SHADOW BANK RUNS*

David Andolfatto
University of Miami

Ed Nosal
Federal Reserve Bank of Atlanta

April 16, 2022

Abstract

Peck and Shell (2003) describe conditions under which constrained-efficient risk-sharing arrangements may be run-prone. Their result hinges critically on the existence of a sequential service constraint. While sequential service is ubiquitous in retail settings, it is notably absent in wholesale settings. Since shadow banks live in the wholesale sector, an implication is that an efficient risk-sharing shadow bank is always run-proof. We demonstrate that this need not be the case when there are fixed costs of intermediation. Efficient risk-sharing shadow banks are run-proof when the fixed costs of intermediation are sufficiently small but are potentially run-prone when these costs are sufficiently large and the propensity of a run is sufficiently small. If the probability of a bank run is sufficiently high, it may be optimal to render a shadow bank run-proof at the expense of some risk-sharing. We discuss how the perceived propensity of a bank run may evolve over time in a manner consistent with Minksy (1992). In particular, a

*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, University of Madison, Wisconsin 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 Bank of Atlanta, or the Federal Reserve System. JEL Codes: G01, G21, G28

period of financial stability may engender complacency over run risk, a phenomenon that may result in alternating cycles of less and more stringent financial sector regulations.

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 allegedly rests uncomfortably on the state of creditor confidence. When confidence wanes, the structure collapses. But if this is the case, what accounts for the widespread use of such arrangements? 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.[1]

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 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 insights supplied by Diamond and Dybvig (1983) applies equally to the financial arrangements we observe in the wholesale shadow banking sector. But the application is not so obvious. In particular, as demonstrated by Peck and Shell (2003), the fragility of an optimal banking arrangement in the Diamond and Dybvig (1983) model relies critically on the assumption that depositor withdrawal requests are processed on a first-come, first-served basis. This so-called sequential service constraint

[1] Although this may seem surprising to some readers, the main message in Diamond and Dybvig (1983) is that banking arrangements, when optimally designed, are stable. In their model without aggregate risk, the optimal arrangement entails the use of suspension of payments and in their model with aggregate risk, the optimal arrangement entails insuring deposit liabilities.

is notably absent in wholesale settings where shadow banks typically operate. And absent sequential service, uninsured short-term financing is always stable—not subject to runs—in the Diamond and Dybvig (1983) model of banking (Green and Lin, 2003).[2]

It is of some interest to note that the empirical literature on shadow bank instability does not appeal to sequential service in the models used to interpret the data; see, for example, Chen, Goldstein and Jiang (2010), He and Xiong (2012), Martin, Skeie and von Thadden (2014), Schroth, Suarez and Taylor (2014), Morris, Shim and Shin (2017) and Foley-Fisher, Narajabad and Verani (2020). In those set-ups, strategic complementaries that give rise to instability are determined by *exogenous*, albeit empirically-motivated, contractual arrangements. While assuming a particular contractual form—relative to deriving it—has its uses, the approach has some drawbacks as well. In particular, as stressed by Green and Lin (2000), it prevents one from knowing whether banking instability is an unavoidable consequence of the economy's underlying structural characteristics or whether it is simply the by-product of an *ad hoc* contractual arrangement.[3] Our view is that it would be both prudent and wise for policymakers to consider the former possibility when contemplating the nature of an optimal regulatory framework.

In this paper, we use a mechanism design approach to study a finite-trader version of the Diamond and Dybvig (1983) model similar to Green and Lin (2003) and Peck and Shell (2003). This approach allows us to characterize risk-sharing arrangements that are optimal relative to the environment. Our environment differs from standard approaches in two ways: first, we abandon the sequential service constraint and second, we introduce a fixed cost of intermediation.[4] As it turns out, this fixed cost generates a type of increasing returns to scale for banks—a property that is consistent with the evidence

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

[3]The equilibrium bank run described by Diamond and Dybvig (1983) in the first part of their paper is entirely the by-product of a sub-optimal contractual arrangement. Indeed, they explain how an optimal contract eliminates the bank run equilibrium. In the second part of their paper, they introduce aggregate risk and characterize an efficient run-proof contractual arrangement, although their solution is not without controversy (Wallace 1988).

[4]To be clear, fixed cost may reinforce sequential service as a fundamental source of instability. Our main results demonstrates that sequential service is not necessary for bank instability.

in Mester (2008), Wheelock and Wilson (2017) and Corbae and D'Erasmo (2018). Banks with this property are in a position to offer long-term depositors attractive returns when fund redemption rates remain low. But when redemption activity is elevated, the return on any remaining capital *net of the fixed cost* declines. The use of short-term debt to finance portfolios with this property is potentially unstable. In particular, if confidence vanishes and investors call their loans, the scale of reinvestment collapses, unit costs rise, and the net return on the remaining portfolio 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.

We find that the presence of fixed costs in intermediation need not imply that an efficient risk-sharing shadow banks is run-prone. They are potentially a contributing factor only if they are sufficiently large in a well-defined sense. And even when they are sufficiently large, a run-prone, efficient risk-sharing shadow bank emerges only if the propensity of a bank run is sufficiently low in a well-defined sense. If the propensity of a bank run is sufficiently high, then it is optimal to run-proof the shadow bank at the expense of some risk-sharing.

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. In July 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 to reduce heavy redemption activity.[5] 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

[5] A liquidity fee is a payment that the investor incurs to withdrawl 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).

more favorable view of liquidity fees and redemption gates.[6] However, the strains exhibited by prime money mutual funds in March 2020 suggest we were too optimistic on this score.[7] 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 some shadow banks as *de facto* depository institutions. But our theoretical analysis cautions against the notion that any regulatory reform must necessarily render shadow banks absolutely run-proof.

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). This asset can be reasonably interpreted to be a portfolio of U.S. Treasury securities. If so, then the Diamond-Dybvig model suggests that government money funds and narrow banks are potentially subject to runs because of sequential service. In contrast, our model suggests that government money funds and narrow banks should be run-proof because the fixed costs of intermediating a portfolio of U.S. Treasury securities is relatively low.

2 The environment

Our model is based on the Green and Lin (2000, 2003) version of Diamond and Dybvig (1983). There are two *ex post* dates, $t = 1, 2$ 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 at $t = 1$ that determines their type: *impatient* or *patient*. An impatient individual only values c_1 while a patient individual only values c_2. Let $Au(c_1)$ denote the utility payoff associated with consuming early, where A is a preference parameter and $u(c) = c^{1-\sigma}/(1-\sigma)$ with $\sigma > 1$. The utility payoff associated with consuming later is $u(c_2)$. An

[6] Our 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.

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

individual is impatient with probability π. The probability that there are $0 \leq n \leq N$ impatient individuals is π_n. We assume that types are *i.i.d.*, which means that $0 < \pi_n < 1$ for all n (the distribution has full support) and that $\pi_n = \binom{N}{n}\pi^n(1-\pi)^{N-n}$. Each individual is endowed with y units of output. The endowment can be costlessly stored across time. In autarky, an individual attains the expected utility payoff

$$W^A = N^{-1}\sum_{n=0}^{N}\pi_n\left[nAu\left(y\right) + (N-n)u\left(y\right)\right]. \tag{1}$$

Because preference shocks have an idiosyncratic component, there is an incentive to pool risk. We refer to a contract that is designed to pool this risk as a *bank* because the arrangement pools all of the individuals' endowments in exchange for liabilities that embed an early-redemption option (at date 1). Because early redemptions are not subject to sequential service, we think of the arrangement as a *shadow bank* instead of a conventional retail bank. There is a second reason to pool resources. In particular, we assume that the bank has access to a higher return storage technology. Specifically, pooled resources that are not liquidated at date 1 remain invested in a capital project that yields a gross rate of return $R > 1$ per unit invested at date 2. Resources that are liquidated and paid out early may either be consumed or stored at a unit rate of return by depositors who wish to redeem their claims early.

Assume, for the moment, that depositor type is publicly observable. Because there is no sequential service constraint, the bank contract—a time and state-contingent allocation—takes 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 individuals at $t = 1$. The *ex ante* utility payoff associated with allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is given by

$$W(\mathbf{c}_1, \mathbf{c}_2) = N^{-1}\sum_{n=0}^{N}\pi_n\left[nAu\left(c_1(n)\right) + (N-n)u\left(c_2(n)\right)\right]. \tag{2}$$

An important part of the environment is a "travel itinerary" that describes the restrictions on communications between depositors and the bank. The timing of events is as follows. *Ex ante*, individuals begin by choosing whether or not to participate in the risk-sharing arrangement, i.e., whether or not to deposit y at the shadow bank. The payoff to not participating is given by (1). Individuals participate by depositing their endowments y

with the bank and accepting the terms of the bank contract (allocation). Following this decision, individuals disperse to remote locations where they remain incommunicado until they return to (communicate with) the bank. An individual can only return to the bank only once, either at $t = 1$ or $t = 2$. One interpretation of this latter assumption is that depositors are "rationally inattentive" in the sense that it is not economical to be in constant touch with one's bank.[8] Note that when depositors visit the bank in a given period, they are serviced at the same time—there is no sequential service constraint.

Our main innovation relative to the literature is the specification of fixed costs related to the business of banking. Let κ denote the fixed cost incurred by the bank at dates $t = 1$ and $t = 2$. Assuming that all N individuals become depositors, the value of deposit liabilities issued at $t = 1$ and cannot exceed $Ny - \kappa$. Consequently, there is the resource constraint at date $t = 1$,

$$0 \leq nc_1(n) \leq Ny - \kappa \tag{3}$$

for all n. The resources remaining after $t = 1$ redemptions is given by $k(n) = [Ny - nc_1(n) - \kappa]$. These resources, which remain invested with the bank, return $Rk(n) - \kappa$ units of output at $t = 2$. Hence, there is another resource constraint at date $t = 2$,

$$(N - n)\, c_2(n) = R[Ny - nc_1(n)] - (1 + R)\kappa \tag{4}$$

for $n = 0, 1, ..., N - 1$. Note that the fixed cost κ is incurred at date 2 only if the bank remains in operation after $t = 1$. In the event that $n = N$, all funds are withdrawn at $t = 1$ so that $k(N) = 0$. The bank effectively shuts down at the end of $t = 1$ when $n = N$ and so does not incur the fixed cost a $t = 2$.

3 Efficient risk-sharing

An *efficient allocation* $(\mathbf{c}_1, \mathbf{c}_2)$ maximizes maximizes (2) subject to (3) and (4). To simplify the analysis, assume $A = R$. This restriction has no bearing on the qualitative nature of the results we report below. Let $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ denote

[8] The concept of "rational inattention" has been employed widely in monetary theory; see, for example, Sims (2010).

the efficient allocation for a given κ. Given our CES preference specification—and assuming that $\kappa < RNy/(1+R)$—the solution is given by[9]

$$\begin{aligned} c_1^{\kappa}(n) &= \frac{RNy-(1+R)\kappa}{N+n(R-1)} &(5)\\ c_2^{\kappa}(n) &= c_1^{\kappa}(n) &(6)\end{aligned}$$

for $n = 0, 1, 2, ..., N-1$ and

$$\{c_1^{\kappa}(N), c_2^{\kappa}(N)\} = \{y - \kappa/N, 0\}. \tag{7}$$

Note that (7) differs qualitatively from (5)-(6) in that the fixed cost κ is incurred only in the early period when $n = N$; the bank is effectively shut down as a going concern at the end of $t = 1$ when $n = N$.

The restriction $A = R$ is a simplification that serves to equate consumption across periods on a state-by-state basis. Note that (5)-(6) reveal that consumption at both dates is decreasing in the number of early redemptions. This property reflects the fact that less funding is available for the higher-return investment as more depositors withdraw their funds early. Efficient risk-sharing implies that both the short and long rates of return on deposits must decline.

Moreover, note that the following is also true

$$c_1^{\kappa}(n+1) < c_2^{\kappa}(n) \tag{8}$$

for all $n = 0, 1, 2, ..., N-2$. Notice that (8) need not apply for $n = N-1$. When $n = N-1$, $c_1^{\kappa}(n+1)$ incurs the fixed cost for one period only, while $c_2^{\kappa}(n)$ incurs a fixed cost in both periods. Depending on the size of the fixed cost, we can have

$$c_2^{\kappa}(N-1) = \frac{RNy-(1+R)\kappa}{N+(R-1)(N-1)} \gtrless c_1^{\kappa}(N) = y - \kappa/N. \tag{9}$$

Note that $c_2^0(N-1) > c_1^0(N)$ and that $c_2^{\kappa}(N-1)$ declines more rapidly than $c_1^{\kappa}(N)$ as $\kappa \nearrow 0$. Evidently, there exists a unique κ_0 that solves

$$c_2^{\kappa_0}(N-1) = c_1^{\kappa_0}(N). \tag{10}$$

[9] If $\kappa = RNy/(1+R)$, the fixed cost is so large that it completely exhausts the pooled endowment Ny.

The solution is given by,

$$\kappa_0 = \left[\frac{R-1}{1-(R-1)/N} \right] y > 0. \tag{11}$$

The implication here is that when the fixed costs of intermediation are sufficiently large, $\kappa > \kappa_0$, a single patient depositor, $n = N-1$, is left materially worse off than those who withdrew earlier. If it was possible to do so, it would be desirable to alert this patient depositor and have them withdraw funds early along with everyone else. In this way, the fund can be shut down at the end of $t = 1$, thereby avoiding the need to incur the $t = 2$ fixed cost necessary to keep the fund operating. The ability to communicate in this manner, however, is rendered impossible by the fact that depositors are not in constant contact with their bank—recall our discussion above about depositors dispersing to remote locations. Conditions (8)-(11) imply the following result,

Lemma 1 For all $\kappa < RNy/(1+R)$ we have: (i) $c_2^\kappa(n) \geq c_1^\kappa(n+1)$ *for* $n = 0, 1, 2, ..., N-2$ and (ii) $c_1^\kappa(N-1) > c_2^\kappa(N)$ *for* $\kappa \in [0, \kappa_0)$ *and* $c_2^\kappa(N-1) < c_1^\kappa(N)$ *for* $\kappa > \kappa_0$.

For the bank to attract depositors, the fixed costs of operations cannot be too large, otherwise autarky would be preferable. The expected utility payoff associated with $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is given by

$$W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) = N^{-1} \sum_{n=0}^{N} \pi_n \left[(R-1)n + N \right] u[c_1^\kappa(n)] \tag{12}$$

Since $c_1^\kappa(n)$ are all strictly decreasing in κ, the value function $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ inherits this property. Since $W(\mathbf{c}_1^0, \mathbf{c}_2^0) > W^A$—as defined by (1) when $A = R$—it follows that there exists a $0 < \overline{\kappa} < \infty$ such that $W(\mathbf{c}_1^{\bar{\kappa}}, \mathbf{c}_2^{\bar{\kappa}}) = W^A$. In what follows, we restrict attention to fixed costs in the range $\kappa \in [0, \overline{\kappa}]$ since $\kappa > \overline{\kappa}$ implies autarky is the best outcome.

Lemma 2 *There exists a unique* $0 < \overline{\kappa} < \infty$ *defined by* $W(\mathbf{c}_1^{\bar{\kappa}}, \mathbf{c}_2^{\bar{\kappa}}) = W^A$ *such that* $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) > W^A$ *for* $\kappa \in [0, \overline{\kappa})$ *and* $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) < W^A$ *for* $\kappa > \overline{\kappa}$.

There remains a question regarding the size of κ_0 relative to $\overline{\kappa}$. We can demonstrate that $0 < \kappa_0 < \overline{\kappa}$ for empirically plausible parameters.[10] In what follows, we restrict attention to parameters in this region of the parameter space so that the following lemma holds,

Lemma 3 $0 < \kappa_0 < \overline{\kappa}$.

3.1 Incentive-compatibility

Let's now invoke the standard assumption that depositor type is private information. Unlike other forms of private information that are conceivably verifiable by a third party at some expense, the cost of verifying a depositor's liquidity preference is typically assumed prohibitively high. For this reason, depositor type must be learned from the depositor through some message or action. In the context of a banking arrangement, as opposed to a conventional insurance arrangement, deposit liabilities are demandable at date $t = 1$. That is, a demand deposit grants the depositor the right to withdraw funds on demand at their discretion. The act of withdrawing funds early in the context of our model may signal the depositor's unobserved "impatience." The question is what motivates depositors to signal their type truthfully?

Let $m \in \{0, 1, ..., N\}$ 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 $(\mathbf{c}_1, \mathbf{c}_2) \equiv \{c_1(m), c_2(m)\}_{m=0}^{N}$.

Once depositors have dispersed to their remote locations, they learn their types at the beginning of $t = 1$. They then play the following *withdrawal game*. Each depositor $j \in \{1, 2, ..., N\}$ 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 his own type when choosing t_j.

[10] If R is empirically plausible, e.g., $R < 1.5$, then it is trivial to show that for any $N \geq 3$, $c_2^{\kappa_0}(n) > y$ for all $n \leq N - 3$ and $c_1^{\kappa_0}(n) < y$ for $n = N, N-1, N-2$. If N is large, then the expected utility of allocation $(\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0})$ exceeds that of autarky, W^A, since the state and date contingent payoff associated with allocation $(\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0})$ exceeds that of autarky, y, in the vast majority of states (precisely because N is large). Therefore, $\bar{\kappa} > \kappa_0$.

A *strategy profile* $\boldsymbol{t} \equiv \{t_1, t_2, ..., t_N\}$ implies an $m \in \{0, 1, ..., N\}$, 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 $m = n$.

Given an allocation $(\mathbf{c}_1, \mathbf{c}_2)$, a 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 $\mathbf{t}_{-j} \equiv \{t_1, ..., t_{j-1,} t_{j+1}, ..., t_N\}$ for all $j \in \{1, 2, ..., N\}$. 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, i.e., impatient depositors visit at date $t = 1$ and patient depositors at date $t = 2$.

Note that it is a strictly dominant strategy for impatient depositors to visit the bank at $t = 1$ since they do not value consumption $t = 2$. A patient depositor, on the other hand, may have an incentive to misrepresent their type. (Recall that an early withdrawal of funds can be carried into the next period at a unit rate of return). If all patient depositors are expected to visit the bank at $t = 2$, then a patient depositor will not have an incentive to deviate from the truth-telling strategy if the following incentive-compatibility condition holds,

$$\sum_{n=0}^{N-1} \Pi^n u\left(c_2(n)\right) \geq \sum_{n=0}^{N-1} \Pi^n u\left(c_1(n+1)\right) \tag{13}$$

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

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

Lemma 4 *If types are i.i.d. and N is large, the efficient risk-sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is incentive-compatible.*

Lemma 4 can be demonstrated as follows. Consider first the set of allocations $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when $\kappa \leq \kappa_0$. By Lemma 1, $c_2^\kappa(n) \geq c_1^\kappa(n+1)$ for every $n \neq N$, which means that the incentive-compatibility condition (13) is trivially satisfied. Consider next the set of allocations $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when $\kappa \in (\kappa_0, \bar{\kappa})$.

Again, by Lemma 1 we have $c_2(n) \geq c_1(n+1)$ holds for every $n \neq N$ except $n = N-1$. The incentive constraint (13) can be rewritten as

$$\sum_{n=0}^{N-2} \binom{N-1}{n} (1-\pi)^{N-n} \pi^n [u(c_2(n)) - u(c_1(n+1))] \geq$$
$$(1-\pi)^N \pi^{N-1} [u(c_1(N)) - u(c_2(N-1))] \quad (14)$$

Note that as N gets large, the right side approaches zero since $(1-\pi)^N \pi^{N-1} \to 0$. The inequality will strictly hold since $u(c_2(n)) - u(c_1(n+1)) > 0$ for $n \neq N-1$ and $\sum_{n=0}^{N-2} \binom{N-1}{n} (1-\pi)^{N-n} \pi^n$ approaches 1 as N gets large. Intuitively, there are $N-1$ positively weighted terms on the left-hand side and one term, multiplied by a very small weight when N is large and finite, on the right-hand side.

3.2 Run-proof and run-prone banking

We label an allocation $(\mathbf{c}_1, \mathbf{c}_2)$ that satisfies (3), (4) and (13) *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 $\boldsymbol{t} \equiv \{1, 1, ..., 1\}$. That is, a run strategy profile is one in which all depositors visit the bank at $t = 1$ (in particular, all patient depositors misrepresent themselves as impatient). We say that an incentive-feasible allocation is *run-prone* if it admits a run strategy profile 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,

$$c_2(N-1) < c_1(N) \quad (15)$$

Note that condition (15) can hold at the same time (13) is true. Since $(\mathbf{c}_1, \mathbf{c}_2)$ is incentive-feasible, truthtelling is an equilibrium of 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 $\boldsymbol{t} = \mathbf{1}$. If depositor j visits the bank early, $t_j = 1$, then his payoff is $c_1(N)$ since everyone else visits at $t = 1$. If, instead, depositor j visits the bank at later, $t_j = 2$, his payoff is $c_2(N-1) < c_1(N)$. Patient depositor j will therefore choose to visit at $t = 1$, which implies the existence of a run equilibrium.[11] It follows that an allocation $(\mathbf{c}_1, \mathbf{c}_2)$ is run-proof if it has the property

$$c_2(N-1) \geq c_1(N). \tag{16}$$

4 Shadow banking when fixed costs are low

We say that fixed costs κ are "low" when $\kappa \in [0, \kappa_0]$. Lemma 1 and (16) lead us directly to

Proposition 1 *When fixed costs κ are low, the allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is run-proof.*

Our model nests the version of the Green and Lin (2000, 2003) model without sequential service when $\kappa = 0$. By Proposition 1, the efficient risk-sharing allocation $(\mathbf{c}_1^0, \mathbf{c}_2^0)$ is run-proof—a result proved in Green and Lin (2000). Our proposition generalizes Green and Lin (2000) by showing that the efficient risk-sharing allocation continues to be run-proof when fixed costs of intermediation are sufficiently low, as defined above.

Note that the allocations $\{(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) : \kappa \in [0, \kappa_0]\}$ are run-proof even though liquidity preference is not observable and depositors can withdraw on demand. Intuitively, the efficient risk-sharing allocation places increasingly stringent limits on the amount that can be withdrawn at the end of $t = 1$ as the number of depositors requesting early withdrawal, m, increases. That is, the maximum early-withdrawal amount $c_1^\kappa(m)$ is decreasing in m in a manner that leaves sufficient resources, $R[Ny - mc_1^\kappa(m) - \kappa] - \kappa$, for those who would rather withdraw their funds at a later date.

The result here is consistent with Diamond and Dybvig's (1983) conclusion that an optimally-designed banking system is run-proof. To make the mapping between their paper and ours, note that our mechanism-design approach takes no stand on how economic activity is divided across private and public sectors. This is in contrast to Diamond and Dybvig (1983) who assume

[11] Interestingly, 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.

that while banks are subject to sequential service, the government is not similarly constrained. It follows that the government should use its comparative advantage to support, or even nationalize, the banking sector. They demonstrate that a federal deposit insurance program helps render the efficient risk-sharing banking 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 queues to withdraw their funds resonates for those of us familiar with the way people behaved in historical retail bank runs. Sequential service, however, seems a poor description of how exchange occurs at the wholesale level. Money mutual funds, for example, 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.[12]

Since some types of mutual funds exhibit the symptoms of being run-prone (even those with NAV-pricing protocols), some characteristic other than sequential service must be responsible this apparent instability. We offer up fixed costs as an alternative characteristic.

5 Shadow banking when fixed costs are high

We say that fixed costs κ are "high" when $\kappa \in (\kappa_0, \bar{\kappa}]$. Lemma 1 and (15) lead us directly to

Proposition 2 *When fixed costs are high, the efficient risk-sharing allocation* $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ *is run-prone.*

Interestingly, a "high" fixed cost need not be very large, say, relative to the size of the bank's deposits Ny. To see this, consider the expression for κ_0 in (11). Notice that as N grows large, κ_0 approaches $(R-1)y$. This

[12] The 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.

means that the high fixed cost necessary to render the efficient risk-sharing allocation run-prone in a large bank is approximately equal to the total net return associated with the deposit of *one* depositor. This latter number is *tiny* relative to the size of the bank's balance sheet when N is large.

We have established that the efficient risk-sharing allocation is run-prone when fixed costs κ are high. This result, however, is not enough to conclude that a run-prone shadow bank can exist in equilibrium. First, it is not clear whether a shadow bank would be able to attract depositors if it offered a run-prone risk-sharing allocation. Second, even if depositors might find the proposition attractive relative to autarky, it is not clear whether an efficient risk-sharing, run-prone shadow bank is the better than an arrangement that is rendered run-proof at the expense of some risk-sharing.

5.1 Can a run-prone shadow 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 expected risk-sharing services delivered in non-run states of the world dominates the expected costs associated with an infrequent run event. 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.

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.[13] The sunspot, if it occurs, is observed after individuals deposit their endowments but before they learn their types.

A *sunspot equilibrium* is characterized by the incentive-feasible risk-sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$, a probability θ, and a set of withdrawal-game strategy

[13] A sunspot is a theoretical device that coordinates depositors' beliefs about how all other depositors are expected to behave in the withdrawal game. There is nothing here that fundamentally determines the value of θ. It may or may not correspond to the actual probability of observing the sunspot. One interpretation of θ is that it indexes 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. Or, it may be that θ evolves over time as the history of financial crises unfolds. We consider this latter interpretation in a follow-up section below.

profiles contingent on the occurrence of the sunspot. In particular, when the sunspot is not observed, each depositor believes that the other $N-1$ depositors play truthfully. When the sunspot is observed, each depositor believes that the other $N-1$ depositors will visit the bank at $t=1$. Because $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is incentive-compatible, truth-telling is an equilibrium when the sunspot is not observed. By Proposition 2, type-misrepresentation by patient depositors is an equilibrium when the sunspot is observed.[14]

The level of consumption for each depositor when the sunspot is observed and depositors run is equal to $y - \kappa/N$. Let Z^κ denote the expected utility payoff associated with a run, i.e.,

$$Z^\kappa = N^{-1} \sum_{n=0}^{N} \pi_n \left[nRu\left(y - \kappa/N\right) + (N-n)u\left(y - \kappa/N\right)\right],$$

which can be rewritten as

$$Z^\kappa = u\left(y - \kappa/N\right) \sum_{n=0}^{N} \pi_n \left[n(R-1) + N\right]/N. \tag{17}$$

Notice that Z^κ monotonically decreasing in κ. Comparing (17) to the autarkic payoff (1) when $A = R$, we can conclude that

Lemma 5 *$W^A > Z^\kappa$ for $\kappa \in (0, \overline{\kappa}]$* and *$W^A = Z^0$.*[15]

Let $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa)$ denote the expected utility associated with allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when it is assumed that depositors play the sunspot strategies described above, i.e.,[16]

$$V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa) \equiv (1-\theta) W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) + \theta Z^\kappa. \tag{18}$$

Since both $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ and Z^κ are strictly decreasing in κ, then so is $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa)$. And since $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) > W^A$ when $\kappa < \overline{\kappa}$ (Lemma 2) and $W^A > Z^\kappa$ for $\kappa > 0$

[14] Notice that if κ is low, $\kappa \in [0, \kappa_0)$, there cannot exist a sunspot equilibrium.

[15] Here we are comparing payoffs associated with playing certain strategies. Although depositors do not, in equilibrium, play the sunspot strategies described above when $\kappa < \kappa_0$, nothing prevents us from calculating and comparing sunspot strategies even if they are not played in equilibrium.

[16] Again, in equilibrium, the sunspot strategies will not be played when $\kappa < \kappa_0$. Nevertheless, it is still possible to calculate the expected utility associated with playing these strategies.

(Lemma 5), it follows that $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa)$ is strictly decreasing in the run propensity, θ, for any $\kappa \in (0, \bar{\kappa})$. It also follows that there exists a critical value $\widetilde{\theta}(\kappa)$ such that $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \widetilde{\theta}(\kappa), \kappa) = W^A$. If a run-prone shadow bank is sufficiently stable in the sense of $\theta < \widetilde{\theta}(\kappa)$, it will be able to attract depositors because the expected utility associated with allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when depositors play sunspot strategies exceeds that of autarky. Furthermore, since $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa)$ is decreasing in κ, the critical probability $\widetilde{\theta}(\kappa)$ must also be decreasing in κ. This implies that a higher-κ shadow bank must also be a more stable shadow bank in the sense that it must be associated with a lower θ if it is to attract depositors, compared to a lower-κ shadow bank. The following summarizes these results,

Lemma 6 *For any $\kappa \in [0, \overline{\kappa}]$, there exists a $0 \leq \widetilde{\theta}(\kappa) \leq 1$ such $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \widetilde{\theta}(\kappa), \kappa) = W^A$, with $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa) > W^A$ for $\theta < \widetilde{\theta}(\kappa)$ and $V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa; \theta, \kappa) < W^A$ for $\theta > \widetilde{\theta}(\kappa)$. Moreover, $\widetilde{\theta}(\kappa)$ is strictly decreasing in κ, with $\widetilde{\theta}(\overline{\kappa}) = 0$ and $\widetilde{\theta}(0) = 1$.*

We now characterize the *equilibrium* outcomes for various combinations of (κ, θ). By Proposition 1, when $\kappa \leq \kappa_0$, the unique equilibrium is the run-proof allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ and this is independent of the value of θ. The set of these equilibria for various combinations of (κ, θ) is illustrated in Figure 1. By Proposition 2, allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is run-prone when $\kappa \in (\kappa_0, \bar{\kappa})$ and by Lemma 6, allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is preferred to autarky if $\theta < \tilde{\theta}(\kappa)$ when $\kappa \in (\kappa_0, \bar{\kappa})$. The set of equilibrium run-prone allocations—where depositors play the sunspot strategies—is identified by the triangular region in Figure 1. Finally, by Lemmas 2 and 6, autarky is the equilibrium outcome when $\theta > \tilde{\theta}(\kappa)$ and $\kappa \in (\kappa_0, \bar{\kappa})$ or when $\kappa > \bar{\kappa}$, and is illustrated in Figure 1.

The discussion above is predicated on the assumption that depositors can only choose between the efficient allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ and autarky (y, y). There is, in fact, another possibility to consider. In particular, it is possible to construct a run-proof allocation that offers inefficient risk-sharing in some state(s). One can interpret such an allocation as arising from a regulatory intervention designed to render the shadow bank sector more stable. We can then ask if such a regulatory intervention is optimal in the sense of improving depositor welfare.

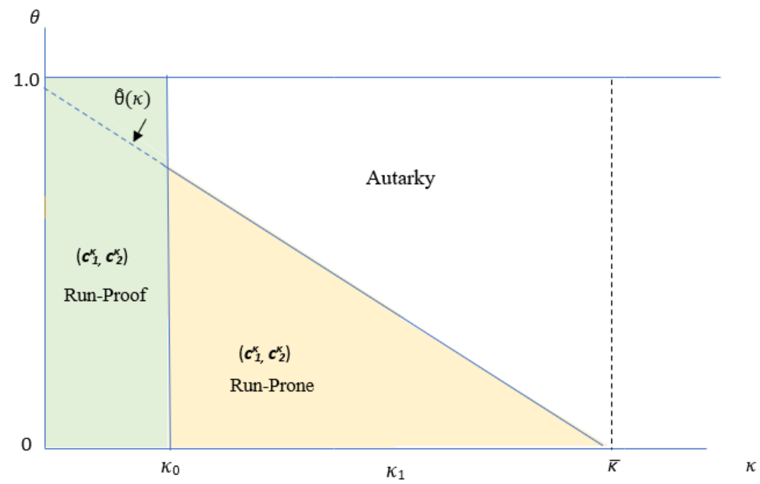

Figure 1: Equilibria

5.2 Choosing between run-prone and run-proof shadow banks

A run-prone, efficient risk-sharing allocation can attract depositors if the propensity to run is not too high, $\theta < \tilde{\theta}(\kappa)$. In this section, we ask whether a less-efficient but run-proof allocation might be preferred to the efficient run-prone allocation.

The efficient allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is run-prone—$c_2^\kappa(N-1) < c_1^\kappa(N)$—when $\kappa \in (\kappa_0, \overline{\kappa})$. This allocation can be rendered run-proof by modifying the payoff to the late arriving depositor in state $n = N-1$ to $\hat{c}_2^\kappa(N-1) \geq c_1^\kappa(N) = y - \kappa/N$. A run-proof allocation that does the least harm to risk-sharing is one which sets the modified payoff equal to $y - \kappa/N$.[17] Of course, making such a modification for the late arriving depositor necessarily decreases the payoff to those who withdraw early, meaning that the modified payoff to early arriving depositors must be reduced from $c_1^\kappa(N-1)$. The important question is whether the implied loss of risk-sharing from this modification is worth incurring if it makes the shadow bank run-proof.

[17] One might interpret this modification 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).

Let's be more specific. Consider the run-proof allocation $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ where $(\hat{c}_1^{\kappa}(m), \hat{c}_2^{\kappa}(m)) = (c_1^{\kappa}(m), c_2^{\kappa}(m))$ for $m \neq N-1$ and

$$\begin{aligned} \hat{c}_1^{\kappa}(N-1) &= \frac{RNy-(1+R)\kappa-(y-\kappa/N)}{R(N-1)} < y-\kappa/N \\ \hat{c}_2^{\kappa}(N-1) &= y-\kappa/N = \hat{c}_1^{\kappa}(N) \end{aligned} \tag{19}$$

The spirit of the deposit contract $(\mathbf{c}_1^{\hat{\kappa}}, \mathbf{c}_2^{\hat{\kappa}})$ is to penalize early redemptions in heaviest redemption state $n = N-1$. The penalty here has the flavor of a "liquidity fee" reminiscent of what prime institutional money funds in the United States are sometimes permitted to apply at the discretion of fund managers. If the liquidity fee is credible (something that may have to be based on legislation, rather than the discretion of management), then it ensures that sufficient resources will be made available for those who are 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 $t=2$. This property, in turn, implies that allocation $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ can be implemented as a unique equilibrium.

The *ex ante* welfare associated with allocation $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$, denoted as $W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$, is given by,[18]

$$\begin{aligned} W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa}) \equiv\; & N^{-1}\{\sum_{n=0}^{N-2} \pi_n \left[(R-1)n+N\right] u[c^{\kappa}(n)] \\ & +\pi_{N-1}\{(N-1)Ru\left[\hat{c}_1^{\kappa}(N-1)\right] + u(y-\kappa/N)\} \\ & +\pi_N NRu(y-\kappa/N)\}. \end{aligned} \tag{20}$$

Since allocation $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ does not provide efficient risk-sharing in state $n = N-1$, it is necessarily the case that $W(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}) > W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ for all κ, which implies that $W(\hat{\mathbf{c}}_1^{\bar{\kappa}}, \hat{\mathbf{c}}_2^{\bar{\kappa}}) < W^A$. Since each element of $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ is decreasing in κ, there exists a $\kappa_1 < \bar{\kappa}$ such that $W(\hat{\mathbf{c}}_1^{\kappa_1}, \hat{\mathbf{c}}_2^{\kappa_1}) = W^A$. Hence,

Lemma 7 *There exists a* $0 < \kappa_1 < \bar{\kappa}$ *such that* $W(\hat{\mathbf{c}}_1^{\kappa_1}, \hat{\mathbf{c}}_2^{\kappa_1}) = W^A$, with $W(\hat{\mathbf{c}}_1^{\kappa_1}, \hat{\mathbf{c}}_2^{\kappa_1}) > W^A$ *for* $\kappa \in [0, \kappa_1)$ *and* $W(\hat{c}_1^{\kappa_1}, \hat{c}_2^{\kappa_1}) < W^A$ *for* $\kappa \in (\kappa_1, \bar{\kappa}]$.

[18] Depositors always play their truth-telling strategies for the run-proof allocation.

Since allocation $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa_1})$ is identical to allocation $(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa})$ in all states but $n = N-1$, if we restrict parameters so that R is empirically plausible and N is sufficiently large, as we did in our discussion of $\bar{\kappa}$, then we can conclude that[19]

Lemma 8 $\kappa_0 < \kappa_1$.

For the discussion in this paragraph, we restrict our attention to $\kappa \in (\kappa_0, \kappa_1)$. We know that allocation $(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ is preferred to autarky (Lemmas 7 and 8) and that allocation $(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa})$ is run prone (Proposition 1). Furthermore, since $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \theta = 0, \kappa) > W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ and $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \theta = 1, \kappa) = Z^{\kappa} < W^A < W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$, there exists a $\hat{\theta}(\kappa)$ that satisfies $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \hat{\theta}(\kappa), \kappa) = W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$, and using (18), we can solve

$$\hat{\theta}(\kappa) = \left[\frac{W(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}) - W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})}{W(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}) - Z^{\kappa}} \right]. \tag{21}$$

Notice that the numerator of (21) represents the benefit of risk-sharing that is lost under the run-proof arrangement. The higher this benefit, the more tolerant depositors are to run risk. The denominator of (21) represents the value of risk-sharing relative to the value of a run. If the payoff associated with the run state Z^{κ} increases relative to $W(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa})$, then the tolerance for run risk $\hat{\theta}(\kappa)$ increases.

From the discussion above, we have the following result,

Lemma 9 When $\kappa \in [\kappa_0, \kappa_1)$, $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \theta, \kappa) > W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ *for* $\theta < \hat{\theta}(\kappa)$ *and* $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \theta, \kappa) < W(\hat{\mathbf{c}}_1^{\kappa}, \hat{\mathbf{c}}_2^{\kappa})$ *for* $\theta > \hat{\theta}(\kappa)$.

Recall that $\widetilde{\theta}(\kappa)$ is defined by $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \widetilde{\theta}(\kappa), \kappa) = W^A$. Since $W(\hat{\mathbf{c}}_1^{\kappa_1}, \hat{\mathbf{c}}_2^{\kappa_1}) = W^A$, $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \hat{\theta}(\kappa_1), \kappa_1) = W(\hat{\mathbf{c}}_1^{\kappa_1}, \hat{\mathbf{c}}_2^{\kappa_1})$ and $V(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa}; \widetilde{\theta}(\kappa_1), \kappa_1) = W^A$, it follows that

[19] The discussion in footnote 9 for allocation $(\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0})$ can be applied directly here. In particular, if R is empirically plausible, autarky delivers a higher state contingent expected utilty than allocation $(\hat{\mathbf{c}}_1^{\kappa_0}, \hat{\mathbf{c}}_2^{\kappa_0})$ in states $n = N, N-1, N-2$, and allocation $(\hat{\mathbf{c}}_1^{\kappa_1}, \hat{\mathbf{c}}_2^{\kappa_1})$ and delivers higher state contingent expected utility than autarky in all states $n \leq N-3$. If N is large, then the expected utility of allocation $(\mathbf{c}_1^{\kappa_0}, \mathbf{c}_2^{\kappa_0})$ exceeds autarky, which implies that $\kappa_1 > \kappa_0$.

Lemma 10 $\hat{\theta}(\kappa_1) = \widetilde{\theta}(\kappa_1)$.

We can now provide a complete characterization of the equilibrium outcomes in (θ, κ) space. Since $W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa) < W^A$ for $\kappa > \kappa_1$ and $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) > W(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ for all κ, we have already established (in see Figure 1) that: (i) when $\kappa \in [0, \kappa_0)$ the equilibrium is the run-proof allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$; (ii) when $\kappa > \bar{\kappa}$, the equilibrium is autarky; and (iii) when $\kappa \in (\kappa_1, \bar{\kappa})$, the equilibrium is the run-prone allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when $\theta < \tilde{\theta}(\kappa)$ and autarky when $\theta > \tilde{\theta}(\kappa)$. These equilibrium allocations are displayed in Figure 2. From Lemmas 9 and 10, when $\kappa \in (\kappa_0, \kappa_1)$ the equilibrium is the run-prone allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ when $\theta < \hat{\theta}(\kappa)$ and the run-proof allocation $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$ when $\theta > \hat{\theta}(\kappa)$. Figure 2 provides the complete characterization of equilibrium outcomes.[20]

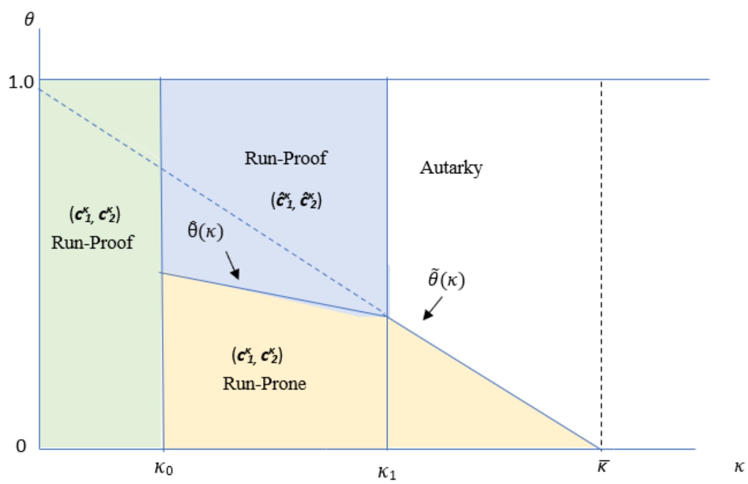

Figure 2: Equilibria

Our analysis is meant to gently push back against an uncritical inclination among economists and policy makers that it is always desirable to eliminate unstable outcomes, such as bank runs, if there is a way to do so. Our analysis, which is summarized in Figure 2, indicates that in some circumstances, it may be better not to completely extinguish the possibility of bank runs. We find

[20]Notice that Figure 2 shows that $\tilde{\theta}(\kappa) > \hat{\theta}(\kappa)$ for all $\kappa \in (\kappa_0, \kappa_1)$. Since $W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) > W^A$ for all $\kappa \in (\kappa_0, \kappa_1)$, this relationship is obvious.

that if either fixed costs associated with intermediation are neither trivially small nor extremely large, and the propensity to run is not "too high," i.e., $\theta < \hat{\theta}(\kappa)$, then the best shadow banking arrangements are not run-proof.

6 Central bank liquidity provision

Our mechanism design approach treats the economy as a closed system with contractual arrangements designed optimally relative to the environment describing that closed system. The allocation in such a closed system is supported in one way or another by a variety of private and public institutions, with the division of responsibilities between private and public sectors indeterminate. The mechanism design approach focuses more on identifying the principles of good governance, not who is responsible for governance.

In this section, we want to think of the shadow bank sector as distinct from a central bank or government treasury. To do so, we must introduce another actor into the model that interacts with the shadow bank sector, but in a limited manner. We want to think of an interaction that resembles a lender-of-last-resort policy that potentially operates only in high-redemption states. The limited nature of the intervention can be thought of modeling the central bank's desire to minimize its "financial footprint." We want to explore how a limited policy of this sort may be used to eliminate bank runs that are optimal relative to the closed system. We are motivated here by the events of March 2020 when the Federal Reserve's emergency provision of liquidity was designed to prevent runs (instability) on mutual money funds.

Suppose that the run-prone, efficient risk sharing allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is optimal in the closed system. Suppose further that the central bank (possibly with support from the government treasury) seeks to eliminate bank runs by providing liquidity, via a repo arrangement, to the shadow bank. In this situation, the central bank will provide resources $x > 0$ at date 1 to the shadow bank when $m = N - 1$; for all other realizations $m \neq N - 1$, $x = 0$. (The central bank minimizes its footprint by providing resources only when $m = N - 1$.) The repo contract is "unwound" at date 2 when the shadow bank repurchases its claim from the central bank with x units of date 2 output.

To ensure that the allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ is now run-proof and that the shadow bank is able to repay x at date 2, the date 1 liquidity provision x must be such that the total date 2 investment outcome of the shadow bank is at least equal to $y - \kappa/N + x + \kappa$ when $m = N - 1$. That is, the single depositor arriving at date 2 receives a payoff equal to $y - \kappa/N$ and the central bank is repaid x, with sufficient resources remaining to cover the fixed cost, κ. The minimum level of liquidity provision x that satisfies all these requirements can determined as follows: Since the amount of resources invested at date 1 equals $Ny + x - (N-1)c_1^\kappa(N-1)$, in order to ensure that x is repaid to the central bank, the date 2 fixed cost κ is covered and that the depositor arriving at date 2 receives $y - \kappa/N$, it must be the case that $R[Ny + x - (N-1)c_1^\kappa(N-1)] = y - \kappa/N + x + \kappa$, which implies that

$$\tilde{x} = \frac{y - \kappa/N + \kappa - R[Ny - (N-1)c_1^\kappa(N-1)]}{R-1}.$$

Define the allocation with central bank liquidity provision in state $m = N-1$ as $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$, where $(\tilde{c}_1^\kappa(m), \tilde{c}_2^\kappa(m)) = (c_1^\kappa(m), c_2^\kappa(m))$ for all $m \neq N-1$ and $(\tilde{c}_1^\kappa(N-1), \tilde{c}_2^\kappa(N-1)) = (c_1^\kappa(N-1), y - \kappa/N)$. The expected utility associated with allocation $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ is given by

$$\begin{aligned} W(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa) \equiv\ & N^{-1}\{\sum_{n=0}^{N-2} \pi_n \left[(R-1)n + N\right] u[c^\kappa(n)] \\ & + \pi_{N-1}\{(N-1)Ru\left[c_1^\kappa(N-1)\right] + u(y - \kappa/N)\} \\ & + \pi_N NRu(y - \kappa/N)\} \end{aligned}$$

Note that $W(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa) > V(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) = (1-\theta)W(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa) + \theta Z^\kappa$, which is the expected utility associated with the run-prone, efficient risk sharing allocation. The former exceeds the latter along two dimensions. First, since allocation $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ is run-proof, it avoids the low payoff, Z^κ, associated with a run. Second, $\tilde{c}_2^\kappa(N-1) = y - \kappa/N > c_2^\kappa(N-1)$ while $\tilde{c}_1^\kappa(N-1) = c_1^\kappa(N-1)$, meaning that the expected utility for $m = N-1$ is greater for allocation $(\tilde{\mathbf{c}}_1^\kappa, \tilde{\mathbf{c}}_2^\kappa)$ than $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$. Hence, by providing liquidity—without subsidy—to the shadow bank, a central bank is able to prevent bank runs while, at the same time, increasing the welfare of shadow bank stakeholders.

7 Endogenous complacency

Suppose that 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, $t = 0, 1, 2, ...\infty$. People enter into the initial period with a prior believe 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.,

$$\theta_t = \Pr\left[\theta \mid s_{t-1}, s_{t-2}, ...s_1, \theta_0\right] \tag{22}$$

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 $\kappa_0 < \kappa \leq \kappa_1$ 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, $(\hat{\mathbf{c}}_1^\kappa, \hat{\mathbf{c}}_2^\kappa)$. 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.[21] 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.

The analysis above demonstrates a difficulty in designing appropriate legislation. If the true run propensity is unknown, then the best society can do is estimate it based on experience. In reality, this likely takes the form of estimating run propensities based on the frequency of recent financial crises. As a run-proof financial system is not likely to generate crises, it seems reasonable to suppose that investors may become increasingly complacent

[21] Any learning rule failing to display this property would arguably mean that people fail to learn from experience.

over time. If so, then this growing complacency may eventually manifest itself as calls for regulatory reforms that effectively render the financial system run-prone. This type of cycle is reminiscent of the Minsky (1992) "stability breeds instability" hypothesis. It is not entirely clear what the solution here might be, especially as people are likely to differ on their assessment of actual run propensities. But at least the framework here provides a framework from which on might understand the predicament.

8 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 potentially run-prone. An indispensable element of their theory is 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).

Shadow banks in the wholesale sector do not follow sequential service protocols in the manner of their retail counterparts. The Diamond and Dybvig (1983) model therefore implies that shadow banks should be immune to runs. This conclusion is unwarranted. When the fixed costs of intermediation are embedded in an otherwise standard Diamond and Dybvig (1983) model, we find that shadow banks may be run-prone and, moreover, that such an outcome is neither inevitable or undesirable. Our approach permits contractual arrangements (broadly defined to include regulatory measures) to tailor themselves to the economic environment in a way that preserves risk-sharing with run-proof demandable debt structures. Nevertheless, our theory identifies regions in the parameter space where investors face a trade-off between risk-sharing and stability. In a region where the fixed costs of intermediation are sufficiently high and the probability of coordination failure is low, investors are willing to expose themselves to fragile shadow bank arrangements.

The theory laid out above suggests that a degree of humility is in order for both economists and policymakers. It should not be taken for granted that appropriate policy action always entails a perfectly run-proof, stable financial structure. To take an extreme example, financial autarky would be one way to achieve that result but few, if any, would view this as desirable.

Nevertheless, it is difficult to make conclusive statements about the merits of any given contractual/policy arrangement without knowing approximately where the economy is situated 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 largely on the recent history of financial crisis. A long period of financial stability is likely to engender a degree of complacency over run risk (Minsky 1992). Since perceptions over the likelihood of run risk—the parameter θ in the model above—are likely to vary across the population, perhaps the only practical and legitimate solution is to have people vote over an acceptable run tolerance. If so, then one should expect changes to legislation over time that broadly reflect recent experience. In particular, one would expect rules to become more stringent soon after a crisis and to become less stringent as a period of financial stability extends itself. If this broadly understood when a run occurs, then less energy might be wasted in political wrangling, leaving more energy to deal with any crisis and its aftermath in a rational manner.

9 References

1. Andolfatto, David and Ed Nosal (2018). "Bank Runs without Sequential Service." Federal Reserve Bank of Atlanta, Working Paper 2018-6.

2. 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.*

3. Chen, Qi, Itay Goldstein and Wei Jiang (2010). "Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows," *Journal of Financial Economics*, 97: 239–262.

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. Corbae, Dean and Pablo D'Erasmo (2018). "Capital Requirements in a Quantitative Model of Banking Industry Dynamics," unpublished manuscript.

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

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

8. Fisher, Irving (1936). "100% Money and Public Debt." *Economic Forum*, Spring Number, April-June: 406-420.

9. Foley-Fischer, Nathan, Borghan Narajabad and Stephane Verani (2020). "Self-fulfilling Runs: Evidence from the US Life Insurance Industry," *Journal of Political Economy*, 128 (9), 3520–3569.

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. He, Zhiguo and Wei Xiong (2012). "Dynamic Debt Runs," Review of Financial Studies, 25(6): 1799–1843.

15. Martin, Antoine, David Skeie and Ernst-Ludwig von Thadden (2014). "Repo Runs," *Review of Financial Studies*, 27(4): 957–989.

16. Mester, Loretta (2008). "Optimal Industrial Structure in Banking," Chapter 5 in the *Handbook of Financial Intermediation and Banking*, 133–162.

17. Minsky, Hyman P. (1992). "The Financial Instability Hypothesis," Levy Economics Institute of Bard College Working Paper No. 74.

18. Stephen Morris, Stephen, Ilhyock Shim and Hyun Song Shin (2017). "Redemption risk and cash hoarding by asset managers," *Journal of Monetary Economics*, 89: 71–87.

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

20. Schroth, Enrique, Gustavo A.Suarez and Lucian A.Taylor (2014). "Dynamic Debt Runs and Financial Fragility: Evidence from the 2007 ABCP Crisis," *Journal of Financial Economics*, 112: 164-189.

21. 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.

22. Sims, Christopher A. (2010). Rational inattention and monetary economics. Handbook of Monetary Economics, 3, pp.155-181.

23. Wallace, Neil (1988). "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model With Sequential Service

Taken Seriously," Federal Reserve Bank of Minneapolis *Quarterly Review*, 12(4): 3-16.

24. 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.

25. Zeng, Yao (2017). "A Dynamic Theory of Mutual Fund Runs and Liquidity Management," University of Washington Working Paper.

10 Appendix: Incentive compatibility

We cannot find model parameters where allocation $(\mathbf{c}_1^\kappa, \mathbf{c}_2^\kappa)$ violates (13). To provide some insight for this outcome, rewrite constraint (13) as

$$\sum_{n=0}^{N-2} \Pi^n \{u\left[c_2^\kappa(n)\right] - u\left[c_1^\kappa(n+1)\right]\} \geq \pi^{N-1}(1-\pi)\{u(y) - u\left[c_2^\kappa(N-1)\right]\}. \quad (23)$$

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 as the right-side weight is very small relative to the sum of weights on the left side, incentive constraint (13) is satisfied. We now formalize this intuition by example.

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

$$\sum_{n=0}^{N-2} \Pi^n \{u\left[c_2^\kappa(n)\right] - u\left[c_1^\kappa(n+1)\right]\} >$$
$$[1 - \pi^{N-1}(1-\pi)]\frac{1}{\sigma-1}[(\frac{(N-1)R+1}{RNy-\kappa})^{\sigma-1} - (\frac{(N-2)R+2}{RNy-\kappa})^{\sigma-1}], \quad (24)$$

where the right side of this equality is $u\left[c_2^\kappa(N-2)\right] - u\left[c_1^\kappa(N-1)\right]$, which is the smallest difference on the left-side of (23), multiplied by

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

Using our CES utility function (23) can be rewritten as

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

and re-arranged to

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

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 (23) holds for this value of κ, then it holds for all $\kappa \in (\kappa_0, \kappa_1)$ where the incentive constraint (23) 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 (23) is maximized (if 23 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

$$[N^N - (N-1)^{N-1}]\{(N-1)R+1-[(N-2)R+2]\} \geq$$
$$(N-1)^{N-1}\{(N-1)R+1-[RNy-(R-1)(N-1)]\} \quad (25)$$

holds, (23) holds. (25) is obtained by substituting the right side of (24) for the right side of (23), setting $\pi = (N-1)/N$ and $\kappa = (R-1)(N-1)$, and then rearranging. (25) 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 (23) in a way that "works against" inequality (23) holding. Yet we we still find that this highly restricted inequality–and hence, inequality (23)–is satisfied.

Suppose that we have somehow overlooked a reasonable model parameterization that implies allocation $(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa})$ violates (13). In this case the efficient contract is determined by maximizing (2) subject to (3), (4) and (13). 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 $c_2^{\kappa}(m)$, $m = 0, ..., N-1$ and decreasing $c_1(m)$ from $c_1^{\kappa}(m)$, $m = 1, ..., N-1$, until (13) 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 (13)–then the efficient incentive-compatible contract will also be run-prone. This is important because, qualitatively speaking, our results—which are derived on the basis that $(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa})$ satisfies (13)—remain valid for the efficient incentive compatible allocation if $(\mathbf{c}_1^{\kappa}, \mathbf{c}_2^{\kappa})$ does not satisfy (13).