Economic Brief

March 2018, EB18-03

Preventing Bank Runs

By Renee Haltom and Bruno Sultanum

Banking can be defined as the business of maturity transformation, or "borrowing short to lend long." Economists and policymakers have long viewed banking as inherently unstable, that is, prone to runs. This *Economic Brief* reviews the intuition and theory behind bank runs and the most popular proposed solutions. It also explores new research suggesting that runs might be prevented by creating a new, low-cost type of deposit contract that eliminates the incentive to run.

Banking can be defined as the business of maturity transformation, or "borrowing short to lend long." For example, commercial banks take deposits from households and businesses — short-term liabilities that depositors can withdraw on demand — and use them to make longer-term loans to other households and businesses. Banking activity does not take place only within commercial banks; entities ranging from money market mutual funds to investment banks perform this bank-like function in what has become known as the "shadow-banking" sector.

Economists and policymakers have long grappled with the question of how to promote a stable banking sector. The core issue in this debate is that banking activity is viewed as inherently fragile — that is, prone to runs. A run is when many depositors withdraw their funds to avoid losing those funds if the bank becomes insolvent. Notably, a run can occur whether the bank is insolvent or not — that is, the fear of a run may be enough to produce one.

Runs can be very costly for economic activity; in fact, many economists view the 2007–08 financial crisis as having resembled a traditional bank run across many markets.[1] Given the potential costs of runs, preventing them is of interest to policymakers and the general public.

This *Economic Brief* reviews the basic theory of runs and describes how the economics profession has explored potential solutions. It also presents a new solution proposed in a recent model by one author of this brief (Sultanum) along with David Andolfatto of the St. Louis Fed and Ed Nosal of the Atlanta Fed.

What Makes Banking Unstable?

To understand why banking may be inherently fragile, it is helpful to articulate its fundamental purpose. Banks solve an important problem for households and businesses by connecting savers (direct lenders) and borrowers. Direct lenders would like a safe place to put their money that offers a return, and borrowers would like to finance productive endeavors. Absent banks, the timing of a direct lender's desired investment may not match up with the duration of the project that the borrower wants to fund. Also, direct lenders may struggle to monitor borrowers' projects. Finally, direct lenders may want the option to withdraw their funds on demand to account for uncertainty in their liquidity needs. This last issue, the uncertainty of when the need for liquidity will arise, is a fundamental issue in banking instability and the subject to which we now turn.

Banks address uncertainty in the liquidity needs of depositors by pooling the funds of many depositors together. If a bank's base of depositors becomes large enough, the bank could be fairly certain about the fraction of depositors who will need to withdraw their funds. This enables the bank to set aside a portion of liquid reserves and use its remaining deposits to make loans. This arrangement means that banks would not be able to meet the demands of all depositors at any given time and leaves them prone to runs. If enough depositors think, for any reason, that too many other depositors are about to withdraw funds, they may try to withdraw first, before the bank runs out of reserves. This behavior can be costly to economic activity because if borrowers have to repay their loans early, productive projects may not be completed.

Bank runs occurred frequently before 1933. The creation of the Federal Reserve as lender of last resort in 1913 was partly a response to this instability, as was the creation of deposit insurance in 1933. Deposit insurance should in principle eliminate depositors' need to run: just as the fear of runs can be self-fulfilling, having deposit insurance should ensure it never needs to be used.

Deposit insurance comes with problems of its own, however. It significantly reduces depositors' incentive to care about the bank's health, which may lead banks to fund riskier projects than they otherwise would, with potential negative consequences for shareholders and society. This generally is the reason why banks are supervised and regulated in a way that most other industries are not. But regulation is not a trivial task. For example, banks' incentives may not be aligned with regulators' incentives, and banks have a comparative advantage over regulators in evaluating loans. In part because of the difficulty of aligning incentives once deposit insurance is in place, economists and policymakers have explored other possible ways of ensuring banking stability.

One alternative is "narrow banking," in which banks offer demand deposits but hold only very safe, liquid assets — such as treasuries — as opposed to loans.[2] In principle, this approach eliminates the possibility of runs because banks would virtually always be solvent. In this scheme, risky, information-intensive projects would be funded by capital markets — so this framework potentially involves higher costs of providing loans and less economic output. Another class of solutions involves suspending depositors' ability to withdraw, the subject of the remainder of this *Economic Brief*.

The Diamond-Dybvig Model of Bank Runs

Economists have formally modeled bank runs to identify conditions that might lead to runs and to explore potential solutions. The first two models of bank runs were provided by Bryant (1980) and Diamond-Dybvig (1983), the latter serving as the basis for many formal analyses of runs.[3] Diamond-Dybvig (D-D) is a model of banking in which a collection of savers pools its funds in an entity, which can be thought of as a bank that invests the funds in a longer-term project.

The model's basic mechanism can be seen in a very simplified example. Imagine you are one of just two depositors. You put \$100 into a bank account that earns an expected return of 10 percent at the end of one year. You also know that if the other depositor withdraws early, the bank's resources will be depleted such that your payout is only \$90. The other depositor faces the same conditions. However, the probability that the other depositor will need the resources early is low, so as long as both depositors withdraw only when they truly need liquidity, the expected return of keeping the money in the bank is close to 10 percent. One possible outcome is exactly this: you both keep your money invested and earn \$110 at the end of the year when the investment project is complete and pays out. However, if either of you expects the other to withdraw early — even if there is no fundamental need — the other will as well, both receiving only \$90.

In the D-D model, there are two types of agents. "Impatient" agents will discover they need to withdraw their investments prematurely and consume today. On the other hand, "patient" agents can consume a greater amount at the end of the investment period, or they can withdraw early if they suspect a run. Once an agent realizes its type, this remains private information — only the agent knows it. This feature is important because the possibility of runs gives patient agents some incentive to misrepresent their type. Similarly, depositors in the real world who do not "need" their funds may choose to withdraw anyway if they anticipate a run.

This simple example illustrates two core insights from the D-D model. The first insight is that at least two equilibria are possible: a "run" equilibrium where both types of agents, patient and impatient, announce they are impatient (by withdrawing early) and one where patient agents reveal their true type (by not withdrawing). The second insight is that the equilibrium where agents reveal their true type yields a higher payout than a bank-run equilibrium. That is, an ideal depositor contract would create incentives for patient agents, who do not truly need their funds, to not withdraw.

Potential Solutions to Runs

Overall, the D-D framework suggests three main factors that may lead to bank runs: maturity transformation, each agent having private information about its own need for liquidity, and sequential service (first come, first served — as in actual runs).

This framework also implies a solution to runs: embedding deposits with a suspension clause that says when redemptions exceed a specific threshold, the bank will suspend all payments until the end of the investment period. If the clause triggers suspension once the number of withdrawals exceeds the number of impatient agents, then the run will never occur because all parties will know that the bank will not become insolvent. Put differently, this arrangement allows for withdrawals based on a fundamental need for liquidity but prevents those based purely on fear.

However, this solution is only possible when there is certainty about the total number of impatient agents. On the other hand, if there is *aggregate uncertainty*, the suspension may not occur at the right time, and that would be costly: suspending prematurely may leave some impatient agents without needed resources — hurting economic output — while suspending too late may leave the bank with insufficient funds to avoid a run. Aggregate uncertainty arguably is the more realistic scenario. This could be solved if the bank could collect information about liquidity needs before making payments. However, sequential service, both in the model and in the real world, means that it is not possible for the bank to collect information about total liquidity needs before determining payouts. Thus, the requirement of aggregate certainty for the suspension solution to work is a limitation of the D-D model.

It has proven difficult to modify the D-D framework with aggregate uncertainty to identify run solutions. Green and Lin (2003) were the first to do so, and their model provides a unique equilibrium, one in which runs do not occur.[4] Key to their model is the structure of sequential service: they allow the payments that impatient withdrawers receive to depend on payments from past withdrawers. This "partial suspension" structure preserves some resources for later withdrawers, which is clearly preferable from a social perspective to forcing the bank's insolvency earlier. Partial suspension has historically been present in bank contracts and invoked in panic episodes. By contrast, in the D-D setup, withdrawal sizes are not restricted, and thus the bank's resources are necessarily exhausted in a run.

The notable result in this setup is that backward induction avoids a bank-run equilibrium. Since resources are preserved for later withdrawers, the last agent in line knows that he or she can either withdraw early and earn a discounted payout or hold the investment until maturity and earn what is left in the bank. The latter will always be greater because the investment has been allowed to mature, so if the agent's type is patient, the incentive is not to run. The penultimate agent understands that the final agent will only withdraw if truly impatient. Similarly, all agents know that all depositors who come after them face incentive to report their type truthfully, providing incentive to report truthfully themselves and never withdraw early.

Runs would be prevented. At the same time, their model has a very particular structure in which the shocks that agents face are uncorrelated. Ennis and Keister (2009) generalize the model by allowing the shock that one agent faces to reveal information about the aggregate state of the economy and thus the liquidity needs likely to be faced by other agents.[5] This can cause the Green-Lin mechanism to break down, once again enabling a run equilibrium: while the last agent still will report truthfully, earlier agents may withdraw early.

Peck and Shell (2003) also generalize the Green-Lin (2003) model but in a different way.[6] When agents in their model withdraw, they don't know what other agents have done — the game becomes simultaneous, not sequential as in Green-Lin. Since they do not know their place in queue, they cannot backward induct to a no-run equilibrium. Arguably this is a more realistic portrayal; bank depositors may know some events that are underway — based, for example, on news reports — but they will not necessarily have full certainty about other depositors' liquidity demand.[7]

One takeaway from this discussion is that several models have proposed solutions to runs, but the robustness of the solution depends very much on the assumptions of the model. Even to the extent that any one set of assumptions matches the real world, the real world is liable to change with new information and innovations in the financial industry. Thus, a surefire solution to runs remains elusive.

A New Proposal

Suspension clauses are one of the most common prescriptions for enhancing the stability of demand deposits. For example, they have been part of the reform of money market mutual funds (MMMFs), which experienced turbulence during the financial crisis. A 2014 rule adopted by the Securities and Exchange Commission recommended that MMMF boards of directors have the discretion to impose penalty redemption fees and redemption gates — effectively to suspend payments — in times of heavy redemption activity.[8] This proposal is consistent with D-D (1983), Green and Lin (2003), and Peck and Shell (2003). But the fact that bank-run equilibria remain possible in these models suggests that suspensions alone may not fully prevent runs or that suspensions can indeed prevent runs but at too high a cost.[9]

As noted, the core question is when to suspend payments. The answer depends on whether the run is based on fundamentals or fear. If withdrawals are driven by fundamentals — that is, depositors truly needing their funds for productive uses — then suspension is not desirable. But if withdrawals are driven by fear, it is desirable to suspend payments. The challenge is that only depositors know what is driving their withdrawals. If those agents had incentive to provide that information, the bank could in principle design a suspension clause with a trigger based on that information, eliminating the run equilibrium.

Much of the literature focuses on a particular direct mechanism that depositors could use to signal their individual liquidity needs: they simply withdraw their deposits. But this is not effective at signaling fears of a run. For example, the volatility of redemption rates varies across classes of MMMFs.[10] How are directors of those funds to determine whether withdrawals signal a run based on fear or fundamentals? This suggests that some *indirect* mechanism that contains information about *why* depositors are withdrawing is needed to prevent runs.

This information could be conveyed by a new type of financial instrument that rewards the depositor for delivering such a message when a run is occurring. In a 2017 paper, Andolfatto, Nosal, and Sultanum model such a mechanism.[11]

In their model, depositors have the option to pay a small fee to have their funds held in a priority account. Each depositor faces a choice between the priority account or a normal account from which depositors could withdraw early by forfeiting all interest. If a run occurred, the return on the priority account would exceed the return on the normal account (with or without early withdrawal from the latter). But if a run did not occur, the return on the normal account would exceed the return on the priority account.

In other words, choosing the priority account could be interpreted as a communication of an impending run; the fraction of depositors who choose the priority account over normal accounts informs the bank on the degree to which depositors expect a fear-based run. If depositors believe a run is occurring, they would be better off choosing priority accounts. If enough depositors exercise that option, the bank would know that a fear-based run is occurring. The suspension clause could then be conditional on this fraction hitting a certain threshold. Since depositors know suspensions would be triggered before fear-based redemptions affect their payouts, they would have no reason to actually exercise the priority option, and this option would simply never be used.

Conclusion

In general, understanding the sources of banking fragility helps explain why banks are supervised, the origins of the 2007–08 financial crisis, and the regulatory response to the crisis. Modeling fragility in the way described in this *Economic Brief* also provides guidance for designing regulations and institutional features that might prevent runs.

The mechanism provided by Andolfatto, Nosal, and Sultanum suggests that suspension clauses should be conditioned on information relating to depositor beliefs about the likelihood of a run. That information could be elicited through a modification to the deposit contract that would provide incentive for depositors to reveal that information.

How feasible is this proposal? Since it is difficult to run experiments in financial markets, there is little empirical information on which to assess its efficacy. One can imagine depositors as already having behaved as if they had this priority option, for example, when massive withdrawals from MMMFs were channeled into highly safe and liquid assets like Treasury securities. Regulators and fund managers widely interpreted this to be an indication of a run — for example, this type of occurrence triggered many of the Fed's unprecedented actions during the crisis to backstop certain markets. In this case, there was no priority account to keep investors from running, but it suggests that the necessary information may be available for the taking if depository institutions create arrangements for receiving it. ■

Renee Haltom is the editorial content manager and Bruno Sultanum is an economist in the Research Department at the Federal Reserve Bank of Richmond.

Endnotes

[1] For example, see Gary Gorton and Andrew Metrick, "Securitized Banking and the Run on Repo," *Journal of Financial Economics*, June 2012, vol. 104, no. 3, pp. 425–451.

[2] This is the essence of a proposal put forth by John Cochrane (2014) to tax or legislate demandable debt out of existence. See "Toward a Run-Free Financial System," in *Across the Great Divide: New Perspectives on the Financial Crisis* (Martin Neil Baily and John B. Taylor, eds.), Stanford, Calif.: Hoover Institution Press, 2014.

[3] For a detailed yet fairly nontechnical description of the Diamond-Dybvig model and later extensions, see Huberto Ennis and Todd Keister, "On the Fundamental Reasons for Bank Fragility," Federal Reserve Bank of Richmond *Economic Quarterly*, First Quarter 2010, vol. 96, no. 1, pp. 33–58. This discussion draws from theirs. See John Bryant, "A Model of Reserves, Bank Runs, and Deposit Insurance," *Journal of Banking and Finance*, December 1980, vol. 4, no. 4, pp. 335–344; and Douglas W. Diamond and Philip H. Dybvig, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, June 1983, vol. 91, no. 3, pp. 401–419.

[4] See Edward J. Green and Ping Lin, "Implementing Efficient Allocations in a Model of Financial Intermediation," *Journal of Economic Theory*, March 2003, vol. 109, no. 1, pp. 1–23.

[5] See Ennis and Keister, "Run Equilibria in the Green-Lin Model of Financial Intermediation," *Journal of Economic Theory*, September 2009, vol. 144, no. 5, pp. 1996–2020.

[6] See James Peck and Karl Shell, "Equilibrium Bank Runs," *Journal of Political Economy*, February 2003, vol. 111, no. 1, pp. 103–123.

[7] Ennis and Keister (2010) provide another important contribution to this literature in a model without commitment. One downside of suspensions is that some agents who actually need their reserves are not able to withdraw. In the Ennis and Keister framework, a bank that need not commit to implementing suspensions would ex post not want to maintain the suspension but instead make payments to agents, hoping to catch those who really have an urgent need for liquidity. Because depositors anticipate this scenario, they don't believe

suspensions would be implemented, making the suspension clause ineffective at preventing runs. See "Banking Panics and Policy Responses," *Journal of Monetary Economics*, May 2010, vol. 57, no. 4, pp. 404–419.

[8] See U.S. Securities and Exchange Commission, "SEC Adopts Money Market Fund Reform Rules," Press Release, July 23, 2014.

[9] For a nontechnical overview of why stemming runs in the MMMF case is so difficult, see Ennis and Haltom, "Reforming Money Market Mutual Funds: A Difficult Assignment," Federal Reserve Bank of Richmond *Economic Brief* No. 14-02, February 2014.

[10] See Lawrence Schmidt, Allan Timmermann, and Russ Wermers, "Runs on Money Market Mutual Funds," *American Economic Review*, September 2016, vol. 106, no. 9, pp. 2625–2657.

[11] See David Andolfatto, Ed Nosal, and Bruno Sultanum, "Preventing Bank Runs," *Theoretical Economics*, September 2017, vol. 12, no. 3, pp. 1003–1028.