of output at $t = 0$, which is invested. Any fraction of the investment can be liquidated and consumed at $t = 1$ at a unit rate of return. The remaining fraction of investment that is not liquidated at $t = 1$ yields rate of return $R > 1$ at $t = 2$.

Since individuals are risk averse, they have an incentive to pool risk. In what follows, we refer to a risk-sharing arrangement where sequential service is absent as a *shadow bank*.⁷ The timing of events is as follows. Individuals deposit their endowments at the bank at $t = 0$ in exchange for a contract that specifies an allocation. The allocation specifies (state-contingent) payoffs for $t = 1, 2$. Depositors learn their type between dates $t = 0$ and $t = 1$. We

⁶These preferences are equivalent to Diamond and Dybvig's (1983) preferences

$$\pi u(c_1) + (1 - \pi)\rho u(c_1 + c_2),$$

with $\rho = R^{-1} < 1$.

⁷That is, as distinct from a *retail bank*, which we think of as employing sequential service protocols.

assume that depositors can visit the bank only once, either at $t = 1$ or at $t = 2$, and that they arrive simultaneously (there is no sequential service). After the $t = 1$ redemption requests are processed, k units of investment remain. At $t = 2$, the maturing investment (interest and principal) equals Rk and is distributed to depositors arriving at date 2.

An individual is impatient with probability π . The probability that there are $0 \leq n \leq N$ impatient individuals is π_n . If types are *i.i.d.*, then

$$\pi_n = \binom{N}{n} \pi^n (1 - \pi)^{N-n},$$

where $0 < \pi_n < 1$ for all n . That is, the distribution of individuals who are impatient has full support. Given $R > 1$, it is never optimal for a patient individual to consume (and therefore visit the bank) at date 1. In addition, if type is observable, then the time and state-contingent allocation will take the form $(\mathbf{c}_1, \mathbf{c}_2) \equiv \{c_1(n), c_2(n)\}_{n=0}^N$, where n denotes the number of impatient depositors visiting the bank at date 1. Altogether, this implies that *ex ante* preferences ranking allocations can be expressed as,

$$\sum_{n=0}^N \pi_n [nAu(c_1(n)) + (N - n)u(c_2(n))]. \quad (2)$$

Feasibility requires that aggregate withdrawals at date 1 cannot be negative and cannot exceed the aggregate endowment, i.e.,

$$0 \leq nc_1(n) \leq Ny, \quad (3)$$

for all n . The resources that remain invested after early redemptions are satisfied is given by $k(n) = [Ny - nc_1(n)]$. Feasibility also requires that aggregate redemptions at date 2 cannot exceed $Rk(n)$. We assume, without loss, that all of the maturing investment is paid out to depositors at date 2, so that feasibility also respects,

$$(N - n)c_2(n) = R[Ny - nc_1(n)]. \quad (4)$$

2.2 Efficient risk-sharing

An efficient risk-sharing arrangement is an allocation $(\mathbf{c}_1, \mathbf{c}_2)$ that maximizes (2) subject to (3) and (4). Given our CES preference specification, the solu-

tion is

$$c_1^*(n) = \frac{RNy}{mR + (N - n)(R/A)^{1/\sigma}} \quad (5)$$

$$c_2^*(n) = (R/A)^{1/\sigma} c_1^*(n), \quad (6)$$

for $n = 1, 2, \dots, N - 1$, with

$$\{c_1^*(0), c_2^*(0)\} = \{0, Ry\} \quad (7)$$

$$\{c_1^*(N), c_2^*(N)\} = \{y, 0\}. \quad (8)$$

It will prove convenient to assume, with little loss of generality, $A = R$. In this case, it is optimal to equate consumption across periods on a state-by-state basis. Conditions (5) and (6) imply,

$$y < c_1^*(n) = c_2^*(n) < Ry \quad (9)$$

$$c_1^*(n + 1) < c_2^*(n), \quad (10)$$

for $n = 1, 2, \dots, N - 1$. Conditions (9) and (10) imply that the withdrawal amounts in both periods are strictly decreasing in the number of early redemptions.

2.3 Private information and the withdrawal game

We now add the restriction that type is private information. Let $m \in \{0, 1, \dots, N\}$ be the number of depositors visiting the bank at $t = 1$. Because type is private information, m is conceptually distinct from n , the true number of impatient depositors. As is standard in this literature, we restrict attention to direct mechanisms that condition $t = 1$ payouts on m , which is observable. The allocation now takes the form $(\mathbf{c}_1, \mathbf{c}_2) \equiv \{c_1(m), c_2(m)\}_{m=0}^N$.

After depositors learn their types, they play the following *withdrawal game*. Each depositor $j \in \{1, 2, \dots, N\}$ simultaneously chooses an action $t_j \in \{1, 2\}$, where t_j denotes the date depositor j visits the bank. Depositor j only knows the structure of the economy and their own type when choosing t_j . A *strategy profile* $\mathbf{t} \equiv \{t_1, t_2, \dots, t_N\}$ implies an $m \in \{0, 1, \dots, N\}$, the number of depositors that visit the bank at date $t = 1$. A *truth-telling strategy* is a strategy profile in which impatient depositors visit the bank at $t = 1$ and

patient depositors at visit at $t = 2$. If all depositors play a truth-telling strategy, then $m = n$.

An allocation $(\mathbf{c}_1, \mathbf{c}_2)$ and strategy profile \mathbf{t} constitutes a Bayes-Nash *equilibrium* to the withdrawal game if $\mathbf{t}_j = \{t_j\}$ is a best response for depositor j against $\mathbf{t}_{-j} \equiv \{t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_N\}$ for all $j \in \{1, 2, \dots, N\}$. An allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is said to be *incentive-compatible* if the truth-telling strategy is an equilibrium of the withdrawal game.

Note that it is a strictly dominant strategy for impatient depositors to visit the bank at $t = 1$ since they do not value consumption $t = 2$. If all patient depositors are expected to visit the bank at $t = 2$, then each patient depositor has no incentive to deviate from the truth-telling strategy if the following incentive-compatibility condition holds,

$$\sum_{n=0}^{N-1} \Pi^n u(c_2(n)) \geq \sum_{n=0}^{N-1} \Pi^n u(c_1(n+1)), \quad (11)$$

where Π^n is the conditional probability that there are n impatient individuals given there is at least one patient individual. Since type realizations are *i.i.d.*, we have

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

Note that condition (10) continues to hold when n is replaced by m , i.e.,

$$c_1^*(m+1) < c_2^*(m) \text{ for } m = 1, 2, \dots, N-1. \quad (12)$$

Condition (12) implies that regardless of how other depositor report their type via their choice of withdrawal date, a patient depositor will always receive a higher level of consumption by withdrawing at $t = 2$. This implies that allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ satisfies the incentive-compatibility condition (11) with strict inequality.

Conclusion 1 *The efficient risk-sharing arrangement is incentive-compatible.*

2.4 Run-proof and run-prone allocations

We label an allocation $(\mathbf{c}_1, \mathbf{c}_2)$ that satisfies (3), (4) and (11) *incentive-feasible*. It is clear that for any incentive-feasible allocation, there exists

an equilibrium where all depositors play the truth-telling strategy. There *may*, however, exist other equilibrium outcomes associated with the withdrawal game. Of particular interest is an equilibrium where depositors play a *run strategy* as defined by the strategy profile $\mathbf{t} \equiv \{1, 1, \dots, 1\}$. That is, a run strategy implies that all depositors visit the bank at $t = 1$. We say that an incentive-feasible allocation is *run-prone* if it admits a run strategy as an equilibrium and is *run-proof* if it does not.

Consider an incentive-feasible allocation $(\mathbf{c}_1, \mathbf{c}_2)$ with the property,

$$c_2(N - 1) < c_1(N). \quad (13)$$

It is straightforward to show that such allocations exist. Since $(\mathbf{c}_1, \mathbf{c}_2)$ is incentive-feasible, truth-telling is an equilibrium for the withdrawal game. But the allocation is also run-prone. To see this, consider patient depositor j 's best response to a proposed run strategy profile $\mathbf{t} = \mathbf{1}$. If j visits the bank early, $t_j = 1$, then his return is y since everyone else visits at $t = 1$. If, instead, j visits the bank at later, $t_j = 2$, his return is $c_2(N - 1) < c_1(N)$. Depositor j will therefore choose to visit at $t = 1$, which implies that a run equilibrium exists.⁸ It follows that a run-proof allocation must satisfy,

$$c_2(N - 1) \geq c_1(N). \quad (14)$$

2.5 The efficient risk-sharing arrangement is run-proof

While a large set of incentive-feasible allocations are run-prone, we want to establish the stability properties of the efficient incentive-feasible allocation. As it turns out, Green and Lin (2000, 2003) have already shown that the efficient risk-sharing arrangement here is run-proof.

To see this, consider the solution presented in (5)–(6) and, in particular, the property of this solution as described by condition (12). This latter condition implies that truth-telling is a strictly dominant strategy, so that the allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ strictly satisfies the incentive-compatibility condition (11). In particular, note that condition (12) combined with the solution

⁸Interestingly, incentive-feasible allocations can be run-prone *even in the absence of sequential service*. Note that an inefficient allocation (contractual arrangement) is not necessarily run-prone. The autarkic allocation, for example, is both inefficient and run-proof.

implies,

$$c_2^*(N-1) = \left[\frac{RN}{(N-1)R+1} \right] y > c_1^*(N) = y,$$

where the inequality follows because $R > 1$. This latter condition is the requirement for stability, (14). It therefore follows that $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ can be uniquely implemented as a truth-telling equilibrium.

Note that the allocation $(\mathbf{c}_1^*, \mathbf{c}_2^*)$ is run-proof even though liquidity preference is not observable and depositors can withdraw on demand. Intuitively, the efficient allocation places increasingly stringent limits on the amount that can be withdrawn at date 1 as the number of depositors requesting withdrawal (m) increases. That is, the withdrawal amount $c_1^*(m)$ is decreasing in m so that sufficient resources, $R[yN - c_1^*(m)m]$, remain those who would rather withdraw at a later date. Evidently, patient depositors are assured of superior payoffs by waiting so that (10) does not bind.

Conclusion 2 *Absent fixed costs of production and sequential service, the efficient risk-sharing arrangement is run-proof.*

The result above supports Diamond and Dybvig's (1983) conclusion that a banking system with deposit insurance is run-proof. To make the mapping between their paper and ours, note that our mechanism-design approach takes no stand on how economic organization is divided across private and public sectors. This is in contrast to Diamond and Dybvig (1983) who assume that while banks are subject to sequential service, the government is not. Because banks are subject to sequential service, the efficient and run-proof risk-sharing arrangement we describe above is not feasible for banks. If banks attempted to replicate the efficient allocation, sequential service would render the arrangement run-prone. The public sector in their model is not subject to sequential service. In particular, taxes can be used to recover funds that depositors have already withdrawn (sequential service implies that a private bank has no such recourse). In this way, a tax-supported deposit insurance scheme can support private banks in a way that renders the efficient risk-sharing arrangement run-proof. Our mechanism-design approach effectively consolidates their private-public sector risk-sharing arrangement.

The Diamond and Dybvig (1983) model is appealing, in part, because the vision of desperate depositors forming physical queues to withdraw their funds resonates for those of us familiar with the way people behaved in

historical retail-level bank runs. Sequential service, however, seems a poor description of how exchange occurs at the wholesale level. For example, money mutual funds trade once at the end of every business day. Since all buy and sell orders for the fund are simultaneously executed and all sell orders receive the same per share payout, there is no sequential service associated with withdrawal requests submitted in a given period.⁹

Since some types of mutual funds exhibit the symptoms of being run-prone (even those with NAV-pricing protocols), some property other than sequential service must be responsible this apparent instability.

3 Banking with fixed costs

Fixed costs are a salient feature of virtually every organization and production process, including financial intermediation. In what follows, we assume that production is subject to a fixed cost.

3.1 The efficient risk-sharing arrangement

We model the fixed cost in production as follows. An investment of $k \geq 0$ units of output at date 1 now delivers $\max\{Rk - \kappa, 0\}$ units of output at date 2, where $\kappa \geq 0$ is the fixed cost parameter, where $\kappa < RNy$.¹⁰ The date 2 feasibility constraint, formerly (4), is now given by

$$(N - n)c_2(n) = R[Ny - nc_1(n)] - \kappa. \quad (15)$$

The efficient risk-sharing arrangement is now an allocation that maximizes (2) subject to (3) and (15). Let $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) \equiv \{c_1^\kappa(m), c_2^\kappa(m)\}_{m=0}^N$ denote the efficient risk-sharing arrangement when $m = n$. This allocation inherits

⁹The per unit payout that sellers receive may depend on how many other fund investors want to sell. Nevertheless, whatever that amount is, each seller receives the same payout per share sold. Our model of shadow banking is consistent with this pricing protocol. It is also consistent with the way short-term repo arrangements work. Specifically, the repo lender is repaid either in cash or collateral, without any sequential service consideration.

¹⁰If $\kappa > RNy$, investment will never be undertaken. In this case the economy is characterized by autarky, where all individuals consume y at date 1.

the same qualitative form as (5) and (6) which is given by

$$c_1^\kappa(m) = \frac{RNy - \kappa}{mR + N - m} \quad (16)$$

$$c_2^\kappa(m) = c_1^\kappa(m), \quad (17)$$

for $m = 1, 2, \dots, N - 1$, together with

$$\{c_1^\kappa(0), c_2^\kappa(0)\} = \{0, Ry - \kappa/N\} \quad (18)$$

$$\{c_1^\kappa(N), c_2^\kappa(N)\} = \{y, 0\}. \quad (19)$$

Note that $c_1^\kappa(m)$ and $c_2^\kappa(m)$ as defined in (16)-(17) are strictly decreasing in the number of early redemptions for $m = 1, 2, \dots, N - 2$. This, in turn, implies

$$c_1^\kappa(m + 1) < c_2^\kappa(m) \quad (20)$$

for $m = 1, 2, \dots, N - 2$. Condition (20) is similar to condition (12) except that the former condition does not *necessarily* hold for $m = N - 1$. In particular, for $m = N - 1$ we have

$$c_2^\kappa(N - 1) = \frac{RNy - \kappa}{(R - 1)(N - 1) + N} \gtrless c_1^\kappa(N) = y, \quad (21)$$

where the direction of the inequality above is determined by the size of the fixed cost, κ . Define κ_0 as the solution to

$$c_2^{\kappa_0}(N - 1) = y, \quad (22)$$

which is easily calculated to be

$$\kappa_0 = (R - 1)y > 0. \quad (23)$$

Note that (23) can be rearranged to $y > Ry - \kappa_0$. This inequality has an important implication. In particular, a patient individual in autarky that chooses between consuming y at date 1 or investing y and consuming $Ry - \kappa$ at date 2, will choose to consume at date 1 when $\kappa > \kappa_0$.

Is the efficient risk-sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ incentive compatible? It follows that for small fix costs, $\kappa < \kappa_0$, condition (20) is identical to condition (12) and the efficient risk-sharing arrangement is incentive-compatible on a state-by-state basis. Hence, the incentive-compatibility condition (11) is

satisfied. When fixed costs are large, $\kappa > \kappa_0$, the state-by-state incentive-compatibility conditions are violated (only) in state $m = N - 1$. We show in an Appendix that incentive-compatibility condition (11) is not likely to bind for economically relevant parameters when fixed costs are large, $\kappa > \kappa_0$.

Conclusion 3 *The efficient risk-sharing arrangement $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is incentive-compatible.*

We now identify the range of fixed costs for which risk-sharing, i.e., shadow banking, remains an economically viable proposition. Intuitively, there should exist a cut-off value $\kappa_1 > \kappa_0$ such that autarky is weakly preferred to the efficient risk-sharing arrangement $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ for any $\kappa \geq \kappa_1$. When $\kappa > \kappa_0$, the payoffs associated with autarky are $Ru(y)$ and $u(y)$ for impatient and patient individuals, respectively (since a patient individual prefers to consume y at date 1 than $Ry - \kappa$ at date 2). For convenience we denote the autarky allocation as $(\mathbf{c}_1^a, \mathbf{c}_2^a)$. The expected (date 0) payoff associated with autarky, $V(\mathbf{c}_1^a, \mathbf{c}_2^a)$, is

$$V(\mathbf{c}_1^a, \mathbf{c}_2^a) \equiv V^A \equiv \sum_{n=0}^N \pi_n [(R-1)n + N] u(y). \quad (24)$$

The critical cut-off κ_1 is determined by equating V^A with $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$, the expected utility associated with the efficient allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when depositors play truth-telling strategies.¹¹ The latter is defined by,

$$V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) \equiv \sum_{n=0}^N \pi_n [(R-1)n + N] u[c^\kappa(n)], \quad (25)$$

where $c^\kappa(n) = c_1^\kappa(n) = c_2^\kappa(n)$ for $n = 1, 2, \dots, N - 1$, $c_1^\kappa(0) = 0$ and $c_2^\kappa(0) = Ry - \kappa/N$ and $c_1^\kappa(N) = y$ and $c_2^\kappa(N) = 0$. Since $V(\mathbf{c}_1^0, \mathbf{c}_2^0) > V^A$ and $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is monotonically decreasing in κ , there exists a $0 < \kappa_1 < RNy$ that satisfies $V(\mathbf{c}_1^{\kappa_1}, \mathbf{c}_2^{\kappa_1}) = V^A$. If depositors play truth-telling strategies, allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is preferred to autarky only if $\kappa \leq \kappa_1$.

We now establish that $\kappa_0 < \kappa_1$. By construction, $c^{\kappa_0}(N-1) = c^{\kappa_0}(N) = y$ and $c^{\kappa_0}(m) > y$ for all $m \leq N - 1$. Therefore, $V(\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0}) > V^A$. Since

¹¹The operator $V(\cdot)$ evaluates the expected utility of an incentive-feasible allocation assuming that agents play truth-telling strategies.

$V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is strictly decreasing in κ and $V(\mathbf{c}_1^{\kappa_1}, \mathbf{c}_2^{\kappa_1}) = V^A$, it follows that $\kappa_0 < \kappa_1$. In what follows, we restrict attention to fixed costs less than κ_1 since $\kappa > \kappa_1$ implies that autarky is the equilibrium outcome.

Conclusion 4 *The efficient risk-sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is strictly preferred to autarky for any fixed cost $\kappa < \kappa_1$, where $\kappa_0 < \kappa_1 < RNy$ satisfies $V(\mathbf{c}_1^{\kappa_1}, \mathbf{c}_2^{\kappa_1}) = V^A$.*

To this point, we have demonstrated that the efficient risk-sharing arrangement delivers gains to trade and is incentive-compatible for all fixed costs in the range $0 \leq \kappa < \kappa_1$. Next, we check the stability properties of the allocation. As it turns out, some of the work has already been done in this regard. In particular, since $c_1^\kappa(N) = y$ for all κ , $c_2^{\kappa_0}(N-1) = y$ by condition (22) and $c_2^\kappa(N-1)$ is decreasing in κ , it follows that $c_2^\kappa(N-1) \geq c_1^\kappa(N)$ for all $\kappa \leq \kappa_0$ and that $c_2^\kappa(N-1) < c_1^\kappa(N)$ for all $\kappa > \kappa_0$. These conditions correspond to the run-proof (14) and run-prone (13) conditions, respectively. Hence,

Conclusion 5 *When fixed costs are small, $0 \leq \kappa \leq \kappa_0$, the efficient risk-sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is run-proof. When fixed costs are large, $\kappa_0 < \kappa \leq \kappa_1$, the efficient risk-sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is run-prone.*

3.2 Can a run-prone bank attract depositors?

Diamond and Dybvig (1983, pp. 409-410) suggest that investors may be willing to fund run-prone banks if run risk is sufficiently small because the risk-sharing services delivered in the non-run event dominates what is available in autarky (direct investing). They go on to suggest that this explains why such arrangements are used in spite of the danger of runs. The same rationale applies in our analysis if we assume depositors must choose between a run-prone arrangement and autarky.

In what follows, we adopt the “sunspot” equilibrium concept described in Peck and Shell (2003) and also alluded to in Diamond and Dybvig (1983, pg. 410). That is, assume there exists an extrinsic event—a “sunspot”—that is observed by all depositors with some probability θ . A sunspot is a theoretical device that coordinates depositors’ beliefs about how all other depositors are

expected to behave in the withdrawal game.¹² The sunspot, if it occurs, is observed after individuals deposit their endowments but before they learn their types.

A *sunspot equilibrium* is characterized by an incentive-feasible allocation $(\mathbf{c}_1, \mathbf{c}_2)$, a probability θ , and a set of strategy profiles contingent on the occurrence of the sunspot, with the following properties. When the sunspot is not observed, each depositor believes that the other $N - 1$ depositors play truthfully: since $(\mathbf{c}_1, \mathbf{c}_2)$ is incentive-compatible, in equilibrium, impatient depositors visit the bank at date 1 and patient depositors at date 2. When the sunspot is observed, each depositor believes that the other $N - 1$ depositors will visit the bank period 1. If allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is characterized by $c_2^\kappa(N - 1) < c_1^\kappa(N) = y$, then, in equilibrium, depositors will play run strategies.

Let $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta)$ denote the expected utility associated with allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when depositors play the sunspot strategies just described, i.e.,

$$W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta) \equiv (1 - \theta)V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) + \theta V^A. \quad (26)$$

Notice that allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ generates an expected utility that is higher than the autarkic allocation $(\mathbf{c}_1^a, \mathbf{c}_2^a)$ since $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta) > V^A$ for all $\theta < 1$ with $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa, 1) = V^A$.¹³ It follows that the efficient risk-sharing arrangement will attract depositors *regardless run risk* because, in this environment at least, the most pain a run can inflict on depositors is the autarkic outcome.

Conclusion 6 *The efficient risk-sharing arrangement $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ will attract depositors even if the allocation is run-prone, independent of the perceived probability of a run.*

The discussion above is predicated on the assumption that depositors can only choose between the efficient risk-sharing arrangement and autarky.

¹²There is nothing here that fundamentally determines the value of θ . It may or may not correspond to the actual probability of observing the sunspot. We interpret θ as indexing the cultural propensity of a society—depositors in this case—to lose faith in their fellow citizens. That is, $1 - \theta$ measures the degree to which a community is confident that each will “do the right thing” when the time comes.

¹³This result is in contrast to Peck and Shell (2003), where in their model if θ is sufficiently large individuals strictly prefer autarky. This is because their model assumes sequential service, which implies that the expected utility associated with playing the run equilibrium is strictly less than autarky.

There is, in fact, another possibility to consider. In particular, for fixed costs in the range $\kappa_0 < \kappa \leq \kappa_1$, it is feasible to construct a run-proof allocation at the expense of some risk-sharing. In what follows, we interpret such an allocation as arising from a regulatory intervention designed to render the shadow bank sector more stable.

3.3 Run-proof shadow banks

The best run-proof risk-sharing arrangement is an allocation that maximizes (2) subject to (3), (11), (14) and (15). The solution to this problem, $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$, is given by $\{\hat{c}_1^\kappa(m), \hat{c}_2^\kappa(m)\} = \{c_1^\kappa(m), c_2^\kappa(m)\}$ for all $m \neq N - 1$ and

$$\begin{aligned}\hat{c}_1^\kappa(N - 1) &= \frac{RNy - \kappa - y}{R(N - 1)} < y \\ \hat{c}_2^\kappa(N - 1) &= y.\end{aligned}\tag{27}$$

The allocation in state $m = N - 1$, given by (27), penalizes those who wish to withdraw funds early. As such, it can be thought of as a “liquidity fee,” equal to $y - (RNy - \kappa - y)/[R(N - 1)]$, similar to what prime institutional funds are now permitted to apply at the discretion of fund managers. If the application of the liquidity fee is credible (something that may need to rely on legislation, rather than the discretion of management), then it ensures that sufficient resources will be made available for those not in dire need of liquidity. Notice that allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ is characterized by $\hat{c}_2^\kappa(m) \geq \hat{c}_1^\kappa(m + 1)$ for all $n \leq N - 1$, which implies that it is a dominant strategy for patient depositors to visit the bank at date 2. This property, in turn, implies that allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ can be implemented as a unique equilibrium.

There remains the question of whether such an intervention might actually improve depositor welfare. To explore this possibility, let $V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ denote the expected utility payoff associated with the allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ when depositors play truthtelling strategies, where

$$\begin{aligned}V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) &\equiv \sum_{n=0}^{N-2} \pi_n [(R - 1)n + N] u[c^\kappa(n)] \\ &\quad + \pi_{N-1} \{(N - 1)Ru[\hat{c}_1^\kappa(N - 1)] + u(y)\} + \pi_N NRu(y).\end{aligned}\tag{28}$$

Since allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ departs from optimal risk-sharing in state $n = N - 1$ only, we necessarily have: (i) $W(\mathbf{c}_1^{\kappa_1}, \mathbf{c}_2^{\kappa_1}; \theta) = V^A > V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ and (ii)

$V(\hat{\mathbf{c}}_1^{\kappa_0}, \hat{\mathbf{c}}_2^{\kappa_0}) = V(\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0}) > V^A$ since, by construction, $(\hat{\mathbf{c}}_1^{\kappa_0}, \hat{\mathbf{c}}_2^{\kappa_0}) = (\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0})$. Because $V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ is decreasing in κ , there exists a fixed cost $\hat{\kappa}$ that satisfies $V(\hat{\mathbf{c}}_1^{\hat{\kappa}}, \hat{\mathbf{c}}_2^{\hat{\kappa}}) = V^A$, where $\kappa_0 < \hat{\kappa} < \kappa_1$. It follows that for $\kappa \in (\kappa_0, \hat{\kappa})$, we have $V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) > V^A$ and for $\kappa \in (\hat{\kappa}, \kappa_1)$, we have $V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) < V^A$. The best run-proof allocation therefore delivers expected utility $V^{RP} = \max\{V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa), V^A\}$, where the best allocation is $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ when $\kappa \in (\kappa_0, \hat{\kappa}]$ and $(\mathbf{c}_1^a, \mathbf{c}_2^a)$ when $\kappa \in (\hat{\kappa}, \kappa_1)$.

Conclusion 7 *The best run-proof risk-sharing arrangement $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ is preferred to autarky for fixed costs in the range $\kappa_0 < \kappa \leq \hat{\kappa}$, and is autarkic for fixed costs $\hat{\kappa} < \kappa \leq \kappa_1$, where $\hat{\kappa}$ solves $V(\hat{\mathbf{c}}_1^{\hat{\kappa}}, \hat{\mathbf{c}}_2^{\hat{\kappa}}) = V^A$.*

The result above implies that the intervention harms depositor welfare if fixed costs are in the range $\hat{\kappa} < \kappa \leq \kappa_1$. This is because first, the best run-proof allocation is autarkic in this range of fixed costs and second, run-prone allocations generically dominate autarky. However, for fixed costs less than $\hat{\kappa}$, the run-proof allocation does offer a considerable degree of risk-sharing together with the benefit of ensured stability. It is therefore possible that the intervention improves welfare over some regions of the parameter space. We now demonstrate that this is indeed the case.

Consider now fixed cost is in range $\kappa \in (\kappa_0, \hat{\kappa}]$. Then we know $V^{RP} = V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) > V^A$, which implies that the best run-proof allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ is preferred to autarky. Since $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \theta)$ is strictly decreasing in θ and since $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; 0) > V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) > W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; 1) = V^A$, there exists a critical probability $\hat{\theta}(\kappa)$ such that $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \hat{\theta}(\kappa)) = V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$. If the probability of a run is less than this critical value, $\theta < \hat{\theta}(\kappa)$, then $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \theta) > V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$. In this situation, depositors prefer the run-prone allocation, so that the intervention reduces welfare. However, if the probability of a run is higher than the critical cutoff value, $\theta > \hat{\theta}(\kappa)$, then $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \theta) < V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$. In this situation, the intervention resulting in the run-proof allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ improves depositor welfare.¹⁴

Conclusion 8 *A run-proof shadow bank improves depositor welfare for run risks θ and fixed costs κ satisfying $\kappa_0 < \kappa \leq \hat{\kappa}$ and $\theta > \hat{\theta}(\kappa)$, where $\hat{\theta}(\kappa)$*

¹⁴We have chosen to label this situation as a government intervention but, of course, it may also be an arrangement that emerges voluntarily from within the shadow bank. Either way, what is important is the efficient contractual form and not on its source (i.e., whether the regulatory protocols emerge from within or from without the organization).

satisfies $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \hat{\theta}(\kappa)) = V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$. For any κ in this range, a run-proof shadow bank lowers depositor welfare if the run risk is sufficiently small, $\theta < \hat{\theta}(\kappa)$.

Let us examine the properties of $\hat{\theta}(\kappa)$, which is defined by $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \hat{\theta}(\kappa)) = V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ or

$$[1 - \hat{\theta}(\kappa)]V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) + \hat{\theta}V^A = W[\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \hat{\theta}(\kappa)] \quad (29)$$

over the range $\kappa \in [\kappa_0, \hat{\kappa}]$. Using (29), we can solve explicitly for this critical value,

$$\hat{\theta}(\kappa) = \frac{V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) - W[\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa; \hat{\theta}(\kappa)]}{V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) - V^A} \quad (30)$$

It is straightforward to show that $\hat{\theta}(\kappa)$ is increasing in κ when $\kappa \in [\kappa_0, \hat{\kappa}]$. In other words, as the fixed cost is increased over the range $\kappa \in [\kappa_0, \hat{\kappa}]$, the perceived run risk has an ever-higher hurdle to clear before run-proofing the risk-sharing arrangement improves depositor welfare. As we established earlier, for fixed costs beyond $\hat{\kappa}$, run-proofing the risk-sharing arrangement reduces depositor welfare for any $\theta < 1$.

3.4 Endogenous complacency

Suppose that a society is characterized by a “true” θ but that the true value is unknown. Suppose further that the economy described above is repeated over time, $t = 0, 1, 2, \dots, \infty$. People enter into the initial period with a prior belief of the value of θ , say, θ_0 . Let $s_t \in \{0, 1\}$ denote the sunspot variable observed at date t , where $s_t = 1$ indicates that a sunspot appeared. Then it is reasonable to expect people to form posterior beliefs that depend on the history of sunspot realizations together with their initial belief, i.e.,

$$\theta_t = \Pr[\theta \mid s_{t-1}, s_{t-2}, \dots, s_1, \theta_0] \quad (31)$$

In case that θ does not change over time (unlikely), and that prior beliefs are passed on to future generations in a perfect manner (unlikely), then society can reasonably be expected to learn of the true propensity to run over a sufficiently long period of time. If the true θ is close to zero, even in this optimistic scenario, learning could take a very long time.

The choice of the best allocation described above now takes place with θ_t replacing θ . Imagine a scenario at some date t in which $\kappa_0 < \kappa \leq \hat{\kappa}$ and $\theta_t > \hat{\theta}(\kappa)$, which is a parameter configuration where depositors are made better off by choosing the efficient risk-sharing run-proof arrangement, $(\hat{c}_1^\kappa, \hat{c}_2^\kappa)$. Next, imagine that depositors do not observe sunspots for several subsequent periods. Under any reasonable learning protocol, the posterior belief θ_t will decline over time. After a sufficient period of tranquility, it is possible that θ_t falls below the threshold $\hat{\theta}(\kappa)$. At that time depositors will lobby to have the onerous regulation relaxed. If successful, the run-proof shadow bank becomes run-prone, as run risk is perceived to be sufficiently small. If and when a sunspot with its associated run is actually observed, θ_t may very well jump back over the threshold $\hat{\theta}(\kappa)$, leading to calls for stricter regulation.

4 Central bank liquidity provision

As already mentioned, the intervention we described above resembles a voluntary/mandatory liquidity fee and NAV pricing stipulations similar to what is presently imposed on institutional money market funds. As we also mentioned in the introduction, these provisions seemed insufficient to stem the growing financial panic associated with COVID-19 pandemic. In March of 2020, the U.S. Federal Reserve intervened to restore calm.

To think about how a central bank might be modeled in this economy, we step beyond the “closed” system studied above and introduce another actor into the model that we interpret as a central bank or treasury. The central bank is endowed with resources that can be used to provide “liquidity” to the shadow bank at date 1. The way this is done is via a repo contract with the shadow bank. For simplicity, we assume that the value of the repo contract to the central bank is zero. Therefore, a repo contract specifies that the central bank exchanges x units of output at date 1 for a claim on the shadow bank’s output at date 2 of equal value, x . The contract is unwound at date 2 when the shadow bank repurchases its claim for x units of output.

The resources the central bank brings to the financial market can potentially help shadow banks manage risk over all possible states of the world. But here, we assume, realistically, that the central bank’s only goal is to prevent banks runs with the minimal “footprint” possible. Specifically, subject to eliminating fragility, the central bank minimizes both the number of states

m it provides liquidity support to the shadow bank and the level of liquidity provision x_m .

As a result, the central bank never intervenes when the shadow bank is run-proof. In our model, the shadow bank is run-proof if: (i) the fixed cost κ is less than the critical value κ_0 ; or (ii) $\kappa_0 < \kappa < \hat{\kappa}$ and the equilibrium is characterized by the run-proof allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$. Regarding (ii), the equilibrium allocation is $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ whenever the probability of a sunspot θ is less than $\hat{\theta}(\kappa)$. If conditions (i) or (ii) are not met, then the equilibrium—in the absence of central bank intervention or regulation—is characterized by the run-prone allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$. This allocation is run-prone because $c_2^\kappa(N-1) < y$.

When the equilibrium is characterized by allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$, the shadow bank does not offer the run-proof allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ because the expected utility associated with it, $V(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$, is less than the payoff associated with the run-prone allocation, $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta)$. The central bank may be able to eliminate shadow bank fragility by providing liquidity $x_m \geq 0$ to the shadow bank when m depositors withdraw at date 1. Intuitively, the provision of liquidity may increase the expected utility associated with a run-proof contract to the point where it equals or exceeds that of the run-prone allocation. We now investigate this possibility.

An allocation is run-proof if the payoff to the single depositor who visits the shadow bank at date 2 satisfies $c_2(N-1) \geq y$. Consider the best run-proof allocation when the central bank provides liquidity $x > 0$ if $m = N-1$ via a repo contract, with $x = 0$ otherwise. Denote this allocation as $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ and assume that $[\tilde{c}_1^\kappa(m), \tilde{c}_2^\kappa(m)] = [c_1^\kappa(m), c_2^\kappa(m)]$ for all $m \neq N-1$. To ensure that this allocation is run-proof, the date 1 investment must produce output at date 2 equal to at least $y + x + \kappa$. That is, depositors arriving at date 2 must be promised at least y , with x repaid to the central bank and κ to cover the fixed cost. The allocation for $m = N-1$ can be determined as follows. Since the amount of resources that will be invested at date 1 equals $Ny + x - (N-1)\tilde{c}_1(N-1)$, to ensure that x is repaid, fixed cost κ is covered and that depositors arriving at date 2 receive y , it must be the case that $R[Ny + x - (N-1)\tilde{c}_1(N-1)] = y + x + \kappa$. Hence we have

$$\begin{aligned}\tilde{c}_1^\kappa(N-1) &= \frac{RNy + (R-1)x - \kappa - y}{R(N-1)} \\ \tilde{c}_2^\kappa(N-1) &= y.\end{aligned}\tag{32}$$

Notice that the date 1 payoff $\tilde{c}_1^\kappa(N-1)$ is strictly increasing in x and that if $x = 0$, then $\tilde{c}_1^\kappa(N-1) = \hat{c}_1^\kappa(N-1)$, i.e., compare (32) with (27).

Suppose the central bank chooses x so that the expected utility associated with the (run-proof) allocation $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ equals the expected utility associated with run-prone allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$, i.e., $V(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa) = W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta)$.¹⁵ Since: (i) $V(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa) = W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta)$; (ii) $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) > V(\mathbf{c}_1^a, \mathbf{c}_2^a)$; (iii) $[\tilde{c}_1^\kappa(m), \tilde{c}_2^\kappa(m)] = [\hat{c}_1^\kappa(m), \hat{c}_2^\kappa(m)]$ for all $m \neq N-1$; and (iv) $\tilde{c}_2^\kappa(m) = y > \hat{c}_2^\kappa(m)$, we necessarily have $\tilde{c}_1^\kappa(N-1) < \hat{c}_1^\kappa(N-1)$. This latter result is important because it implies that allocation $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ is state-by-state incentive-compatible. More specifically, since we constructed $\tilde{c}_1^\kappa(N-1)$ to ensure that $\tilde{c}_2^\kappa(N-1) = y$, we have to check that $\tilde{c}_2^\kappa(N-2) > \tilde{c}_1^\kappa(N-1)$.¹⁶ This inequality is valid since $\tilde{c}_2^\kappa(N-2) = c_2^\kappa(N-2) > c_1^\kappa(N-1) > \tilde{c}_1^\kappa(N-1)$. Hence, a footprint-minimizing central bank can eliminate shadow bank fragility by providing liquidity $x > 0$ only in state $m = N-1$, where x is chosen to equate $V(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ with $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta)$.

5 Summary and conclusion

While short-term debt provides creditors with the flexibility they desire, it exposes debtors to the possibility of runs. Diamond and Dybvig (1983) provide a theory that simultaneously explains the benefit of liquidity transformation and why banks are run-prone. The fundamental properties of the environment responsible for instability are private information and sequential service. While it is always possible to render risk-sharing arrangements run-proof, doing so may come at a cost (Peck and Shell, 2003).

Our paper explores whether there are properties other than sequential service that may be responsible for banking instability. We were motivated by the fact that shadow banks in the wholesale sector often appear to display the fragility highlighted by Diamond and Dybvig (1983) but without the sequential service protocol characteristic of retail-level banking. We identified a heretofore neglected property of real-world organizations; namely, the existence of fixed costs of production. When fixed costs are present, a large

¹⁵The central bank intervention here is structured in a way to eliminate the possibility of a run without benefiting the shadow bank or its investors.

¹⁶By construction, all of the other states are state-by-state incentive-compatible.

contraction in the level of output increases unit costs of production and lowers the net return investors can expect on the underlying project. If a bank's asset portfolio consists of loans to organizations subject to fixed costs, then the use of short-term financing can be run-prone. If loans are called, production collapses, unit costs rise, and net returns decline. In this way, a collapse in investor confidence can become a self-fulfilling prophecy.¹⁷

We think our explanation of run-prone financial arrangements compares favorably to Diamond and Dybvig (1983) because in our model, unlike in theirs, a bank run leads to an actual deterioration in asset quality. As we already mentioned, the asset in Diamond and Dybvig (1983) and Peck and Shell (2003) could be interpreted as a U.S. Treasury security. The runs that occur in those models are largely about depositors scrambling for a share of an asset whose quality remains invariant to the volume of early redemptions.

Nevertheless, our theory is broadly consistent with the policy implications that emerge from Diamond and Dybvig (1983) and Peck and Shell (2003) and indeed, as we argued above, these implications could be extended to a broader set of intermediaries than commonly considered. Interventions that render run-prone financial arrangements run-proof may or may not be consistent with improving depositor welfare. As is often the case, it is difficult to make conclusive statements about the merits of any given intervention without knowing approximately where we are in the parameter space. It is also possible, as explained above, that our perceptions of where the economy is located in the parameter space evolves over time. In particular, our assessment of run risk may depend mostly on whether we experienced financial crisis in recent history. This latter possibility adds an additional complicating factor to an already difficult regulatory problem.

¹⁷Although a scale economy assumption may in itself imply multiple equilibria in a game where *ex ante* individuals decide *how much to deposit* at the bank, as in Cooper and Corbae (2002), this sort of multiple equilibrium/instability result is conceptually different from our withdrawal game. Multiple equilibria in our withdrawal game relies on type-misrepresentation *after* funds are deposited in the bank.

6 References

1. Andolfatto, David and Nosal, Ed (2018). “Bank Runs without Sequential Service.” Federal Reserve Bank of Atlanta, Working Paper 2018-6.
2. Andolfatto, David, Nosal, Ed and Bruno Sultanum (2017). “Preventing Bank Runs,” *Theoretical Economics*,12: 1003-1028.
3. Bernanke, Ben S. (2009). “Opening Remarks: Reflections on a Year of Crisis,” *Federal Reserve Bank of Kansas City’s Annual Economic Symposium, Jackson Hole, Wyoming, August 21*.
4. Cochrane, John (2014). “Toward a Run-free Financial System,” in *Across the Great Divide: New Perspectives on the Financial Crisis*, Chapter 10, pp. 197-249, edited by Martin N. Baily and John B. Taylor, Hoover Institution Press, Stanford University, Stanford, California.
5. Cooper, Russell and Dean Corbae (2002). “Financial Collapse: A Lesson from the Great Depression,” *Journal of Economic Theory*, 107: 159–190.
6. Corbae, Dean and Pablo D’Erasmus (2018). “Capital Requirements in a Quantitative Model of Banking Industry Dynamics,” unpublished manuscript.
7. Diamond, Douglas and Philip Dybvig (1983). “Deposit Insurance and Liquidity,” *Journal of Political Economy*, 91(June): 401–419.
8. Ennis, Huberto M. and Todd Keister (2010). “On the Fundamental Reasons for Bank Fragility,” FRB Richmond *Economic Quarterly* 96, First Quarter: 33-58.
9. Fisher, Irving (1936). “100% Money and Public Debt.” *Economic Forum*, Spring Number, April-June: 406-420.
10. Gorton, Gary (2010). *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press.
11. Gorton, Gary and Metrick, Andrew (2010). “Regulating the Shadow Banking System,” *Brookings Papers on Economic Activity*, Fall: 261-312.

12. 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.
13. Green, Edward and Ping Lin (2003). “Implementing Efficient Allocations in a Model of Financial Intermediation,” *Journal of Economic Theory*, 109(1); 1–23.
14. 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.
15. Mester, Loretta (2008). “Optimal Industrial Structure in Banking,” Chapter 5 in the *Handbook of Financial Intermediation and Banking*, 133–162.
16. Peck, James and Karl Shell (2003). “Equilibrium Bank Runs,” *Journal of Political Economy*, 111(1): 103–123.
17. Sengupta, Rajdeep and Xue, Fei (2020). “The Global Pandemic and Run on Shadow Banks.” *Main Street Views*, Federal Reserve Bank of Kansas City, May 11.
18. Wheelock, David and Paul Wilson (2017). “The Evolution of Scale Economies in U.S. Banking,” Federal Reserve Bank of St. Louis working paper 2015-021C.

7 Appendix: Incentive compatibility

We can not find model parameters where allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ violates (11). To provide some insight for this outcome, rewrite constraint (11) as

$$\sum_{n=0}^{N-2} \Pi^n \{u[c_2^\kappa(n)] - u[c_1^\kappa(n+1)]\} \geq \pi^{N-1}(1-\pi)\{u(y) - u[c_2^\kappa(N-1)]\}. \quad (33)$$

Notice that the differences on the left and right sides are strictly positive and the difference on the right side is weighted by $\pi^{N-1}(1-\pi)$ while the aggregate weight on the left side is $1 - \pi^{N-1}(1-\pi) \gg \pi^{N-1}(1-\pi)$. If, for example, N is reasonably large—and “large” can be as small as $N = 2$ —then since the right-side weight is very small compared to the sum of weights on the left side, incentive constraint (11) is satisfied. We now formalize this intuition by example.

It is straightforward to show that when $u(c) = c^{1-\sigma}/(1-\sigma)$, $u[c_2^\kappa(n)] - u[c_1^\kappa(n+1)]$ is decreasing. This implies that

$$\begin{aligned} & \sum_{n=0}^{N-2} \Pi^n \{u[c_2^\kappa(n)] - u[c_1^\kappa(n+1)]\} > \\ & [1 - \pi^{N-1}(1-\pi)] \frac{1}{\sigma-1} \left[\left(\frac{(N-1)R+1}{RNy-\kappa} \right)^{\sigma-1} - \left(\frac{(N-2)R+2}{RNy-\kappa} \right)^{\sigma-1} \right], \quad (34) \end{aligned}$$

where the right side of this equality is $u[c_2^\kappa(N-2)] - u[c_1^\kappa(N-1)]$, which is the smallest difference on the left-side of (33), multiplied by

$$\sum_{n=0}^{N-2} \Pi^n = 1 - \pi^{N-1}(1-\pi).$$

Using our CES utility function (33) can be rewritten as

$$\begin{aligned} & \sum_{n=0}^{N-2} \Pi^n \frac{1}{1-\sigma} \left\{ \left(\frac{RNy-\kappa}{nR+N-n} \right)^{1-\sigma} - \left(\frac{RNy-\kappa}{(n+1)R+N-(n+1)} \right)^{1-\sigma} \right\} \geq \\ & \pi^{N-1}(1-\pi) \frac{1}{1-\sigma} \left\{ y^{1-\sigma} - \left(\frac{RNy-\kappa}{(N-1)R+N-1} \right)^{1-\sigma} \right\} \end{aligned}$$

and re-arranged to

$$\sum_{n=0}^{N-2} \Pi^n \frac{1}{1-\sigma} \{(nR + N - n)^{\sigma-1} - ((n+1)R + N - (n+1))^{\sigma-1}\} \geq \\ \pi^{N-1} (1-\pi) \frac{1}{1-\sigma} \left\{ \left(\frac{RNy - \kappa}{y} \right)^{\sigma-1} - [(N-1)R + N - 1]^{\sigma-1} \right\}.$$

Notice that the right side is strictly increasing in κ . We will choose κ large; in particular we will choose κ so that the expected utility of allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is less than or equal to autarky for any probability weight π . This necessarily implies that $\kappa \geq \kappa_1$. If (33) holds for this value of κ , then it holds for all $\kappa \in (\kappa_0, \kappa_1)$ where the incentive constraint (33) is relevant. We set $y = c_1^\kappa(1) = c_2^\kappa(1)$, which implies that the expected $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is strictly less than autarky (the best consumption state, $n = 1$, provides the autarky payoff, meaning that all other states provide less than autarky). If, for convenience, we set $y = 1$, then $y = c_1^\kappa(1) = c_2^\kappa(1)$ implies

$$\kappa = (R-1)(N-1).$$

We will choose π so that the probability on the right side of (33) is maximized (if 33 holds for this π , it will hold for any π). The probability $\pi^{N-1}(1-\pi)$ is maximized for

$$\pi = \frac{N-1}{N}.$$

For convenience let $\sigma = 2$. Then if

$$[N^N - (N-1)^{N-1}] \{(N-1)R + 1 - [(N-2)R + 2]\} \geq \\ (N-1)^{N-1} \{(N-1)R + 1 - [RNy - (R-1)(N-1)]\} \quad (35)$$

holds, (33) holds. (35) is obtained by substituting the right side of (34) for the right side of (33), setting $\pi = (N-1)/N$ and $\kappa = (R-1)(N-1)$, and then rearranging. (35) can be further simplified to

$$[N^N - (N-1)^{N-1}](R-1) > (N-1)^{N-1}(R-1)(N-2)$$

or

$$N^N - (N-1)^{N-1} > (N-1)^{N-1}(N-2)$$

which is a valid inequality. Hence, we have chosen κ and π and replaced with left side of (33) in a way that “works against” inequality (33) holding. Yet

we still find that this highly restricted inequality, and hence, inequality (33), is satisfied.

Suppose that we have somehow overlooked a reasonable model parameterization that implies allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ violates (11). In this case the efficient contract is determined by maximizing (2) subject to (3), (11) and (15). The solution to this problem necessarily implies that the allocation is run-prone. Intuitively, the efficient contract is determined by increasing $c_2(m)$ from $c_2^\kappa(m)$, $m = 0, \dots, N - 1$ and decreasing $c_1(m)$ from $c_1^\kappa(m)$, $m = 1, \dots, N - 1$, until (11) is just satisfied with equality. This necessarily implies that if $c_2^\kappa(N - 1) < c_1^\kappa(N) = y$, the efficient incentive compatible contract will also be characterized by $c_2(N - 1) < c_1(N) = y$, i.e., the efficient contract is run-prone. The important point here is that if the run-prone allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is not incentive compatible, i.e., does not satisfy (11), then the efficient incentive-compatible contract will also be run-prone. This is important because, qualitatively speaking, our results—which are derived on this basis that $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ satisfies (11)—remain valid for the efficient incentive compatible allocation if $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ does not satisfy (11).