Can Redemption Fees Prevent Runs on Funds?

Xuesong Huang
Lingnan College
Sun Yat-sen University

Todd Keister
Rutgers University

October 2024

Broad motivation

- Recurrent phenomenon: runs on banks and related institutions
  - Spring 2023: *Silicon Valley Bank (SVB)*, Signature, First Republic
  - Spring 2020: Money Market Mutual Funds (MMFs)
  - Fall 2008: investment banks, repo markets, MMFs, many more
- Much discussion and policy reforms on how to prevent runs
  - government guarantees, lender of last resort, capital requirements, liquidity regulation, etc.
- We look at one approach: redemption fees
  - adjust payments based on redemption/withdrawal demand
  - recent reforms to MMFs in the U.S. provide a concrete laboratory
    - but the ideas potentially apply much more broadly

Runs on MMFs

- Sept. 2008: runs on institutional prime MMFs
- July 2014: SEC modified the rules governing these MMFs
  - allowed to impose gates and redemption fees …
  - … when a fund's ratio of liquid to total assets falls below a threshold
- Interpretation: allow funds to operate as usual in normal times
  - but react to "unusually" high redemption demand
  - hope to put these events *off the equilibrium path of play*
- March 2020: runs on institutional prime funds again

  ⇒ the 2014 reform was ineffective

Recent reforms

- July 2023: SEC finalized new rules
  - removed the liquid-asset threshold and the option to use gates
  - impose fees based on *current redemption demand*

    "*A mandatory fee is charged to redeeming investors when the fund has net redemptions above 5% of net assets*."

- Interpretation: apply redemption fees more often
  - on the equilibrium path (when no run is occurring)

    "*We estimate that an average of 3.2% of institutional funds would cross a 5% net redemption threshold on a given day.*"

- Will the new reform work? What is the optimal fee policy?
  - how should the size of the fee and the threshold be set?

This paper

- Develop a model to study MMF redemption-fee policies
- Show: using fees only in extraordinary times is ineffective
  - fund is often susceptible to a preemptive run (~March 2020)
- Derive the best run-proof fee policy
  - can be complex, depends on difficult-to-measure parameters
  - but illustrates general principles for effective fee policies
- Derive the best simple, robust run-proof fee policy
- Compare to the 2023 reform
  - current approach is vulnerable when market liquidity may worsen
  - best policy has smaller fee that applies more often

Related literature

- Existing models of *preemptive* bank runs
  - Engineer (1989), Cipriani et al. (2014), Voellmy (2021)
- Runs on MMFs and patterns of redemptions at mutual funds more broadly
  - Chen et al. (2010), Schmidt et al. (2016), Parlatore (2016), Goldstein et al (2017), Zeng (2017), Cipriani & La Spada (2020), Alvados & Xia (2021), Jin et al. (2022), Li et al. (2021), and others
- Policy papers on MMF reform
  - Ennis (2012), McCabe et al. (2013), President's Working Group Report (2020), Ennis, Lacker and Weinberg (2023), and others
- Our contribution: if the goal is to prevent runs …
  - what *principles* should determine MMF redemption fees?

Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

4) Robust run-proof policies

- best policy vs. the 2023 reforms

5) Concluding remarks

Environment

- Investors: $i \in [0,1]$
  - endowed with one unit of good at $t = 0$, nothing later
- $t = 0,1,2,3$
- Technologies:
  - storage yields gross return of 1 in any period
  - investment at $t = 0$ yields: $\begin{Bmatrix} r_1 < 1 \\ r_2 < 1 \\ R > 1 \end{Bmatrix}$ at $\begin{Bmatrix} t = 1 \\ t = 2 \\ t = 3 \end{Bmatrix}$
    - R is known
    - r_2 may be random
- Utility: $\begin{Bmatrix} u(c_1) \\ u(c_1 + c_2) \\ u(c_1 + c_2 + c_3) \end{Bmatrix}$ if investor is $\begin{Bmatrix} \text{type 1} \\ \text{type 2} \\ \text{patient} \end{Bmatrix}$ ← "impatient" (type 1 and type 2)
  - focus on: $u(c) = \ln(c)$

Key assumptions

- Fraction of impatient investors (types 1 & 2) is known: π
- Fraction of type 1 investors is random: $\pi_1 \sim F[0, \pi]$
  - no uncertainty about *total* early redemption demand
  - but uncertainty about the *timing* of that demand
- Investors learn their type gradually
  - at $t = 1$, only learn whether or not they are type 1
- A fraction $\delta \in (0,1]$ of non-type 1 investors can redeem at $t = 1$
  - the remaining $1 - \delta$ are inattentive ("don't see the sunspot")
  - role: limits size of a potential run in period 1
  - assume δ is known (for now)

Efficient allocation

- A planner with full information would:
  - pay type 1 and 2 investors using goods in storage
  - pay type 3 investors using matured investment
- Log utility ⇒ planner will set: $c_1 = c_2 = 1$
  $c_3 = R$
  - portfolio: π in storage, $(1-\pi)$ invested

⇒ The same allocation as in a two-period model

Q: How might this allocation be decentralized …

- … when preference types are private information?

MMFs

- Suppose investors pool endowments, set up a *fund* that:
  - follows the planner's portfolio $(s, 1-s) = (\pi, 1-\pi)$
  - allows investors to choose when to redeem ($\Rightarrow$ a game)
- At $t = 1,2$: fund observes redemption demand m_t
  - then pays all redeeming investors

no
sequential service
within a period

- A *policy* specifies:
  - $c_1(m_1)$
  - $c_2(m_1, m_2)$
  - $c_3(m_1, m_2)$
- Easy to implement the planner's allocation as <u>an</u> equilibrium
  - example: set $c_1 = c_2 = 1$ for all (m_1, m_2) ("pay at par")
- But … is the fund susceptible to a run?

Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

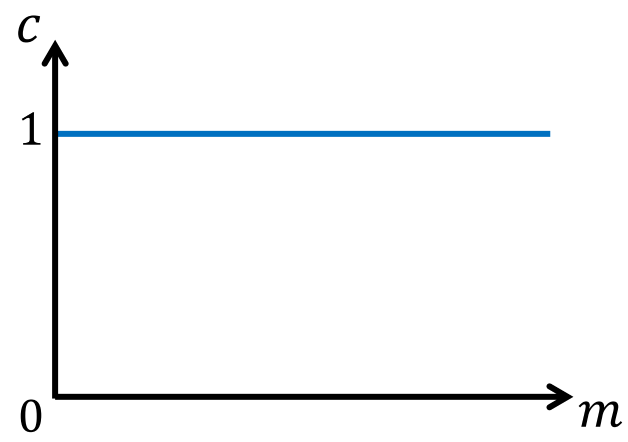
4) Robust run-proof policies

- best policy vs. the 2023 reforms

5) Concluding remarks

Classic runs

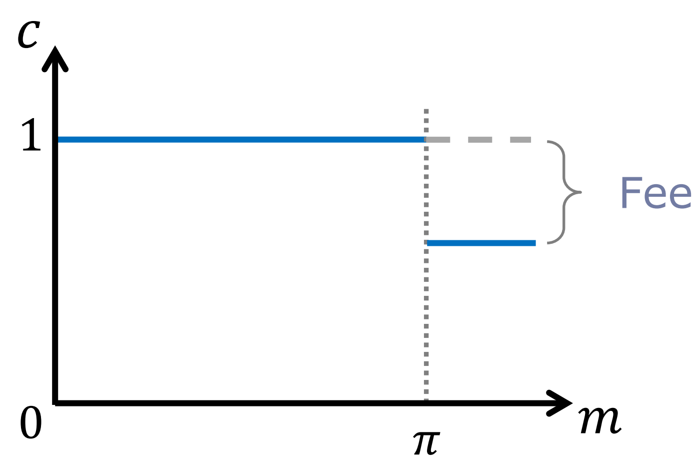
- Suppose the fund sets $c_1 = 1$ and $c_2 = 1$ whenever possible
- $t = 3$: remaining investors get pro-rata share of matured investment
  - if no run: dividend $= R - 1$
  - resembles pre-2014 rules for MMFs



- If investment is perfectly liquid …
  - … there is no bank run equilibrium (→ log utility)
- If r_1 is small enough: a bank run equilibrium exists …
  - … for the "classic" Diamond-Dybvig reason
    - one way of thinking about the runs on MMFs in 2008

Redemption fees

- Now suppose fund imposes a redemption fee …
  - if net redemptions are "extraordinary" (only consistent with a run)
  - here: $m_1 > \pi$ or $m_1 + m_2 > \pi$
- Aim: remove the incentive to run …
  - with fees that are *off-equilibrium* when there is no run (→ no cost)



Q: How to set the fee? 100%?

- Require the policy to satisfy *time consistency*
  - if m indicates a run, redemption fee must be ex-post efficient
    - in the spirit of Ennis and Keister (2009, 2010)

details

- Would still prevent runs in a two-period model. But …

Preemptive runs

- Suppose a non-type 1 investor expects a run at $t = 1$
- Compares the expected utility of:

$$m_1 = \pi_1 + \delta(1 - \pi_1)$$
$$m_2 = (1 - \delta)(\pi - \pi_1)$$

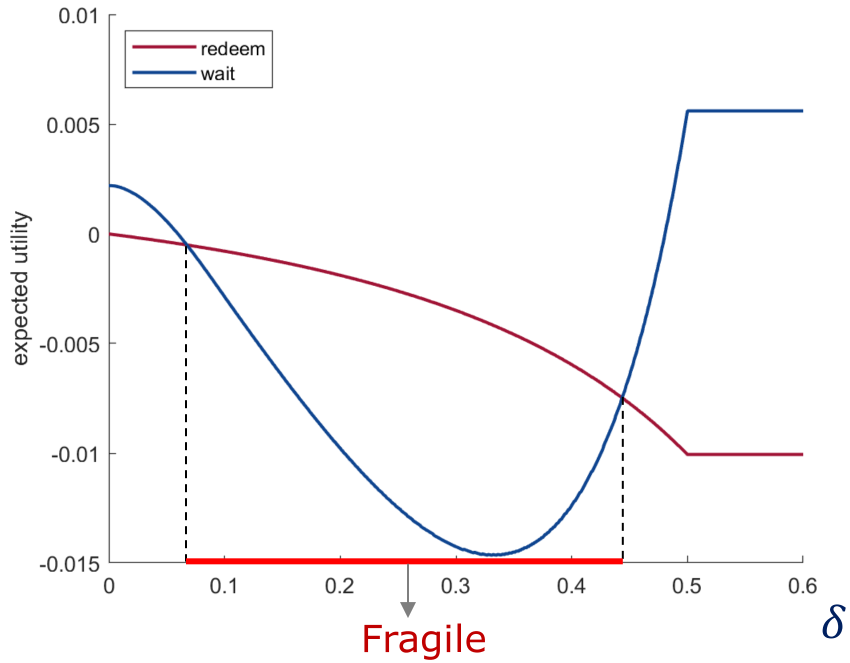
redeem: $$\int_0^{\pi} u\big(c_1(m_1)\big) f_n(\pi_1) d\pi_1$$

wait: $$\int_0^{\pi} \big[p_n u\big(c_2(m_1, m_2)\big) + (1 - p_n) u\big(c_3(m_1, m_2)\big)\big] f_n(\pi_1) d\pi_1$$

- If π_1 is large enough: $m_1 > \pi$ and run is detected immediately
  - fee imposed at $t = 1 \rightarrow$ no incentive to join the run
- Worry: if π_1 is small, run will not be detected until $t = 2$
  - a fee will be imposed then – and I might need to redeem
  - generates an incentive to redeem preemptively (today)

An example

- Compare $EU(wait)$ and $EU(redeem)$ as δ varies



Run equilibrium tends to exist …

- … when δ is moderate

When δ is large, a run is likely detected by the fund at $t = 1$

- fee applied at $t = 1$ (and $t = 2$)
- no incentive to redeem early

- When δ is small, a run is small ⇒ fund is in good shape
- In between: a moderate-sized run may initially go undetected
  - in this region: incentive to redeem before the fee is imposed

Takeaways so far

- We argue: the 2014 reforms had this flavor
  - allow funds to operate as usual in normal times
  - take action (fees, gates) if redemption demand is extraordinary
- Such policies can prevent “classic” runs …
- … but are often susceptible to preemptive runs
  - even when there is *no sequential service* within a period
- Danger comes from intermediate values of δ
  - a run that is large enough to cause damage …
  - … but small enough to go undetected in the first period

$\Rightarrow$ To prevent runs: need to impose fees in normal times as well

Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

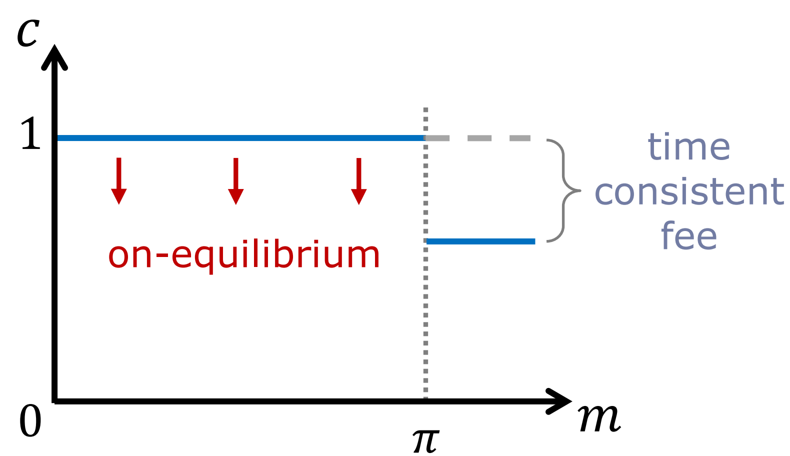
4) Robust run-proof policies

- best policy vs. the 2023 reforms

5) Concluding remarks

Run-proof policies

- If the previous policy is fragile for some value of δ …
  - need to impose fees in normal times $(m < \pi)$ to prevent runs
- What form should the fee policy take?



- The best policy:
  - maximizes investors' expected utility …
    - in equilibrium, where no run occurs
  - subject to the constraint that "wait" is a dominant strategy …
    - for non-type 1 investors at $t = 1$

Best policy

Choose the policy $c_1(m_1)$ for $m_1 \leq \pi$ to solve:

expected utility with no run

$$\max_{\{c_1(m_1)|m_1 \leq \pi\}} \int_0^\pi \left\{ \begin{array}{c} \pi_1 u\big(c_1(\pi_1)\big) + (\pi - \pi_1) u\big(c_2(\pi_1, \pi_2)\big) \\ +(1-\pi) u\big(c_3(\pi_1, \pi_2)\big) \end{array} \right\} f(\pi_1) d\pi_1$$

- subject to the run-proof constraint:

redeem $$\int_0^\pi u\big(c_1(m_1)\big) f_n(\pi_1) d\pi_1 \leq$$

if I expect all others to run

wait $$\int_0^\pi \big[p_n u\big(c_2(m_1, m_2)\big) + (1 - p_n) u\big(c_3(m_1, m_2)\big)\big] f_n(\pi_1) d\pi_1$$

- where $c_2(m_1, m_2)$ and $c_3(m_1, m_2)$ are:

  (i) feasibile

  (ii) chosen optimally for $m_1 + m_2 = \pi$

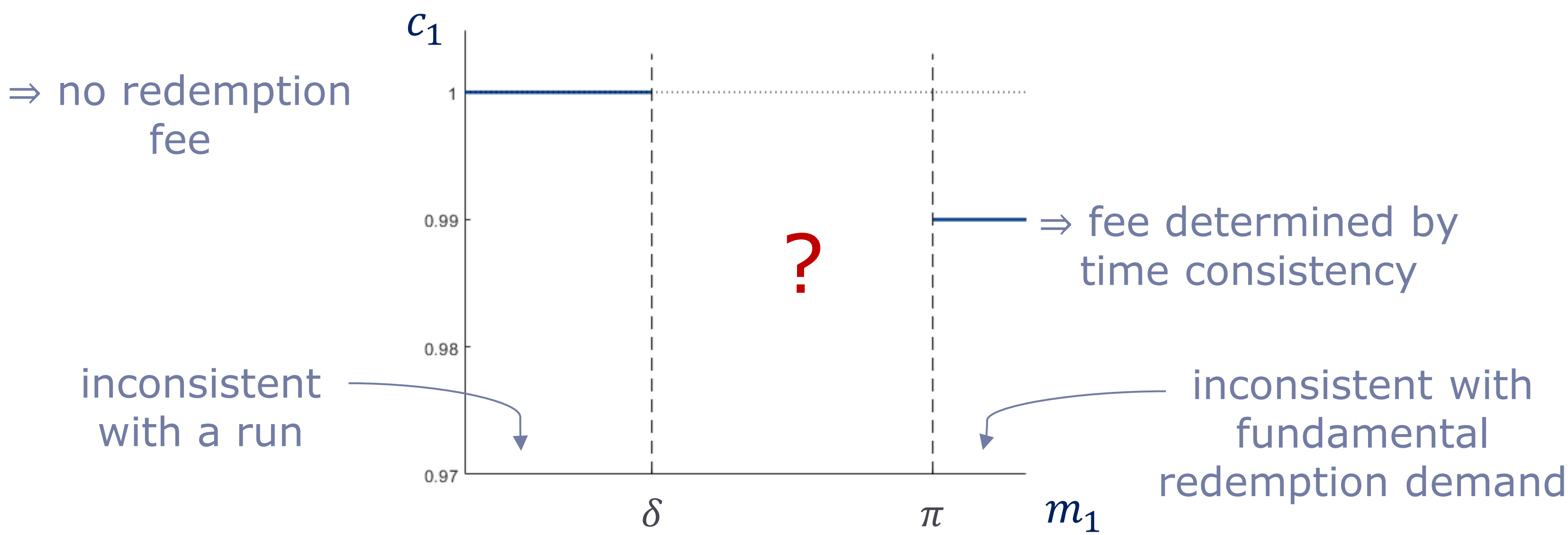
  (iii) time consistent for $m_1 + m_2 > \pi$

same functions in objective and constraint

but evaluated at *different points*

Three regions

- The best run-proof contract:



- When $m_1 < \delta$, fund is sure there is no run ⇒ no fee
- When $m_1 > \pi$, fund knows a run is underway

  ⇒ sets the time-consistent fee
- In between …

A general principle

- Optimal payout in the middle region depends on the ratio:

likelihood of m_1:

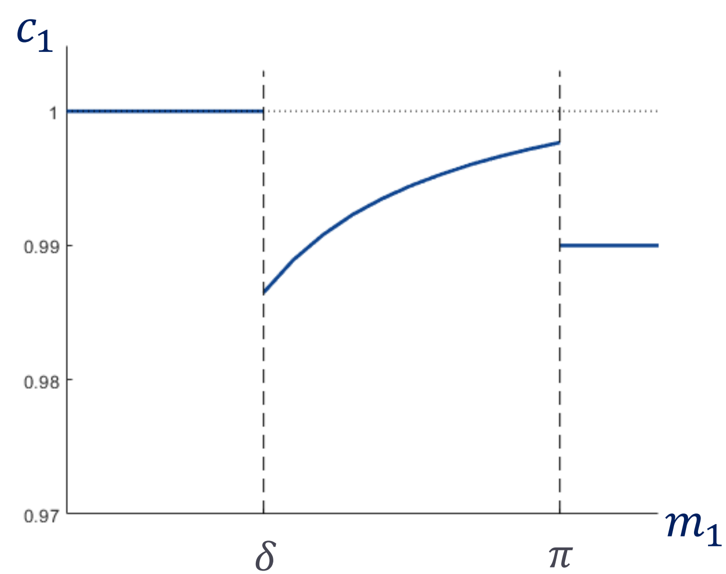
number of investors who pay the fee

conditional on no run

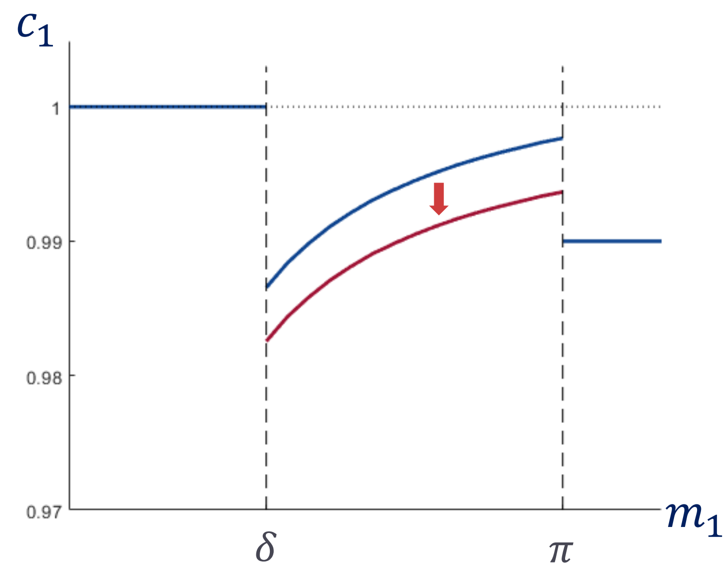
conditional on a run (if I am non-type 1)

$$\frac{f(m_1)\cdot m_1}{f_n\left(\frac{m_1-\delta}{1-\delta}\right)} \approx \frac{\text{decrease in objective}}{\text{benefit in meeting the constraint}}$$

- Overall shape depends on f
- But fee tends to *decrease* in this region (counterintuitive?)
  - costly to impose fees when many investors (truly) need the money



- Optimal fee schedule depends on:
  - current liquidation cost (r_1)
  - dist. of future liquidation cost (r_2)
    - or, investors' beliefs about r_2
  - the size of a run (if one were to occur →



- Concerns:
  - the optimal fee schedule is complex; could it be implemented?
  - may be difficult to measure beliefs of r_2 and incorporate into fee
  - may be difficult to measure δ
    - could use past run episodes (2008, 2020), but …
    - may change (ex: group of investors join same Slack channel)

Remaining steps

- We deal with these concerns in two steps
- First: restrict attention to *simple* policies
  - fee in the middle region can be zero or a constant
  - derive the best simple run-proof policy
- Second: study *robust* policies
  - require the policy to be run-proof for a range of r_2, δ
  - derive the best robust, simple run-proof policy
- Compare this policy to the 2023 reforms

Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

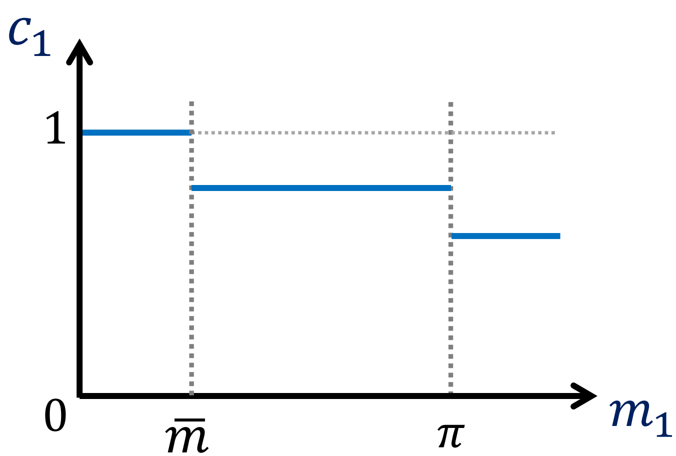
4) Robust run-proof policies

- best policy vs. the 2023 reforms

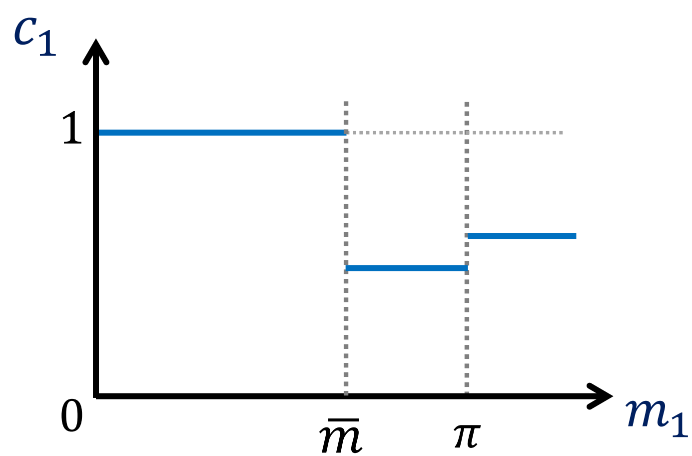
5) Concluding remarks

Simple policies

- A *simple* policy is characterized by two numbers:
  - $\bar{m}$: threshold below which no fee is applied
  - $\bar{c} < 1$: payment between the threshold $\bar{m}$ and π (fee $= 1 - \bar{c}$)



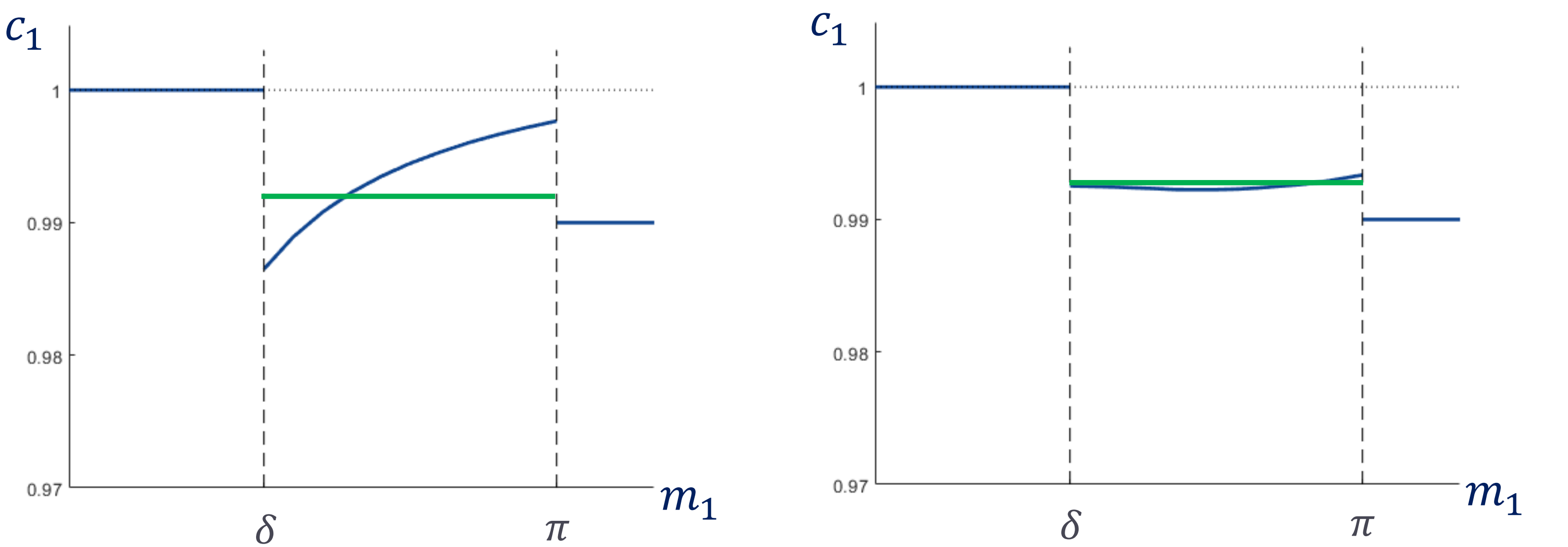
or



- Time-consistent fee still applies when $m_1 > \pi$
  - recall: lies off the equilibrium path
  - represents extraordinary actions (perhaps closing the fund)

Best simple policy

- Intuitively: best simple policy is an average …
  - … of the fee in the best general policy over $[\delta, \pi]$

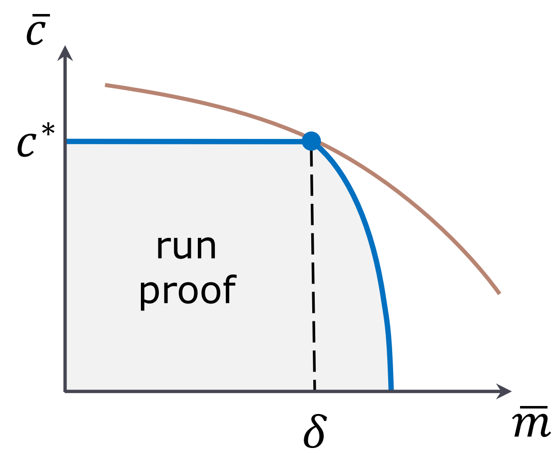


- Best simple policy tends to set $\bar{m} = \delta$
  - always: if best general policy is increasing (ex: if f is uniform)

$\Rightarrow$ Apply a fee whenever redemptions are consistent with a run

Another view

- Ask: what combinations of $(\bar{m}, \bar{c})$ are run proof?
- Suppose $\bar{m} = \delta$. Ask: what c^* would make policy run-proof?
- For $\bar{m} < \delta$: boundary is flat
  - because fee will always apply in run
- For $\bar{m} > \delta$: boundary is slopes down
  - larger $\bar{m} \rightarrow$ higher probability a run will not be detected until $t = 2$
  - requires a higher fee



- Best policy is often at the kink point (set threshold $= \delta$)
- But: optimal policy still depends on r_2, δ, so …

Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

4) Robust run-proof policies

- best policy vs. the 2023 reforms

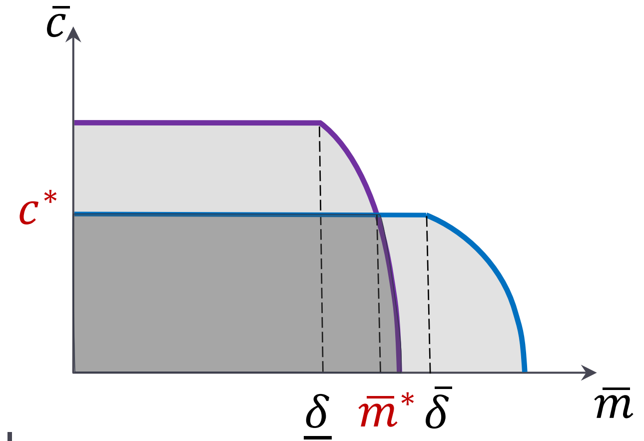
5) Concluding remarks

Robust policy

- So far, the optimal policy relies on knowing δ and r_2
  - which can easily change over time; difficult to monitor
- Robust approach: policy must be run-proof …
  - for <u>all</u> $\delta \in [0,1]$ and for <u>all</u> $r_2 \in [\underline{r}, 1]$
- The run-proof condition is monotone in r_2
  - lower (distribution of) r_2 always makes running more attractive
  - focus on the worst-case scenario: $r_2 = \underline{r}$ with probability 1
- The run-proof condition is <u>not</u> monotone in δ
  - recall: the danger is a run that is "medium-sized"
  - what is the worst-case scenario for δ?

Graphically

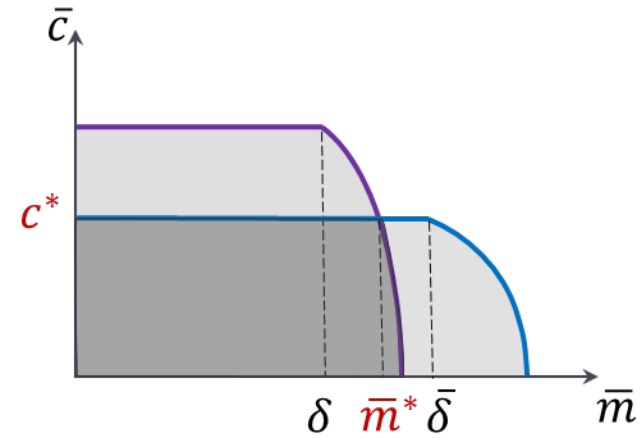
- Look at the intersection of the run-proof set for all $\delta \in [0,1]$
- Focus on two cases:
  - a large-ish $\bar{\delta}$
    - optimal threshold and fee are high
  - a smaller $\underline{\delta}$
    - both threshold and fee are both smaller



- Can show: there exits a unique (m^*, c^*) such that:
  - the robust run-proof boundary is flat up to (m^*, c^*)
    - then downward sloping
  - optimal robust policy is often (m^*, c^*) (always true if f is uniform)

In other words

- Model offers a *theory* of how $(\bar{m}, \bar{c})$ should be set
- Fee: set $\bar{c}$ to guard against "large" runs
  - a large run will very likely trigger the fee (⇒ $\bar{m}$ not important)
  - find worst-case large run $(\bar{\delta})$ → set fee to remove run incentive



- Threshold: set $\bar{m}$ to guard against "smaller" runs
  - a small run may or may not trigger the fee
  - find worst-case small run $(\underline{\delta})$ → set threshold to remove run incent.

Q: How does (m^*, c^*) how does it compare to the 2023 reforms?

Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

4) Robust run-proof policies

- best policy vs. the 2023 reforms

5) Concluding remarks

2023 reforms

- New rules require:
  - threshold: $\bar{m} = 5\%$
  - fee: determined by a "vertical slice rule"

"The size of the fee generally is determined by … costs the fund would incur if it were to *sell a pro rata amount of each security* in its portfolio to satisfy the amount of net redemptions."

- In our model:

$$\bar{c} = \pi + r_1 (1 - \pi) \quad \text{for } m_1 \in [\bar{m}, \pi]$$

  - note: equal to the time-consistent fee for $m > \pi$

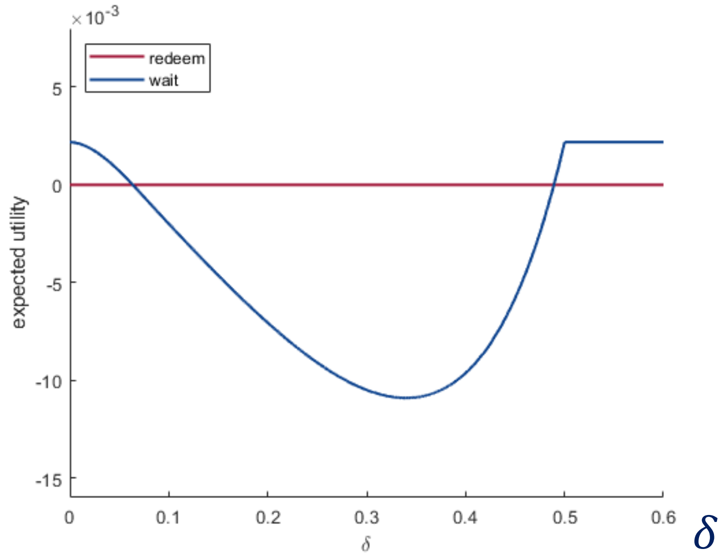
- Justification: "removes the first-mover advantage"
  - true in a sense. But …

Effective?

Q: Is this policy robust run-proof in our model?

A: No.

- suppose $r_1 = 1$, but r_2 may be < 1
- vertical slice rule sets fee = 0
  - ≈ the first policy we studied



- Model shows where current rules are vulnerable
  - if investors worry that market conditions may deteriorate …
  - a redemption fee based on <u>current</u> liquidation values is too small
    - investors fear the fee will increase → run preemptively

Efficient?

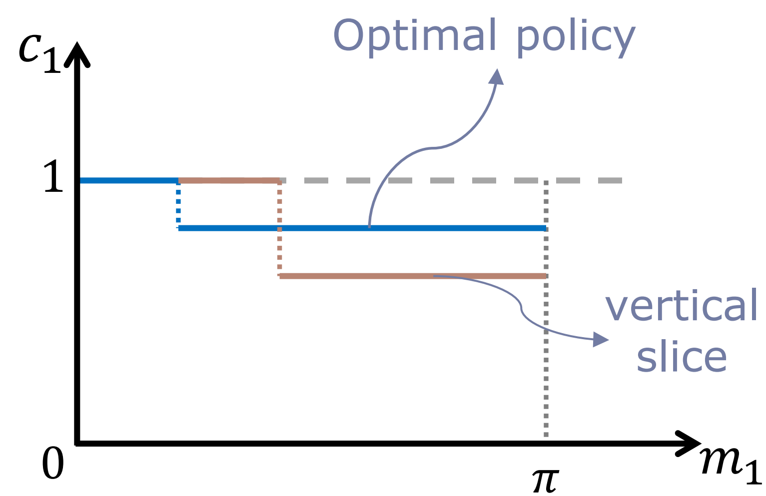
- One fix: use $\underline{r}$ in the vertical slice rule

$$\bar{c} = \pi + \underline{r}\,(1 - \pi) \quad \text{for } m_1 \in [\overline{m}, \pi]$$

  - price according to the worst-case scenario for r_t
  - set $\overline{m}$ to the maximum value that is run proof for all δ

- This policy is robust run-proof … but too harsh
  - large fee in states where many investors need to redeem

- Optimal fee is smaller
  - threshold is also smaller
    - fee is imposed more often, but fewer on investors



Outline

1) Model

2) Run equilibria

- classic vs. preemptive runs

3) Run-proof policies

- general principles; simple policies

4) Robust run-proof policies

- best policy vs. the 2023 reforms

5) Concluding remarks

Concluding remarks

Q: Can redemption fees prevent runs on funds?

- in a robust way, using a "simple" policy?

A: Yes

- plus: model illustrates how the fee and threshold should be set

- Would MMFs be useful/viable under this policy?
  - it depends … especially on $\underline{r}$
- Note: a backstop facility would set a floor for $\underline{r}$
  - could make this approach more viable/attractive
- We think this approach could also be applied more broadly
  - corporate bond mutual funds, and beyond?

Appendix

Time consistency in period 2

- What information does the fund have in period 2?
  - redemption demand in periods 1 and 2: (m_1, m_2)
  - remaining portfolio: (s_2, i_2)
- The time-consistent allocation (c_2, c_3) solves

$$\max_{\{c_2,c_3\}} \quad m_2 u(c_2) + (1 - m_1 - m_2) u(c_3)$$

$$m_2 c_2 + e_2 = s_2 + r\ell_2 \qquad \mathrm{e}_2 \geq 0$$

$$(1 - m_1 - m_2) c_3 = R(i_2 - \ell_2) + e_2 \qquad \ell_2 \geq 0$$

  - solution has $c_2 \leq c_3 \Rightarrow$ no incentive to run in period 2
- When $m_1 \leq \pi$ and $m_1 + m_2 > \pi$, $s_2 = \pi - m_1$ and $i_2 = 1 - \pi$
  - solution has $c_2 < 1 \rightarrow$ fee imposed in period 2

Time consistency in period 1

- If $m_1 > \pi$, the fund can forecast m_2
  - assumes a run is underway $\Rightarrow m_1 = \pi_1 + \delta(1 - \pi_1)$
    - observing $m_1 > \pi$ allows the bank to infer π_1
  - no run at $t = 2$ $\Rightarrow m_2 = (1 - \delta)(\pi - \pi_1)$
- Time consistency at $t = 1$ requires (c_1, c_2, c_3) to solve:

$$\max_{\{c_1, c_2, c_3\}} \quad m_1 u(c_1) + m_2 u(c_2) + (1 - m_1 - m_2) u(c_3)$$

$$m_1 c_1 + m_2 c_2 = s + r\ell$$

$$(1 - m_1 - m_2) c_3 = R(i - \ell) \qquad \ell \geq 0$$

  - solution has $c_1 = c_2 < c_3$ and $c_1 = c_2 < 1 \rightarrow$ fee imposed in period 1
- Note: redemption fee removes the incentive to run **if** the run is detected right away

return