ISSN 0956-8549

Market-Based Liquidity Transformation

By

Kathy Yuan

FINANCIAL MARKETS GROUP DISCUSSION PAPER NO. 978

September 2026

Market-Based Liquidity Transformation

Kathy Yuan*
LSE, CEPR, FMG

August 31, 2026

Abstract

We develop a theory of liquidity transformation performed by market participants. An in-kind wrapper (e.g., a corporate bond ETF), issued by a fee-collecting sponsor, is a portfolio of illiquid OTC securities trading in the secondary market. Market makers clear the wrapper, dealers the OTC market, and authorised participants arbitrage between the two. The wrapper price is a public signal on which privately informed holders coordinate selling in stress. The informed sponsor values the wrapper off the same price, so a run affects its decision to close the wrapper and sell the portfolio at a fire-sale discount, leading to possibly self-fulfilling failure. We decompose failure into fundamental-driven and panic-driven components. We evaluate new stability instruments more relevant than deposit-era ones.

Keywords: Liquidity Transformation, Fire Sales, Non-Bank Intermediation, Global Games, Financial Fragility
JEL: G23, G14, G01, G28

*email:k.yuan@lse.ac.uk

1 Introduction

The financial system is transitioning from bank-based to market-based intermediation. Non-bank financial intermediation reached $257 trillion in 2024, 51% of global financial assets, and grew at twice the pace of the banking sector (Financial Stability Board, 2025). The transition has a technological driver. Electronic trading platforms, algorithmic market-making, and big-data analytics on dealer inventories and order flow have sharply reduced search and matching frictions in over-the-counter markets, lowering the cost of basket assembly and unwinding for dealer intermediaries. In the US corporate bond market, the electronic share of investment-grade trading rose from about 20% in 2015 to about 48% in 2025 (Coalition Greenwich, 2025).

Liquidity transformation, the economic function that Diamond and Dybvig (1983) characterised for banks, is increasingly performed by markets. Investors with heterogeneous liquidity needs hold market-priced wrappers, most prominently exchange-traded funds backed by a basket of illiquid assets, instead of bank deposits, and dealer-authorised-participants (AP) running inventory books on their own balance sheets play the role that deposit-taking institutions played under the bank-based regime. The shift extends liquidity transformation to a wider class of underlying assets and a wider class of investor clienteles, but it also relocates the source of fragility: the binding constraints on intermediation are the dealer's balance-sheet capacity and whether the secondary-market price acts more as a coordination signal for runs or as a penalty on coordinated selling, not the intermediary's liquid reserves.

To study this, we provide a model of liquidity transformation in which the secondary-market price does what no bank deposit can: it aggregates the dispersed private information of investors and serves as the public signal on which unshocked patient holders condition their hold-or-sell decision, the dealer-APs anchor their arbitrage of the wrapper against a basket of illiquid assets, and the sponsor conditions the valuation that decides closure. Bank-deposit architectures have no market price. Existing wrapper models give the price a clearing role but not an informational one. We let the same price perform both at once, which to our knowledge the literature has not done.

These wrappers share a common institutional architecture, and each of its elements enters our model. Every wrapper is created and operated by a *sponsor*, a fund manager such as BlackRock, State Street, or Invesco, which defines the underlying basket, collects management fees, and retains the right to close the fund. Redemption is in-kind through dealer-APs, which run inventory books linking the wrapper's secondary market to the underlying OTC market and absorb flow imbalances on their own balance sheets, rather than first-come-first-served sequential service on a reserve pool. The wrapper itself is market-priced rather than redeemable at par. And when the sponsor closes the fund, the basket is sold into the OTC market, whose depressed prices feed back into the wrapper's NAV.

The paper's scope is the in-kind ETF wrapped around illiquid securities: corporate bond ETFs (LQD, HYG, JNK), bank-loan ETFs (BKLN), collateralised-loan-obligation ETFs (JAAA, CLOZ), and a growing pipeline of private-credit ETFs. Related architectures with different redemption promises, dealer portfolio trading of corporate bonds (Li et al., 2025), closed-end CLOs (Diamond et al., 2025), and stablecoins, par claims on a reserve pool rather than in-kind claims on a basket (Ma et al., 2024), are treated as extensions and discussed in the architecture section and the conclusion.

Three empirical observations motivate the failure structure we model, and each maps to a model element. ETFs trade at premiums and discounts to NAV that are larger and more persistent than costless AP arbitrage predicts (Petajisto, 2017; Pan and Zeng, 2019); in the model the discount is the price of limited absorption capacity. In March 2020, LQD traded at a 5% discount and HYG at 7–8% (Falato et al., 2021; Haddad et al., 2021) as bid-ask spreads on individual bonds widened and dealer balance sheets strained (O'Hara and Zhou, 2021), making creation and redemption de facto more expensive, and the Fed's Secondary Market Corporate Credit Facility (SMCCF) closed the discounts within weeks; in the model this is strained AP capacity, and the backstop that caps the fire-sale loss. The Grayscale Bitcoin Trust (GBTC) traded at 20–50% discounts over 2021–2023 as a closed-end trust without AP creation or redemption, and the discount closed upon spot-ETF conversion in January 2024; in the model this is the wrapper with its primary market shut.

These episodes indicate two regimes of wrapper failure: a *mild* regime with sustained discount but basket intact, of which GBTC is the limiting case and the March 2020 dislocation the acute one (discounts widened while primary-market activity fell relative to trading, and no large bond ETF closed), and a *severe* regime with sponsor liquidation and basket fire-sale; the population of small and specialised funds closed in 2020–2022 after sustained primary-market outflows documents the ingredients of the severe mechanism, closure reached through the primary market, though not the dealer-capacity-breach event itself. The severe-regime theory is therefore a theory of the closure-exposed segment of the wrapper universe: the stress of the large benchmark funds enters as the mild regime, and becomes severe only where closure risk is active.

The severe regime is sponsor-driven liquidation: the sponsor maintains an internal assets-under-management (AUM) cutoff, and when assets persist below it, closure stops creations, the basket is sold through the AP, and proceeds are distributed pro-rata. Closure is common. US ETF closures run at roughly 100–300 per year against a universe of about 3,500 funds, with sponsor AUM thresholds typically in the \$25M–\$100M range[1] and bond ETFs over-represented in the 2020–2022 stress window. The liquidation channel, not cash-redemption dilution, is what exposes patient wrapper holders to the basket fire-sale loss. We find, in the full population of US bond-ETF closures over 2008–2024, that among the closures that lost assets before delisting, shares outstanding account for 95 percent of the median decline, and that trailing net redemptions predict the closure announcement conditional on the sponsor's fee revenue, the signature of the continuation decision we model (Appendix D).

Based on these empirical observations, we propose a model of liquidity wrappers and study the economic mechanism underlying their fragility. We define a security as liquid if it can be sold before maturity at a small discount to its expected value. In our model, a continuum of illiquid securities, each with a systematic component θ and an idiosyncratic component ε_i, is wrapped into

[1]Closure counts are from Morningstar (January 2021); the threshold range is from industry reporting of sponsor batch closures (ETF.com, December 2019).

a liquid portfolio. By the law of large numbers this liquid portfolio bears only θ risk. The model has three dates. At $T = 0$, heterogeneous investors with private liquidity-shock probabilities ψ_h choose between holding individual illiquid securities directly or holding the liquid wrapper. At $T = 1$, shocked investors must sell for cash, and unshocked patient wrapper holders play a global game in which they choose between holding to $T = 2$ and selling the wrapper share at the secondary-market price. Holders who hold to $T = 2$ receive NAV if the wrapper survives, and the depressed liquidation proceeds if the wrapper has been forced to close.

A small number of dealer-APs arbitrage the wrapper-vs-NAV discount by buying wrapper shares, redeeming them in kind, and selling the basket's illiquid securities to OTC dealers.[2] The APs are mechanical: like the arbitrage desks they represent, they trade the observed gap between the price and the basket's public valuation, the estimate anyone can form from the price alone, without a view on fundamentals, and their capacity is limited by trading-book and warehouse charges and by the OTC dealers' balance sheets, a common pool their unwind flows share. The sponsor is the informed institution: it observes a private signal, learns from the wrapper price, sets the fund's NAV, and closes the wrapper when its valuation is low relative to the redemptions the fund has absorbed.

Two regimes of wrapper failure emerge endogenously, matching the two empirical regimes above. In the *mild* regime, the wrapper trades at a sustained discount; the basket is intact, so $T = 1$ sellers absorb the loss through the discount and patient holders receive NAV at $T = 2$. In the *severe* regime, the sponsor's valuation reaches the floor, by redemptions that shrink the share count, by a falling NAV that devalues every remaining share, or by both; the wrapper is liquidated, the residual basket is sold wholesale into the OTC bond market, and combined OTC flow exceeds dealer capacity. The resulting fire-sale lowers the basket's value by an amount D that patient holders bear. The two regimes share a common equilibrium structure but differ in who bears the loss: $T = 1$ sellers in the mild regime, patient holders in the severe regime. The severe regime endogenises Diamond–Dybvig strategic complementarity in a market-based setting: more selling raises closure probability, lowers the expected hold payoff, and induces more selling.

In this theory of liquidity transformation, the secondary-market price of the wrapper plays a dual role. The price clears the wrapper market by equating mechanical and strategic selling against the wrapper-side absorption that APs and market makers jointly provide, and so it determines the proceeds the shocked seller receives. This clearing role generates the substitution effect that is standard in noisy rational-expectations models (Yuan, 2005): selling has price impact, so each unit of strategic selling clears at a wider discount, and the discount discourages further selling.

The same price is also a public signal: patient holders observe it before they choose between holding to maturity and selling at $T = 1$, and the sponsor observes it before valuing the fund whose closure it decides. This signal role generates an information effect: a lower price leads holders to infer that others have received worse fundamental information, which strengthens their incentive to sell, and it leads the sponsor to lower its valuation of the fund.

[2] Since our focus is on the fragility of wrappers, we model the wrapper-discount side rather than the premium side.

In this model, the possibility of failure induces a strategic complementarity among holders' selling decisions: holders want to sell when others sell, and wrapper failure can be self-fulfilling, as in a bank run (Diamond and Dybvig, 1983; Goldstein and Pauzner, 2005). Unlike a bank run, in which withdrawals deplete the bank's liquid reserves, coordinated selling puts pressure on the absorption capacity of the market makers and the APs, and on the sponsor's continuation decision, through two channels that generate information feedback. First, the market maker's demand is downward sloping and its depth is limited, and the APs' redemptions respond to the arbitrage wedge mechanically. When others sell, the price falls, the wedge widens, the absorbed flow fills the APs' warehouse, and redemption accelerates, depleting shares outstanding. Second, the sponsor closes the wrapper when AUM falls below the level at which its fee revenue covers the fixed cost of operating the fund. The accelerated redemption raises the valuation threshold at which the sponsor closes, and because the sponsor infers the basket's value partly from the price, the states with heavy selling are the states with low valuations: closure risk rises for everyone who holds. Consequently, facing the higher closure risk, more holders sell.

This feedback, coordinated by an endogenous public price, is the mechanism through which public information typically restores multiple equilibria in coordination models (Angeletos and Werning, 2006; Hellwig et al., 2006). In our model the information feedback is weaker when the marginal holder retains more uncertainty about the sponsor's closure decision, because the sponsor's valuation embeds a private signal about the wrapper fundamental that no market observable transmits. It is weaker still when the fire-sale loss is small relative to this valuation dispersion.

Ozdenoren and Yuan (2008) show that the equilibrium is nevertheless unique when the substitution effect dominates the information effect, which occurs there when limited market depth gives selling a large price impact. We show that in our setting the substitution effect does not come from the market maker's depth, which is a fixed market parameter in Ozdenoren and Yuan and in the existing global-games models. It comes from an institutional feature unique to liquidity wrappers, the AP sector's absorption, whose strength is set by the APs' endogenous disposal of absorbed shares and varies across price regimes. Absorption by the APs, whether the absorbed shares are warehoused or redeemed, reduces the price impact of coordinated and noise selling alike, and so weakens the substitution penalty on selling. The two disposal modes differ in their effect on closure. A warehoused share only absorbs selling pressure. A redeemed share also retires shares outstanding and moves the fund toward the sponsor's floor, so the redeemed part of absorption strengthens the coordination incentive that the price penalty restrains. And because redeeming into a possible closure is costly, the APs' demand itself recedes as closure risk rises: the substitution effect is weakest in calm states and steepens where closure risk is high.

Uniqueness in our model therefore depends on a comparison: the information feedback against the state-dependent substitution effect that the APs provide with their limited capacities. Market depth does not enter this comparison. Depth determines how far one unit of selling moves the price, and the same price movement drives both sides, the penalty on selling and the fear of closure. A deeper market scales both down in the same proportion, so depth cancels. What decides whether self-fulfilling failure is possible is the APs' redemption response against the sponsor's valuation dispersion. Market makers cannot redeem, so market-maker depth is not required for uniqueness; it enters only as further discipline. Absorption capacity, market-maker depth plus the AP's trading-book and warehouse capacities, instead governs the size and likelihood of the failure regimes. The trading book has two opposing effects: it narrows the discount, and it speeds the depletion of shares.

Our first set of results isolates what the price's information role adds to the run. We present a series of benchmarks, each of which removes one force from the model, to illustrate how the informative market price contributes to fragility. Without the severe-state closure loss, $d_H \bar{P} + D = 0$, the model is a standard noisy rational-expectations setting: the substitution effect dominates the information effect, the equilibrium is unique, and there is no run. Without an informative price, the coordination channel and the valuation feedback are absent, but the mechanical redemption margin is at its widest, because price movements with no fundamental content still pull baskets out of the fund; the equilibrium is unique, disciplined by the sponsor's unlearnable signal. An informative price coordinates holders, partially reveals the closure decision, and transmits selling back into the sponsor's valuation; the same informativeness anchors the arbitrage to value and narrows the redemption response. With a perfect price the wedge closes outside the closure region, and closure reduces to a fundamental valuation event there; what survives at that extreme is a narrow band of states near the closure boundary, small in our computations, where the classic bank-run indeterminacy returns through the APs' access advantage. Away from the band, self-fulfilling failure requires something else, a sponsor valuation mechanical enough that holders can reconstruct the closure event from the price, so that coordinated selling produces the closure it anticipates.

We then decompose the failure probability into a fundamental-driven component, the failures that fundamentals, noise trading, and arbitrage flows produce on their own, and a panic-driven component, the failures that would not occur without the fear of closure. The panic-driven component is the analogue of the panic-based runs of Goldstein and Pauzner (2005), and it requires an intermediately informative price: informative enough to coordinate on, and noisy enough to leave a wedge for arbitrage to redeem against. The benchmarks bound the panic component from both sides. When the price carries no information, the panic component is zero even though failure is frequent: noise selling keeps the redemption margin wide open, and every failure would have occurred without the fear of closure. When the price is nearly perfect, the panic component is small: the wedge closes outside the closure region, and what remains is the narrow boundary band. In our computations the panic component peaks between the two informational extremes, near the base parameter values. Figure 1 plots the decomposition across the range.

Three further results sharpen the analysis. The first separates AP balance-sheet strain from the information channel. The APs choose their redemptions and their warehousing at every state; what is parametric is the cost of their capacity. Because the APs are mechanical arbitrageurs between the wrapper price and the OTC prices of the basket's individual securities, their balance-sheet pa-

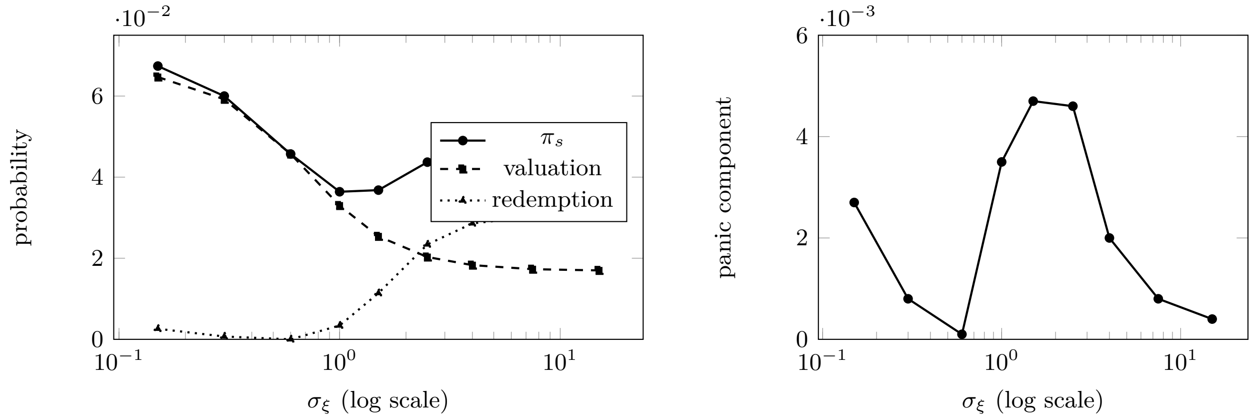

Figure 1: The decomposition of wrapper failure against price informativeness. The horizontal axis of both panels is the volatility of noise demand in the wrapper market, σ_ξ, on a log scale; a smaller σ_ξ means noise moves the price less, so the price is more informative about the fundamental. The base value is $\sigma_\xi = 1.5$, and all other parameters are at the base values of Table 1. In the left panel the vertical axis is a probability: the solid line is the severe-failure probability π_s, the probability that the sponsor closes the wrapper and the basket is sold at a fire-sale loss; the dashed line is its valuation component, the probability of closure with no redemptions, in which the sponsor's valuation alone reaches the floor; the dotted line is the redemption component, π_s minus the valuation component, the part of failure produced by redemptions retiring shares outstanding. In the right panel the vertical axis is the panic-driven component of π_s, defined in (45): the failures that would not occur without the fear of closure.

rameters do not affect the information content of prices. All effects of capacity are mechanical, and their direction is the opposite of the usual presumption. In our computations, more trading-book capacity raises total closure risk and nearly triples its panic-driven component, because capacity converts discounts into redemptions faster. AP strain, falling capacity, lowers closure risk at the cost of deeper discounts.

The second result concerns the structure of the AP sector. Each AP's redemption widens the OTC discount all APs receive and moves the fund's valued AUM toward its floor; each internalises only its own share of the damage. Fewer, larger APs therefore redeem less against the same state, and in our computations fragmentation raises both closure risk and its panic-driven share. The observed concentration of primary-market activity in a handful of APs is, through the model, a stabiliser, and, holding aggregate capacity fixed, recruiting a broader AP base would weaken it.

The final result endogenises the wrapper architecture itself, a choice the existing literature takes as given. The same assets can reach investors through an open-end wrapper that redeems in kind, through a closed-end fund, or not at all, and the choice belongs to the sponsor, which pays the setup and operating costs and collects the fee on AUM. We add this choice as a stage before trading begins. The closed-end fund is the wrapper with its primary market permanently shut: no shares are created or redeemed. Shutting the primary market closes the redemption margin, so the equilibrium is unique for any fire-sale loss, but the valuation margin stays open, and with no arbitrage anchoring the price to NAV the price is untethered, as GBTC's years of deep discounts illustrate.

Under risk neutrality the two forms compete for the same liquidity-demanding clientele, because the untethered closed-end price is costless in expectation; pricing its dispersion, which a shocked seller bears at the moment of need, assigns the liquidity-needers to the wrapper and the patient types who trade on their information to the closed-end fund. What only the open-end wrapper can serve is the mandate-driven institutional base, index funds and insurance portfolios, that requires a price tied to NAV. The sponsor weighs the wrapper's larger fee base against its closure risk and chooses the open-end form only while closure risk stays below a threshold. The threshold organises the cross-section of architectures: assets with broad benchmark demand are wrapped open-end, and assets with thin index demand and heavy fire-sale exposure, such as the leveraged loans held in CLOs, are wrapped closed-end. Anticipated fragility therefore destroys liquidity transformation at the launch decision, before any run occurs: an economy with high closure risk loses liquidity transformation through two channels, the wrappers that launch and later fail, and the wrappers that sponsors, anticipating the risk, never launch.

Our comparative statics deliver empirically testable hypotheses on how balance-sheet capacities, the sponsor's floor, and price informativeness move discounts and failure probabilities. We then ask the policy question in positive terms: which tools stabilise the wrapper, in the sense of relaxing the uniqueness condition and shrinking the failure regimes? Most of the deposit-era toolkit loses its objects here: there is no par claim for insurance to guarantee and no reserve pool that meets redemptions, because redemption delivers the basket itself. The instrument that does carry over is suspension of convertibility, whose primary-market analogues, gates and redemption fees, stabilise at the cost of the liquidity transformation the wrapper exists to provide. The model has implications for the instruments that act on market-based liquidity transformation directly, on the fire-sale loss, the primary market, the sponsor's fixed cost and valuation process, and the dealers' balance sheets. Section 8 evaluates them through their effects on the uniqueness condition and the failure probabilities. As liquidity transformation migrates from banks to markets, the tools of stability policy must move with it.

The remainder of the paper is organised as follows. Section 2 relates to the literature. Section 3 lays out the setting: securities, markets, the wrapper architecture, and information. Section 4 introduces the agents, including market makers, dealer-APs, and the sponsor, whose closure rule is derived from its participation constraint, together with the noise-demand shock. Section 5 specifies wrapper secondary-market clearing and the OTC bond market. Section 6 defines the monotone rational-expectations equilibrium and proves existence and uniqueness. Section 7 conducts the equilibrium analysis, including the failure decomposition and the analysis of strained AP capacity. Section 8 develops the policy analysis in positive terms, as effects on the uniqueness condition and the failure probabilities. Section 9 summarises the architecture-choice implication, the sponsor's choice between the open-end wrapper and a closed-end fund, with the full analysis in the online appendix. Section 10 concludes. The appendix collects a notation table, the proofs, the robustness analysis of the market maker's schedule, and case histories of sponsor closures. An online appendix, for online publication, contains the full architecture analysis, the welfare analysis, and the robustness of the comparative-statics signs.

2 Related literature

Our modelling framework is closely related to the literature on bank runs, global games, and learning from prices. Diamond and Dybvig (1983) provide the canonical model of bank-based liquidity transformation; there is no market price, and coordination runs through the sequential-service rule. Carlsson and van Damme (1993) introduce global games as an equilibrium-selection device; Morris and Shin (1998) apply the selection argument to currency attacks; Goldstein and Pauzner (2005) extend it to bank runs. In these models the public signal is either absent or exogenous to market clearing. Angeletos and Werning (2006) and Hellwig et al. (2006) endogenise the public signal as a market price and show that the price restores multiple equilibria. Yuan (2005) identifies the opposing substitution effect, selling has price impact, and Ozdenoren and Yuan (2008) show that the equilibrium is unique when the substitution effect dominates the information effect of the price.

Goldstein and Pauzner (2005) also show that the unique equilibrium separates failures into two kinds: fundamental-based failures, which occur even if no depositor runs, and panic-based failures, which lie between the fundamental threshold and the run threshold and would not occur without the run. Our decomposition of failure into fundamental-driven and panic-driven components (Section 7.2) is the counterpart of theirs. The departure is the channel that governs the panic component. In their framework coordination runs through the sequential-service rule, and the deposit contract disciplines the panic. In ours coordination runs through the market price, and the panic-driven component is governed by the price's informativeness: it vanishes along the noisy limit, is maximal in between, and at the informative extreme is confined to a narrow boundary band; self-fulfilling failure requires not an informative price but a sponsor valuation close enough to a function of the price alone. In addition, our fundamental benchmark is an affine equilibrium rather than a dominance region, because failure can be caused by noise trading and arbitrage flows as well as by the fundamental.

Ozdenoren and Yuan (2008), Goldstein et al. (2011), and Goldstein et al. (2013) study how intermediaries and speculators learn from prices when their own actions feed back into fundamentals, generating informational complementarities. The price plays a dual role in their setting: it clears the market and aggregates information. Our contribution is to put the dual role of the secondary-market price at the centre of non-bank liquidity transformation. In-kind redemption changes the trade-off between the two effects of the endogenous public signal. AP absorption weakens the price penalty on coordinated selling. The absorbed shares that the APs redeem also retire shares outstanding and move the fund toward the sponsor's floor, so this part of absorption strengthens the incentive to coordinate. Furthermore, because redeeming into a possible closure is costly to the APs, absorption itself recedes in stress. The substitution effect therefore varies across price regimes rather than being a fixed parameter. This variation is what lets the model map the region of unique equilibrium, the region where self-fulfilling failure is possible, and the panic-driven component of failure. It also relocates the uniqueness question: market depth, which decides uniqueness in that setting, is not required in ours, and what decides it is the APs' redemption response against the dispersion of the sponsor's valuation. The sources of multiplicity are institutional, the in-kind redemption system and the closure rule, not market depth. The modelling apparatus is contemporaneous with Lehar et al. (2026) on liquid staking: both papers use a rational-expectations equilibrium with binary action and informational complementarity. The application differs. We study a wrapper that holds illiquid underlying securities, mediated by mechanical arbitrage desks and closed by an informed sponsor whose valuation rule is the locus of coordination.

Petajisto (2017) documents persistent ETF premiums and discounts. Pan and Zeng (2019) are the closest predecessor for our AP-side mechanism. They model a dealer-AP whose balance sheet supports both OTC inventory-making in the underlying bonds and arbitrage of the ETF-vs-NAV spread through creation and redemption. Because the two activities share capacity, AP arbitrage is state-dependent: the discount widens when inventory is full or OTC unwinding is costly. Their model rationalises the persistent NAV deviations of Petajisto (2017) and the March 2020 corporate-bond ETF dislocation. In Pan-Zeng the wrapper price functions only as a clearing object: patient holders are not modelled, the AP responds to the level of the discount mechanically rather than extracting information from it, and the price plays no informational role. Their persistent-discount equilibrium maps directly into our *mild* regime, where the wrapper trades below NAV because the AP cuts arbitrage while the basket is intact. Our contribution is the dual role of the wrapper price: the same price that clears AP arbitrage also serves as the public signal that patient holders observe in their hold-or-sell decision and the sponsor observes in valuing the fund it can close. This coupling delivers two new objects that are absent in Pan-Zeng: a closure margin driven by the AP's redemption response to the discount and by the sponsor's price-based valuation, with a uniqueness condition governed by the holders' residual uncertainty about the sponsor's closure decision, and a *severe* regime in which redemptions and a falling valuation push AUM below the sponsor's floor, the residual basket is sold wholesale into the OTC market, and patient holders bear the fire-sale loss. Todorov (2021) studies AP basket selection. Dannhauser (2017) and Goyal et al. (2025) link ETFs to underlying bond-market liquidity. Li et al. (2025) document that dealer portfolio-trade pricing is jointly sensitive to portfolio size and prior inventory, supporting a unified balance-sheet-capacity treatment of AP arbitrage and OTC market-making.

An emerging literature studies market-based financial intermediation in stress, building on the pecuniary-externality and fire-sale frameworks of Lorenzoni (2008) and Stein (2012). Chen et al. (2010) and Goldstein et al. (2017) document run-like dynamics in open-ended mutual funds: concave flow-to-performance generates payoff complementarities among redeeming investors, the empirical counterpart to the strategic complementarity our model formalises. Schmidt et al. (2016) document money-market-mutual-fund runs around the 2008 Reserve Primary Fund episode. Within the ETF wrapper specifically, Falato et al. (2021) and Haddad et al. (2021) document the March 2020 bond-

ETF dislocation, and O'Hara and Zhou (2021) the illiquidity of the underlying corporate-bond market and the dealer balance-sheet constraints behind it. Ma et al. (2024) develop a Diamond–Dybvig model of stablecoin runs with stochastic intermediary capacity. Diamond et al. (2025) study CLOs as fire-sale insulation through non-runnable architecture; we treat their CLO design as the closed-end counterpart to our open-end wrapper in Section 9.

3 Setting

3.1 Securities and the wrapper basket

A continuum of illiquid securities indexed by $i \in [0,1]$. Each security pays at $T = 2$

$$v_i \ = \ \mu + \theta + \varepsilon_i,$$

with systematic factor $\theta \sim N(0, 1/\tau_\theta)$ and idiosyncratic component $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ i.i.d. across securities, both independent of all other shocks. By the LLN, the equally-weighted basket at $T = 2$ pays $\mathrm{NAV}_2 = \mu + \theta$. The $T = 1$ NAV is $\mathrm{NAV}_1 = \mu + E_1[\theta]$, with E_1 the sponsor's posterior, whose information set Section 4.5 states.

3.2 Markets and prices

Each security trades over-the-counter (OTC) at mid-price $P_{i,t}$. A forced seller at $T = 1$ receives $P_{i,1} - d_H\bar{P} - D\,\mathbf{1}\{\text{severe}\}$, where $d_H\bar{P}$ is the baseline OTC discount in dollar terms and D is the additional basket fire-sale loss in the severe state; $P_{i,1}$ settles at the security's fundamental value, a settlement convention under which the date-1 price content is carried by the discounts rather than by differences in agents' posteriors. The wrapper trades on a continuous exchange at price P_E^t. Risk-free rate $r_f = 0$. We normalise to $\bar{P} \equiv E_0[\mathrm{NAV}_1] = \mu$, using $\bar{\theta} = E[\theta] = 0$.

3.3 Wrapper architecture and closure

The wrapper is in-kind: dealer-APs create and redeem wrapper shares against the underlying basket. APs that redeem take physical delivery of the underlying securities, subject to the AP sector's aggregate trading-book capacity χ, of which each of the n APs operates χ/n; they do not force the sponsor to liquidate basket assets to pay cash. Patient holders therefore retain a pro-rata claim on the underlying basket as long as the wrapper continues to operate.

The wrapper does not have unlimited life. The sponsor closes the fund when its estimate of the fundamental is low relative to the redemptions the fund has absorbed. Both parts of that comparison are economic. A low fundamental means a low valuation of the basket, poor prospects for the fee revenue the fund generates, and a low price at which the fund could be sold to another sponsor; accumulated redemptions raise the cost of running the fund, since each in-kind redemption consumes administrative resources and a shrinking base carries the fixed overhead on fewer shares.

When the sponsor closes, it sells the residual basket into the OTC market through APs and distributes the cash proceeds pro-rata to remaining holders. Section 4.5 states the continuation test formally; its observable counterpart is the AUM threshold that sponsors publish, and Appendix D documents both the thresholds and the closure process.

Closure is the institutional mechanism through which patient holders are exposed to basket fire-sale loss in an in-kind wrapper: while the wrapper survives holders keep the basket exposure $\mu + \theta$, but on closure they receive the OTC sale price of the residual basket, which is depressed when total OTC flow exceeds dealer capacity.

Two quantities enter the continuation test, and both matter for the closure rule. The first is accumulated redemptions. Secondary-market sales transfer wrapper shares from one investor to another, or from an investor to a market maker; they do not retire shares and so add nothing to the cost of running the fund, unless the selling depresses the price. Shares are retired only through the primary market, when a dealer-AP delivers accumulated shares to the sponsor and takes out the redemption basket; the AP's redemption decision is derived in Section 4.4. The second is the sponsor's valuation of the basket, its NAV conditional on price, an administered number: the sponsor computes it from its own information about the basket, which includes what it learns from the wrapper price. The sponsor closes when this valuation is low relative to accumulated redemptions; in asset terms, when valued AUM is at the floor.

The sponsor's valuation matters because it opens a second margin through which the fund can reach the floor. Redemptions shrink the share count; a falling price lowers the sponsor's assessment of NAV, which shrinks the value of every remaining share, so the fund can be revalued down to its floor while the primary market is silent. Because the sponsor learns from the price, the two margins are connected: the price that heavy selling depresses is the same signal from which the sponsor estimates the basket, so low prices and low valuations arrive together, and price movements with no fundamental content depress the valuation directly. The formal closure condition, and the probability agents attach to it, are derived in Section 5.2.

3.4 Information

At $T = 1$ a wrapper holder either is hit by a liquidity shock, in which case she must sell for cash, or is not, in which case she chooses whether to sell at $T = 1$ or hold to $T = 2$. We call the latter the patient holders. Section 4.1 specifies the shock probabilities and the portfolio choice they induce at $T = 0$.

Each patient wrapper holder h at $T = 1$ receives a private signal

$$\tilde{s}_h \;=\; \theta + \sigma_s \tilde{\varepsilon}_h, \qquad \tilde{\varepsilon}_h \sim N(0, 1) \text{ i.i.d.}, \qquad \tau_s \;=\; 1/\sigma_s^2.$$

The Gaussian-noise convention makes the Bayesian-conjugate posterior formula exact and is consistent with the rest of the model's information structure.

The sponsor receives a private signal

$$\tilde{s}_{sp} \;=\; \theta + \sigma_{sp}\tilde{\varepsilon}_{sp}, \qquad \tilde{\varepsilon}_{sp} \sim N(0,1), \qquad \tau_{sp} = 1/\sigma_{sp}^2, \tag{1}$$

observed only by the sponsor at $T = 1$, with $\tilde{\varepsilon}_{sp}$ independent of θ, $\tilde{\xi}$, and the holder shocks. The sponsor combines this signal with what it learns from the wrapper price when it sets the fund's NAV; the signal enters the model only through that valuation and the closure rule applied to it.

The dealer-APs receive no signal. They act on market observables alone, and Section 4.4 derives their behaviour as mechanical arbitrage of the wrapper price against the OTC value of the redemption basket. This is the institutional description of ETF arbitrage desks, which trade the gap between the price and the basket without a view on fundamentals.

The wrapper price P_E^1 is public; from it agents extract a public signal $y(P_E^1)$.

3.5 Timeline

$T = 0$. Investors choose between Direct (one security) and the Wrapper. Population is mass 1 with private liquidity-shock probabilities $\psi_h \sim F$ on $[0, 1]$.

$T = 1$. Shocks are realised. The systematic factor θ is partially learned. Shocked investors must liquidate. Unshocked patient wrapper holders observe $(\tilde{s}_h, P_E^1)$ and choose HOLD or SELL. Market makers post price elastic demand schedules. Noise traders trade. The n dealer-APs arbitrage the wrapper price against the OTC value of the redemption basket: each chooses its secondary-market absorption and how to dispose of it, redeeming part of the position (which retires wrapper shares and delivers the basket in kind) and warehousing the remainder to $T = 2$. Wrapper and OTC markets clear. The sponsor observes $(\tilde{s}_{sp}, P_E^1)$, sets the fund's NAV, and closes the wrapper if its continuation test (Section 4.5) is violated.

$T = 2$. Idiosyncratic shocks ε_i realise. Wrapper holders receive $\mathrm{NAV}_2 - (d_H\bar{P} + D)\cdot\mathbf{1}\{\text{severe failure at } T = 1\}$. Direct holders receive v_i.

4 Agents

4.1 Investors at $T = 0$

Two clienteles hold the wrapper. The first is a buy-and-hold institutional base of mass Ω_B: insurance companies, pension target-date funds, and index-mandated vehicles that hold the wrapper for reasons outside the marginal-type calculation, which include index mandates, regulatory accounting, and diversification under risk-management constraints. These holders neither sell in the secondary market at $T = 1$ nor redeem. Their shares enter shares outstanding and AUM, and they bear the closure loss when the wrapper fails, but they take no decision in the model: like the market makers and noise traders, they are parametric. They are not the source of fragility we study.

The second clientele is a flexible investor base of mass m, each investor endowed with one unit of capital at $T = 0$ and risk-neutral over $T = 2$ wealth.[3] In this paper the wrapper provides a liquidity service, raising the expected sale price of a shocked investor, but it carries a fire-sale exposure, the cost the patient holder bears in the severe regime.

Flexible investor h has private liquidity-shock probability $\psi_h \in [0, 1]$ drawn from continuous distribution F with positive density f. With probability ψ_h she is shocked at $T = 1$ and must liquidate; with probability $1 - \psi_h$ she consumes at $T = 2$.

Investor h chooses $a_h^0 \in \{\text{Direct}, \text{Wrapper}\}$, where Direct means one security i, to maximise expected payoff. The payoffs are specified as follows:

$$\text{Payoff}\left(a_h^0 = \text{Direct}\right) = \begin{cases} \mu + \theta + \varepsilon_i & \text{if not shocked} \\ \mu + \theta - d_H \bar{P} - D \cdot \mathbf{1}\{\text{severe}\} & \text{if shocked} \end{cases} \tag{2}$$

$$\text{Payoff}\left(a_h^0 = \text{Wrapper}\right) = \begin{cases} \text{NAV}_2 - (d_H \bar{P} + D) \cdot \mathbf{1}\{\text{severe}\} - \phi & \text{if not shocked} \\ P_E^1 - \phi & \text{if shocked} \end{cases} \tag{3}$$

The term $(d_H \bar{P} + D) \cdot \mathbf{1}\{\text{severe}\}$ in (3) is the patient-holder loss in severe failure. Severe failure here corresponds to sponsor-driven closure of the wrapper under the sponsor's continuation rule. On closure the basket is sold in the OTC market at the depressed sale price $P_{i,1}^{\text{sale}} = \mu + \theta - d_H \bar{P} - D$, and patient holders receive these proceeds pro-rata rather than NAV. Their loss decomposes into the baseline OTC discount $d_H \bar{P}$ and the additional fire-sale lump-sum D. This is the in-kind analogue of cash-redemption dilution: the patient holder bears the basket fire-sale loss because the wrapper is forced to liquidate at depressed prices, but the trigger is sponsor closure rather than cash-redemption pressure on liquid reserves.

We denote the wrapper's secondary-market price by P_E^1. The unconditional expected wrapper price is $E_0[P_E^1] = \mu - \Delta_0$, where Δ_0 is the ex-ante wrapper-vs-NAV spread characterised below in equilibrium. A shocked $T = 1$ wrapper seller receives $\mu - \Delta_0$ in expectation. She does not face an additional discrete fire-sale haircut: the closure loss D enters the patient-holder closure proceeds and the direct-holder OTC sale price, not the wrapper resale price, which is a smooth function of state variables.

Under risk-neutrality, the idiosyncratic basket noise ε_i enters expected payoffs with mean zero and contributes nothing to the wrapper-vs-Direct comparison. One more object enters the payoff of a wrapper holder who is not shocked. The not-shocked case of (3) is the payoff to holding through $T = 2$, but she is not committed to holding: she sells at $T = 1$ when her signal is bad enough. Her expected payoff therefore exceeds the always-hold expectation by the option value of the $T = 1$ decision,

$$\text{OV} \;\equiv\; E_0\big[\max\{\text{HOLD}, \text{SELL}\}\big] \;-\; E_0[\text{HOLD}] \;\geq\; 0, \tag{4}$$

[3] Risk neutrality keeps the patient holder's fire-sale loss a deterministic deduction conditional on the severe state, rather than a binary-lottery risk premium. The wrapper's two effects both enter expected payoffs directly, so neither is lost.

evaluated at the $T = 1$ equilibrium. OV does not depend on the holder's type ψ_h, because the $T = 1$ decision does not. The wrapper advantage is

$$\Delta(\psi_h) \;=\; \psi_h\big(d_H\bar{P} + D\pi_s - \Delta_0\big) \;-\; (1-\psi_h)\big[(d_H\bar{P} + D)\pi_s - \mathrm{OV}\big] \;-\; \phi. \tag{5}$$

The first term is the shocked-investor's wrapper benefit. The wrapper saves the OTC discount $d_H\bar{P}$ that a direct sale would absorb in all regimes, plus the expected fire-sale loss $D\pi_s$ that a direct seller bears in severe, less the wrapper-vs-NAV spread Δ_0 paid into the secondary market on average. Because the wrapper price is smooth across regimes (it does not jump by D in severe), the shocked seller is insulated from the discrete OTC fire-sale even when severe failure occurs. The second term is the patient-holder severe-state loss $(d_H\bar{P} + D)$ weighted by π_s, borne via sponsor closure, net of the option value of selling ahead of it. The third term is the wrapper fee, a per-share amount: ϕ is the present value of fees and participation costs over the holding period, not an annual expense ratio.

Setting $\Delta(\psi^*) = 0$ and solving the linear equation:

$$\psi^* \;=\; \frac{\phi + (d_H\bar{P} + D)\pi_s - \mathrm{OV}}{d_H\bar{P} + D\pi_s - \Delta_0 + (d_H\bar{P} + D)\pi_s - \mathrm{OV}}. \tag{6}$$

Under the sorting condition $d_H\bar{P} + D\pi_s - \Delta_0 + (d_H\bar{P} + D)\pi_s - \mathrm{OV} > 0$, Δ is strictly increasing in ψ_h and the marginal type partitions the population: $\psi_h < \psi^*$ choose Direct, $\psi_h > \psi^*$ choose Wrapper.

The sorting condition is a clientele condition, not a technical regularity. The wrapper attracts liquidity-demanders, the high-ψ types, when the displayed slope is positive: the baseline liquidity advantage and the severe-state loss terms together outweigh the option value. Closure risk reinforces the sorting but is not necessary for it. When the slope condition fails, Δ decreases in ψ, the partition reverses, and the wrapper attracts patient types who value the option to trade on their information rather than investors who need the liquidity service. Lemma A.4 in Appendix B.7 exhibits both cases.

$T = 0$ **aggregates.** Each wrapper holder holds one share. Shares outstanding at the start of $T = 1$ are the institutional base plus the flexible wrapper share,

$$\Omega_W \;=\; \Omega_B \;+\; m\big(1 - F(\psi^*)\big). \tag{7}$$

Unshocked patient wrapper-holder mass and mechanical wrapper-selling mass are

$$\kappa \;=\; m\int_{\psi^*}^{1}(1-\psi)\,dF(\psi), \qquad \lambda \;=\; m\int_{\psi^*}^{1}\psi\,dF(\psi), \tag{8}$$

respectively, with $\Omega_W = \Omega_B + \kappa + \lambda$, and direct-holder shocked OTC selling is $\lambda_b = m\int_0^{\psi^*}\psi\,dF(\psi)$. Because the institutional base neither sells nor redeems, the shares available for AP absorption and redemption come from the flexible clientele and the noise traders; noise selling beyond the flexible float is supplied through securities lending out of custody accounts, which the model leaves implicit, and the float-capped variant noted with the share-accounting constraint bounds absorption by the float and changes no conclusion. We impose no restriction on the value of the institutional base relative to the floor: under the NAV-based rule closure is not the exhaustion of the price-sensitive clientele, because the fund can be revalued down to its floor with the flexible float intact, and redemptions and valuation reach the same boundary. When the institutional base valued at par exceeds the floor, $\mu\,\Omega_B > \bar{A}$, redeeming the entire flexible float does not reach the floor on its own, and closure requires a fall in the sponsor's valuation.

4.2 Patient wrapper holders at $T = 1$

Patient holders (mass κ) observe $(\tilde{s}_h, P_E^1)$ and choose $a_h^1 \in \{\text{HOLD}, \text{SELL}\}$. We adopt risk-neutrality at $T = 1$ over the binary action (Morris and Shin, 1998; Goldstein and Pauzner, 2005; Goldstein et al., 2011): with binary action and discrete failure event, strategic complementarity is delivered cleanly through Pr(severe) without a posterior-variance correction. We specify the following:

$$\text{Payoff} = \begin{cases} \mu + \theta - (d_H \bar{P} + D) \cdot \mathbf{1}\{\mathcal{C}\} & \text{if } a_h^1 = \text{HOLD} \\ P_E^1 & \text{if } a_h^1 = \text{SELL} \end{cases} \tag{9}$$

where $\mathcal{C}$ is the sponsor-closure event, the event that the sponsor's valuation of the fundamental is low relative to accumulated redemptions, computed once the AP's redemption decision has been derived; closure triggers the basket fire-sale at the OTC dealer. In severe failure the wrapper is liquidated and the basket is sold at the OTC sale price $P_{i,1}^{\text{sale}} = \mu + \theta - d_H \bar{P} - D$, so patient holders receive $\mu + \theta - d_H \bar{P} - D$ as their effective consumption. Their loss relative to the $T = 2$ NAV they would have received is the baseline OTC discount $d_H \bar{P}$ plus the fire-sale lump-sum D.

The holder observes the realised price, so her decision can condition on it. Let $y(P_E^1)$ denote the public signal extracted from the price (Section 6.2 verifies that the equilibrium price is invertible in y), and define the holder's posterior mean of θ,

$$z(\tilde{s}_h, y) \;=\; \frac{\tau_s \tilde{s}_h + \tau_p y}{\hat{\tau}}, \qquad \hat{\tau} = \tau_\theta + \tau_s + \tau_p, \qquad k \;\equiv\; \tau_p / \tau_s,$$

where τ_p is the precision of y and k is the Bayesian weight on the public signal relative to the private signal. Conditional on $(\tilde{s}_h, y)$, the posterior of θ is $N(z, 1/\hat{\tau})$. Because the SELL payoff is the realised price itself while the HOLD payoff rises with the holder's posterior, the holder's best response is a threshold in her private signal at every realised price, and the threshold generally varies with the price (we verify this monotonicity below, once the severe event is specified). A strategy is therefore a price-contingent cutoff function $s^*(\cdot)$:

$$a_h^1 = \text{SELL} \quad \Longleftrightarrow \quad \tilde{s}_h \;\leq\; s^*(y). \tag{10}$$

Each patient holder i draws a private signal $\tilde{s}_h^i = \theta + \sigma_s \tilde{\varepsilon}_h^i$ with $\tilde{\varepsilon}_h^i \sim N(0,1)$ i.i.d. across holders. By (10), holder i sells iff

$$\theta + \sigma_s \tilde{\varepsilon}_h^i \;\leq\; s^*(y) \quad \Longleftrightarrow \quad \tilde{\varepsilon}_h^i \;\leq\; \frac{s^*(y) - \theta}{\sigma_s}.$$

Conditional on (θ, y), the fraction of patient holders who sell is therefore

$$\mathcal{A}(\theta, y) \;\equiv\; \Phi\left(\frac{s^*(y) - \theta}{\sigma_s}\right), \tag{11}$$

by the standard normal CDF and the law of large numbers across the continuum of holders. The strategic-selling mass that hits the wrapper secondary market at $T = 1$ is $\kappa \mathcal{A}(\theta, y)$. $\mathcal{A}$ is decreasing in θ (better fundamentals lower the fraction selling), with slope $-\Phi'(\cdot)/\sigma_s$ per unit of θ, scaling inversely with private-signal noise. Its response to the public signal is proportional to the equilibrium sensitivity of the cutoff function $s^*(\cdot)$, with the same factor $\Phi'(\cdot)/\sigma_s$.

The institutional channel through which patient holders bear the total severe-state loss $d_H \bar{P} + D$ in an in-kind ETF is sponsor-driven closure of the wrapper. Closure occurs when the sponsor's continuation test fails; to first order this is AUM, shares outstanding valued at the sponsor's NAV, falling below the floor $\bar{A}$, and Section 4.5 derives the exact form. A holder's sale on the secondary market moves her shares to the market maker's book and leaves shares outstanding unchanged; shares are retired only when the dealer-AP redeems. Strategic selling therefore drives closure through the price: heavier selling clears at a wider wrapper discount, a wider discount raises the AP's redemption flow, and redemptions deplete shares outstanding and raise the level of the sponsor's valuation at which closure occurs. The closure event $\mathcal{C}$ is stated formally in Section 5.2, after the AP's redemption decision is derived; there we also derive the holder's conditional closure probability $\Pr(\mathcal{C} \mid \tilde{s}_h, y)$, which uses the private signal and the public signal separately, and verify that cutoff strategies are the right class: the individual HOLD-minus-SELL difference is strictly increasing in the private signal, with no side condition.

On closure, the sponsor's mandate is to liquidate the residual basket, the holdings backing the remaining shares outstanding, at once, not in small quantities, so the closure-induced wholesale sale exhausts dealer capacity, and we model the fire-sale loss D as triggered directly by the closure event $\mathcal{C}$. The indicator $\mathbf{1}\{\mathcal{C}\}$ in (9) therefore captures the joint event of closure plus fire-sale-priced basket sale.

4.3 Market makers

Market makers (MM) have no private signal of the fundamental and act as the sole intermediary on the wrapper secondary market: they absorb all wrapper sales (mechanical and strategic) and route AP arbitrage flow against their own book. Their quoted demand schedule is

$$D_m(P_E^1) \;=\; \kappa\, \Phi\Big(-\alpha(P_E^1 - \mu) + \sigma_\xi \tilde{\xi} + W\Big) \;+\; \lambda, \tag{12}$$

where $\alpha > 0$ measures the market depth provided by the dealer/MM and hence the price elasticity, $\tilde{\xi} \sim N(0,1)$ is the noise-demand shock, $\sigma_\xi \tilde{\xi}$ is the noise-demand component, and W is the aggregate AP order flow received by the MM (derived in Section 4.4). In practice, the largest bond-ETF dealers run market-making and AP desks inside the same firm (Citadel Securities, Jane Street, Flow Traders, and the bank-affiliated dealers all have both desks): the MM and AP share an internal book and the AP's intended creation/redemption quantity is observable to the MM in real time. The functional form of MM's demand mirrors the supply side: the $\kappa\Phi(\cdot)$ piece matches the strategic-selling mass $\kappa\Phi((s^*(y)-\theta)/\sigma_s)$, and the additive λ matches the mechanical-selling baseline. Because MM demand does not depend on θ, market makers do not contribute fundamentals information to the price, and neither does the aggregate AP flow W, which is a function of market observables; the sellers' private signals enter the price through the strategic-selling mass on the supply side. The two Φ terms are identical, so they cancel exactly when the market clears, with no Taylor approximation. The state then enters the clearing condition only through the single Gaussian statistic y, and the equilibrium price is strictly increasing in it. Because y is Gaussian, the posteriors of the holders and the sponsor are closed-form.

The $\kappa\Phi(\cdot)$ piece is a reduced-form modelling choice that delivers this tractability. The economic content behind it is the assumption that the MM understands the structure of strategic holder selling well enough to quote against it, plausible because liquid wrappers (e.g., ETFs) MMs are the same firms that intermediate redemptions and observe order flow continuously. With a standard linear quote $D_m = -\alpha(P_E^1 - \mu) + \sigma_\xi \tilde{\xi}$ the Φ on the supply side would not cancel, clearing would be nonlinear in price, and a first-order linearisation around the cutoff would be needed to solve the model. The mechanical implication is also weaker risk-sharing in stress: each unit of additional redemption pushes the price down by a constant slope $1/\alpha$ rather than being absorbed by the matching Φ-piece, so price crashes are sharper and the closure-fear channel correspondingly amplified. We interpret $\kappa\Phi(\cdot)$ as a proxy for dealer balance sheets, internalisation, and dynamic inventory management, which lets us capture these stabilising forces without modelling them explicitly.

Additionally, one might be concerned about an asymmetry in the reduced form of the market maker's quoted demand (12). The AP's disposal problem, specified in the next subsection, prices closure risk: the AP pays a closure charge on the shares it redeems and on those it warehouses. The market maker's quoted demand carries no such charge, although shares on the MM's book bear the closure loss like any others. The asymmetry is the price of the schedule convention: the MM quotes a demand curve rather than solving a disposal problem, and a closure-sensitive MM would contribute a destabilising term parallel to the trading book's. The statements the paper makes about market-maker depth are therefore statements about this reduced form, and we maintain it as one: the matching quote is a modelling choice, not a microfounded market maker, and the variants below are computational robustness rather than alternative microfoundations. We demonstrate the robustness of the model to closure-sensitive and linear variants of the market maker's schedule[4] in Appendix C.

[4] Under a closure-sensitive quote, one that carries the same closure charge as the AP's warehouse, the destabilising

4.4 Dealer-Authorised Participants

There are n dealer-APs, indexed by j. Each absorbs wrapper shares in the secondary market and disposes of them in one of two ways: redeeming them in kind through the primary market, which retires shares outstanding, or warehousing them to $T = 2$. We write Q for the AP sector's aggregate redemption, w for its aggregate warehousing, and $W = Q + w$ for its aggregate absorption. Each AP has trading-book capacity χ/n and warehouse capacity χ_w/n, so that the AP sector's quadratic position costs aggregate to $W^2/(2\chi) + w^2/(2\chi_w)$ at symmetric play and aggregate capacity is held fixed as n varies. Two objects used in this subsection are constructed afterwards, once the AP's behaviour is in hand: the public closure probability $\Pi_c(y, Q)$ with its boundary $\theta_c(Q)$ (Section 5.2), and the OTC dealer bid (Section 5.3). At this point the reader needs only that Π_c is a smooth function of the state and of aggregate redemptions, increasing in Q. The APs receive no signal: each acts on the market observables alone, arbitraging the wrapper price against the OTC value of the redemption basket. This mechanical description has an informational consequence. The equilibrium price is a strictly increasing function of a single Gaussian statistic of the state,

$$p \;\equiv\; P_E^1 - \mu \;=\; \mathscr{P}(y), \qquad y \;=\; \theta + \sigma_s\sigma_\xi\,\tilde{\xi}, \qquad \tau_p \;=\; \frac{1}{(\sigma_s\sigma_\xi)^2}, \tag{13}$$

where the price map $\mathscr{P}(\cdot)$ is strictly increasing but not linear in general (verified in Section 6.2), and the noise coefficient requires no fixed point: a demand schedule that responds only to the price responds to θ and $\tilde{\xi}$ in the proportion the price already carries, so no behaviour of the AP sector, however nonlinear, can tilt the ratio. The AP sector is informationally neutral. It adds no noise to the price and extracts nothing private from it, and the public-signal precision τ_p is a primitive, common to the holders and the sponsor alike. Observing the price is equivalent to observing y, and the public estimate of the fundamental is

$$E[\theta \mid y] \;=\; \rho_y\, y, \qquad \rho_y \;=\; \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_s^2\sigma_\xi^2}. \tag{14}$$

Redeemed baskets are unwound in the OTC dealer market. Dealers quote the basket off the same public valuation everyone sees: the per-share bid for redemption baskets is

$$B(y, Q) \;=\; \mu + \rho_y y \;-\; d_o \;-\; \lambda_o Q, \tag{15}$$

a constant dealer margin d_o on an institutional block, plus a linear impact $\lambda_o \geq 0$ of the aggregate unwind flow Q on dealer balance sheets. The assumption embedded in B is that dealer quotes carry no independent information about θ, which is the institutional reading of matrix pricing: dealers price the basket from traded prices. The AP's access advantage over households is captured by the

term is present and small: the expected discount widens from 0.008 to 0.011 and the severe-failure probability rises from 0.037 to 0.038, while equilibrium uniqueness, the monotone price map, and every comparative-statics sign are unchanged. Under a plain linear quote the price no longer transmits the sellers' private information, and the economy behaves as it does when the price is uninformative.

level of this schedule: $d_o < d_H\bar{P}$, the dealer margin on a block against the outsider's discount on a direct sale. The advantage is capacity-limited, not absolute: the marginal unwind cost $d_o + \lambda_o Q$ reaches the household discount at the flow $\bar{Q} = (d_H\bar{P} - d_o)/\lambda_o$, so in deep stress dealer capacity erodes the intermediation advantage. The wrapper's liquidity transformation runs through an access advantage that is itself a scarce dealer resource.

Each AP runs an intra-period arbitrage book between the wrapper and the basket, and it makes two choices: how many wrapper shares to absorb from the secondary market, and how to dispose of them. An absorbed share can be redeemed now, delivering the share to the sponsor, taking out the basket in kind, and selling it at the OTC bid (15); redemption retires the share and is the only channel through which shares outstanding fall. Or it can be warehoused to $T = 2$ to collect NAV, paying the warehouse charge: term funding, haircuts, and capital charges of keeping inventory on the balance sheet across periods. Both disposal modes carry closure exposure. A redeemed basket caught in a closure is unwound in the same fire sale as the residual basket and bears the additional discount D; a warehoused share caught in a closure receives the liquidation proceeds and loses $L = d_H\bar{P} + D$, the outsider's fire-sale loss, the worst position of all. Writing $\Pi_c(y, Q)$ for the public closure probability, derived in Section 5.2, and

$$\Delta_A \;\equiv\; \mu + \rho_y y - P_E^1 \;=\; \rho_y y - p \tag{16}$$

for the public wedge, AP j chooses redemption $q_j \geq 0$ and warehousing $w_j \geq 0$, with $W_j = q_j + w_j$, to solve

$$\max_{q_j,\, w_j \geq 0} \; q_j\left[\Delta_A - d_o - \lambda_o Q - D\,\Pi_c(y, Q)\right] + w_j\left[\Delta_A - L\,\Pi_c(y, Q)\right] - \frac{n\,w_j^2}{2\chi_w} - \frac{n\,W_j^2}{2\chi}, \tag{17}$$

taking the others' redemptions in $Q = \sum_k q_k$ as given, and subject to the share-accounting constraint $\sum_k (q_k + w_k) \leq \Omega_W$: redeemed and warehoused shares are shares, and the AP sector cannot dispose of more than are outstanding. The constraint does not bind in equilibrium, and stating it makes Q and w bounded by Ω_W, which the regularity conditions use.[5] Each AP internalises its own flow's effect on the dealer bid and on the closure probability. At the symmetric equilibrium the aggregate first-order conditions are

$$w: \quad \Delta_A - L\,\Pi_c \;-\; \frac{w}{\chi_w} \;-\; \frac{W}{\chi} \;=\; 0, \qquad \text{so} \qquad W \;=\; \chi(\Delta_A - L\,\Pi_c) - \frac{\chi}{\chi_w}\,w, \tag{18}$$

[5] Our numerical analysis verifies that no state attains the bound. Two tighter versions of the constraint deserve note. That the AP sector dispose only of shares it has absorbed holds by construction, since $q_j + w_j = W_j$ splits the position AP j takes in clearing between the redemption and the warehouse. A float version, requiring every redeemed share to be sourced from the flexible float rather than from the institutional base, holds except in deep-crash states, where noise selling, which the model does not tie to the initial share allocation, supplies the AP sector with shares beyond the float. Imposing it in those states lowers the severe-failure probability and leaves the discount and the qualitative results unchanged, so the maintained specification is the conservative one for closure risk.

$$q: \quad \Delta_A - d_o - \lambda_o Q - D\,\Pi_c \underbrace{-\ \frac{Q}{n}(\lambda_o + D\,\Pi_c') \ - \ \frac{w}{n}\,L\,\Pi_c'}_{\text{internalisation}} - \frac{W}{\chi} \ = \ 0, \tag{19}$$

with $\Pi_c' = \partial\Pi_c/\partial Q$ and $W = Q + w$. Given (y, p), the system is a scalar fixed point in Q, smooth, with the corners $Q \geq 0$ and $w \geq 0$ as its only kinks. The premium region, with creations replacing redemptions, is analogous but not symmetric, because the closure charges are one-sided; we focus on the discount side, where the fragility of interest operates.

The internalisation terms in (19) are the economics of a common pool, and they are where the number of APs matters. Each AP's redemption widens the OTC discount all APs receive and moves the fund toward the sponsor's floor, damaging every exposed book; each AP internalises only its own share of the damage, which is a $1/n$ fraction of the aggregate. Fewer, larger APs therefore redeem less against the same state: they do not compete the dealer margin away against themselves, and they do not race the fund's buffer down. Primary-market activity is in fact concentrated in a small number of APs.

The two capacity parameters divide the AP's balance sheet by holding horizon. The trading-book capacity χ governs the intra-period charge on the gross position and the price elasticity of AP demand on the wrapper secondary market, hence the AP sector's absorption capacity in clearing. The warehouse capacity χ_w governs the disposal split: how much absorbed flow is stored rather than redeemed, and so the law of motion of shares outstanding. In the data the warehouse is the dominant disposal mode at daily frequency (most bond-ETF trading days see no primary-market activity), with redemptions concentrated in episodes of wide discounts.

Because the APs are mechanical arbitrageurs between the wrapper price and the OTC price of the basket's constituents, their demand tracks the public wedge: heavier selling by holders pushes the price down, the wedge $\Delta_A = \rho_y y - p$ widens, and the AP sector absorbs and redeems more. There is no inference by the AP to manipulate.

4.5 The sponsor

The sponsor is the fund company that operates the wrapper: it lists the fund, licenses the index, administers creations and redemptions, computes and publishes the NAV, and decides whether the fund continues to operate. Its revenue is the fee ϕ (Section 4.1) accruing on AUM.[6] Its cost of operating the wrapper, listing, index licensing, administration, custody, and audit, is largely fixed: a fund-level operating expense in the hundreds of thousands of dollars per year, small next to the assets involved. A fund charging 20 basis points needs roughly \$25 million under management for each \$50,000 of annual cost. We write c_f for this fixed cost.

At $T = 1$, after the wrapper and OTC markets clear, the sponsor observes the wrapper price and its private signal about the fundamental, $\tilde{s}_{sp} = \theta + \sigma_{sp}\tilde{\varepsilon}_{sp}$ from (1), with noise $\tilde{\varepsilon}_{sp} \sim N(0, 1)$

[6] At the par normalisation $\bar{P} = \mu = 1$, the per-share fee of Section 4.1 and the rate on AUM coincide. The fee is a transfer with two sides: the investors' payoffs charge the constant per-share ϕ, ignoring the fee times the valuation deviation, a product of two small quantities, while the sponsor's continuation value in (21) keeps the valuation term, because there it is the object of interest, not a rounding.

independent of all other shocks and precision $\tau_{sp} = 1/\sigma_{sp}^2$. Its information set is $(\tilde{s}_{sp}, P_E^1)$. The private signal stands for what the fund company sees and the market does not, redemption-desk traffic, custodial flows, and the pricing-service inputs behind its NAV computation. The fund's published NAV aggregates the valuations supplied by pricing services and the sponsor's own inputs; we represent it as the sponsor's posterior valuation of the basket. Its posterior mean on the fundamental combines the price with its private signal,

$$m_{sp} \;=\; E[\theta \mid y, \tilde{s}_{sp}] \;=\; b_y\, y + b_s\, \tilde{s}_{sp}, \qquad b_y = \frac{\tau_p}{\tau_\theta + \tau_p + \tau_{sp}}, \quad b_s = \frac{\tau_{sp}}{\tau_\theta + \tau_p + \tau_{sp}}, \tag{20}$$

and the NAV is $\mu + m_{sp}$. AUM is shares outstanding times this valuation, with the law of motion of shares outstanding derived in Section 5.2.

Two valuation objects should be kept distinct, because the price enters each differently. The published NAV is the posterior valuation $\mu + m_{sp}$: its dependence on the price is the learning weight b_y, and its institutional counterpart is the valuations that pricing services supply for the NAV computation, which for an illiquid basket are matrix prices built from traded prices. The continuation decision below depends on the franchise value, which is proportional to the same posterior at franchise scale and inherits the same two components. In both objects the price-based part is $b_y\, y$, and the discipline is $b_s\, \tilde{s}_{sp}$, the component no market observable transmits; Section 6.3 studies the configuration in which this private component vanishes and the valuation becomes computable from the price.

The sponsor then decides whether to continue operating the fund. The fund is a transferable franchise: sponsors adopt funds from one another, so the value of continuing is the value of the future fee stream the franchise generates, whether the sponsor operates it or sells it. We take this franchise value to be proportional to the sponsor's assessment of the quality of the underlying product at the scale of the franchise, $\phi\, \Omega_W\, (\mu + m_{sp})$: a better fundamental means a more valuable basket, better prospects for the fee base, and a higher price a buyer would pay for the fund. Continuing also carries the cost of operating, which rises with the redemptions the fund has absorbed, $c_f + c_Q\, Q$: the fixed overhead c_f, plus a per-share redemption cost c_Q that stands for the administrative burden of in-kind redemption, basket delivery, custody transfers, and rebalancing, and for the franchise damage of visible outflows. Closing, or failing to find a buyer, yields zero: the residual basket is sold and the proceeds are distributed pro-rata to holders, on whose assets the sponsor has no claim. The sponsor therefore operates while

$$\phi\, \Omega_W\, (\mu + m_{sp}) \;\geq\; c_f \;+\; c_Q\, Q, \tag{21}$$

and closes otherwise: it closes when its estimate of the fundamental is low relative to the redemptions the fund has absorbed. The fixed component of the cost defines the floor,

$$\bar{A} \;=\; \frac{c_f}{\phi}, \tag{22}$$

the level of assets at which the fee covers the overhead alone; at typical fee rates this delivers the \$25M–\$100M range of observed sponsor thresholds. The test also explains why a sponsor cannot credibly promise to operate through adverse shocks: below the floor, such a promise is a commitment to run losses without bound. We maintain the calibration $c_Q = c_f/\Omega_W$, under which a redeemed share takes its overhead allocation with it, and we treat $\bar{A}$ as a primitive in the analysis, with (22) as its content; Appendix D documents the floor and the closure process empirically.[7]

The closure rule of the model is a static reduced form of the process the appendix documents. In practice sponsors close after assets persist below the floor, sometimes waiving fees, charging less than the stated fee for a period, to keep a marginal fund alive, and a wind-down can stretch over months: the median closed fund spent sixteen consecutive months below an illustrative break-even before announcing, and four of five funds that fell below it recovered without closing (Appendix D). The model collapses this into a one-shot trigger and a lump-sum loss because the coordination economics uses only two features of the institution: closure becomes more likely as valued AUM approaches the floor, and closure imposes a discrete loss on the holders who remain. Persistence, waivers, and stretched liquidation move the level of the floor and the size of D; they remove neither feature. The sponsor's option to wait would smooth the trigger ex post, but in holders' beliefs the trigger is already smooth, because the closure probability is continuous in the state, and it is this smoothed object that the run economics uses. The omitted dynamics point in a direction the static model already prices: a discount that persists functions like an accumulation of public information, strengthening the coordination channel, a force the comparative statics in price informativeness describe in reduced form and a formal treatment of which needs a dynamic signal structure the static model does not carry; a persistence requirement thus lets the public signal sharpen while the fund waits at the floor. This is the risk the comparative statics in price informativeness describe. A dynamic closure rule with a persistence requirement would therefore alter the timing of the mechanism, not its structure: the run economics use the smoothed closure probability, and persistence smooths it further while the informativeness channel sharpens it.

The closure rule that the rest of the paper applies, close when the sponsor's valuation is low relative to accumulated redemptions, is therefore a decision rule, the continuation test (21) evaluated at the sponsor's posterior, and it inherits the sponsor's information. Because the posterior (20) depends on the price, coordinated selling can move the valuation that decides closure. Because it depends on the private signal, part of the closure decision is unlearnable from market observables, and Section 6.3 shows that this residual is what disciplines coordination. The sponsor takes one further decision, the choice of the wrapper's architecture at launch, which Section 9 summarises and Online Appendix OA.1 analyses using the same objects: the fee revenue, the operating cost, and the closure risk of each form.

[7] The linear form of value in the posterior and of cost in redemptions is the reduced-form choice: any smooth continuation value increasing in the one and decreasing in the other delivers a boundary of this shape locally. Only the ratio $c_Q/(\phi\,\Omega_W)$ enters the closure boundary, so the calibration fixes one number; Section 5.2 shows that it makes the continuation test coincide with the observed AUM rule to first order in redemptions. The test governs an operating fund: a fund whose base has been entirely redeemed has nothing to operate and closes trivially.

5 Markets and clearing

5.1 Wrapper secondary-market clearing

Total wrapper sales at $T = 1$ are mechanical λ plus strategic $\kappa\mathcal{A}(\theta, y)$, with $\mathcal{A}(\theta, y)$ from (11). The MM is the sole counterparty on the buy side (Section 4.3); the aggregate AP order flow W enters through the MM's quote in (12) rather than as a separate buyer. Setting $D_m(P_E^1)$ equal to total selling, the wrapper-clearing condition in the wrapper-discount region is

$$\kappa\,\Phi\Big(-\alpha(P_E^1 - \mu) + \sigma_\xi\tilde{\xi} + W\Big) + \lambda \;=\; \lambda + \kappa\,\Phi\Big(\frac{s^*(y) - \theta}{\sigma_s}\Big). \tag{23}$$

The λ's on the two sides cancel, and the $\kappa\Phi(\cdot)$ structures are identical. Equating the Φ arguments gives the exact clearing equation

$$-\alpha(P_E^1 - \mu) + \sigma_\xi\tilde{\xi} + W \;=\; \frac{s^*(y) - \theta}{\sigma_s}, \tag{24}$$

with no Taylor approximation. The aggregate AP demand W is the solution of the AP sector's first-order conditions (18)–(19) at the observed price, a function of (y, p) alone. This is what makes the clearing equation aggregate so cleanly: every term other than θ and $\tilde{\xi}$ is measurable with respect to the price, so clearing writes the price as a function, strictly increasing under the monotonicity verified in Proposition 1, of the single statistic $y = \theta + \sigma_s\sigma_\xi\tilde{\xi}$ in (13), and the cutoff term $s^*(y)/\sigma_s$ determines the level of the price map rather than entering the extraction of y from the price. No signal-coefficient fixed point arises: the mechanical AP sector contributes no noise of its own to the price.

5.2 Shares outstanding and sponsor closure

We can now state the law of motion of shares outstanding, the sponsor's valuation, and the closure event. Shares outstanding start at $\Omega_W = \Omega_B + \kappa + \lambda$ and fall only through aggregate AP redemptions Q from (19):

$$\Omega_W^1 \;=\; \Omega_W \;-\; Q. \tag{25}$$

Secondary-market sales, mechanical or strategic, do not appear in (25): they are absorbed by the market maker and the AP sector, and they retire shares only to the extent that the APs redeem.

The sponsor values the fund by (20) and applies the continuation test (21) of Section 4.5: it closes when its posterior on the fundamental is low relative to accumulated redemptions. Dividing (21) by $\phi\,\Omega_W$ and using $\bar{A} = c_f/\phi$ with the maintained calibration $c_Q = c_f/\Omega_W$, the closure event is a threshold event in the sponsor's posterior,

$$\mathcal{C} \;=\; \Big\{ m_{sp} \;<\; \theta_c(Q) \Big\}, \qquad \theta_c(Q) \;=\; \Big(\frac{\bar{A}}{\Omega_W} - \mu\Big) \;+\; \kappa_Q\,Q, \qquad \kappa_Q \;=\; \frac{\bar{A}}{\Omega_W^2}, \tag{26}$$

where we normalise $\mu = 1$ in the boundary. The boundary is affine in redemptions because the continuation test is: revenue is linear in the posterior and cost is linear in redemptions.[8] The slope has a simple reading: $\kappa_Q = c_Q/(\phi\,\Omega_W)$ is the cost a redemption adds relative to the fee a share of the base earns, and at the maintained calibration it equals $(\bar{A}/\Omega_W)\cdot(1/\Omega_W)$, the per-share floor times the fraction of shares outstanding that one unit of redemption retires. The boundary (26) displays the two margins through which the fund reaches the floor. Redemptions raise $\theta_c(Q)$: the coordination-sensitive margin. A low sponsor posterior reaches the boundary with no redemptions: the valuation margin, which no AP decision touches.

The closure probabilities that enter behaviour follow from (20) and (26). By iterated expectations, $E[m_{sp} \mid y] = \rho_y y$: the public expectation of the sponsor's valuation is the price-based estimate, regardless of the sponsor's precision. For the APs, whose information is the price, the residual uncertainty is the sponsor-signal innovation, Gaussian with standard deviation $\sigma_g = b_s\sqrt{\mathrm{Var}(\theta \mid y) + \sigma_{sp}^2}$, so the public closure probability is

$$\Pi_c(y, Q) \;=\; \Phi\!\left(\frac{\theta_c(Q) - \rho_y\, y}{\sigma_g}\right), \tag{27}$$

smooth and increasing in Q with slope $\Pi_c' = \phi(\cdot)\,\kappa_Q/\sigma_g$. This is the object in the AP first-order conditions (18)–(19). It is a posterior probability of the sponsor's valuation outcome, not a randomisation: closure is uncertain at the moment agents act because the valuation embeds a signal they do not observe.

A patient holder knows more than the price: its private signal informs θ, and θ informs the sponsor's signal. The holder's conditional closure probability therefore depends on the private posterior z,

$$\Pr\left(\mathcal{C} \mid \tilde{s}_h, y\right) \;=\; \Phi\!\left(\frac{\theta_c(Q(y)) - b_y\, y - b_s\, z(\tilde{s}_h, y)}{\sigma_{g,h}}\right), \qquad \sigma_{g,h} \;=\; b_s\sqrt{1/\hat{\tau} + \sigma_{sp}^2}, \tag{28}$$

where $Q(y)$ is the equilibrium redemption at the observed price. It depends on the public signal through the boundary, the valuation, and the posterior, and on the private signal through z alone. A holder with a better signal assigns a higher fundamental, hence a higher sponsor valuation, hence a lower closure probability: the closure-fear term reinforces the monotonicity of the holder's problem in its signal. Strategic selling drives closure through the price level: a higher aggregate cutoff lowers $\mathscr{P}(y)$ at every y, widens the wedge, and raises equilibrium redemptions $Q(y)$, which raise the boundary $\theta_c(Q(y))$ in (28). The sponsor's valuation itself does not move: the sponsor recovers y from the price whatever cutoff the holders use, so its posterior is invariant to the shift.

[8] The observed institutional form of the rule is the AUM test: close when the assets the fund holds, $(\Omega_W - Q)(\mu + m_{sp})$, fall below the floor $\bar{A}$. Expanding that test to first order in Q around $Q = 0$ gives (26) exactly, so the continuation test and the AUM test coincide where redemptions are small, which is the modal state, since most bond-ETF trading days see no primary-market activity; they separate as redemptions accumulate, the AUM test's threshold rising faster. Solving the model under the exact AUM test raises the severe-failure probability by about an eighth and leaves the equilibrium properties and every comparative-statics sign unchanged, with the clearing root unique at every state: the results do not depend on which form of the rule is used, and the maintained form is the conservative one for closure risk.

This is the strategic complementarity of the $T = 1$ game, and it operates through the APs: selling widens the wedge the APs redeem against, redemptions raise the valuation threshold at which the sponsor closes, and the higher closure probability is what each holder fears.

Cutoff strategies remain the right class: at any realised y, the individual HOLD-minus-SELL difference is strictly increasing in the private signal,

$$\frac{\partial}{\partial \tilde{s}_h}\Big[\, z - (d_H \bar{P} + D)\ \Pr(\mathcal{C} \mid \tilde{s}_h, y) - (P_E^1 - \mu)\Big] \ = \ \frac{\tau_s}{\hat{\tau}}\Big[1 + (d_H \bar{P} + D)\,\frac{\phi(\cdot)\, b_s}{\sigma_{g,h}}\Big] \ > \ 0, \qquad (29)$$

with no side condition: under the NAV-based rule the closure-fear term works with the signal, not against it, and monotone best responses are automatic.

5.3 OTC market and the fire-sale link

Three components of basket-selling flow reach the OTC market at $T = 1$: the unwound redemption baskets Q from (19) (the short leg of the APs' redeemed positions, scaling up as the wrapper-vs-NAV spread widens); direct-holder shocked sales λ_b; and the sponsor's closure flow $\mathcal{L} = \Omega_W^1$, the basket backing the remaining shares outstanding, which hits the market when the closure condition (26) is met. The closure flow is a forced sale of the residual basket at once, so it dominates the discretionary flows: the redemption baskets Q are redeemed share by share and unwound at the dealer bid (15), while closure sells the entire remaining basket Ω_W^1 at once. The APs price their exposure to this event on both books: the expected fire-sale charge $D\,\Pi_c$ on redeemed baskets and $L\,\Pi_c$ on warehoused shares appear directly in the first-order conditions (18)–(19). The sequencing matters. Redemptions deplete shares outstanding gradually as the wedge widens; the fire-sale event occurs at the moment the sponsor closes, when the depletion crosses the floor and the residual basket hits the market wholesale. The redeemed book bears the charge $D\,\Pi_c$ in (19) because the in-kind unwind settles inside the window in which closure strikes: the AP is exposed through settlement, and the dealer's bid on a basket unwound in the closure window carries the fire-sale price, of which the charge is the reduced form.

The model collapses the wind-down into the $T = 1$ closure event; empirically the liquidation is spread over weeks, with a median of 54 days between announcement and delisting (Appendix D). The collapse preserves the key reduced-form object, because a longer wind-down reallocates the liquidation cost rather than avoiding it; what the collapse abstracts from is timing, of who bears the cost within the window and of what arrives as information while it runs. Suppose the sponsor sells the residual basket over a window of length k. The per-share price concession is convex in the selling rate, $\delta(B/k)$ with δ increasing and $\delta(B) = D$ the wholesale concession at the capacity breach; and the window suspends the wrapper's liquidity service, since the fund is closed to creation and redemption from the announcement, so a holder shocked during the window faces the unanchored sale or the wait, an exposure cost that accumulates with the window's length, εk per residual share. Any liquidation policy therefore bears $c(k) = \delta(B/k) + \varepsilon k$, whose minimum over k is strictly positive whenever the underlying is illiquid: stretching the sale substitutes time for price and cannot make the residual basket liquid. The lump-sum D is this minimised cost in reduced form. The acute severe event of the model is the short-window corner, forced when a redemption wave carries assets through the floor inside the stress window; the extended liquidations of Appendix D, FM's seven months, are interior choices in which the cost is taken as delay and exposure rather than price. A long wind-down preserves the same economic cost in reduced form, though it changes its timing and incidence, and the cost disappears only in the limit of a liquid underlying, where the model predicts no severe regime to begin with. The severe-failure event is therefore a closure-driven fire-sale, on the redemption margin preceded by a redemption wave, while on the valuation margin closure arrives with the primary market silent; the redemption-margin pattern is consistent with the bond-ETF closure pattern documented in 2020–2022, in which closures followed sustained primary-market outflows.

The OTC dealer absorbs the discretionary flows at a baseline discount, but is overwhelmed when the basket is sold wholesale under sponsor closure. We model this by tying the fire-sale loss directly to the closure event: the dealer's quoted sale price is

$$P_{i,1}^{\text{sale}} \;=\; \mu + \theta \;-\; d_H \bar{P} \;-\; D \cdot \mathbf{1}\{\mathcal{C}\}, \tag{30}$$

$$\text{basket fire-sale loss} \;=\; D \cdot \mathbf{1}\{\mathcal{C}\} \;=\; \begin{cases} 0 & \text{if not } \mathcal{C}, \\ D & \text{if } \mathcal{C}. \end{cases} \tag{31}$$

Throughout, $d_H \in (0, 1)$ is the baseline OTC discount expressed as a fraction of $\bar{P}$, $D \geq 0$ is the fire-sale loss in dollar units, and the absolute-dollar baseline discount is $d_H \bar{P}$. Treating D as a fixed lump-sum penalty rather than as a continuous function of excess flow keeps D statistically a constant once severe failure occurs, so $E[D \cdot \mathbf{1}\{\text{severe}\}] = D \cdot \pi_s$ is exact. The lump-sum specification captures the discrete nature of dealer-capacity breach: dealers either absorb the discretionary flow at the baseline discount or refuse to make markets once the sponsor sells the residual basket wholesale, in which case sellers face a step-function loss.

The OTC side of the model is a reduced form, and it is worth stating plainly what it does and does not contain. Three schedules summarise it: the household seller's price (30), the AP's dealer bid (15), and the closure sale at the household price. Only the AP's unwind flow moves the dealer bid; household flow and the closure flow enter as level effects, through $d_H \bar{P}$ and D. We do not model the OTC dealer's balance sheet: the schedules stand for its capacity, D is the reduced form of the capacity breach at the wholesale sale, and the common pool that the internalisation results price is the pool of AP unwind flow.

5.4 Failure regime triggers

The model has two failure regimes, each defined by a state event in $(\theta, \tilde{\xi}, \tilde{\varepsilon}_{sp})$:

Mild trigger. The realised wrapper discount exceeds the OTC baseline discount:

$$\Delta_E(\theta, \tilde{\epsilon}) \; > \; d_H \bar{P}, \tag{32}$$

where $\Delta_E = \mu - P_E^1$ is the realised wrapper discount and $\tilde{\epsilon} = (\tilde{\xi}, \tilde{\varepsilon}_{sp})$ collects the noise shocks. AP arbitrage is no longer tight enough to hold the wrapper near NAV; sellers absorb the wrapper discount at $T = 1$, but the basket remains intact and patient holders' payoff is unaffected. Because the price is strictly increasing in the public signal, the mild trigger is a lower-threshold event in y: it holds iff $y < y_m$, where y_m solves $\mathscr{P}(y_m) = -d_H \bar{P}$.

Severe trigger. The sponsor's closure condition is met:

$$m_{sp} \; < \; \theta_c(Q), \tag{33}$$

the closure event of (26). When it is met, the fund's valued AUM has reached the floor, by redemptions, by valuation, or by both, the sponsor closes, the residual basket $\mathcal{L} = \Omega_W^1$ is sold wholesale, and the dealer-capacity breach lump-sum D is incurred. The fire sale is an event of the closure itself: what breaches dealer capacity is the volume of the wholesale sale, regardless of the wrapper's secondary-market discount at that moment.

The mild trigger classifies the wrapper-market regime, and the two triggers are measured on different information: the mild trigger on the market price, the severe trigger on the sponsor's valuation, which embeds a signal no market observable carries. Neither therefore implies the other state by state; the sponsor can close a fund whose discount is inside the band when its private signal is bad enough. Because the sponsor learns from the price, however, the two events move together, and closure at a moderate discount requires the private signal to contradict the price, a rare configuration.[9] The severe-failure event is a coordination event transmitted through the wrapper discount, the APs' redemption response, and the sponsor's valuation, and completed by sponsor closure.

Both triggers are state events, but evaluating their probabilities and the implied θ-thresholds requires the equilibrium objects: the price map $\mathscr{P}(\cdot)$, the cutoff function $s^*(\cdot)$, and the redemption function $Q(\cdot)$.

5.5 The assumptions that matter

Five modelling choices matter for the results, and each is examined at a stated place. First, the market maker's quote responds to the structure of order flow; this is what lets the price aggregate the sellers' information, and the panic channel requires it: with a quote that transmits little order-flow information the economy operates near the no-learning corner, where the panic-driven component is small. Appendix C computes two alternatives, a quote that prices closure risk and a plain linear

[9] In our computations, closure with the discount inside the baseline band has probability 0.0007 against a closure probability of 0.037: conditional on closure, the discount exceeds the band 98 percent of the time.

quote, and locates each in the model's informativeness range. Second, the APs are mechanical: their demand responds to the observed wedge and to nothing else. The results use only that AP demand is measurable with respect to the price, so that no AP parameter enters the information structure (Lemma A.1); this holds for any demand schedule that brings no independent signal, and it is the counterpart of arbitrage desks trading the observed basis. Third, the sponsor's valuation combines the price with an independent private input; Section 6.3 studies both the discipline this input provides and the configuration in which it vanishes. Fourth, closure is a one-shot test at $T = 1$, while observed practice is deliberative; Section 4.5 states the relation to the observed AUM test and the sense in which the maintained form is the conservative one. Fifth, the basket fire-sale loss D is a parameter: the model endogenises the probability of the fire-sale, not its depth, and Online Appendix OA.4 recomputes the directional results across the loss plane.

6 Equilibrium

6.1 Definition

Definition 1 (Monotone rational-expectations equilibrium)**.** A monotone rational-expectations equilibrium is a tuple

$$(\psi^*,\ k,\ s^*(\cdot),\ \mathscr{P}(\cdot),\ W(\cdot),\ w(\cdot),\ Q(\cdot),\ \kappa, \lambda, \lambda_b)$$

such that:

(E.1) *Marginal investor at $T = 0$.* $\Delta(\psi^*) = 0$ from (5), and aggregate masses $\kappa, \lambda, \lambda_b$ are the integrals over the resulting partition (8).

(E.2) *Patient holders at $T = 1$.* Holders follow the cutoff strategy (10) with the Bayesian weight $k = \tau_p/\tau_s$ in the posterior z, and the cutoff function satisfies pointwise indifference: for every realised y,

$$\mathscr{P}(y) \ = \ z(s^*(y), y) \ - \ (d_H \bar{P} + D) \cdot \Pr\left(\text{severe} \mid \tilde{s}_h = s^*(y),\, y\right). \tag{34}$$

(E.3) *APs at $T = 1$.* At every realised price, the aggregate absorption, warehouse, and redemption $(W(y), w(y), Q(y))$ constitute the symmetric equilibrium of the n-AP game (17): the first-order conditions (18)–(19) hold with the public closure probability (27).

(E.4) *Wrapper clearing.* (23) holds pointwise in $(\theta, \tilde{\xi}, \tilde{\varepsilon}_{sp})$.

(E.5) *Shares outstanding, the sponsor's valuation, and the OTC market.* Shares outstanding follow the law of motion (25), the sponsor values the fund by (20) and closes on the event (26), its continuation test from Section 4.5, and prices and the basket fire-sale loss follow (30) and (31).

(E.6) *Rational expectations.* The conjectured price map $\mathscr{P}(\cdot)$ used in (E.2)–(E.3) equals the realised one from (E.4), and $\mathscr{P}(\cdot)$ is strictly increasing, so the price reveals its sufficient statistic

$y = \theta + \sigma_s \sigma_\xi \tilde{\xi}$.

Condition (E.2) is the holder's individual optimality state by state: the SELL payoff on the left-hand side is the realised price at the observed y, not its expectation, and the severe probability on the right-hand side conditions on the marginal private signal and the public signal separately through (28). Because the severe event is the closure event, $\Pr(\text{severe} \mid \tilde{s}_h, y)$ is the conditional closure probability (28) itself.

6.2 Existence and uniqueness

What the price reveals is fixed by primitives before any equilibrium object is solved. The price's sufficient statistic is $y = \theta + \sigma_s \sigma_\xi \tilde{\xi}$ from (13), with precision

$$\tau_p \;=\; \frac{1}{\sigma_s^2\,\sigma_\xi^2}, \tag{35}$$

common to holders and the sponsor, and the Bayesian weight $k = \tau_p/\tau_s$ is a function of parameters alone. The reason is that the holders' cutoff term $s^*(y)/\sigma_s$ and the AP demand are functions of the observed price, so they move to the price side of clearing and do not affect the extraction of y. What remains to solve at $T = 1$ is the cutoff function $s^*(\cdot)$ and the redemption function $Q(\cdot)$, each a scalar equation state by state.

Substituting the AP sector's first-order conditions into (24), the price map is

$$\mathscr{P}(y) \;=\; \frac{\tilde{Y} y \;-\; s^*(y)/\sigma_s \;-\; \chi L\, \Pi_c(y, Q(y)) \;-\; (\chi/\chi_w)\, w(y)}{\Lambda}, \qquad \tilde{Y} = \frac{1}{\sigma_s} + \chi \rho_y, \quad \Lambda = \alpha + \chi, \tag{36}$$

affine in y apart from three state-dependent terms: the cutoff term $s^*(y)/\sigma_s$, the AP sector's closure-risk charge, and the cost of carrying its warehoused inventory. The last two rise in stress, so the closure margin widens the discount where closure risk is high, which is the feedback amplification of the NAV-based rule.

Proposition 1 (Monotone equilibrium)**.** *Suppose:*

1. *The fire-sale-loss-versus-valuation-dispersion condition holds:*

$$(d_H \bar{P} + D)\, \kappa_Q\, \bar{S}_Q \;<\; \sqrt{2\pi}\, \sigma_{g,h} \Big(1 + \frac{\sigma_s \alpha\, \tau_s}{\hat{\tau}}\Big), \tag{37}$$

where $\sigma_{g,h} = b_s \sqrt{1/\hat{\tau} + \sigma_{sp}^2}$ is the dispersion of the sponsor's valuation from the marginal holder's viewpoint and $\bar{S}_Q$ is the maximal redemption response to the price, a closed-form function of the primitives given in (47) *of Appendix B.1; with $\bar{S}_Q$ so defined,* (37) *is an inequality in primitives.*

2. *The AP sector's regularity condition* (46) *of Appendix B.1 holds; using the share-accounting bounds $Q, w \le \Omega_W$, it too is a closed-form inequality in primitives. The amount by which it*

holds, $\rho_R > 0$, bounds the AP sector's redemption response: $\bar{S}_Q$ in condition (1) is inversely proportional to ρ_R.

3. *Assumption T holds: the AP sector's demand is downward-sloping in the price and the price map is strictly increasing at the equilibrium.*
4. *The sorting condition of Section 4.1 holds.*
5. *The $T = 0$ feedback satisfies the contraction condition* (50) *in Appendix B.1, which bounds the per-round gain to the strategic-complementarity feedback $\kappa \to \pi_s \to \psi^* \to \kappa$ strictly below one.*

Then there exists a unique monotone rational-expectations equilibrium, with the following characterisation:

- *The price map is* (36)*; under Assumption T it is strictly increasing in y, so the price reveals y and the conjectured information structure is consistent.*
- *Cutoff function $s^*(\cdot)$: at each realised y, $s^*(y)$ is the unique solution of the pointwise indifference equation* (34)*, a scalar equation state by state with no coupling across states; redemption $Q(y)$ and warehousing $w(y)$ are the unique solution of the AP sector fixed point at the resulting price.*
- *Expected wrapper-vs-NAV spread*

$$\Delta_0 \;\equiv\; E_0[\Delta] \;=\; -E_0[\mathscr{P}(y)] \;=\; \frac{\bar{c}^*/\sigma_s \;+\; \chi L\, E[\Pi_c] \;+\; (\chi/\chi_w)\, E[w]}{\Lambda}, \qquad \bar{c}^* \;\equiv\; E[s^*(y)], \tag{38}$$

  which compensates the marginal buyer for average residual strategic-selling pressure, the AP sector's expected closure-risk charge, and the expected cost of carrying warehoused inventory to $T = 2$. The mass parameters λ and κ cancel in clearing and enter Δ_0 only through the share count Ω_W in the expected closure charge.
- *The marginal investor ψ^* is given by* (6)*.*

Proof. See Appendix B.1. □

Conditions (1)–(3) govern the $T = 1$ equilibrium, which determines uniqueness and the possibility of self-fulfilling failure: (1) and (2) are inequalities in primitives, and (3) collects the technical monotonicity properties, which are properties of equilibrium objects rather than primitives; primitive sufficient conditions for them exist but are conservative, and Lemma A.4 verifies the properties directly on the state grid. The $T = 0$ participation fixed point exists with no condition, since $T(\Omega_W) = \Omega_B + m(1 - F(\psi^*(\Omega_W)))$ is continuous on the compact interval $[\Omega_B, \Omega_B + m]$; condition (5) delivers its uniqueness, and condition (4) fixes the direction of the clientele partition.

Holders' selling decisions are strategic complements through the closure channel: heavier strategic selling clears at a lower price, the wider wedge raises the AP sector's redemption flow, and the redemptions raise the valuation threshold at which the sponsor closes, moving the fund toward the floor, and the higher closure probability lowers the HOLD payoff for those who continue to hold, tipping the marginal holder further toward selling. This complementarity is the multiplicity force that global-games models must regularise to obtain a unique threshold. In our setting the channel runs through the endogenous price, which plays two distinct roles, as in Ozdenoren and Yuan (2008).

Information role. The price aggregates strategic-selling pressure into a public signal $y(P_E^1)$ that all holders observe. When private signals are precise (small σ_s), the strategic-selling argument $(s^*(y) - \theta)/\sigma_s$ enters the price with large weight, the price-extracted signal precision τ_p rises, and the price becomes a strong common-knowledge device that lets holders coordinate on a self-fulfilling run. The same public signal moves the sponsor's valuation, so the coordination device and the closure trigger share a common input. Through this channel the endogenous price amplifies the complementarity. Noisier private signals (larger σ_s) weaken it by lowering the dependence of strategic selling on the price. This is the endogenous-public-signal mechanism of Angeletos and Werning (2006), Hellwig et al. (2006), Ozdenoren and Yuan (2008), Goldstein et al. (2011), and Lehar et al. (2026), distinct from Morris and Shin (1998) and Carlsson and van Damme (1993) where the public signal is exogenous.

Substitution role. The price also moves *against* sellers: heavier strategic selling forces the wrapper to clear at a wider discount, lowering the realised SELL payoff and dissuading holders from running together. The strength of this channel is governed by wrapper-side absorption $\Lambda = \alpha + \chi$: small Λ means each unit of strategic selling induces a sharp price drop, so substitution is strong; large Λ flattens the price response and weakens the penalty on sellers. Under the NAV-based closure rule, the same price movement that penalises sellers also accelerates the run: a lower price widens the wedge, raises the redemption flow, and moves the closure boundary. The substitution penalty and the closure-acceleration channel enter the marginal holder's calculus with the same slope $1/(\sigma_s \Lambda)$, so they offset in scale.

This is where the model departs from Ozdenoren and Yuan (2008), and the difference is the AP system. There the substitution effect is market depth, a fixed parameter, and its balance against the information effect decides uniqueness, so depth enters the condition directly. Here in-kind redemption splits absorption into two parts with opposite effects: all absorption flattens the price response and weakens the substitution penalty, while the redeemed part moves the closure boundary, so the APs' disposal choice places the redemption response on the destabilising side, and the common price scaling cancels from the comparison. What survives of the two capacities is asymmetric: market-maker depth reappears only on the stabilising side, because market makers cannot redeem, while the trading book reappears only inside the redemption response $\bar{S}_Q$, on the destabilising side. What survives the cancellation is condition (37), and it is a comparison of magnitudes. On the destabilising side is the coordination force the AP system transmits: the fire-sale loss, scaled by the boundary slope and the AP sector's maximal redemption response $\bar{S}_Q$. On the stabilising side is the dispersion of the sponsor's valuation from the marginal holder's viewpoint: the noise the price cannot carry, because the valuation embeds a signal nobody else observes. Uniqueness therefore rests not on the existence of sponsor noise but on its magnitude relative to the AP redemption response.

Unlike in Ozdenoren and Yuan (2008), market-maker depth α is not required for uniqueness: condition (37) can hold at $\alpha = 0$, and α enters only through the stabilising factor on the right side, so greater depth can only relax the condition, widening the parameter region on which Proposition 1 guarantees a unique equilibrium. Because (37) is sufficient rather than necessary, its failure at low α does not by itself imply multiplicity. The direct effects of depth are on the level of the wedge and the probabilities of the failure regimes, a property of the reduced-form market maker of Section 4.3, whose quote carries no closure charge. Appendix C computes the closure-sensitive variant: uniqueness and the monotone price map survive, and depth remains stabilising on net.

The dispersion $\sigma_{g,h} = b_s\sqrt{1/\hat{\tau} + \sigma_{sp}^2}$ has a floor along any path on which the price does not become perfect: as $\sigma_s \to 0$ with the price precision $\tau_p = 1/(\sigma_s\sigma_\xi)^2$ held fixed (the product $\sigma_s\sigma_\xi$ constant), it converges to $b_s\,\sigma_{sp} > 0$. However precise the holders' information about the fundamental, the sponsor's valuation retains an idiosyncratic component that nobody can learn from prices, and that component disciplines coordination. When instead $\sigma_s \to 0$ at fixed σ_ξ, the price becomes perfect, $b_s \to 0$, and the dispersion vanishes, the perfect-price limit of Section 6.3; uniqueness fails only when the sponsor's signal becomes public or degenerate, the limits studied there.

Setting the severe-state loss to zero, $d_H\bar{P} + D = 0$, isolates the source of the nonlinearity; Section 6.3 develops this case as the Grossman–Stiglitz benchmark. Both closure exposures vanish from the AP sector's problem, the AP sector's demand is affine, the pointwise indifference equation (34) is linear in (s, y), and the unique cutoff and price map are exactly affine:

$$s^*(y) \;=\; s_0 + s_1 y, \tag{39}$$

with intercept and slope given explicitly in Appendix B.1. All nonlinearity in the model enters through the single severe-loss term in (34). The slope s_1 shows the two roles of the price competing at the marginal holder's decision: the information role contributes $-\tau_p/\hat{\tau}$ (good public news deters selling), while the substitution role contributes $+\tilde{Y}/\Lambda$ (a higher price is a better exit). When absorption is large the substitution term dominates and the cutoff rises with the price: holders sell more when the price is high, which flattens the equilibrium price response. In the run region the severe-loss term steepens the cutoff sharply as the price falls, concentrating strategic selling near the closure boundary and steepening the price map there.

We also conduct numerical illustrations for the equilibrium results. Table 1 collects the base parameter values.[10] These numerical exercises serve three purposes. First, conditions (1) and (2) of Proposition 1 are inequalities in primitives, and the base values demonstrate that they define a non-empty region: (37) holds as $0.50 < 0.73$ (Lemma A.4 in Appendix B.7). Second, Assumption T concerns equilibrium objects rather than primitives, so it can only be checked on the solved equilibrium; the same lemma verifies it state by state. Third, the offsets and decompositions that

[10] Here $\sigma_\theta \equiv \tau_\theta^{-1/2}$. The $T = 0$ stage is instantiated by F uniform on $[0, 1]$, flexible mass $m = 0.3$, institutional base $\Omega_B = 0.45$, and fee $\phi = 0.124$, which deliver $\psi^* = 0.5$ and $\Omega_W = 0.6$; the base value of ϕ is a holding-period participation cost pinned by this sorting construction, pricing the wrapper's embedded trading option as well as its liquidity service, and is not comparable to an annual expense ratio. The feedback map from the float to the participation decision is a contraction with slope 0.43 in our computations.

Table 1: Base parameter values. All numerical illustrations use these values, varying only the parameters stated in each illustration; the values are chosen for transparency and are not calibrated to data. $\sigma_\theta \equiv \tau_\theta^{-1/2}$. The $T = 0$ stage is instantiated by F uniform on $[0, 1]$; the fee ϕ delivers $\psi^* = 0.5$, hence $\Omega_W = \Omega_B + m/2$.

Parameter	Value	Object
$\sigma_\theta,\ \sigma_s,\ \sigma_{sp}$	0.5	prior, holder-signal, and sponsor-signal noise
σ_ξ	1.5	noise-demand volatility
n	3	number of dealer-APs
$\chi,\ \chi_w$	$1.5,\ 0.75$	trading-book and warehouse capacity
α	1	market-maker depth
$d_o,\ \lambda_o$	$0.04,\ 0.2$	dealer margin and OTC flow impact
$d_H\bar{P},\ D$	$0.13,\ 0.3$	household OTC discount and fire-sale loss
$\bar{A}$	0.15	sponsor's AUM floor
$\Omega_B,\ m$	$0.45,\ 0.3$	institutional base and flexible mass
ϕ	0.124	holding-period participation fee

drive the results are matters of relative magnitude, and the illustrations demonstrate signs and mechanisms at a transparent set of base values. The values are chosen for transparency and are not calibrated to data; every illustration varies only the parameters it states, holding the rest at the base values.

6.3 The role of learning from prices

To demonstrate the information role of prices, we present a series of benchmarks, each of which removes one force from the model. The Grossman–Stiglitz benchmark removes the fire-sale loss. The no-learning benchmark removes the information in the price. The Diamond–Dybvig benchmark removes the noise in the holders' private signals. The uninformed-sponsor benchmark removes the sponsor's private information, and the perfectly-informative-price benchmark removes the noise in the price. The Diamond–Dybvig and perfectly-informative-price benchmarks produce the same limit, because the price's informativeness is powered by the precision of the signals that drive trading: as either noise vanishes, the price becomes perfect. What that limit looks like under the NAV-based closure rule is one of the results of this section, and it is not the bank-run indeterminacy. The indeterminacy arises at the uninformed-sponsor end. Together, the benchmarks separate the sources of the failure regimes from what learning from prices adds.

The Grossman–Stiglitz benchmark keeps learning and removes the fire-sale loss: set $d_H\bar{P} + D = 0$. The benchmark is a standard noisy rational-expectations setting in the tradition of Grossman and Stiglitz (1980), except that holders follow threshold strategies, and the equilibrium is the affine one in (39). It is unique without any condition, and there is no run. The reason is as follows. In equilibrium, holders invert the price to recover the public signal y: the price performs an information role. This role typically generates a complementarity, because holders want to trade in the direction of the price. When others sell, the price falls, the marginal holder infers weaker fundamentals, and selling becomes more attractive. There is an opposing force. Others' selling lowers the price the marginal holder receives when she sells, which makes selling less attractive: the price also performs a substitution role. Through this channel, selling decisions are strategic substitutes. Yuan (2005) shows that in the standard setting the substitution effect dominates the information effect, and the equilibrium is unique. The same is true in this benchmark.[11] In the full model, selling also widens the wedge and raises the AP sector's redemption flow, which raises the probability of closure. This channel scales with the fire-sale loss $d_H\bar{P} + D$, and it makes selling more attractive: investors want to sell before closure happens, which is the run. Through this additional channel, selling decisions are strategic complements, and when the fire-sale loss is large relative to the dispersion of the sponsor's valuation the complementarity dominates and self-fulfilling failure becomes possible (condition (37)). The Grossman–Stiglitz benchmark removes this channel, and the equilibrium is unique. Hence the strategic complementarity in the baseline model comes predominantly from the fire-sale loss transmitted through the closure channel. The informational role of the price is dominated by its substitution role.

The no-learning benchmark removes the information from the price itself. Let noise trading become extreme, $\sigma_\xi \to \infty$. The public precision vanishes, holders stop learning from the price, and the sponsor's valuation stops responding to it ($b_y \to 0$): the feedback complementarity through the valuation is absent, and so is the coordinating public signal. What survives is a run game in which the price only clears, together with a mechanical force. The APs are mechanical: their wedge is $\rho_y y - p$, and as the price loses information content ($\rho_y \to 0$) the wedge tracks the price movement one for one. A noisy price therefore keeps the redemption margin wide open: price declines with no fundamental content still pull baskets out of the fund. In our computations, moving σ_ξ from its base value to ten times it raises the redemption component of the failure probability from 0.011 to 0.031 while the equilibrium stays unique throughout. An informative price, by contrast, anchors the wedge to the public valuation and narrows the primary market's flow, which vanishes only in the perfectly informative limit outside the closure region. Learning from prices thus plays two opposing roles for closure risk: it coordinates holders, which destabilises, and it disciplines the mechanical arbitrage, which stabilises. Neither the coordination channel nor the anchoring channel requires the other, and the benchmarks that follow separate them. Stating the results formally requires the limit of the marginal holder's residual uncertainty about the closure event, and the following exact decomposition organises all of them.

Lemma 1 (Decomposition of the valuation dispersion)**.** *The conditional variance of the sponsor's*

[11] For holders, the dominance of the substitution effect is exact. A change in the aggregate cutoff moves the level of the price but does not change the signal that holders extract from it, so others' selling affects the marginal holder only through the price she receives. However, there is a further feedback. The sponsor forms its valuation partly from the price; within the monotone equilibrium it recovers the statistic y whatever the holders play, so coordinated selling reaches the valuation only when the price map is no longer invertible. There the feedback induces complementarities and resembles the trading frenzy of Goldstein et al. (2013), in which trades move the beliefs of a capital provider who learns from the price; here the capital provider is the sponsor and the belief is the NAV. With the fire-sale loss at zero the feedback is costless and the benchmark remains unique; when the sponsor relies mainly on the price, the closure event becomes price-measurable and self-fulfilling failure possible (Proposition 2(ii) below).

valuation, from the marginal holder's viewpoint, has two components,

$$\sigma^2_{g,h} \;=\; \underbrace{\frac{b_s^2}{\hat{\tau}}}_{\textit{fundamental component}} \;+\; \underbrace{b_s^2\,\sigma^2_{sp}}_{\textit{sponsor-signal component}}, \tag{40}$$

where $b_s = \tau_{sp}/(\tau_\theta+\tau_p+\tau_{sp})$ is the weight of the sponsor's signal in the valuation. Both components are strictly positive for all interior parameters ($0 < \sigma_s, \sigma_\xi, \sigma_{sp} < \infty$), and both vanish exactly when $b_s \to 0$: when the valuation becomes measurable with respect to the price.

Proof. See Appendix B.3. □

The quantity $\sigma_{g,h}$ measures what the marginal holder cannot know about the number that decides closure, after observing her signal and the price. The lemma shows that this uncertainty has exactly two sources: her remaining uncertainty about the fundamental, which enters the valuation through the sponsor's posterior, and the sponsor's signal noise, which no market observable transmits. Uniqueness fails when both sources vanish while the closure event still responds to behaviour. This is the standard condition for uniqueness in global games: the equilibrium is unique when agents retain enough residual uncertainty, and multiplicity returns as that uncertainty vanishes, with perfect common knowledge the limiting case (Carlsson and van Damme 1993; Morris and Shin 2003). When the public signal is an endogenous price, Angeletos and Werning (2006) and Hellwig et al. (2006) show that multiplicity returns already at a sufficiently informative price, without common knowledge. What is specific to our model is the object of the residual uncertainty: the sponsor's closure decision, rather than the fundamental itself.

Proposition 2 (Learning benchmarks)**.** *(i)* No-learning benchmark. *As $\sigma_\xi \to \infty$, the public precision vanishes ($\tau_p \to 0$, $k \to 0$, $b_y \to 0$): no agent learns from the price, and the valuation dispersion converges to*

$$\sigma_{g,h} \;\to\; \frac{\tau_{sp}}{\tau_\theta+\tau_{sp}}\sqrt{\frac{1}{\tau_\theta+\tau_s}+\sigma^2_{sp}} \;>\; 0,$$

strictly positive even as $\sigma_s \to 0$, so the uniqueness condition (37) *defines a non-degenerate region. (ii)* Uninformed-sponsor benchmark. *As $\sigma_{sp} \to \infty$ at fixed σ_ξ, the valuation becomes a function of the price alone ($b_s \to 0$, $\sigma_{g,h} \to 0$) while the redemption margin remains open, and* (37) *fails for any positive fire-sale loss beyond a threshold σ_{sp}: closure approaches common knowledge given the price, the monotone construction of Proposition 1 fails, and coordinated selling can produce the closure it anticipates. (iii)* Perfect-price benchmarks. *As $\sigma_s \to 0$ or $\sigma_\xi \to 0$, the price reveals the fundamental ($\tau_p \to \infty$), the sponsor rationally discards its own signal ($b_s \to 0$), and the valuation becomes price-measurable. On the no-closure region $\{\theta > \theta_c(0)\}$ the same limit closes the arbitrage wedge and redemptions vanish; on $\{\theta < \theta_c(0)\}$ closure occurs at any feasible redemption level and behaviour cannot affect it. On the band $\theta \in (\theta_c(0), \theta_c(Q^{run}))$, where $Q^{run} = (d_H\bar{P} - d_o)/(\lambda_o(1 + 1/n) + 1/\chi)$, a believed closure prices the wrapper at $\theta - (d_H\bar{P} + D)$, the wedge equals the closure loss, redemption remains profitable into the believed closure, and the no-run and run outcomes are* *both self-consistent: a band of classic common-knowledge indeterminacy survives the limit, of width* $\kappa_Q Q^{run}$*, equal to* 0.07 *at the base parameter values.*

Proof. See Appendix B.3. □

Part (i) is the no-learning benchmark introduced above, reached as a limit of the model rather than imposed. The run game survives: holders still sell before closure when their information is bad enough, the AP sector still redeems on the wedge, and the wrapper still fails in bad states. But the marginal holder's closure fear responds smoothly to the price, because the number that decides closure embeds a signal she cannot see, and the smooth response caps the complementarity.

Part (ii) is where the model loses its uniqueness guarantee, and it is a statement about the closure rule, not about what traders know. As the sponsor's own information becomes less precise, it prices off the market: its valuation approaches a known function of the price, the closure event approaches common knowledge given the price, and the redemption wave that moves the boundary is itself a function of the holders' own selling. This is the configuration in which coordinated selling can produce the closure it anticipates: the monotone equilibrium breaks down (in our computations, at $\sigma_{sp} = 10$ the clearing equation acquires multiple roots and the price map is no longer monotone, so the monotone-equilibrium construction fails; we do not construct the non-monotone equilibria). This is the trading-frenzy mechanism of Goldstein et al. (2013) operating on an institution: the manipulated belief is the fund's official NAV. The practical reading is about valuation regimes: a sponsor whose published NAV is a mechanical transformation of last trades, matrix prices with no independent input, runs a closure rule that a run can capture. The quality of the sponsor's independent information is a stability parameter, and Section 8 returns to it.

Part (iii) is the limit that both the Diamond–Dybvig and the perfectly-informative-price benchmarks reach, and under the NAV-based rule most of it is not the bank-run limit. Outside the closure region a perfect price closes the wedge: mechanical arbitrage leaves no discount to trade against, the primary market becomes inactive, and with $Q = 0$ the closure boundary stops responding to anything holders or APs do; closure there is a pure valuation event, the fund valued below its floor because the fundamental is bad, and an event that behaviour cannot move cannot be self-fulfilling. What survives at the limit is the boundary band of the proposition, where the price of a believed closure embeds the closure loss and redemption stays profitable: there the two self-consistent outcomes are exhibited, and the indeterminacy is the classic one. In our computations the redemption component of the failure probability is below half a percentage point in both limits and the band carries prior probability 0.003. The contrast with a share-count closure rule remains instructive: under a rule that closes on the share count alone, perfect information produces common knowledge of the trigger and multiplicity everywhere behaviour can move it; under the NAV-based rule, the same limit disciplines the redemption margin outside the closure region, and what remains of the bank run is the narrow band. The indeterminacy otherwise arises at the uninformed-sponsor end of the parameter space, not at the well-informed one.

The comparison across benchmarks identifies what the informative price adds to the run game. Two forces are present at every level of price informativeness, because they require only clearing: a falling price makes selling less attractive (the substitution role), and a falling price widens the wedge and raises the probability of closure (the closure fear). What an informative price adds comes from learning, and two agents learn from it: the holders and the sponsor. Holders' learning does two things. It coordinates: the price is a public signal, so it moves all holders' posteriors together and concentrates their selling in the same states. And it partially reveals the closure decision: holders use their information to reconstruct part of the sponsor's valuation, which shrinks $\sigma_{g,h}$, steepens the response of the closure fear to the price, and tightens the uniqueness condition (37). The sponsor's learning adds the feedback: holders' selling moves the price, the price moves the valuation, and the valuation moves the boundary. Through this feedback, coordinated selling reaches the trigger it fears. And against all three, the same informativeness anchors the mechanical arbitrage to the public valuation, narrowing the redemption margin that a noisy price leaves open.

These mechanisms do not all push in the same direction, and we make no monotone claim about the effect of learning on the failure probabilities π_m and π_s. In our computations the redemption component of failure falls as the price becomes more informative, from 0.031 in the no-learning limit to below half a percentage point in the perfect-price limit, while the coordination and revelation channels concentrate the remaining failures in the same states and the valuation feedback sharpens their dependence on the price. Section 7.2 quantifies which closures are panic-driven in each configuration.

The benchmarks together answer the question the paper is built around. Both banks and wrappers can experience runs; that parallel is not the point. The defining difference between the two architectures is the market price: a deposit has no price, while a wrapper share trades continuously, every agent conditions on its price, and, under the NAV-based rule, so does the institution that decides closure. Without a fire-sale loss (the Grossman–Stiglitz benchmark) there is no run, regardless of what the price reveals, because the substitution effect dominates. Without an informative price (the no-learning benchmark) the coordination and feedback channels are absent, but the mechanical redemption margin is at its widest. With a perfect price the redemption margin closes and only fundamental closures remain. Multiplicity requires something else: a closure rule mechanical enough to be computed from the price. Whether the equilibrium remains unique depends on whether any part of the closure decision stays hidden from the marginal holder (condition (37)), and the model locates the danger not at informed extremes but at the uninformed-sponsor corner. Learning also delivers the model's positive content: the public precision τ_p, the information term in the cutoff slope (39), and the informational comparative statics of Section 7.3. The same price that aggregates information values the fund that coordinated selling can close. Which force wins is a parameter question, and the model answers it; it cannot be settled by analogy with banks.

7 Equilibrium Analysis

7.1 Three regimes and basket fire-sale loss

The equilibrium delivers three regimes indexed jointly by the realised wrapper discount $\Delta_E = \mu - P_E^1$ and the sponsor-closure indicator $\mathbf{1}\{\mathcal{C}\}$. Figure 2 illustrates how the equilibrium prices and payoffs move across the three regimes as the fundamental θ varies.

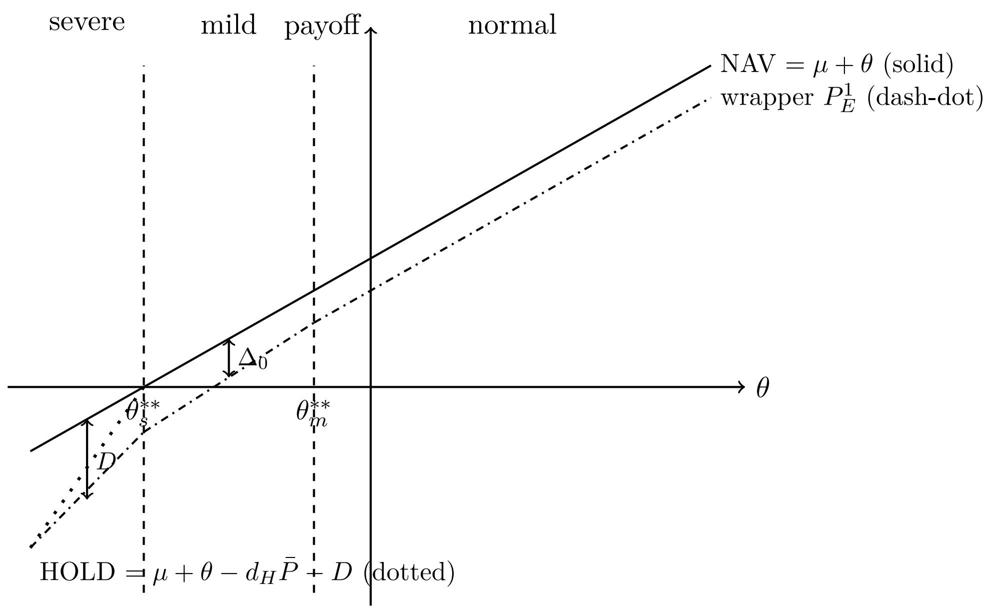

Figure 2: Equilibrium prices and payoffs across the three regimes as θ varies. Solid: NAV $\mu + \theta$. Dash-dot: wrapper price P_E^1. Dotted: HOLD payoff in severe failure, $\mu + \theta - d_H \bar{P} - D$ (in the normal and mild regimes HOLD coincides with NAV). Thresholds θ_m^{**} and θ_s^{**} are given by (41) and (44).

In the *normal regime* ($\Delta \le d_H \bar{P}$, no closure), arbitrage holds the wrapper near the public valuation and the redemption flow, if any, leaves AUM well above the floor. SELL pays the equilibrium wrapper price P_E^1; HOLD pays $\mu + \theta$; the basket is intact. In the *mild-failure regime* ($\Delta > d_H \bar{P}$, no closure), the wrapper price has decoupled from NAV but the sponsor's valued AUM remains above the floor. Sellers absorb the wrapper discount; holders still receive $\mu + \theta$. In the *severe-failure regime* (the closure event $\mathcal{C}$ of (26)), the sponsor's valuation, depressed by redemptions, by the fundamental, or by both, has fallen below the floor; the residual basket $\mathcal{L} = \Omega_W^1$ is sold into the OTC market wholesale, dealer capacity is exhausted, and the basket bears a fire-sale loss $D > 0$. SELL pays a depressed P_E^1; HOLD pays closure proceeds $\mu + \theta - d_H \bar{P} - D$.

We now characterise the threshold for the mild failure regime. By Proposition 1 the price is strictly increasing in the public signal, so the mild trigger (32) is the lower-threshold event $y < y_m$, where y_m is the unique solution of $\mathscr{P}(y_m) = -d_H \bar{P}$, that is, of

$$\tilde{Y} y_m \;-\; \frac{s^*(y_m)}{\sigma_s} \;-\; \chi L\, \Pi_c(y_m, Q(y_m)) \;-\; \frac{\chi}{\chi_w}\, w(y_m) \;=\; -\Lambda\, d_H \bar{P}. \tag{41}$$

Since $y = \theta + \tilde{u}$ with $\tilde{u} \sim N(0, 1/\tau_p)$ independent of θ, the conditional failure probability is

$$\Pr(\text{mild} \mid \theta) \;=\; \Phi\!\left(\frac{\theta_m^{**} - \theta}{\sigma^{\text{mild}}}\right), \tag{42}$$

where $\sigma^{\text{mild}} = \sigma_s \sigma_\xi = 1/\sqrt{\tau_p}$ and the threshold on θ at zero noise is $\theta_m^{**} = y_m$ (at $\tilde{u} = 0$ the public signal equals the fundamental). In the affine benchmark (39) the threshold is in closed form; in general it is implicit but unique by price monotonicity.

We next characterise the threshold for the severe failure regime. The severe trigger (33) is the closure event $m_{sp} < \theta_c(Q)$. Conditional on θ,

$$\Pr(\mathcal{C} \mid \theta) \;=\; \Pr\left(b_y\, y + b_s\, \tilde{s}_{sp} \;<\; \theta_c(Q(y)) \,|\, \theta\right), \tag{43}$$

a Gaussian integral over the noise $(\tilde{\xi}, \tilde{\varepsilon}_{sp})$. At zero noise both signals equal the fundamental, so the valuation is $(b_y + b_s)\theta$ and the closure condition is $(b_y + b_s)\,\theta = \theta_c(Q(\theta))$. The left side is strictly increasing in θ and the right side is non-increasing (a better fundamental narrows the wedge and lowers redemptions), so the crossing is unique. The crossing defines the zero-noise closure threshold θ^{**}_{AUM}, with closure at zero noise for $\theta < \theta^{**}_{\text{AUM}}$. The severe threshold at zero noise is the closure crossing,

$$\theta_s^{**} \;=\; \theta^{**}_{\text{AUM}}. \tag{44}$$

For the mild regime to be a meaningful intermediate region we impose the regime-defining condition $\theta^{**}_{\text{AUM}} < \theta_m^{**}$: the closure threshold lies below the mild threshold. The ordering is a restriction on parameters, not a consequence of the model, and we maintain it throughout; it holds at the base values. Severe failure is then strictly nested inside mild at zero noise; with noise the nesting is approximate, as quantified at the trigger definitions. The regime structure on θ at zero noise is

$$\theta < \theta_s^{**} : \text{severe} \;<\; \theta < \theta_m^{**} : \text{mild only} \;<\; \theta > \theta_m^{**} : \text{normal}.$$

In states where the severe-loss term is negligible, the pointwise indifference condition (34) reduces to $\mathscr{P}(y) = z(s^*(y), y)$, a pure rational-expectations consistency between the realised price and the marginal holder's posterior: coordination is purely informational and the cutoff is locally affine as in (39). In the run region, the term $(d_H \bar{P} + D) \cdot \Pr(\text{severe} \mid \tilde{s}_h, y)$ delivers Diamond–Dybvig strategic complementarity.

The balance-sheet parameters now affect closure risk on both sides, and the two capacities pull in different directions. Neither capacity is required for uniqueness (Section 6.2); what the capacities move is the discount and the failure probabilities, and stabilising in this section means lowering these. Trading-book capacity χ stabilises the wrapper-vs-NAV spread: higher χ raises AP absorption, narrowing Δ_0 through (38) (because $\Lambda = \alpha + \chi$), and lowering the mild-failure probability. But the same capacity raises the closure boundary $\theta_c(Q)$ in (26): a higher-capacity AP redeems more shares per unit of wedge, so a given discount depletes the share buffer faster and moves the boundary up. Whether more AP trading capacity shrinks or expands the severe region depends on whether the wedge-narrowing effect dominates the faster depletion, a comparative-statics question that the model makes explicit rather than assumes. Warehouse capacity χ_w is unambiguously stabilising for closure risk: a deeper warehouse lowers the closure boundary $\theta_c(Q)$ by absorbing more of the selling wave without retiring shares. The equilibrium feedback through the cutoff function $s^*(\cdot)$ operates on top of both direct effects, since the cutoff moves the price level and hence the wedge state by state. This two-sided role of AP capacity follows from the law of motion of shares outstanding (25): capacity supports the price, and it also makes redemption easier.

The three regimes map onto the three capacities in a simple way. In the normal and mild regimes the object that matters is the discount, and the discount is an absorption statistic: state by state, the price map divides selling pressure by $\Lambda = \alpha + \chi$, so market-maker depth and the AP trading book support the price interchangeably, dollar for dollar, and warehouse capacity plays no role, because a stored share and a redeemed share are equally good bids for the flow. In the severe regime the object that matters is the depletion of shares outstanding, and there the composition of absorption matters. Market-maker depth is unambiguously stabilising: it narrows the wedge and retires no shares, because market makers have no access to the primary market. AP trading capacity is two-sided, as above. Warehouse capacity is unambiguously stabilising. In short, absorption capacity prices the mild regime; the composition of absorption capacity, how much of it can redeem and how much of what it absorbs it can store, governs the severe regime.

This mapping separates two kinds of wrapper failure that a single-trigger specification conflates. If the AP sector steps back from arbitrage, because its trading capacity is small or because closure risk makes redemption expensive (the $D\,\Pi_c$ charge in (19)), a wide wedge can stand while redemptions stay small, and the wrapper survives at a persistent discount: a deep mild regime. Collapse requires the opposite behaviour: a wide public wedge and an AP sector redeeming at scale. The severe regime is reached not when arbitrage is abandoned but when it is exercised through the primary market faster than the fund's valued AUM can bear. This matches the empirical pattern: sponsor closures follow sustained primary-market outflows, while funds whose redemption channel is shut trade at wide and persistent discounts without closing. In the 2008–2024 population of US bond-ETF closures (Appendix D), the shares-outstanding component accounts for 95 percent of the median asset decline among the closures that lost assets before delisting, and trailing net redemptions predict the closure announcement conditional on the sponsor's fee revenue.

The episode evidence lines up with this mapping once it is read through the primary market. In March 2020, the large corporate bond ETFs traded at discounts of 5% and more (Falato et al., 2021; Haddad et al., 2021) while the underlying bond market dislocated and dealer balance sheets strained (O'Hara and Zhou, 2021), and secondary volume surged as creations and redemptions fell relative to trading;[12] no large bond ETF closed, and the discounts disappeared within weeks of the SMCCF announcement. Through the model this is a deep mild regime: absorption continued

[12]Bank of Canada Staff Analytical Note 2020-27 documents fewer fixed-income ETF creations and redemptions in March 2020 alongside a 67 percent rise in exchange volume for corporate-bond ETFs.

to price the selling wave while the redemption margin narrowed, so the discount widened and the share buffer survived. The severe regime's empirical counterpart is instead the population of small and specialised funds closed in 2020–2022, where sustained primary-market outflows pushed AUM below the sponsor's floor. The model separates these outcomes through a single decision, the AP's disposal choice, rather than treating a wide discount as a step toward liquidation. Appendix D documents the closure rule and walks through two recent liquidations whose mechanics map into the model's objects.

7.2 Panic-driven and fundamental-driven failure

Uniqueness is only one dimension of fragility. Condition (37) determines whether self-fulfilling failure is possible; where it fails, the monotone-equilibrium construction no longer rules out states in which coordinated selling alone decides between survival and closure. Within the unique-equilibrium region, fragility has a second measure: how much of the failure probability is caused by the run, as opposed to failure that fundamentals, noise trading, and arbitrage flows would produce on their own. A no-closure-fear benchmark provides the decomposition. In the benchmark, holders follow the cutoff of the Grossman–Stiglitz benchmark of Section 6.3: they ignore the severe-state loss in their decisions, while the AP sector's problem, including its closure charges, is unchanged; the benchmark therefore differs from the affine economy (39), which also removes the AP's closure charges. The physical closure trigger operates unchanged: selling responds to fundamentals and the price, the AP redeems on the wedge, and closure occurs when redemptions cross the buffer. Failure in the benchmark is fundamental-driven in this sense. Write π_m^b, π_s^b for the benchmark failure probabilities and define the panic-driven components

$$\pi_r^{\text{panic}} \;\equiv\; \pi_r \;-\; \pi_r^b \;\geq\; 0, \qquad r \in \{m, s\}. \tag{45}$$

Proposition 3 (Panic-driven failure)**.** *Under* (37) *and* (49)*, with* $d_H\bar{P}+D > 0$*: (i) the equilibrium cutoff lies strictly above the benchmark cutoff at every state,* $s^*(y) > s^b(y)$*; (ii) the equilibrium price lies strictly below the benchmark price at every state; (iii) the mild and closure events of the equilibrium contain those of the benchmark state by state, so* $\pi_m \geq \pi_m^b$ *and* $\pi_s \geq \pi_s^b$*, with strict inequality whenever the corresponding boundary carries positive density; and (iv) the panic-driven components* (45) *are strictly increasing in the fire-sale loss* $d_H\bar{P}+D$ *and vanish as the loss vanishes.*

Proof. See Appendix B.4. □

The proposition formalises the sense in which the run is an amplification mechanism: the fear of closure raises the cutoff at every state, the heavier selling clears at lower prices, the wider wedges accelerate redemption, and the failure regions expand beyond what fundamentals and arbitrage would deliver. The panic-driven component is the model's measure of coordination-induced fragility, and it prices the fire-sale loss: by (iv), π^{panic} scales with $d_H\bar{P}+D$, which is why a fire-sale backstop acts on the panic component directly (Section 8).

The benchmark cutoff has a dominance interpretation. Given the price, a holder sells for any beliefs about others when her posterior is below the price, $z < \mathscr{P}(y)$, and holds for any beliefs when $z > \mathscr{P}(y) + d_H\bar{P} + D$, since the closure loss is bounded by $d_H\bar{P} + D$. The benchmark cutoff is the first boundary, and the equilibrium cutoff lies in the band between the two: the proof of Proposition 4 bounds the gap $s^*(y) - s^b(y)$ by $(d_H\bar{P} + D)/[\tau_s/\hat{\tau} + 1/(\sigma_s\Lambda)]$. The panic-driven component is therefore the probability mass that the dominance band contributes to failure.

To help with the interpretation, we provide three clarifications. First, the strict dominance boundary at the realised equilibrium price lies below $s^b(y)$, because the equilibrium price is lower than the benchmark price and a lower price makes selling attractive to fewer holders. We use $s^b(y)$ rather than that boundary because the benchmark is the equilibrium of a well-defined economy in which no holder's decision includes the closure loss, while a no-fear selling rule evaluated at the equilibrium price is not an equilibrium of any economy. The choice is conservative: everything between the two boundaries is classified as fundamental-driven, so the panic-driven component, if anything, understates panic.

Second, even fundamental-driven failures destroy value because the closure rule itself triggers the liquidation that realises the discount. The fire-sale discount is avoidable: no property of the asset's cash flows requires it, and it is not incurred if the wrapper does not close. The decomposition therefore classifies the cause of a failure, whether the closure fear in holders' decisions was needed for it. It does not classify the loss, which in both components is created by the institution's liquidation rather than by the asset.

Third, the benchmark retains the AP sector's closure charges: the APs still price the fire-sale exposure of redeemed and warehoused shares. This too is deliberate. The APs' closure fear is a price, and it restrains the redemption race: removing the charges raises failure rather than lowering it. In our computations, a fully fearless benchmark, with the AP sector's charges removed as well, has severe probability 0.035 against the benchmark's 0.032, so the panic component measured against it is 0.002 against the reported 0.005. The holders' closure fear is a coordination motive, and it intensifies the run. The decomposition removes the destabilising fear and retains the disciplining one: the reported panic component is the coordination contribution measured against the economy whose intermediation technology, including its pricing of closure risk, is held fixed, and the two-benchmark comparison brackets what reclassifying the AP sector's priced fear would change.

The decomposition parallels the panic-based runs of Goldstein and Pauzner (2005). In their bank-run model, the unique equilibrium has a run threshold above the fundamental threshold, and failures between the two are panic-based: the bank would have survived had depositors not run. Our panic-driven component is the analogue. By part (i), the equilibrium cutoff lies above the benchmark cutoff at every state, and the failures in between would not have occurred without the closure fear. Two differences follow from the architecture. Because our state is multidimensional, the fundamental-driven component is defined by the affine benchmark rather than by a dominance region on the fundamental, and it includes failures caused by noise trading and arbitrage flows. And because coordination runs through the price rather than through a sequential-service rule, the size of the panic component is governed by price informativeness: it vanishes along the noisy limit, is maximal in between, and at the informative extreme is confined to the narrow boundary band (Proposition 4), and indeterminacy requires a closure event that holders can reconstruct and behaviour can move (Proposition 2(ii), and (iii) at the perfect-price extreme). In their framework the deposit contract disciplines the panic; here the price and the sponsor's valuation do.

The decomposition also answers the question of why the coordinating role of an informative price is the object of study. In the no-learning benchmark of Proposition 2(i) the price is so noisy that no agent learns from it; at the opposite extreme the price is nearly perfect. The next proposition shows that the panic-driven component vanishes along the noisy limit and is confined to the boundary band at the informative one.

Proposition 4 (Panic requires an intermediately informative price)**.** *Fix the other primitives, with the fire-sale loss small enough that* (37) *holds along the noisy limit. (i) As* $\sigma_\xi \to \infty$*, the panic-driven components* π_m^{panic} *and* π_s^{panic} *converge to zero, while the benchmark failure probabilities remain bounded away from zero. (ii) As* $\sigma_\xi \to 0$*, equilibrium redemptions vanish outside the band of Proposition 2(iii), closure converges to the valuation event there, and the panic-driven components converge to at most the prior probability of the band,* 0.003 *at the base parameter values. Panic-driven failure away from the band requires a price informative enough to coordinate and noisy enough to leave a wedge for arbitrage to redeem against.*

Proof. See Appendix B.4. □

The intuition for the noisy limit is as follows. The panic-driven component of failure comes from the closure fear in the holders' cutoff, and this fear has limited force. The closure probability in a holder's payoff is at most one, and it enters multiplied by the fire-sale loss $d_H\bar{P} + D$, so the fear raises her selling trigger by at most a fixed amount, at every state and at every level of price informativeness. A bounded rise in the cutoff lowers the price, and so moves the sponsor's boundary, by a bounded amount as well. The run can therefore cause a closure only in states where the valuation would have been within a fixed distance of the floor anyway: coordinated selling can push the wrapper over the boundary, but only from nearby. The states in which behaviour can move redemptions occupy a window of fixed width in the public state, and as σ_ξ grows, the noise-demand shock spreads the state over a widening range, so the window makes up a shrinking fraction of outcomes, and the panic-driven component vanishes. The failures themselves do not vanish: heavy noise selling keeps the mechanical redemption margin wide open, and the benchmark failure probabilities stay large. Every failure in this limit would have occurred even if no holder had feared closure.

The intuition for the informative limit is different, and it is two-sided. Outside the closure region a nearly perfect price closes the wedge: mechanical arbitrage finds no discount, the primary market becomes inactive, and the closure boundary $\theta_c(Q)$ stops moving with behaviour; closure fear is fear of an event that selling cannot influence, so the benchmark and equilibrium cutoffs move together and the gap between them closes. Panic requires that some part of the closure event respond to the coordinated action, and that part is the redemption margin, which an informative price shuts wherever closure is not itself in prospect. Angeletos and Werning (2006) and Hellwig et al. (2006) show that an informative price restores multiplicity without restoring common knowledge; here, as the price becomes perfect, the closure event approaches common knowledge itself. Common knowledge of a trigger produces multiplicity only if behaviour can move the trigger, and outside the closure region the same limit shuts the redemption margin, so there is nothing left for coordinated selling to move. What survives is the band of Proposition 2(iii): close enough to the boundary, the price of a believed closure embeds the closure loss, the wedge this opens is exactly what keeps the APs' redemption profitable, and the classic indeterminacy returns on a set of states whose width is set by the APs' access advantage $d_H\bar{P} - d_o$. The band is narrow in our computations, with prior probability 0.003 against $\pi_s = 0.037$, and the panic component in Figure 1 turns up again at the far left of the range as the band opens.

Between the limits the panic-driven component is positive and the model quantifies it. In our computations, moving σ_ξ across the range, the severe-regime panic component is 0.001 at σ_ξ five times the base value, 0.005 at the base value, and 0.001 at one fifth of the base value, against total severe probabilities of 0.048, 0.037, and 0.060; the mild-regime panic component is negligible at the base values. The total failure probability is non-monotone in informativeness because the mechanical redemption channel and the coordination channel move in opposite directions, and the decomposition separates them: informativeness strengthens the coordination channel and narrows the redemption margin. Figure 1 plots the decomposition across the range.

We now collect what the propositions say as the price becomes more informative. When the price is uninformative, the equilibrium is unique and the panic-driven component is zero: failure is frequent but fundamental-driven, carried by the mechanical redemption margin. As informativeness rises, the coordination and revelation channels switch on, the panic-driven component turns positive, and the uniqueness condition (37) tightens as holders reconstruct more of the sponsor's valuation. When the price is informative enough, the wedge closes outside the closure region, the panic component falls, and most remaining failure is the valuation event: the sponsor values the fund down to its floor on a bad fundamental, with no redemptions to move the boundary. At the informative extreme itself a narrow band of classic indeterminacy survives near the boundary (Proposition 2(iii)). Uniqueness, under the NAV-based rule, holds along the noisy limit and away from the boundary band at the informative one; the indeterminacy of the model arises chiefly along a different axis, the sponsor's information (Proposition 2(ii)), reached not by making traders better informed but by making the closure rule mechanical. Any claim that transparency stabilises or destabilises the wrapper must therefore say which object it refers to, the total failure probability, the panic-driven component, or the possibility of self-fulfilling failure, and which channel it acts on, the information in the price or the information in the sponsor's valuation.

7.3 Comparative statics

All comparative-statics results in this section, the policy analysis of Section 8, and the welfare analysis of Online Appendix OA.2 apply within the unique-equilibrium region of Proposition 1, that is, on the parameter set where the fire-sale-loss-versus-valuation-dispersion condition (37) and the price-monotonicity condition (49) hold. Outside that region the equilibrium is not unique and the derivatives below are not well-defined as parameter perturbations. Lemma A.4 in Appendix B.7 verifies that the maintained conditions hold jointly and strictly at the base parameter values, and exhibits a second parameter point for the regularities of Corollary 2.

The comparative-statics statements below are stated as *direct effects*: derivatives of the failure thresholds and the equilibrium spread with respect to a primitive parameter, computed holding the marginal-investor equilibrium objects $(\kappa, s^*(\cdot), \psi^*, \lambda, \lambda_b)$ fixed at their equilibrium values and the informational objects $(k, \tau_p, \rho_y, b_y, b_s)$ fixed. By Lemma A.1 the informational objects are exact closed-form functions of the primitive parameters, so any parameter perturbation moves them mechanically; that movement is the learning channels displayed next, with τ_p and ρ_y subsumed in (k, b_y, b_s) through the closed forms. The full REE total derivative therefore decomposes as

$$\frac{d\mathcal{V}}{d\theta_j} = \underbrace{\frac{\partial \mathcal{V}}{\partial \theta_j}}_{\text{direct effect (this section)}} + \underbrace{\frac{\partial \mathcal{V}}{\partial k}\frac{dk}{d\theta_j} + \frac{\partial \mathcal{V}}{\partial b_y}\frac{db_y}{d\theta_j} + \frac{\partial \mathcal{V}}{\partial b_s}\frac{db_s}{d\theta_j}}_{\text{learning channels}} + \underbrace{\int \frac{\partial \mathcal{V}}{\partial s^*(y)}\frac{ds^*(y)}{d\theta_j}\, dy}_{\text{closure-fear feedback through } s^*(\cdot)}$$

for $\mathcal{V} \in \{\theta_m^{**}, \theta_s^{**}, \Delta_0\}$ and θ_j a primitive parameter. The three terms map to the forces of Section 6.3. The direct effect operates through clearing and the closure boundary: it is the substitution role of the price and the physical movement of the closure boundary, and it is the sign of the parameter's first-order impact on the threshold conditions (41)–(44) or on the expected spread (38). The second term operates through the learning channels: a parameter that moves the precisions moves the holders' learning (the coordination and revelation mechanisms, through the Bayesian weight k and the valuation dispersion $\sigma_{g,h}$) and the sponsor's learning (the valuation feedback, through b_y and b_s), by the precision mapping of Lemma A.1. The third term operates through the closure fear: the parameter moves the cutoff function state by state through the pointwise indifference condition (34), and this is where the panic-driven component of failure forms. We isolate the direct effect because it has a clean structural interpretation; we discuss the learning and closure-fear channels qualitatively after each corollary, and note explicitly when they could overturn the direct sign. Proofs are in Appendix B.2.

The corollaries follow the same order as the forces. Corollary 1 is the substitution side: the equilibrium spread is a clearing object, set by absorption capacity. Corollary 2 is the closure side: the balance-sheet and adoption parameters move the redemption threshold and the closure boundary. The informational channels and Corollary 3 are the learning side: the information parameters move the two learning channels. Corollary 4 combines the channels on the severe regime, and Section 7.4 separates them with the failure decomposition.

Corollary 1 (Equilibrium discount and wrapper-transformation gain)**.** *The equilibrium wrapper-vs-NAV spread* (38) *is*

$$\Delta_0 \;=\; \frac{\bar{c}^*/\sigma_s + \chi L\, E[\Pi_c] + (\chi/\chi_w)\, E[w]}{\Lambda}, \qquad \bar{c}^* = E[s^*(y)], \qquad \Lambda \;=\; \alpha + \chi.$$

The wrapper's expected transformation gain is positive iff $\Delta_0 \le d_H \bar{P}$. In the normal regime, a shocked investor who sells the wrapper at $T = 1$ receives $\mu - \Delta_0$ in expectation, while a shocked direct holder receives $\bar{P}(1 - d_H) = \mu - d_H \bar{P}$. The wrapper-transformation gain per shocked share is therefore

$$g \;\equiv\; d_H \bar{P} - \Delta_0,$$

strictly positive iff $\Delta_0 < d_H \bar{P}$. The spread satisfies

$$\frac{\partial \Delta_0}{\partial \alpha} \;<\; 0, \qquad \frac{\partial \Delta_0}{\partial d_o} \;>\; 0,$$

where the dealer margin is passed through to the spread by the warehousing cost. Trading capacity χ is two-sided: it deepens absorption through Λ and it scales the closure charge and the warehousing cost, and in our computations the two effects nearly offset.

Note that the mechanical-mass parameter λ does not affect Δ_0 because the additive λ on MM's quote in (12) absorbs mechanical sales pointwise; the strategic-mass parameter κ does not affect Δ_0 through clearing because the $\kappa\Phi(\cdot)$ structures cancel exactly. Both masses do enter Δ_0 indirectly, through the share count Ω_W in the expected closure charge $E[\Pi_c]$.

The spread is the substitution role of the price in comparative-statics form: it is a clearing object, and no learning parameter enters it except through the average cutoff. It is the price the AP sector charges to clear the average strategic-selling pressure ($\bar{c}^*/\sigma_s$, the average level at which strategic selling enters the exact clearing equation (24)) plus the AP sector's expected closure-risk charge and warehousing cost, divided by the combined wrapper-side absorption capacity ($\Lambda = \alpha + \chi$). When AP capacity or market-maker depth is large relative to residual flow, the spread is tight and the wrapper closely tracks NAV. A more pessimistic cutoff function (a uniform upward shift raising $\bar{c}^*$) widens the spread; a higher dealer margin (d_o) is passed through to the spread.

The wrapper transformation raises the expected sale price of shocked investors, and the gain g rises with market-maker depth; trading capacity's effect on the spread is two-sided, as the corollary states, and in our computations the two effects nearly offset. This pins down what the wrapper achieves when arbitrage works, while the failure-threshold corollaries below pin down when it breaks.

Corollary 2 (Direct effects on failure thresholds)**.** *In the parameter region of Appendix B.2 ($\theta_m^{**} > 0$; non-empty by Lemma A.4(ii)), the failure thresholds* (41)–(44) *satisfy*

$$\frac{\partial \theta_m^{**}}{\partial \alpha} < 0, \qquad \frac{\partial \theta_s^{**}}{\partial \bar{A}} > 0, \qquad \frac{\partial \theta_s^{**}}{\partial \Omega_W} < 0, \qquad \frac{\partial \theta_s^{**}}{\partial \chi_w} < 0, \qquad \frac{\partial \theta_r^{**}}{\partial \varepsilon} > 0 \text{ for } r \in \{m, s\},$$

where $\partial/\partial\varepsilon$ denotes the derivative with respect to a uniform upward shift of the cutoff function, $s^*(\cdot) + \varepsilon$, and $\Omega_W = \Omega_B + \kappa + \lambda$ is the share count, so the Ω_W-derivative applies to Ω_B, κ, and λ alike. The direct effect of trading-book capacity χ on θ_s^{**} is ambiguous: it tightens the equilibrium wedge (stabilising) and raises the closure boundary $\theta_c(Q)$ through faster redemption (destabilising).

The role of the mechanical mass λ is specific to the primary-market trigger: λ cancels in clearing and has no effect on the price or on the mild threshold, but it enters the severe threshold through the share count, because every share outstanding, whoever holds it, adds to the buffer the redemption wave must exhaust. The direct-holder OTC-flow parameter λ_b has no direct effect on θ_s^{**}, because the trigger is primary-market redemption of wrapper shares.

More wrapper-side absorption capacity (χ) tightens the AP-arbitrage spread Δ_0, lowering the mild-failure threshold. A higher sponsor AUM floor ($\bar{A}$) raises the closure boundary $\theta_c(Q)$ pointwise and expands the severe region. A larger wrapper share count (Ω_W) lowers the boundary and shrinks the severe region. Deeper warehouse capacity (χ_w) lowers equilibrium redemptions, and with them the boundary, and shrinks the severe region. A more pessimistic cutoff function (a uniform upward shift) lowers the price at every state, widening both the realised discount and the AP's wedge, pushing both thresholds up.

There is no learning channel behind $\partial\theta_m^{**}/\partial\chi$. The price precision τ_p is primitive (Lemma A.1), so capacity moves the mild threshold through absorption alone, and the equilibrium feedback of the cutoff function operates through the price level in the pointwise condition (34), not through what anyone learns.

The informational channels are monotone in the public precision $\tau_p = 1/(\sigma_s^2\sigma_\xi^2)$, a primitive common to holders and the sponsor and decreasing in both noises: the sponsor's price weight b_y and the public valuation weight ρ_y rise with τ_p, and the valuation dispersion $\sigma_{g,h}$ falls. A more informative price therefore does three things at once. It strengthens the holders' coordination and revelation channels, shrinking $\sigma_{g,h}$ and tightening the uniqueness condition (37). It strengthens the valuation feedback, since a sponsor that trusts the price (b_y large) writes the NAV down faster when selling depresses it. And it anchors the mechanical arbitrage, narrowing the public wedge and the redemption margin. The first two channels are destabilising and the third stabilising, which is why the panic-driven component of failure is hump-shaped in τ_p (Proposition 4) while the total failure probability is non-monotone.

Corollary 3 (Sponsor-signal precision σ_{sp})**.** *The precision of the sponsor's signal moves two objects in opposite directions:*

1. Closure-frequency channel*: a more accurate valuation (smaller σ_{sp}) tracks the fundamental more closely, and the unconditional closure probability rises toward the fundamental rate* $\Pr(\theta < \theta_c(Q))$*, monotonically along the computed path.*
2. Manipulability channel*: a less accurate valuation (larger σ_{sp}) shifts weight from the sponsor's signal to the price (b_s falls), shrinking the valuation dispersion $\sigma_{g,h}$, tightening the uniqueness condition* (37)*, and reaching the multiplicity of Proposition 2(ii) in the limit.*

The corollary is the model's statement about the quality of the sponsor's independent information, and the two channels price an institutional trade-off. In our computations, moving σ_{sp} from 0.35 to 0.75 around the base value lowers the severe-failure probability from 0.050 to 0.025: a sponsor with poor independent information values the fund off the price, its valuation stays close to the public one, and the fund closes less often. The same movement, however, has an informational cost: the sponsor's valuation becomes reconstructable from the price, coordination sharpens, and beyond a threshold σ_{sp} the monotone equilibrium breaks down (Proposition 2(ii)). Accurate valuation closes funds that should close; price-based valuation closes funds that the market decides to close. The transparency analysis of Section 8 builds on this distinction.

Corollary 4 (Severe-state expected loss)**.** *The expected basket fire-sale loss conditional on the* $T = 0$ *information set is* $E[D \cdot \mathbf{1}\{severe\}] = D\pi_s$*, with* π_s *the probability of the closure event* (26)*. The loss is increasing in the lump-sum penalty* D *at fixed closure probability (the equilibrium response of* π_s *to* D *is not signed analytically) and is strictly increasing in the sponsor's floor* $\bar{A}$ *(which raises the boundary* $\theta_c(Q)$ *at every redemption level), and strictly decreasing in the wrapper share count* Ω_W *and in the warehouse capacity* χ_w *(a deeper warehouse diverts absorbed flow away from redemption).*

Severe failure is driven by the sponsor's closure rule operating on AUM at the sponsor's valuation: the wrapper closes when redemptions and the valuation together push AUM below the floor, and the wholesale sale of the residual basket overwhelms OTC dealer capacity. The expected severe-state loss therefore scales with the size of the fire-sale loss (D) and the size of the closure region. A higher floor $\bar{A}$ makes the sponsor more conservative, expanding the closure region (in our computations, raising $\bar{A}$ from 0.15 to 0.18 raises π_s from 0.037 to 0.050). A larger share count Ω_W thickens the cushion between the starting AUM and the floor. A deeper warehouse delays depletion. Trading capacity χ enters closure risk on both sides, as in Section 7.1; the next subsection separates the two sides with the failure decomposition.

7.4 Strained AP capacity, concentration, and the composition of failure

In the stress episodes that motivate the model, the balance sheets of the dealers that operate as APs contract at the same time as prices fall. This subsection asks what strained AP capacity does to fragility, what the number of APs does, and whether either interacts with the informational role of the price. A distinction matters here. The AP sector's provision of capacity is endogenous throughout the model: how much to absorb and how to dispose of it respond to the public wedge through (18)–(19), and the run advances when this capacity is engaged. What is parametric is the cost side of the balance sheet: the trading-book capacity χ, the warehouse capacity χ_w, the dealer margin d_o, and the number of APs n. We read AP strain as a fall in χ and trace its effect through the failure decomposition of Section 7.2. The next corollary states what strain cannot affect and what it does.

Corollary 5 (Capacity and the information channel)**.** *(i) The public precision τ_p, the valuation weight ρ_y, and the sponsor's posterior weights (b_y, b_s) do not depend on any balance-sheet parameter $(\chi, \chi_w, d_o, \lambda_o, \alpha, n)$. (ii) On the interior region, aggregate redemption $Q(y)$ is strictly increasing in χ at every state at the fixed clearing price; the full-equilibrium path is computed in Section 7.4.*

Proof. See Appendix B.5. □

Part (i) has a simple reason: a mechanical AP sector contributes nothing to the price's information content, so no balance-sheet contraction can degrade what agents learn. AP strain does not affect the information structure. All effects of AP strain are purely mechanical, and part (ii) states the mechanism: capacity converts a given wedge into redemptions faster.

The net effect of capacity on fragility combines two opposing forces, absorption narrows the wedge while redemption capacity moves the boundary, and its sign is not available analytically. We compute the failure decomposition (45) numerically, varying χ (with $\chi_w = \chi/2$) around the base values; the statements that follow are illustrations of signs. As χ rises from 0.5 to 2.5, the total severe-failure probability rises from 0.031 to 0.039, and its panic-driven component nearly triples, from 0.002 to 0.006: at the base values the redemption side dominates, and capacity intensifies both the mechanical depletion and the coordination incentive. AP strain, read as falling capacity, lowers total closure risk and lowers its panic share: under the NAV-based rule a strained AP sector is a stabilised one in the severe regime, at the cost of deeper discounts in the mild regime. The interaction with uncertainty runs opposite to the intuition that strain and turbulence compound. Against a tight prior ($\sigma_\theta = 0.35$), closures are rare ($\pi_s = 0.002$) and nearly a fifth of them are panic-driven; against a diffuse prior ($\sigma_\theta = 0.7$), closures are far more frequent ($\pi_s = 0.117$) but the panic share falls below a tenth. The model's prediction is that the coordination component of fragility is proportionally largest in informationally calm periods, when a wide valuation movement is surprising enough to coordinate on. Figure 3 plots the capacity and concentration paths.

The number of APs is the second parameter we vary, and it works through the internalisation terms of (19).

Proposition 5 (Concentration stabilises)**.** *On the interior region, at fixed price and state, aggregate redemption $Q(y)$, the closure probability $\Pi_c(y, Q(y))$, and hence π_s are increasing in n: the internalisation terms enter the AP sector's first-order conditions scaled by $1/n$, so fewer, larger APs redeem less against the same state. The full-equilibrium ordering, with the price and cutoff adjusting, is verified computationally.*

Proof. See Appendix B.5. □

In our computations, moving from a monopolist AP to ten APs raises the severe-failure probability from 0.034 to 0.038 and its panic-driven component from 0.003 to 0.006: fragmentation raises both the level of closure risk and the share of it that coordination causes. The economics is Cournot competition over two priced commons. Each AP's redemption widens the OTC discount all APs receive and moves the fund's valued AUM toward its floor, damaging every exposed book;

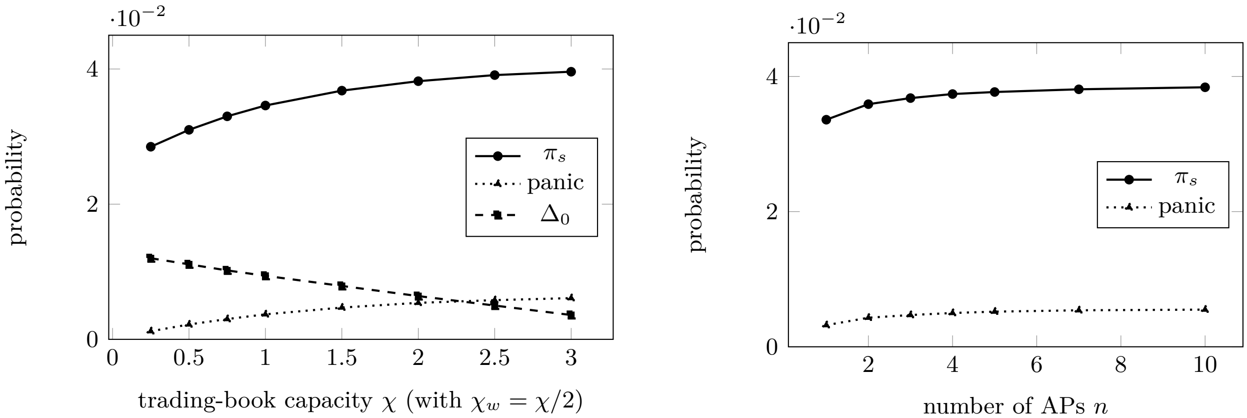

Figure 3: Capacity and concentration paths at the base parameter values of Table 1. Left: the severe-failure probability π_s, its panic-driven component (45), and the expected wrapper-vs-NAV spread Δ_0 as trading-book capacity χ varies with $\chi_w = \chi/2$ and all other parameters fixed; the base value is $\chi = 1.5$, and the path varies the two AP capacities proportionally, a joint capacity experiment rather than a change in χ alone. Right: π_s and the panic-driven component as the number of APs n varies with aggregate capacities held fixed; the base value is $n = 3$.

a concentrated AP sector internalises most of this damage and holds back, while a fragmented one races. The empirical counterpart is the concentration of primary-market activity, measured in the N-CEN filings for our bond-ETF universe: the median fund has four APs with any primary-market activity in a year, the largest AP carries 60 percent of its gross creations and redemptions, and the top three carry nearly all of it. Through the model this concentration is a stabiliser, not a fragility. The prediction runs against the common regulatory instinct that a broader AP base would make redemption capacity more robust: broadening the base weakens internalisation, and in the model it raises both closure risk and its panic component. Because aggregate capacity is held fixed as n varies, the comparison isolates internalisation: a broader base does not deepen absorption, and its full effect is the weakening of the redemption restraint. The result holds aggregate capacity fixed and speaks to the redemption externality alone; operational fragility, the withdrawal of an AP or a balance-sheet shock correlated across a small number of APs, is outside the model, and a broader base may insure against it. Both the capacity and the concentration results are model predictions at transparent parameter values rather than calibrated magnitudes: the signs are the robust content, recomputed across the loss and information planes in Online Appendix OA.4, and the magnitudes are illustrative.

Nothing in the analysis requires capacity to be a constant of the environment. Let χ be realised at $T = 1$ as a publicly observed capacity state, possibly correlated with the fundamental. Conditional on the realisation, the $T = 1$ equilibrium is exactly the one characterised by Proposition 1, with the prior on θ replaced by its conditional counterpart given the observed capacity state; the $T = 0$ participation decision integrates over the capacity distribution.[13] A bad draw then

[13] Because χ is public, its realisation acts as a public signal about θ when the two are correlated; the Gaussian information structure is unchanged with the conditional prior mean and variance in place of $(0, \sigma_\theta^2)$.

Table 2: Policy experiments at the base parameter values. Each row solves the model with one instrument changed from the base values of Table 1. π_s is the severe-failure probability, panic its panic-driven component (45), and Δ_0 the expected wrapper-vs-NAV spread. The gate row is the economy of Proposition 6(i), in which the severe panic-driven component is zero by construction; the swing revaluation raises the boundary slope from κ_Q to $\kappa_Q + s_w$ with $s_w = 0.3$.

Experiment	π_s	panic	Δ_0
Anchor	0.037	0.005	0.008
Primary-market gate ($Q \equiv 0$)	0.025	0	0.007
Flow fee doubled ($\lambda_o = 0.4$)	0.034	0.003	0.008
Dealer margin doubled ($d_o = 0.08$)	0.033	0.004	0.008
Swing revaluation ($s_w = 0.3$)	0.046	0.009	0.011
Floor raised ($\bar{A} = 0.18$)	0.050	0.008	0.013
Warehouse support ($\chi_w = 1.1$)	0.036	0.004	0.007
Depth support ($\alpha = 1.5$)	0.033	0.004	0.007

contracts dealer capacity and lowers expected fundamentals together, which is the joint pattern of the March 2020 episode, and the decomposition of this section attributes to each channel its component state by state: the capacity collapse moves the fundamental-driven component, and the information channel adds the panic-driven component on top.

8 Policy analysis

Whether a tool stabilises the wrapper is a positive question in this model. It is a statement about two objects: the uniqueness condition (37), which determines whether self-fulfilling runs are possible, and the failure probabilities π_m and π_s, which determine how likely the mild and severe regimes are. In this section we evaluate the deposit-era toolkit and the market-era toolkit against these two objects, using only the results already established: Proposition 1, Corollaries 1–5, Lemma 1, and Propositions 2, 3, and 4. None of the conclusions depends on a welfare criterion; the normative analysis is in Online Appendix OA.2. Table 2 collects the computed experiments of this section, evaluated at the base parameter values; the signs, not the magnitudes, are the portable content.[14]

The deposit-era toolkit acts on objects the wrapper does not have. Deposit insurance guarantees redemption at par; a wrapper share carries no par claim, and the holder's loss in failure is the fire-sale discount on the basket, not a broken promise of par. Reserve requirements target the liquid pool from which par redemptions are met; the in-kind wrapper needs no such pool, because redemption delivers the basket itself. The closest analogues of these tools inside the model act on the fire-sale loss and on the primary market, and their effects differ.

Primary-market gates and fees are the suspension-of-convertibility analogue, and their effects

[14] Online Appendix OA.4 recomputes the three directional claims of this section, capacity, concentration, and the floor, across the loss and information planes: no sign reverses at any computed point, and the two redemption-side effects vanish only in the nearly-perfect-price limit where the redemption margin closes.

can be stated exactly.

Proposition 6 (Gates and redemption fees). *(i) Under a primary-market gate, $Q \equiv 0$: the closure boundary is fixed at $\theta_c(0)$, the redemption component of π_s and the severe panic-driven component are zero, and π_s equals the valuation component alone; the mild panic-driven component need not vanish, because closure fear still shifts the holders' cutoff and with it the discount event. The AP sector's demand takes the warehouse form $W = \tilde{\chi}\,(\Delta_A - L\,\Pi_c)$ with $\tilde{\chi} = \chi\chi_w/(\chi + \chi_w)$, absorption falls to $\Lambda_g = \alpha + \tilde{\chi} < \Lambda$, and the equilibrium is unique for any fire-sale loss: the redemption response $\bar{S}_Q$ in* (37) *is zero. (ii) A redemption fee that raises the dealer margin d_o, or a flow-proportional fee that raises λ_o, does not tighten the uniqueness condition* (37) *(the dealer margin leaves it unchanged; a flow-proportional fee relaxes it through the redemption response $\bar{S}_Q$) and lowers equilibrium redemption $Q(y)$ at every state: the closure boundary $\theta_c(Q(y))$ recedes state by state, and π_s and its panic-driven component fall. The direct effect on the expected discount is $\partial\Delta_0/\partial d_o > 0$ (Corollary 1).*

Proof. See Appendix B.6. □

Part (i) is the formal content of suspension, and the NAV-based rule states its limit exactly. A gate that halts creations and redemptions closes the redemption margin, the route through which behaviour reaches the closure boundary: the severe panic-driven component vanishes and uniqueness needs no condition. What the gate cannot remove is the valuation margin. The fund can still be revalued down to closure by a bad fundamental, and in our computations π_s under a gate is the valuation component, 0.025 against the baseline 0.037: a gate removes the coordination-sensitive 0.012, not the closure risk itself. The cost is the closed-end configuration: every absorbed share must be warehoused, absorption falls from Λ to Λ_g, and creation-redemption arbitrage is disabled while the warehouse's convergence demand survives. In our computations the expected spread is little changed, because closure risk falls with the gate; what the gate removes is the primary-market cap on a discount once warehouse capacity fills, the configuration of GBTC before its spot-ETF conversion.

Part (ii) is the interior version, and it is the model's rendering of swing pricing. A fee on redemption charges the transacting party the liquidity cost its flow imposes, which is what swing pricing implements in cash-redemption funds by adjusting the NAV paid to redeemers, and what in-kind redemption already implements in part by making the AP bear its own liquidation costs. The model separates two instruments that the label "swing" bundles, and they act in opposite directions. A swing *fee*, a rise in d_o or λ_o, taxes the redemption flow: in our computations, doubling the flow-proportional component lowers π_s from 0.037 to 0.034 and the panic component from 0.005 to 0.003. A swing *revaluation*, by contrast, adjusts the NAV down more aggressively when outflows are large, raising the redemption slope of the closure boundary (26) from κ_Q to $\kappa_Q + s_w$ with $s_w > 0$: it hands redemptions more leverage over the trigger, and in our computations raises π_s to 0.046 and the panic component to 0.009. Charging redeemers stabilises; revaluing the fund against them destabilises. The empirical literature on swing pricing in open-end bond funds finds that it dampens outflow sensitivity in stress (Jin et al., 2022), consistent with the redemption-fee channel; the theory literature studies its design as an anti-run instrument (Capponi et al., 2020). Gates and fees therefore trade the severe regime for a deeper and more persistent mild regime. They relocate the cost of stress from the patient holders, who bear the fire-sale loss, to the shocked sellers, who bear the discount. This exchange lowers the probability of a basket fire-sale, but it also lowers the liquidity transformation the wrapper provides: the gain from holding the wrapper rather than the underlying is $g = d_H\bar{P} - \Delta_0$ per shocked share (Corollary 1), and a wider expected discount reduces it. If the discount under the instrument exceeds $d_H\bar{P}$, investors do better holding the underlying directly and selling it in the OTC market when the liquidity shock arrives: the wrapper no longer raises the shocked seller's expected sale price, and the $T = 0$ participation decision moves against it.

A fire-sale backstop bounds the loss at which the residual basket clears: a facility that stands ready to buy the basket caps $d_H\bar{P}+D$. This is the one instrument that improves both the uniqueness condition and the severe region while leaving the primary market open and the wrapper's liquidity transformation intact. It relaxes the uniqueness condition (37), whose destabilising side is the product of the fire-sale loss, the closure-boundary slope, and the AP redemption response: capping the loss scales the whole side down, without operating through any equilibrium response. And it shrinks the severe region through behaviour, since the closure boundary itself contains no loss term: the equilibrium cutoff falls at every state when the loss falls, and with it the redemption response that carries selling to the boundary. The Federal Reserve's SMCCF is the institutional analogue, and the model interprets its speed, discounts closing within weeks of the announcement, as the uniqueness condition operating: the facility did not need to absorb the basket, it needed to cap the loss that made the run self-fulfilling.

The last instrument lowers the sponsor's fixed operating cost in stress. Because the floor is the fixed-cost component of the sponsor's continuation test, $\bar{A} = c_f/\phi$ from (22), any policy tool that delivers a stress-contingent reduction of this cost lowers the floor and, at the maintained calibration, the redemption charge with it. The closure boundary $\theta_c(Q)$ then falls at every redemption level, and the severe region shrinks (Corollary 4). The floor enters the price only through the APs' closure charge, so the price effect is small rather than zero; in our computations the reverse movement, raising $\bar{A}$ from 0.15 to 0.18, raises π_s from 0.037 to 0.050 and moves the expected spread from 0.008 to 0.013. The forms of these instruments can be private or public. Sponsors already provide the private form through fee waivers that keep marginal funds alive (Appendix D), and a merger works through the same cost: the acquiring sponsor spreads the fixed cost across a larger platform, so the franchise continues under a lower floor rather than closing, which is the transfer the sellable-franchise reading of Section 4.5 predicts. Exchange or regulatory relief on the listing, licensing, administration, and custody charges that make up c_f is the semi-private form. The public forms are a direct subsidy during a stress window and an organised consortium that takes over troubled funds, the public counterpart of the private merger. In any of these forms, this is the least balance-sheet-intensive intervention in the model: the fixed cost of operating a fund is orders of magnitude below the balance-sheet support the other instruments require, and unlike a gate it keeps the primary market open.

Two clarifications bound what this instrument is. First, it is not loss absorption: no assets are bought, no credit risk is taken, and holders' losses are not socialised; the support removes a self-fulfilling trigger rather than paying for an outcome. Second, the case for the instrument is stress-contingent, for a reason outside the model. In calm periods the AUM floor performs efficient exit, closing unviable funds, and closure is a routine event (Appendix D); the model has no fund heterogeneity and cannot weigh this screening role. It is only in stress, when the same floor becomes the trigger that coordinates the run, that lowering the fixed cost delivers stability.

Balance-sheet support must distinguish three capacities that a single "dealer support" label conflates. Market-maker depth α has two effects, both stabilising: it narrows the discount through absorption and narrows the AP's wedge without any offsetting depletion channel, because market makers cannot redeem. AP trading-book capacity χ has two opposing effects: it narrows the discount, but it converts a given wedge into redemptions faster. Section 7.4 gives the composition consequence: at the base parameter values the redemption side dominates, so trading-book support raises total closure risk and its panic-driven component even as it narrows discounts. AP warehouse capacity χ_w lowers closure risk and leaves the price essentially unchanged: absorbed flow is stored rather than redeemed (in our computations, raising χ_w lowers π_s from 0.037 to 0.036). And the structure of the AP sector is itself a determinant of closure risk: by Proposition 5, broadening the AP base weakens the internalisation that restrains the redemption race, so a policy that recruits more APs to "strengthen arbitrage capacity" raises closure risk through the same channel as trading-book support. A support facility aimed at stability should therefore be directed at warehouse and market-making capacity rather than at the trading book, which converts discounts into redemptions faster; the model's preferred reading of the SMCCF is a transfer of inventory off dealer warehouses. And no capacity is required for the equilibrium to be unique: the common price scaling cancels from condition (37), depth enters only as further discipline, and what balance-sheet support chiefly moves is discounts and failure probabilities. The instruments that act most directly on the possibility of self-fulfilling failure are the ones that shrink the fire-sale loss, cap the APs' redemption response, or strengthen the sponsor's independent information.

Transparency deserves a separate caution, and the results already established state it exactly; no new result is needed. First, let price informativeness rise as σ_ξ falls, other primitives fixed. By Propositions 3 and 4, the panic-driven components of failure vanish along the noisy limit, are strictly positive on the interior region where (37) holds with $d_H\bar{P} + D > 0$, and at the informative extreme are confined to the boundary band of Proposition 2(iii), with prior probability 0.003 at the base values; in our computations they are hump-shaped over most of the range, with a small upturn at the far informative end as the band opens, and the redemption component of π_s falls with price informativeness. The equilibrium is unique along the noisy limit and away from the boundary band at the informative one. Second, let the sponsor's signal precision fall as σ_{sp} rises, other primitives fixed. The valuation dispersion $\sigma_{g,h}$ falls, condition (37) tightens, and beyond a threshold σ_{sp} the equilibrium is not unique (Proposition 2(ii)).

These two results separate two objects that the word transparency conflates: the information in the price and the information in the sponsor's valuation. Transparency of the price is a fall in σ_ξ, and under the NAV-based rule its effects are two-sided but bounded: it strengthens the coordination channel on the interior region, it weakens the mechanical redemption channel everywhere, and near the perfect-price limit it opens the narrow boundary band, the one place where trader-side information alone produces indeterminacy. Transparency of the sponsor's valuation is different in kind. The closure decision is disciplined by the sponsor's private information: it is because the valuation embeds a signal the market cannot reconstruct that coordinated selling cannot fully control the trigger. A valuation process that is mechanical, matrix prices with no independent input, is transparent in the narrow sense that anyone can compute the NAV from traded prices, and it is the process under which the closure rule can be captured (Proposition 2(ii)).

In practice, several disclosures are bundled under the label, and they enter the model differently. Transparency about the underlying, post-trade reporting of trades in the constituent securities and more frequent disclosure of the basket's composition, improves the pricing of the basket and enters as a lower σ_ξ; in so far as it also improves holders' own valuations, it enters as a lower σ_s. Transparency about the wrapper's own market, consolidated reporting of wrapper trading, also lowers σ_ξ. Transparency about the primary market, timely disclosure of creations and redemptions, is nearly redundant: the mechanical AP sector's flow is already inferable from the price. And transparency about the sponsor's valuation process itself, the quality and independence of the inputs behind the NAV, is the disclosure the model newly distinguishes: investing in independent valuation, higher τ_{sp}, is what keeps the closure rule out of the market's hands, at the cost of closing more funds whose fundamentals are bad (Corollary 3). This gives the model's case for a requirement, in listing standards or fund regulation, that the sponsor maintain a valuation function with independent information about the fundamentals of the portfolio securities, either an internal research function or an external pricing agent with inputs of its own. The case is for a minimum standard, and it binds near the mechanical-valuation end. There, more independent information lowers the probability of self-fulfilling failure. Precision is a primitive in the model, so the argument is an externality reading rather than a formal acquisition result: a sponsor choosing precision would internalise the franchise it stands to lose, but the fire-sale losses of a self-fulfilling failure fall on the holders. For a sponsor already well informed, the trade-off runs through the closure-frequency channel of Corollary 3 instead, and the model makes no case for pushing precision higher. The requirement builds on the SEC's fair-value rule, Rule 2a-5 under the Investment Company Act, which already requires a good-faith fair-value determination and active oversight of pricing services rather than mechanical acceptance of matrix prices.

At our illustrative parameters, a more informative price lowers the redemption component of severe failure monotonically while the panic component rises and then falls. Transparency of the price, in short, changes the composition of fragility; it is transparency of the sponsor's valuation, in the narrow price-based sense, that creates the possibility of self-fulfilling failure. The model does not say disclosure is harmful: it says the stability case for it runs through which of the two objects the disclosure informs.

The stability question of the deposit era, is the liquid pool large enough, has no counterpart here. The stability questions of the market era are whether the fire-sale loss is capped, whether the balance sheets behind the price can absorb and store the selling wave, whether the AP base is concentrated enough to restrain the redemption race, and whether the closure rule embeds information the market cannot reconstruct. Online Appendix OA.2 develops the normative counterpart: the wrapper produces pecuniary externalities at the holders' selling margin and the AP's capacity margin, and a redemption tax paired with a warehouse-capacity subsidy implements the planner's allocation under the stated criterion. The criterion there maximises the surplus of the two sides of the liquidity transformation: the investors, who demand liquidity when shocks arrive, and the AP, which supplies it. The criterion therefore measures the gains from liquidity transformation itself, the function the wrapper exists to perform. It does not count the utility of noise traders, wrapper market makers, or OTC dealers: these agents are represented by parametric schedules and take no optimising decision, so they have no utility function to count. For the OTC dealers the exclusion is also substantive: the fire-sale discount is a price below the assets' final value, so part of the sellers' loss is the dealers' gain, and a criterion that counted dealer surplus would treat that part as a transfer rather than a loss (Online Appendix OA.2 discusses this). Nor does it count the sponsor, an optimising agent whose closure and architecture decisions are analysed in Sections 4.5 and 9: its surplus is fee income rather than liquidity transformation, the function the criterion measures. Whether their surplus should enter, and how, is a question we leave open, which is why the analysis of this section is stated in positive terms.

9 The sponsor's choice of wrapper architecture

The paper has so far taken the wrapper's architecture as given; sponsors choose it. The same underlying assets appear in open-end wrappers, in closed-end funds, and in vehicles that are never launched, and the sponsor's choice weighs the open-end form's larger fee base, the mandate-driven institutional base that only an arbitrage-anchored price can serve, against its closure risk. Online Appendix OA.1 states the sponsor's value of each architecture and derives the resulting threshold rule: at the base parameter values the sponsor chooses the open-end wrapper whenever its severe-failure probability stays below $\bar{\pi} = 0.73$, so the wrapper is chosen at $\pi_s = 0.037$, and the choice survives partial attachment of the mandate base up to a fraction 0.94.

Two further results are stated there. Under risk neutrality the two forms compete for the same liquidity-demanding clientele, and the observed split, liquidity-needers in the wrapper and patient option-traders in the closed-end fund, is assigned by the dispersion of the untethered closed-end price, which risk neutrality prices at zero. And anticipated fragility operates at the launch decision: closure risk that never strikes still destroys liquidity transformation by pushing sponsors toward the closed-end form or away from launching, so a regime with high panic-driven closure risk loses liquidity transformation through two channels, the failures that occur and the wrappers that are never launched. The cross-section is the leveraged-loan configuration: assets with broad benchmark demand are wrapped open-end, and assets with thin index demand and heavy fire-sale exposure, such as the leveraged loans held in CLOs, are wrapped closed-end (Diamond et al., 2025).

10 Conclusion

This paper develops a theory of market-based liquidity transformation in which the secondary-market price of the wrapper plays a dual role: it clears the wrapper market, and it is the public signal on which patient holders condition their hold-or-sell decision and the sponsor conditions the valuation that decides closure. The dual role is what distinguishes market-based intermediation from its bank-based predecessor. In Diamond and Dybvig (1983) the intermediary has no market price and coordination runs through the sequential-service rule on the bank's liquid reserves. Here the substitution effect and the information effect of coordinated selling operate through the same price, and the balance between the two determines whether the wrapper is fragile.

The dual role changes what fragility depends on. The price penalises coordinated selling through the discount it must clear at, and the same discount advances the run, through the APs' mechanical redemptions and the sponsor's valuation. Because the two effects move with the same price, market depth cancels from the uniqueness condition. What remains is a comparison of institutional magnitudes: the AP sector's redemption response against the marginal holder's residual uncertainty about the sponsor's valuation. The equilibrium is unique when the fire-sale loss is small relative to that uncertainty. The condition compares a substitution effect with an information effect, as in Ozdenoren and Yuan (2008), but its terms are different: in Ozdenoren and Yuan the substitution effect comes from the market maker's depth; here depth cancels, the substitution is provided by the AP sector's absorption, the unique institutional arrangement of liquidity wrappers, and what stands against the information effect is the redeemed part of that absorption, the AP sector's redemption response. The noise that disciplines the run is the sponsor's private information, which disciplines it at every informativeness short of a perfect price; at the perfect-price limit the sponsor rationally discards it, and a narrow band of classic indeterminacy survives near the closure boundary. Multiplicity chiefly requires the opposite corner: when a sponsor with no independent information values the fund from traded prices alone, the closure event becomes computable from the price, the uniqueness guarantee fails, and the bank-run configuration returns, with the indeterminacy exhibited at the perfect-price boundary band. The sources of fragility are therefore institutional, the in-kind redemption system and the sponsor's information, not the depth of the market or the precision of the traders' information.

The intermediaries that operate the wrapper enter on both sides of this comparison. AP trading capacity narrows the discount but converts a given discount into redemptions faster, and in our computations the redemption side dominates. The AP's warehouse, absorbed flow that is stored rather than redeemed, stabilises unambiguously. And because each AP internalises only its own share of the damage its redemptions do to the common pool of dealer balance sheets and fund headroom, concentration restrains the redemption race that fragmentation intensifies.

The same characterisation organises the policy analysis; the instrument comparisons are mechanism predictions at transparent parameter values, with the signs the robust content, not calibrated policy rankings. Deposit-era tools, insurance and reserve requirements, act on objects the wrapper does not have; the instruments that work are the ones that act on the institutional sources. A fire-sale backstop, of which the Federal Reserve's March 2020 Secondary Market Corporate Credit Facility is the institutional analogue, relaxes the uniqueness condition by capping the fire-sale loss, a direct factor of the condition's destabilising side. Dealer support stabilises in its warehouse form, not in its trading-book form. Lowering the sponsor's fixed operating cost in stress lowers the closure boundary with no balance-sheet support and almost no price effect. A minimum standard of independent valuation keeps the closure event from becoming computable from the price. The tools inherited from the deposit era, gates and redemption fees, trade the severe regime for a deeper and more persistent discount and a smaller liquidity transformation.

The choice of architecture prices the same forces ex ante: the sponsor selects the open-end form only while closure risk stays below a threshold, so a fragile regime loses liquidity transformation in the wrappers that are never launched as well as in the failures that occur. The model also predicts that wrappers backed by deeper underlying markets, with a more concentrated AP sector, should be less prone to discount runs; mapping this prediction to the cross-section of corporate bond, bank-loan, CLO, and private-credit ETFs is a natural empirical exercise.

The scope of the model is the in-kind wrapper on illiquid assets. A stablecoin promises par against a reserve pool, so its fragility is closer to a deposit run (Ma et al., 2024), with the market price adding the informational channel this paper studies; an ETF on a liquid underlying, gold or spot bitcoin, sits at the corner of the model with a small fire-sale loss, where the equilibrium is unique with no run. As liquidity transformation migrates from banks to markets, the price that clears the market also informs every participant who observes it. A model of market-based intermediation that suppresses this informational role abstracts away from what makes markets different from banks; the dual-role framework introduced here is one way to bring it inside the model.

References

George-Marios Angeletos and Ivan Werning. Crises and prices: Information aggregation, multiplicity, and volatility. *American Economic Review*, 96(5):1720–1736, 2006.

Agostino Capponi, Paul Glasserman, and Marko Weber. Swing pricing for mutual funds: Breaking the feedback loop between fire sales and fund redemptions. *Management Science*, 66(8):3581–3602, 2020.

Hans Carlsson and Eric van Damme. Global games and equilibrium selection. *Econometrica*, 61 (5):989–1018, 1993.

Qi Chen, Itay Goldstein, and Wei Jiang. Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, 97(2):239–262, 2010.

Coalition Greenwich. U.S. corporate bond trading in 2024 by the numbers. Technical report, Coalition Greenwich, 2025.

Caitlin D. Dannhauser. The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). *Journal of Financial Economics*, 125(3):537–560, 2017.

Douglas W. Diamond and Philip H. Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419, 1983.

William Diamond, Francesco Falasconi, and Yi Xu. CLOs and run risk in leveraged loans. *Working Paper*, 2025.

Antonio Falato, Itay Goldstein, and Ali Hortaçsu. Financial fragility in the COVID-19 crisis: The case of investment funds in corporate bond markets. *Journal of Monetary Economics*, 123:35–52, 2021.

Financial Stability Board. Global monitoring report on non-bank financial intermediation 2025. Technical report, Financial Stability Board, 2025.

Itay Goldstein and Ady Pauzner. Demand-deposit contracts and the probability of bank runs. *Journal of Finance*, 60(3):1293–1327, 2005.

Itay Goldstein, Emre Ozdenoren, and Kathy Yuan. Learning and complementarities in speculative attacks. *Review of Economic Studies*, 78(1):263–292, 2011.

Itay Goldstein, Emre Ozdenoren, and Kathy Yuan. Trading frenzies and their impact on real investment. *Journal of Financial Economics*, 109(2):566–582, 2013.

Itay Goldstein, Hao Jiang, and David T. Ng. Investor flows and fragility in corporate bond funds. *Journal of Financial Economics*, 126(3):592–613, 2017.

Amit Goyal, Yoshio Nozawa, and Mengxi Qiu. ETF ownership and bond-market liquidity. *Working Paper*, 2025.

Sanford J. Grossman and Joseph E. Stiglitz. On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408, 1980.

Valentin Haddad, Alan Moreira, and Tyler Muir. When selling becomes viral: Disruptions in debt markets in the COVID-19 crisis and the Fed's response. *Review of Financial Studies*, 34(11): 5309–5351, 2021.

Christian Hellwig, Arijit Mukherji, and Aleh Tsyvinski. Self-fulfilling currency crises: The role of interest rates. *American Economic Review*, 96(5):1769–1787, 2006.

Dunhong Jin, Marcin Kacperczyk, Bige Kahraman, and Felix Suntheim. Swing pricing and fragility in open-end mutual funds. *Review of Financial Studies*, 35(1):1–50, 2022.

Alfred Lehar, Christine A. Parlour, and Kathy Yuan. Liquid staking. *Working Paper*, 2026.

Jiacui Li, Maureen O'Hara, Andrew Rapp, and Xing (Alex) Zhou. Portfolio trading in corporate bonds. *Working Paper*, 2025.

Guido Lorenzoni. Inefficient credit booms. *Review of Economic Studies*, 75(3):809–833, 2008.

Yiming Ma, Yao Zeng, and Anthony Lee Zhang. Stablecoin runs and the centralization of arbitrage. *Working Paper*, 2024.

Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, 88(3):587–597, 1998.

Stephen Morris and Hyun Song Shin. Global games: Theory and applications. In Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, editors, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Cambridge University Press, 2003.

Maureen O'Hara and Xing (Alex) Zhou. Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis. *Journal of Financial Economics*, 142(1):46–68, 2021.

Emre Ozdenoren and Kathy Yuan. Feedback effects and asset prices. *Journal of Finance*, 63(4): 1939–1975, 2008.

Kevin Pan and Yao Zeng. ETF arbitrage under liquidity mismatch. *Working Paper*, 2019.

Antti Petajisto. Inefficiencies in the pricing of exchange-traded funds. *Financial Analysts Journal*, 73(1):24–54, 2017.

Lawrence Schmidt, Allan Timmermann, and Russ Wermers. Runs on money market mutual funds. *American Economic Review*, 106(9):2625–2657, 2016.

Jeremy C. Stein. Monetary policy as financial stability regulation. *Quarterly Journal of Economics*, 127(1):57–95, 2012.

Karamfil Todorov. The anatomy of bond ETF arbitrage. *BIS Quarterly Review*, March 2021.

Kathy Yuan. Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. *Journal of Finance*, 60(1):379–411, 2005.

A Notation

Table 3 collects the paper's notation.

Table 3: Notation.

Fundamentals and information		*AP sector*	
$\theta,\ \varepsilon_i$	systematic and idiosyncratic payoff	n	number of dealer-APs
σ_θ	prior noise ($\tau_\theta^{-1/2}$)	$\chi,\ \chi_w$	trading-book, warehouse capacity
$\tilde{s}_h,\ \sigma_s$	holder signal and its noise	$\tilde{\chi}$	$\chi\chi_w/(\chi+\chi_w)$
$\tilde{s}_{sp},\ \sigma_{sp}$	sponsor signal and its noise	$d_o,\ \lambda_o$	dealer margin, OTC flow impact
$\tilde{\xi},\ \sigma_\xi$	noise demand and its scale	q_j, w_j, Q, w, W	redemption, warehousing, totals
$y,\ \tau_p$	price statistic and its precision	Δ_A	public wedge $\rho_y y - p$
$z,\ \hat{\tau},\ k$	holder posterior, precision, weight	L	fire-sale loss $d_H\bar{P}+D$
$\rho_y,\ b_y,\ b_s$	posterior weights on y, $\tilde{s}_{sp}$	$\bar{S}_Q$	redemption-response bound
Participation and masses		*Sponsor and closure*	
$\psi_h,\ F,\ m$	shock probability, its cdf, mass	m_{sp}	sponsor posterior mean
ψ^*	marginal type	$\bar{A},\ c_f,\ \phi$	floor, fixed cost, fee
$\Omega_B,\ \Omega_W$	institutional base, shares outstanding	$\theta_c(Q),\ \kappa_Q$	closure boundary and slope
$\kappa,\ \lambda,\ \lambda_b$	patient, shocked masses	$\mathcal{C},\ \Pi_c$	closure event, public probability
OV	option value of the $T=1$ decision	$\sigma_g,\ \sigma_{g,h}$	valuation dispersions
Markets and equilibrium		*Failure, policy, welfare*	
$P_E^1,\ p,\ \mathscr{P}(\cdot)$	wrapper price, demeaned, map	$\pi_m,\ \pi_s$	mild, severe probability
α	market-maker depth	$\theta_m^{**},\ \theta_s^{**}$	zero-noise thresholds
Λ	absorption $\alpha+\chi$	π_r^{panic}	panic-driven components
$d_H\bar{P},\ D$	household discount, fire-sale loss	$\tau_R,\ s_\chi$	redemption charge, capacity subsidy
$\Delta_0,\ g$	expected spread, liquidity gain	$\Gamma(\chi_w),\ \rho_\chi$	entry value, capacity cost
$s^*(\cdot),\ \mathcal{A}$	cutoff function, patient selling fraction	$\mathscr{W}$	welfare

B Proofs

B.1 Proof of Proposition 1

Throughout, $\Lambda = \alpha + \chi$, $\tilde{Y} = 1/\sigma_s + \chi\rho_y$, $\tilde{\chi} = \chi\chi_w/(\chi+\chi_w)$, $L = d_H\bar{P} + D$, and $\bar{\phi} = 1/\sqrt{2\pi}$. Write $\Pi_c(Q) = \Phi\big((\theta_c(Q) - \rho_y y)/\sigma_g\big)$ for the public closure probability at the state y, and

$$\bar{\Phi}_1 \ \equiv\ \sup_Q \Pi_c'(Q) \ =\ \frac{\kappa_Q\,\bar{\phi}}{\sigma_g}, \qquad \bar{\Phi}_2 \ \equiv\ \sup_Q |\Pi_c''(Q)| \ =\ \Big(\frac{\kappa_Q}{\sigma_g}\Big)^2 \phi(1),$$

using $\sup_x |x|\phi(x) = \phi(1)$. The proof proceeds in four lemmas: the information structure, the AP sector fixed point, the holder cutoff, and the assembly.

Lemma A.1 (Informational neutrality and the price map)**.** *Consider any monotone equilibrium candidate in which the holders follow a measurable cutoff function $s^*(\cdot)$ of the price and the AP sector's aggregate demand W is measurable with respect to the price. Then market clearing implies that the price is a function of the single statistic $y = \theta + \sigma_s\sigma_\xi\tilde{\xi}$ alone, with extraction precision $\tau_p = 1/(\sigma_s\sigma_\xi)^2$; the statistic and its precision are invariant to every balance-sheet and closure parameter $(\alpha, \chi, \chi_w, d_o, \lambda_o, n, \bar{A}, \Omega_W, \sigma_{sp})$. If in addition the map $H(p) \equiv \sigma_s\big[\alpha p - W(p)\big] + s^*(y^e(p))$*

is strictly increasing, the price map $\mathscr{P} = H^{-1}$ is strictly increasing, the conjectured inversion y^{e} is consistent ($y^{e} = y$), and observing the price is equivalent to observing y.

Proof. Multiply the exact clearing equation (24) by σ_s and rearrange:

$$\sigma_s\big[\alpha p - W(p)\big] + s^{*}\big(y^{e}(p)\big) \;=\; \theta + \sigma_s\sigma_\xi\,\tilde{\xi} \;\equiv\; y.$$

The left side is a function of the price alone, because the cutoff term and the AP sector's demand are price-measurable by hypothesis; the right side is the statistic. Hence $H(p) = y$ state by state, and every market-clearing price is a function of y alone, for any demand schedule the AP sector follows: a demand that responds only to the price responds to θ and $\tilde{\xi}$ in the proportion the price already carries and cannot change the ratio. No coefficient fixed point arises, and no parameter of the AP sector or the sponsor enters the statistic. If H is strictly increasing it is invertible; setting $\mathscr{P} = H^{-1}$ delivers $p = \mathscr{P}(y)$, and consistency of the conjectured inversion follows from uniqueness of the clearing price established below. The extraction precision is that of y as a signal of θ: $\mathrm{Var}(\sigma_s\sigma_\xi\tilde{\xi}) = \sigma_s^2\sigma_\xi^2$. □

Lemma A.2 (The AP sector fixed point)**.** *Fix (y, p) and eliminate w from the first-order condition* (18)*: $w(Q) = \tilde{\chi}(\Delta_A - L\,\Pi_c(Q) - Q/\chi)$, truncated at zero. Define the redemption residual*

$$R(Q) \;=\; \Delta_A - d_o - \lambda_o Q - D\,\Pi_c(Q) - \frac{Q}{n}(\lambda_o + D\,\Pi_c'(Q)) - \frac{w(Q)}{n}\,L\,\Pi_c'(Q) - \frac{Q + w(Q)}{\chi}.$$

Suppose the regularity condition

$$\frac{\tilde{\chi}}{\chi}\,L\,\bar{\Phi}_1 + \frac{1}{n}\Big[\Omega_W\,(D+L)\,\bar{\Phi}_2 + \tilde{\chi}\,L\,\bar{\Phi}_1\Big(L\,\bar{\Phi}_1 + \frac{1}{\chi}\Big)\Big] \;<\; \lambda_o\Big(1+\frac{1}{n}\Big) + \frac{1}{\chi+\chi_w}, \tag{46}$$

holds; by the share-accounting constraint $Q, w \le \Omega_W$, every constant in (46) *is a closed-form function of primitives. Then $R' \le -\rho_R < 0$ for an explicit constant ρ_R equal to the difference of the two sides of* (46)*, the equation $R(Q) = 0$ has at most one root on $[0, \Omega_W)$, the AP sector equilibrium is $Q = 0$ if $R(0) \le 0$ and the unique root otherwise, and the redemption response to the price obeys*

$$\Big|\frac{dQ}{dp}\Big| \;\le\; \bar{S}_Q \;\equiv\; \frac{1 + \tilde{\chi}(L\,\bar{\Phi}_1/n + 1/\chi)}{\rho_R}. \tag{47}$$

Proof. Differentiating, with $w'(Q) = -\tilde{\chi}(L\,\Pi_c'(Q) + 1/\chi)$,

$$R'(Q) = -\lambda_o\Big(1+\frac{1}{n}\Big) - D\,\Pi_c'\Big(1+\frac{1}{n}\Big) - \frac{Q}{n}D\,\Pi_c'' - \frac{w}{n}L\,\Pi_c'' - \frac{w'}{n}L\,\Pi_c' - \frac{1+w'}{\chi}.$$

The first two terms are negative. Using $1 + w' \ge 1 - \tilde{\chi}/\chi - \tilde{\chi}L\bar{\Phi}_1 = \chi/(\chi+\chi_w) - \tilde{\chi}L\bar{\Phi}_1$ and the bounds $\bar{\Phi}_1, \bar{\Phi}_2$ on the remaining terms, $R' \le -\rho_R$ with ρ_R the difference of the two sides of (46). Monotonicity gives uniqueness of the root and the corner characterisation; if R remains positive up to the share-accounting boundary $Q + w = \Omega_W$, the sector equilibrium is the boundary allocation, with the disposal split set by equating the two books' marginal payoffs, the response bound below applies on the interior branch, and Lemma A.4 verifies the realised response directly, boundary states included. For (47), the wedge enters R directly with coefficient one and through w with derivative $\partial w/\partial p = -\tilde{\chi}$, so $|\partial R/\partial p| \le 1 + \tilde{\chi}(L\bar{\Phi}_1/n + 1/\chi)$, and the implicit function theorem delivers the bound. At the base parameter values the amount by which (46) holds is $\rho_R = 0.493$, the primitive bound is $\bar{S}_Q = 2.79$, and the realised response never exceeds 0.93. □

Lemma A.3 (The holder root)**.** *Fix (y, p) and the AP sector solution Q. The marginal holder's indifference condition, $z = L\,\Phi\big((\theta_c(Q) - b_y y - b_s z)/\sigma_{g,h}\big) + p$, has exactly one root $z^*(y, p) \in [p,\, p+L]$, and $\partial z^*/\partial p \in (0,\, 1]$ holding Q fixed.*

Proof. The left side minus the right side is $-L\Phi(\cdot) \le 0$ at $z = p$ and $L(1 - \Phi(\cdot)) \ge 0$ at $z = p + L$, and its derivative in z is $1 + L\phi(\cdot)\, b_s/\sigma_{g,h} > 0$, which is (29): under the NAV-based rule the closure-fear term reinforces monotonicity, with no side condition. Uniqueness and the comparative static follow. □

Assembly. Fix y and consider the clearing residual

$$f(p) \;=\; -\alpha p - \frac{s^*(p)}{\sigma_s} + \frac{y}{\sigma_s} + W(p), \qquad s^*(p) = \frac{\hat{\tau}\, z^*(y,p) - \tau_p y}{\tau_s},$$

with $W(p)$ evaluated at the AP sector solution of Lemma A.2. As $p \to -\infty$, $f \to +\infty$, and as $p \to +\infty$, $f \to -\infty$, so a clearing price exists. Differentiating along the equilibrium objects,

$$f'(p) \;=\; -\alpha \;-\; \frac{\hat{\tau}}{\tau_s \sigma_s}\,\frac{dz^*}{dp} \;+\; \frac{dW}{dp}, \qquad \frac{dz^*}{dp} \;=\; \frac{1 + L\,\phi(\cdot)\,\dfrac{\kappa_Q}{\sigma_{g,h}}\,\dfrac{dQ}{dp}}{1 + L\,\phi(\cdot)\,b_s/\sigma_{g,h}}.$$

The denominator exceeds one and $dQ/dp \ge -\bar{S}_Q$, so the numerator of dz^*/dp is at least $1 - L\bar{\phi}\,\kappa_Q \bar{S}_Q/\sigma_{g,h}$. If this bound is non-negative, $dz^*/dp \ge 0$ and $f' \le -\alpha < 0$ directly; if it is negative, dividing by a denominator larger than one raises the ratio, so $dz^*/dp \ge 1 - L\bar{\phi}\,\kappa_Q \bar{S}_Q/\sigma_{g,h}$ in this case as well. Suppose in addition that the AP sector's demand slope is non-positive (Assumption T(i)),

$$\frac{dW}{dp} \;\le\; 0, \qquad \text{for which} \qquad L\,\bar{\Phi}_1 \bar{S}_Q \;<\; \frac{\tilde{\chi}}{\chi} \tag{48}$$

is a primitive sufficient condition (the worst positive contribution of the warehouse feedback is $\chi^2/(\chi + \chi_w)$); the condition is conservative, because $\bar{\Phi}_1$ and the redemption response reach their maxima in the same states, and Lemma A.4 verifies the property directly on the state grid. Then $f' < 0$ whenever

$$\frac{\hat{\tau}}{\tau_s \sigma_s}\Big[1 - \frac{L\,\kappa_Q\,\bar{S}_Q}{\sqrt{2\pi}\,\sigma_{g,h}}\Big] + \alpha \;>\; 0,$$

which is exactly condition (37). Uniqueness of the clearing price for every y follows, and with it the consistency requirement of Lemma A.1. Strict monotonicity of the price map is Assumption

T(ii): differentiating (36) along y,

$$\mathscr{P}'(y) \;=\; \frac{1}{\Lambda}\Big[\tilde{Y} - \frac{s^{*\prime}(y)}{\sigma_s} - \chi L\,\partial_y \Pi_c(y, Q(y)) - \frac{\chi}{\chi_w}\,w'(y)\Big] \;>\; 0, \tag{49}$$

an inequality in equilibrium objects; in the affine benchmark it reduces to a closed-form parameter condition, and Lemma A.4 reports the computed minimum slope at the base values. Condition (46) is condition (2) of the proposition, primitive by the share-accounting bounds; the two properties of Assumption T are condition (3), verified directly in Lemma A.4.

The remaining items of the characterisation follow directly. Averaging (36) over the symmetric distribution of y gives (38), since $E[\tilde{Y}y] = 0$; the mass parameters cancel from clearing as in (23), so they enter (38) only through the share count Ω_W in the expected closure charge and the warehousing cost. The wrapper advantage (5) is linear in ψ_h with slope $d_H\bar{P} + D\pi_s - \Delta_0 + (d_H\bar{P} + D)\pi_s - \mathrm{OV}$; under the sorting condition (iv) the slope is positive and the marginal type (6) is unique. Finally, the $T = 0$ feedback: participation determines the float, $T(\Omega_W) = \Omega_B + m(1 - F(\psi^*(\Omega_W)))$, where $\psi^*(\Omega_W)$ collects the dependence of $(\pi_s, \Delta_0, \mathrm{OV})$ on the float through the boundary θ_c. The joint equilibrium is a fixed point of T on $[\Omega_B, \Omega_B + m]$, and the contraction condition

$$\big|T'(\Omega_W)\big| \;=\; m\,f(\psi^*)\,\Big|\frac{d\psi^*}{d\Omega_W}\Big| \;<\; 1 \tag{50}$$

delivers existence and uniqueness by the contraction mapping theorem; at the base parameter values the computed slope is 0.43. ■

B.2 Proofs of corollaries

Corollary 1. The formula is (38). The derivative in α is negative because α enters only through Λ in the denominator at fixed cutoff function; the dealer margin d_o enters the numerator through the warehousing cost $(\chi/\chi_w)E[w]$, with $\partial w/\partial d_o > 0$ from the AP sector system at fixed (y, p) (raising the unwinding cost shifts disposal toward the warehouse), so the passthrough is positive. Trading capacity enters both the denominator and the closure-charge and warehousing-cost terms of the numerator; the offsetting movements are as computed in the text. The clientele statements follow from $g = d_H\bar{P} - \Delta_0$ and the participation advantage (5). ■

Corollary 2. The mild threshold $\theta_m^{**} = y_m$ solves $\mathscr{P}(y_m) = -d_H\bar{P}$; by price monotonicity, $\partial y_m/\partial\alpha$ has the sign of $-\partial\mathscr{P}/\partial\alpha$ at fixed y, which is positive at negative prices (a deeper book moves the price toward zero), so y_m falls. For the severe threshold, the zero-noise closure condition is $(b_y + b_s)\theta = \theta_c(Q(\theta))$ with the left side increasing and the right side non-increasing in θ; any parameter that raises θ_c pointwise ($\bar{A}$ up, Ω_W down) raises the crossing, and warehouse capacity lowers it through Q. The uniform cutoff shift raises both thresholds by shifting selling pressure into clearing. ■

Corollary 3. With $\tau_{sp} = 1/\sigma_{sp}^2$, $b_s = \tau_{sp}/(\tau_\theta + \tau_p + \tau_{sp})$ is decreasing in σ_{sp}, and $b_s\,\sigma_{sp} = \sigma_{sp}^{-1}/(\tau_\theta + \tau_p + \sigma_{sp}^{-2}) \to 0$ in both limits of channel (ii), so $\sigma_{g,h} \to 0$ as $\sigma_{sp} \to \infty$ and (37) fails beyond a threshold. Channel (i) is the statement that the valuation's coefficient on the fundamental, $b_y + b_s$ at zero noise, rises toward one as σ_{sp} falls, moving the closure event toward the fundamental event $\{\theta < \theta_c(Q)\}$; the monotonicity of π_s along this path is verified in our computations. ■

Corollary 4. $E[D\,\mathbf{1}] = D\pi_s$ is immediate from the lump-sum specification. The direct effect of D holds the closure probability fixed; the equilibrium response of π_s to D is not signed analytically, as the corollary states. The boundary $\theta_c(Q) = \bar{A}/\Omega_W - 1 + (\bar{A}/\Omega_W^2)Q$ is increasing in $\bar{A}$ and decreasing in Ω_W at every Q; warehouse capacity lowers Q at every state by the AP sector system. Each movement shifts Π_c pointwise in the stated direction, hence π_s. ■

B.3 Proofs for Section 6.3

Lemma 1. Conditional on $(\tilde{s}_h, y)$, the only unknown in the sponsor's valuation $m_{sp} = b_y y + b_s \tilde{s}_{sp}$ is $\tilde{s}_{sp} = \theta + \sigma_{sp}\varepsilon_{sp}$, so

$$\mathrm{Var}(m_{sp} \mid \tilde{s}_h, y) \;=\; b_s^2\left[\mathrm{Var}(\theta \mid \tilde{s}_h, y) + \sigma_{sp}^2\right] \;=\; \frac{b_s^2}{\hat{\tau}} + b_s^2\sigma_{sp}^2.$$

Both components are positive on the interior of the parameter space, and both vanish exactly when $b_s \to 0$, which occurs as $\tau_p \to \infty$ or $\tau_{sp} \to 0$: the valuation becomes price-measurable. ■

Proposition 2. (i) As $\sigma_\xi \to \infty$, $\tau_p \to 0$, so $k = \tau_p/\tau_s \to 0$, $b_y \to 0$, $b_s \to \tau_{sp}/(\tau_\theta + \tau_{sp})$, and $\hat{\tau} \to \tau_\theta + \tau_s$, giving the stated limit of $\sigma_{g,h}$, which is positive and remains so as $\sigma_s \to 0$ (only the fundamental component vanishes). The bounds $\bar{\Phi}_1, \bar{\Phi}_2$ and $\bar{S}_Q$ remain finite along the limit, so the right side of (37) is bounded away from zero and the condition holds for fire-sale losses below a strictly positive bound. (ii) As $\sigma_{sp} \to \infty$, $\sigma_{g,h} \le b_s(1/\sqrt{\hat{\tau}} + \sigma_{sp}) \to 0$ by the calculation in the proof of Corollary 3, while ρ_y, the wedge, and hence the redemption response are unchanged, so the left side of (37) is bounded away from zero and the condition fails beyond a threshold σ_{sp}. The computation exhibits multiple statewise clearing roots and a non-monotone price map ($\sigma_{sp} = 10$ at the base values), so the monotone construction fails; we do not construct the non-monotone equilibria. (iii) As $\sigma_s \to 0$ or $\sigma_\xi \to 0$, $\tau_p \to \infty$: $b_s \to 0$, so the valuation becomes price-measurable, and $\rho_y \to 1$. The marginal holder's private signal loses weight, $z^* \to y$ (from $s^* = (\hat{\tau}z^* - \tau_p y)/\tau_s$ with s^* bounded relative to y), and holder indifference $z^* = p + L\,\Phi(\cdot)$ forces $p \to y - L\,\mathbf{1}\{\text{closure believed}\}$. Where no closure is believed, $\Delta_A \to 0$, $R(0) < 0$ eventually, and $Q \to 0$: on $\{\theta > \theta_c(0)\}$ the no-run outcome is self-consistent and behaviour cannot reach the boundary. Where closure is believed, $\Delta_A \to L$, and at $Q = 0$ the marginal redemption return is $L - d_o - D = d_H\bar{P} - d_o > 0$ by the maintained dealer margin, so redemption stays active; with $\Pi_c = 1$ warehousing is dominated, and the sector's first-order condition delivers $Q \to Q^{\mathrm{run}} = (d_H\bar{P} - d_o)/(\lambda_o(1+1/n) + 1/\chi)$. On $\{\theta < \theta_c(0)\}$ closure occurs at any feasible Q; at $\theta > \theta_c(Q^{\mathrm{run}})$ it occurs under no belief. On the band between, both outcomes are self-consistent, and the equilibrium is not unique in the limit. The redemption component of failure outside the band converges to zero, as verified in our computations. ■

B.4 Proof of Proposition 3

The benchmark economy replaces the holders' condition by $z = p$ (the closure loss is ignored in decisions); the AP sector's problem, the sponsor's valuation rule, and clearing are unchanged. Under (37) both economies have unique equilibria.

(i) Fix y and compare the two scalar systems. At any price p, the equilibrium holder root satisfies $z^*(y,p) = L\,\Phi(\cdot)+p \geq p$, with strict inequality wherever the closure probability is positive, so the equilibrium cutoff exceeds the benchmark cutoff pointwise at equal prices. Clearing then forces the equilibrium price below the benchmark price: the equilibrium clearing residual $f(p)$ is the benchmark residual shifted down by $(\hat{\tau}/\tau_s\sigma_s)\,L\Phi(\cdot) \geq 0$, and both residuals are strictly decreasing, so the equilibrium root is smaller. (ii) follows from the same shift. (iii) A lower price widens the wedge and raises $Q(y)$ pointwise (Lemma A.2), and θ_c and Π_c rise with Q; integrating over states gives $\pi_r \geq \pi_r^b$ for both regimes. (iv) The shift is proportional to $L = d_H\bar{P} + D$ and vanishes as $L \to 0$, continuously by the implicit function theorem; monotonicity in L follows from the monotone comparative statics of the shifted system. ■

Proposition 4. (i) The panic components integrate the difference of closure probabilities between the equilibrium and benchmark economies, and the integrand is nonzero only where the two redemption levels differ, $Q(y) \neq Q^b(y)$. That set is a band of fixed width in y. Above an activation threshold the wedge is too narrow to activate redemption in either economy ($R(0) \leq 0$ in both, so $Q = Q^b = 0$); the two economies' thresholds differ by at most a fixed multiple of L, because the cutoff gap is bounded, $z^* - z^b \leq L$ at every state, and the price map's slope in y is bounded below on the maintained region. Below a saturation threshold the wedge is wide enough that the share-accounting constraint $Q + w \leq \Omega_W$ binds in both economies, and at the binding corner the disposal split equates the two books' marginal payoffs, from which the wedge cancels, so the corner allocation coincides across the two economies and $Q = Q^b$ exactly. The interior band between the two thresholds has fixed width in the wedge, because the wedge movement needed to carry Q from zero to the corner is bounded by a fixed function of $(\lambda_o, \Omega_W, \chi, n)$, hence fixed width in y. On the band, the pointwise difference of closure probabilities is at most $(\kappa_Q/\sigma_g)\,\bar{\phi}$ times the redemption gap, itself a fixed multiple of L (Lemma A.2); off the band it is zero. Along $\sigma_\xi \to \infty$ the band sits at a fixed location while $\mathrm{sd}(y) = O(\sigma_\xi)$, so the probability of the band, and with it each panic component, is $O(1/\sigma_\xi)$ and vanishes; the benchmark failure probabilities remain bounded away from zero because the mechanical redemption margin stays open. (ii) Along $\sigma_\xi \to 0$, outside the band of Proposition 2(iii) redemptions vanish in both economies and the closure event converges to the same valuation event in both, so the difference of closure probabilities converges to zero there; on the band the difference is bounded by one, so the panic components are bounded by the band's prior probability, $\Pr\left(\theta \in (\theta_c(0),\, \theta_c(Q^{\text{run}}))\right) = 0.003$ at the base parameter values. ■

B.5 Proof of Corollary 5 and Proposition 5

Part (i) of the corollary is Lemma A.1: the statistic, its precision, and the posterior weights (k, ρ_y, b_y, b_s) are functions of $(\sigma_\theta, \sigma_s, \sigma_\xi, \sigma_{sp})$ alone. Part (ii): at fixed (y, p), $\partial R/\partial\chi = (Q+w)/\chi^2 - (\partial w/\partial\chi)(\cdot) > 0$ on the interior region where the direct capacity effect dominates the disposal shift; with $R' < 0$, the implicit function theorem gives $dQ/d\chi > 0$.

For the proposition, write the internalisation terms of R as $-(1/n)\left[Q(\lambda_o + D\Pi_c') + wL\Pi_c'\right]$. At fixed (y, p, Q),

$$\frac{\partial R}{\partial(1/n)} \;=\; -\Big[Q\big(\lambda_o + D\,\Pi_c'(Q)\big) + w\,L\,\Pi_c'(Q)\Big] \;<\; 0 \quad \text{wherever } Q > 0,$$

so by $R' \le -\rho_R < 0$, $dQ/d(1/n) < 0$: aggregate redemption rises with n state by state. $\Pi_c(y, \cdot)$ is increasing, so the closure probability rises pointwise, and integrating over states, π_s rises at fixed price and cutoff. The full-equilibrium ordering is verified computationally along the computed path of Section 7.4. ■

B.6 Proof of Proposition 6

(i) Under a gate, $q_j \equiv 0$ for every AP, so $Q \equiv 0$ and the closure boundary is fixed at $\theta_c(0)$: the closure event is $\{m_{sp} < \theta_c(0)\}$, a valuation event that no action influences. The redemption component of π_s, defined as π_s minus the closure probability at $Q = 0$, is zero by construction. The equilibrium and benchmark economies share the same closure event, so the severe panic component (45) is zero; the mild component need not be, since closure fear shifts the cutoff and with it the discount event. With only warehouse disposal available, the first-order condition (18) with $Q = 0$ gives $W = \tilde{\chi}\,(\Delta_A - L\,\Pi_c)$, hence absorption $\Lambda_g = \alpha + \tilde{\chi} < \Lambda$. Because the redemption response is identically zero, $\bar{S}_Q = 0$ and condition (37) holds for any fire-sale loss; uniqueness follows from the assembly argument of Appendix B.1 with the closure term a fixed function of y.

(ii) The dealer margin d_o enters the residual R of Lemma A.2 only through the constant $-d_o$, so $\partial R/\partial d_o = -1$ and, by $R' \le -\rho_R < 0$, $dQ/dd_o = -1/|R'| < 0$ at every state with $Q > 0$. A flow-proportional fee enters through $-\lambda_o Q(1 + 1/n)$, so $\partial R/\partial\lambda_o = -Q(1 + 1/n) < 0$ and $dQ/d\lambda_o < 0$ likewise. In both cases the boundary $\theta_c(Q(y))$ falls state by state, Π_c falls pointwise, and π_s and the panic components fall by the comparison argument of Appendix B.4. Neither parameter appears in $\sigma_{g,h}$ or κ_Q; λ_o enters (37) only by lowering $\bar{S}_Q$ through ρ_R, so the condition does not tighten. The discount passthrough is Corollary 1. ■

B.7 Verification of the maintained parameter region

Lemma A.4 (The maintained region is non-empty, and the clientele condition binds)**.** *(P1) At the base parameter values of footnote 10, all maintained conditions of Proposition 1 hold jointly* *and strictly. The primitive conditions evaluate in closed form: the AP sector regularity* (46) *holds with* $\rho_R = 0.493$*, giving the primitive bound* $\bar{S}_Q = 2.79$*, and condition* (37) *holds in its stated primitive form as* $0.501 < 0.733$ *(evaluated at the realised redemption response instead of the bound,* $0.166 < 0.733$*). Assumption T is verified directly: the realised response is* $\sup_y |dQ/dp| = 0.93$*, the AP sector demand slope is non-positive at every state, the price map is strictly increasing with minimum slope* 0.47*, and the clearing root is unique at every state on a seven-standard-deviation grid. The sorting slope is* $+0.019$*, and the* $T = 0$ *feedback is a contraction with* $|T'| = 0.43$*. (P2) At the same values with* $d_H\bar{P}$ *lowered to* 0.10*, the positive-analysis conditions continue to hold (unique equilibrium, monotone price map), but the sorting condition fails: the slope of* (5) *is* -0.011*, the participation partition reverses, and the wrapper attracts the patient option-trading clientele. The sorting condition is therefore a substantive restriction, not a redundancy: closure risk and the household discount must be large enough relative to the option value for the wrapper to sort liquidity-demanders.*

Proof. By computation with the equilibrium solver; the state grids, tolerances, and outputs are reported in the replication files. The role of each condition is established in the proofs above. ■ □

C Robustness: the market maker's schedule

This appendix computes the two variants of the market maker's schedule flagged in Section 4.3: a quote that prices closure risk, and a plain linear quote without the matching Φ-piece. The first tests whether the results depend on the asymmetry of the reduced form; the second tests what the matching structure contributes. Solver outputs are in the replication files.

A closure-sensitive market maker prices shares on its book against the same closure exposure as warehoused shares: replace the quoted demand (12) by a quote whose level carries the warehouse closure charge, $-\alpha(P_E^1 - \mu + L_m\,\Pi_c(y, Q))$, with $L_m \in \{L/2, L\}$. The charge depends only on the public state, so the information structure is unchanged and only market clearing moves. At the full charge $L_m = L$, the destabilising term that the asymmetry discussion of Section 4.3 anticipates is present and small: the expected discount widens from 0.008 to 0.011 and the severe-failure probability rises from 0.037 to 0.038, with the panic-driven component at 0.005 against the baseline 0.005. Every maintained property and every comparative-statics sign survives. The clearing root is unique and the price map strictly increasing at every computed configuration. Trading-book capacity raises closure risk, π_s rising from 0.034 to 0.040 across the χ sweep. Concentration stabilises, 0.034 at $n = 1$ against 0.040 at $n = 10$. A floor raised by twenty percent raises π_s to 0.052. And market-maker depth remains stabilising on net, α raised by half lowering π_s to 0.035: the added charge shifts the level of the quote without overturning the absorption the quote provides. The half-charge case lies between the baseline and the full charge on every statistic.

A linear quoter replaces the mirrored form of (12) with $D_m = -\alpha(P_E^1 - \mu) + \sigma_\xi\tilde{\xi}$, with the strategic supply linearised around its mean. The derivation of Lemma A.1 goes through with one substitution: the informational weight of the marginal seller in the price falls from $1/\sigma_s$ to $\kappa\varphi(0)/\sigma_s$, where φ is the standard normal density, so the linear-quoter economy is the baseline with noise-trader volatility inflated by the factor $1/(\kappa\varphi(0)) = 67$, from $\sigma_\xi = 1.5$ to 100. The equivalence is informational: renormalising the clearing equation so the statistic's coefficient is one rescales the price and AP-demand coefficients as well, and the reported numbers come from solving the renormalised economy directly, with the AP flow entering at the observed price as in the baseline. At that point the equilibrium is unique with a monotone price map, total severe failure is 0.047, equal to the uninformative-price limit of the baseline, the panic-driven component is zero to the reported precision, and the price decouples from NAV, with the expected spread no longer near zero. The mirrored form does not manufacture the panic channel; it determines whether the price carries the information the panic channel needs. A market maker that quotes against the structure of order flow transmits the sellers' information into the price; a linear quoter facing a small flexible clientele does not, and the economy then operates at the no-learning benchmark of Section 6.3. The requirement the mechanism places on market making is therefore graded, not binary. The panic channel needs an informative price, and informativeness is a comparative static the paper computes across its whole range (Figure 1): a quote that transmits part of the information in order flow places the economy at an interior point of that range, where the paper's results are stated, and the closure-sensitive variant above is one such imperfect quote, with every sign surviving. What fails under the linear quoter is not the model but the informational premise: the economy operates at the uninformative-price corner the paper itself characterises.

The interpolation can be computed as a one-parameter family. Let the quote carry η times the structure-matching piece of (12) and $1-\eta$ times its linearisation, so η is the fraction of the strategic order-flow information the quote transmits. Both pieces enter the clearing identity linearly, so the marginal seller's informational weight is $[\eta + (1-\eta)\kappa\varphi(0)]/\sigma_s$, and the η-economy is the baseline with σ_ξ scaled to $\sigma_\xi/[\eta + (1-\eta)\kappa\varphi(0)]$; we solve each economy directly, and Table 4 reports the path. The panic channel degrades smoothly along the family, with no threshold, and is zero to the reported precision at the $\eta = 0$ endpoint, which remains weakly informative at its finite effective noise; the equilibrium is unique and the price map monotone at every computed point, and a quote transmitting a quarter of the information in order flow still supports self-fulfilling failure. What moves sharply at low η is not the mechanism but the arbitrage anchor: the expected spread changes sign near $\eta = 0.5$, and the price decouples from NAV as $\eta \to 0$, as in the linear-quoter case. The sign change is economic rather than an artifact of the interpolation. The model's primary market arbitrages the discount side only: redemption removes discounts, and no creation response removes premia. As the quote stops transmitting information, the price is set by noise demand against the market maker's quoted level, and the average spread settles at a premium that nothing in the model removes. Outputs are in the replication files.

The fragility this paper studies belongs to wrappers whose market making transmits order-flow information into the price, which the institutional evidence of Section 4.3 suggests is the empirically relevant case.

Table 4: The partial-transmission family of market-maker quotes. The quote carries η times the structure-matching piece of (12) and $1-\eta$ times its linearisation, so η is the fraction of the strategic order-flow information the quote transmits; each row is a directly solved equilibrium at the base parameter values with noise volatility $\sigma_\xi^{\text{eff}} = \sigma_\xi/[\eta + (1-\eta)\kappa\varphi(0)]$. Columns: the severe-failure probability π_s; its panic-driven component (45); and the expected wrapper-vs-NAV spread Δ_0, positive for a discount.

η	σ_ξ^{eff}	π_s	panic component	Δ_0
1.00	1.50	0.037	0.005	+0.008
0.75	1.99	0.040	0.005	+0.004
0.50	2.96	0.045	0.004	−0.001
0.25	5.74	0.048	0.001	−0.006
0.00	100.3	0.047	0.000	−0.610

D Sponsor closures in practice

The closure rule in the model is institutional: the sponsor retires the wrapper when assets persist below an internal floor, stops creations, liquidates the residual basket, and pays the remaining holders the proceeds pro-rata. This appendix documents the rule and describes three liquidations whose mechanics map directly into the model's objects.

The result in the appendix validates the institution the model builds on: sponsors maintain and exercise a closure rule; the rule is deliberative rather than a sharp trigger; and trailing net redemptions predict the closure announcement conditional on fee revenue, which a rule responding only to the current AUM level would not generate; a dynamic AUM rule with a persistence requirement could give redemptions proxy power, so we read the evidence as consistent with the continuation test rather than dispositive.

It also provides direct evidence on one mechanism prediction: the discount's predictive power for the closure announcement rises with distance to the floor (Table 6), separating the fundamental-driven closures near the floor from the closures the coordination channel can reach.

It does not test the remaining mechanism predictions: the stabilising role of AP concentration, the difference between warehoused and redeemed absorption, the dependence of the panic-driven component on price informativeness, and the discipline of independent sponsor valuation. Those tests require AP-level primary-market data and valuation-input data, and we leave them for future work.

Data sources. The fund panel is built from two CRSP files. The CRSP daily stock file supplies the ETF universe (share code 73), daily prices, shares outstanding, dollar volume, and delistings, with delisting codes 400–490 classified as liquidations and merger delistings excluded. The CRSP Mutual Fund Database, linked fund by fund, supplies daily NAV, expense ratios, and the fixed-income classification; exchange-traded notes and defined-maturity products are screened out. The resulting universe is 827 US bond ETFs over 2008–2024, of which 160 were liquidated.

Announcement dates come from SEC EDGAR: Form 497 and 497K liquidation supplements, matched for 125 of the 160 liquidations. Measured monthly creation and redemption flows come from SEC Form N-PORT, which reports each fund's gross sales and gross redemptions of its own shares monthly from mid-2019; we match 658 of the 737 universe funds alive in the reporting era. Fund-year authorized-participant identities and their gross creation and redemption volumes come from Form N-CEN, annual from fiscal 2019, matched for 661 funds. Where the measured flows overlap the panel, they validate the shares-outstanding flow measure used throughout: the fund-month correlation of the measured net-flow rate with the change in log shares outstanding is 0.84, with a regression slope of 1.10.

Industry closure counts are from Morningstar's ETF database as reported in its January 2021 review, and the threshold range from trade-press coverage of sponsor batch closures. The replication files contain the panel construction and every estimate reported in this appendix.

Closure is a routine exercise of the rule, not an exceptional event. In 2020, 54 US issuers closed a combined 277 ETFs, a record year.[15]

The floor itself is visible in the batch closures that large sponsors conduct. In December 2019 Invesco announced the closure of 42 ETFs with combined assets of \$1.13 billion, completed by February 2020; the funds ranged from \$2.4 million to \$78.8 million in assets, and industry commentary organised itself around the \$50 million level widely treated as the viability threshold, the empirical counterpart of $\bar{A}$.[16]

The rule takes an AUM form because the sponsor's revenue varies with assets while its costs do not. The sponsor's revenue from a fund is the expense ratio charged on assets, so revenue scales one for one with AUM. The costs of operating the fund are largely fixed at the fund level: exchange listing fees, index licensing, fund administration, custody, audit and legal expenses, and board and compliance costs. A fund charging 20 basis points on \$25 million of assets earns \$50,000 a year, which does not cover fixed operating costs commonly put in the several hundreds of thousands of dollars per year.

The sponsor therefore operates while fee revenue covers the fixed cost, and the viability boundary is a level of assets: this is the fixed-cost component of the continuation test, $\bar{A} = c_f/\phi$ of (22). The industry threshold range is consistent with this arithmetic once two features below are counted: part of the fixed cost is shared across the sponsor's fund line, so the incremental fund-level cost is below the stand-alone figures, and specialised funds charge fees well above 20 basis points.

The rule conditions on assets persisting below the floor rather than on the level alone, and the persistence requirement also has an economic reason. AUM fluctuates with prices and flows, and closure is irreversible and costly: the sponsor bears wind-down expenses, the remaining holders realise the liquidation value, and the closure carries a reputational cost for the rest of the sponsor's fund line. A fund below the floor therefore embodies an option: a young fund may still gather assets, and sponsors sometimes waive fees to keep the option alive. Closure follows when persistence extinguishes the option, that is, when the fund is old enough that growth is unlikely and revenue persistently fails to cover cost.

[15] Morningstar, "2020 Was a Record Year for ETF Records," January 2021.
[16] ETF.com, "Invesco Closing 42 ETFs, \$1B in Assets," December 2019.

The fee-revenue origin of the rule has two visible implications. First, the relevant state variable is revenue, the product of the expense ratio and assets, so higher-fee funds can survive at lower asset levels, and the empirical floor should vary inversely with the expense ratio across funds.

Second, part of the fixed cost is borne at the sponsor level, in administration shared across a fund line, which is why closures arrive in batches when a sponsor rationalises its offerings, as in the Invesco episode above, rather than one fund at a time. For the model, what matters is that the floor exists and binds on shares outstanding valued at the sponsor's NAV; (26) translates it into the closure boundary in the sponsor's posterior.

Three closures illustrate the mechanics, and one of them, displayed in Figure 4, can be followed week by week. The IndexIQ high-yield low-volatility ETF (HYLV) was delisted on 31 January 2023, and over the two years before delisting its assets fell from \$108 million to \$16 million. The two panels of the figure decompose the fall into the model's objects. In the left panel, shares outstanding and AUM nearly coincide along the entire path: the NAV component of the decline is small, so the fund was carried to the floor through the primary market, the model's depletion margin, not by a falling valuation. The decline is not gradual: a single redemption week, 84 weeks before delisting, removes nearly two thirds of the shares, and the later declines arrive in discrete steps, the in-kind redemption of creation baskets.

The right panel shows the price side: the small premium of the first year gives way to a persistent discount in the final year, and the redemption steps continue while the discount stands, which is the arbitrage wedge that AP redemption responds to. By the final weeks the fund held \$16 million, inside the industry threshold range, and the sponsor announced liquidation on 9 December 2022, eight weeks before delisting and inside the population's interquartile range, after the persistence the deliberative rule requires.

The episode displays the model's severe mechanism without the stress-window speed: redemptions retire shares against a standing discount, valued assets persist below the floor, and the sponsor closes.

The other two cases show two features that HYLV does not. The wind-down sequencing is visible in the Xtrackers Risk Managed USD High Yield Strategy ETF (HYRM), whose liquidation DWS announced on 17 April 2026:[17] creation orders stop and exchange trading ends shortly after the announcement, so the wrapper becomes a closed-end vehicle; redemption orders stay open while the wind-down retires the remaining shares; and holders who neither sell nor redeem receive the OTC sale value of the residual basket, the model's closure payoff.

The residual-basket cost is visible in the iShares Frontier and Select EM ETF (FM), a roughly \$400 million wrapper on frontier-market equities that BlackRock terminated on 7 June 2024, citing "persistent liquidity challenges" in the underlying markets:[18] rather than sell the basket at once, the fund spent roughly seven months liquidating, selling assets "where possible" and accumulating the proceeds in cash, with \$27.23 per share paid on 9 January 2025. The lump-sum D of the model

[17]DWS press release, "DWS Announces Liquidation of Two Xtrackers ETFs," 17 April 2026.

[18]BlackRock press release, "BlackRock Announces Product Updates," 7 June 2024; iShares Form 497 filings (2024). The asset figure at announcement is from contemporaneous press coverage.

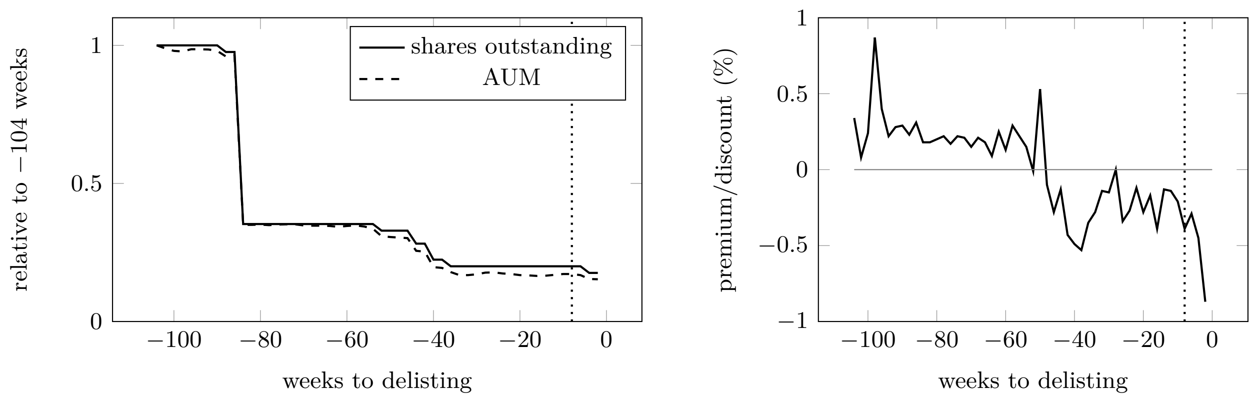

Figure 4: A bond-ETF closure in event time: the IndexIQ high-yield low-volatility ETF (HYLV), delisted 31 January 2023. Left: weekly shares outstanding (split-adjusted) and AUM, each relative to its level 104 weeks before delisting; AUM is shares outstanding times NAV. Right: the weekly premium/discount to NAV in percent, the median over traded days. Data are the CRSP daily stock file and the CRSP Mutual Fund Database NAV. The dotted vertical line in both panels marks the Form 497 liquidation announcement of 9 December 2022; the horizontal axis is weeks to delisting.

is the reduced form of this cost: whether taken in price or in time, the residual-basket sale is costly because the underlying is illiquid, and the cost falls on the holders who remain; Section 4.5 states the reduced form and what it abstracts from.

Read together with the regime mapping of Section 7.1, the three cases display the structure of the severe regime: closure is a sponsor decision under an asset floor, reached through the primary market (HYLV); the wrapper becomes closed-end at the announcement (HYRM); and the residual basket's illiquidity is the source of the terminal loss (FM). What distinguishes the acute severe event the model analyses from these orderly wind-downs is speed: a redemption wave that carries assets through the floor within the stress window forces the liquidation into the same window, which is when the dealer-capacity breach and the fire-sale loss would be incurred. The cases document closure, primary-market depletion, and costly residual-basket liquidation; the joint closure-and-capacity-breach event that defines the model's severe failure is not directly observed in them.

The population of closures displays the same anatomy. From CRSP we assemble the population of US bond-ETF closures over 2008–2024: 827 bond ETFs, of which 160 were liquidated, with announcement dates for 125 collected from Form 497 liquidation supplements on EDGAR; the median announcement precedes delisting by 54 days (interquartile range 29–72 days), the several-week institutional pattern. Writing AUM $= SO \times NAV$, the decline in assets over the two years before delisting decomposes exactly into a shares-outstanding component and a valuation component. Of the 160 closures, 88 lost more than ten percent of assets over the window, and for the median such fund the shares-outstanding component accounts for 95 percent of the decline, a median -0.69 in log shares against -0.04 in log NAV: where assets fall toward the floor, they fall through the primary market, which is the depletion margin of the model. Figure 5 plots the two components in event time for the declining closures. The remaining 72 closures show little pre-closure decline; the

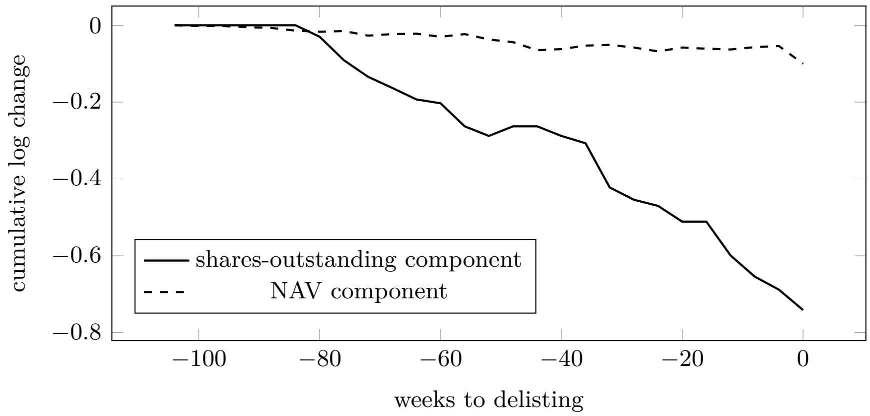

Figure 5: The AUM decomposition in event time for the declining closures. The sample is the 88 US bond-ETF closures of 2008–2024 whose assets fell by more than ten percent over the 104 weeks before delisting (CRSP daily stock file and CRSP Mutual Fund Database). Each line is the pointwise median across events of the cumulated weekly components of $\Delta \log \text{AUM} = \Delta \log SO + \Delta \log NAV$, with shares outstanding split-adjusted and the cumulation starting 104 weeks before delisting.

median closed fund entered the window at $14.5 million, already inside the threshold range, and was retired for persistent smallness, consistent with the fundamental-driven closures of the model's decomposition (the decomposition's counterfactual definition is not itself identified by event-time paths).

Event time before the announcement shows the same split. The median announced closure reaches its announcement with flat shares outstanding, the persistent-smallness pattern, while among the declining half shares outstanding fall by 0.48 log points over the year before the announcement, with dollar volume falling in step. One fact is one-sided across the whole population: only 16 percent of announced closures show any share growth over the final year. Creations stop well before closure, so the announcement arrives at a fund whose primary market is already inactive on the creation side.

The measured flows state this directly: gross creations fall below a tenth of a percent of assets per month over the final three months in 70 percent of announced closures, the announcement month carries the redemption wave, a mean redemption rate of 3.5 percent of assets against creations near 1 percent, and the median announced closure runs its final year with an inactive primary market.

A closure hazard on the fund-month panel connects the two margins. Trailing net redemptions predict the closure announcement ($z = 2.8$) conditional on the sponsor's fee revenue, and fee revenue itself is dominant ($z = -13.5$): the redemption margin operates on top of the fee-revenue economics, in the proportion the model assigns, since at the base parameter values the panic-driven component is 13 percent of severe failure.

The conditional predictive power of redemptions is the continuation test's signature: under a rule conditioning only on the current asset level, redemptions matter only through that level, and conditional on fee revenue they would add nothing; a dynamic AUM rule with persistence could give them proxy power, so the evidence is consistent with the continuation test rather than dispositive. Table 5 reports the baseline hazards across the three specifications.

On the post-2019 subsample with measured flows, the signature survives: trailing measured net outflows predict the announcement conditional on the revenue margin ($z = 2.3$, on 38 announced failures), while gross redemptions alone carry no signal, a sharpening consistent with the model, whose depletion margin is the net retirement of shares rather than gross churn.

The trailing discount predicts announcements for funds whose fee revenue still exceeds the break-even, the funds the redemption margin can still carry to the floor, and adds nothing for funds already below it, whose closure is a matter of time. A direct interaction sharpens the split: dividing the sample at twice the break-even, the discount's hazard coefficient is concentrated in the funds far above the floor ($z = 4.6$, against $z = 0.7$ near it), and a linear interaction of the discount with distance to the floor is positive ($z = 2.2$ to 2.6 across specifications); Table 6 reports these estimates, splitting the discount's effect by distance to the floor. Near the floor, the fee-revenue economics close the fund regardless of the discount, the fundamental-driven exit; the discount carries information for the funds that only the redemption and coordination channels can carry to the floor. The stress-window interaction is positive but imprecisely estimated.

The persistence requirement of the rule is also visible: at an illustrative break-even of \$100,000 per year, the median closed fund spent 16 consecutive months below break-even before announcing, and four of five funds that spent three months or more below it had not closed by the end of the sample. Sponsors wait, as the option reading above predicts.

Merger delistings, the transfer of the franchise to another operator, run in the direction the continuation test implies, although the group is too small for a test (eight events): merged funds show trailing NAV growth as flat as the closures' but growing share counts, and assets an order of magnitude larger, a median \$152 million against \$11 million. Viable franchises are transferred; liquidation is reserved for the small.

Table 5: Closure hazard on the fund-month bond-ETF panel, 2008–2024. The dependent variable is an indicator for the closure event month, estimated by logit; z-statistics in parentheses. Failure is the CRSP delist month in column (1) and the Form 497 announcement month in columns (2) and (3); fund-months after the failure month are dropped. Regressors are measured over a trailing twelve-month window ending one month before the observation month: the discount is the average of the negative part of the weekly premium/discount (in column (3), computed on traded days only); redemptions are the cumulated negative part of weekly log share growth; ln revenue is the log of average fee revenue, the expense ratio times AUM (in column (3), the fee vintage in force at each month; otherwise the latest vintage); stress is an indicator for March–June 2020 and March–October 2022; ln age is the log of the fund's age in months. The discount is in decimal units, so multiplying a discount coefficient by 0.01 gives the log-odds effect of a one-percentage-point wider trailing discount. Fund-clustered, month-clustered, and two-way clustered z-statistics, reported in the replication files, are larger than the classical ones shown for every displayed coefficient, so the reported inference is the conservative choice.

	(1) delist anchored	(2) announcement anchored	(3) announcement, refined measures
discount	−320.5	−408.4	−496.0
	(−1.5)	(−1.8)	(−2.2)
redemptions	+0.43	+0.39	+0.35
	(+4.8)	(+3.4)	(+2.8)
ln revenue	−0.76	−0.76	−0.83
	(−13.5)	(−12.3)	(−13.5)
discount × ln revenue	+32.2	+40.2	+47.4
	(+1.7)	(+2.0)	(+2.4)
stress	−0.37	+0.05	−0.01
	(−0.9)	(+0.1)	(−0.0)
discount × stress	+139.1	+76.1	+68.9
	(+1.1)	(+0.5)	(+0.5)
ln age	+0.78	+0.67	+0.68
	(+5.8)	(+4.6)	(+4.6)
fund-months	50,898	50,720	49,823
failures	126	105	105
pseudo-R^2	0.13	0.12	0.14

Table 6: The discount's predictive power by distance to the floor. Announcement-anchored closure hazards on the fund-month panel of Table 5, columns (2) and (3) conventions (plain and refined trailing measures), with the same controls (redemptions, ln revenue, stress, ln age). Distance to the floor is trailing log fee revenue minus the log of the $100,000 illustrative break-even; near is fee revenue below twice the break-even, far is above. The top panel interacts the trailing discount with the near/far indicators; the bottom panel enters the discount and its interaction with distance. z-statistics in parentheses. Failures split 95 near and 10 far. A median-revenue split gives the same pattern on the plain panel and is quasi-separated on the refined panel (one far-group failure), and is not reported. The discount is in decimal units: multiplying a coefficient by 0.01 gives the log-odds effect of a one-percentage-point wider trailing discount. Clustered z-statistics are larger than the classical ones shown for every coefficient, and a fund-block bootstrap 95 percent interval for the discount × far coefficient is $[+159, +390]$ (replication files).

	plain	refined
discount × near	+13.0	+6.8
	(+0.7)	(+0.3)
discount × far	+256.6	+240.0
	(+4.6)	(+4.6)
discount (at the floor)	+85.7	+85.8
	(+3.4)	(+3.5)
discount × distance	+43.1	+49.9
	(+2.2)	(+2.6)
fund-months	50,720	49,823
failures	105	105

Online Appendix

For online publication. Notation, equation numbers, and references are those of the paper.

OA.1 The sponsor's choice of wrapper architecture: full analysis

The paper has so far taken the wrapper's architecture as given. Sponsors choose it. The same underlying assets appear in open-end wrappers, in closed-end funds, and in vehicles that are never launched, and the choice is the sponsor's: it pays the costs, collects the fees, and selects the form. This section endogenises the choice using objects the paper has already built. The closed-end fund is the permanent-gate economy of Proposition 6(i); the demand for each architecture comes from the participation decision of Section 4.1; and the cost side is the continuation test behind the closure floor.

Add a stage before $T = 0$: a sponsor chooses among three architectures, the open-end wrapper, the closed-end fund, and no launch, paying a setup cost c_s at launch and the operating cost c_f thereafter, and collecting the fee ϕ on AUM. The closed-end fund is the economy of Proposition 6(i), with no creations or redemptions. The difference matters through the closure boundary (26): the fund reaches the floor when AUM $= (\Omega_W - Q) \times$NAV falls below $\bar{A}$, either because redemptions retire shares (Q rises) or because the sponsor's valuation falls. In the closed-end fund the share count never shrinks, $Q = 0$ at every state, so the first route is closed and the equilibrium is unique for any fire-sale loss, with absorption Λ_g. The second route stays open even here, because under the NAV-based rule the sponsor's floor applies to any fund it operates: the fund still closes when the sponsor's valuation falls low enough, with no redemptions involved. At the base parameter values the closed-end closure probability is therefore the valuation component of the wrapper's decomposition, the part generated by the sponsor's valuation alone: $\pi_s^C = 0.025$ against $\pi_s = 0.037$, so the wrapper adds 0.011 of closure risk through the redemption route. This is the cost of the arbitrage that anchors the wrapper's price to NAV. And because the sponsor's valuation moves with the price-based posterior, the closed-end fund escapes the coordination channel, not the informational one. Its price behaves differently from the wrapper's in two ways. The expected discount, 0.007 at the base values, carries the closure premium but no passthrough of unwinding or warehousing costs. And the price is untethered from NAV state by state: no arbitrage anchors it, and its dispersion is 0.46.

The demand for each architecture comes from the same participation comparison. Recall from Section 4.1 that Direct is the outside option of holding a single underlying security outright: a Direct holder who is shocked sells her bond at the household discount $d_H \bar{P}$, and bears the basket fire-sale loss D when her sale falls in the severe state, which in the closed-end economy is the valuation-driven closure of probability π_s^C. For the closed-end fund, the flexible investor's advantage over Direct mirrors (5), with the gate economy's objects in place of the wrapper's:

$$\Delta_C(\psi) \;=\; \psi\left(d_H \bar{P} + D\,\pi_s^C - \Delta_0^g\right) \;-\; (1-\psi)\left[(d_H \bar{P} + D)\,\pi_s^C - \mathrm{OV}_g\right] \;-\; \phi, \tag{51}$$

where Δ_0^g and OV_g are the expected discount and the option value in the gate economy, and the closure terms carry the closed-end fund's own closure risk: the shocked seller sells at the secondary price and escapes the closure loss, the patient holder bears $(d_H\bar{P}+D)\,\pi_s^C$ when the fund is revalued down to its floor, and the shocked Direct seller bears $D\,\pi_s^C$. At the base parameter values $\Delta_0^g = 0.007$ and $\mathrm{OV}_g = 0.128$, comparable to the wrapper's option value, and the coefficient on ψ is $+0.013$. As for the wrapper, the discount saving (0.13) and the forfeited option value (0.128) nearly offset; what remains is the closure incidence, which favours the shocked seller, who escapes the closure loss the patient holder bears. Under risk neutrality, then, the closed-end fund offered on its own against Direct attracts the same liquidity-demanding end of the population as the wrapper: the types above the crossing $\psi_C = 0.47$, giving flexible demand $A_C = m\,(1 - F(\psi_C)) = 0.16$, against the wrapper's $A_W = m(1 - F(\psi^*)) = 0.15$ at the fee's cutoff $\psi^* = 0.5$. The assignment of clienteles between the two forms is therefore not decided by the participation slope, which is small and sign-fragile; it is decided by the object risk neutrality prices at zero, the dispersion of the untethered closed-end price, 0.46 around NAV. A shocked seller forced to trade at that price bears the risk of selling far from NAV, which the wrapper's anchored price removes; any per-unit cost of that dispersion above 0.013 turns the slope negative, reverses the partition, and assigns the liquidity-demanders to the wrapper and the patient option-trading types to the closed-end fund, the observed clienteles. The computations below use the model's own risk-neutral value, $A_C = 0.16$.

Suppose the sponsor could collect fees on the flexible investors alone. The closed-end fund would then be the more valuable architecture: its clientele is no smaller, $A_C = 0.16$ against $A_W = 0.15$, it survives at the higher rate, $1-\pi_s^C$ against $1-\pi_s$, only the wrapper bears the redemption cost, and V_C exceeds V_W. The open-end wrapper, however, has a dedicated client base that the closed-end fund cannot attract, and this reverses the comparison: the institutional base of index funds, insurance portfolios, and benchmark-tracking vehicles, which hold under mandates that are contracts written on NAV. Serving such a contract requires a price tied to NAV; that tie is what the arbitrage anchor of the open-end architecture delivers, and what the closed-end price, with dispersion 0.46 around NAV and no anchor, cannot. We therefore assume the institutional base Ω_B attaches only to the open-end wrapper: the assignment reflects the mandate contracts, not bookkeeping. The sponsor's values are then

$$V_W \;=\; \phi\,(\Omega_B + A_W)(1-\pi_s) \;-\; c_f \;-\; c_Q\,E_0[Q], \qquad V_C \;=\; \phi\,A_C\,(1-\pi_s^C) \;-\; c_f,$$

where each fee base survives its own closure probability: the wrapper can reach the floor through both routes, a fall in the sponsor's valuation and the retirement of shares, while the closed-end fund can reach it only through the first. The wrapper alone bears the expected redemption cost $c_Q\,E_0[Q]$, since the closed-end fund operates no primary market; at the base parameter values the term is 0.001, about 1.5 percent of the wrapper's fee revenue.

Proposition OA.1 (Architecture choice)**.** *The sponsor chooses the open-end wrapper over the* *closed-end fund if and only if*

$$\pi_s \; < \; \bar{\pi} \; \equiv \; 1 - \frac{A_C\,(1-\pi_s^C) + c_Q\,E_0[Q]/\phi}{\Omega_B + A_W},$$

where π_s is the wrapper's severe-failure probability, the ex-ante probability that the sponsor closes the wrapper, and π_s^C is the ex-ante closure probability of the closed-end fund, and launches if and only if $\max\{V_W, V_C\} \geq c_s$. *At the base parameter values $\bar{\pi} = 0.73$ and $\pi_s = 0.037$: the wrapper is chosen. The threshold falls with the closed-end clientele A_C and with the redemption cost, and rises with the mandate base Ω_B and the wrapper clientele A_W; with no mandate base, the closed-end fund is chosen at any closure risk whenever $A_C(1-\pi_s^C) + c_Q E_0[Q]/\phi \geq A_W$.*

Proof. Immediate from the expressions for V_W and V_C. □

The proposition organises the cross-section of architectures. Assets with a large benchmark-driven demand and moderate closure risk are wrapped open-end: the sponsor accepts run risk to collect fees on the mandate base, which only the arbitrage anchor can serve. Assets without index demand and with severe fire-sale exposure are wrapped closed-end: the sponsor gives up the mandate base it could not attract anyway and serves the flexible clientele with no redemption-driven closure risk and no redemption traffic to pay for; valuation-driven closure remains, at probability π_s^C, and the closed-end fee base survives at rate $1 - \pi_s^C$. This is the leveraged-loan configuration: CLOs hold assets with thin benchmark demand and heavy fire-sale exposure, and they are closed-end. Diamond et al. (2025) document that CLOs absorbed the bulk of new leveraged-loan issuance after 2008 and did not run in March 2020 or in the 2022 leveraged-loan correction, while the open-end loan ETF on the same underlying, the Invesco Senior Loan ETF (BKLN), traded at sharp discounts as APs faced strained dealer balance sheets. The persistent coexistence of the two forms in the same underlying, institutional buy-and-hold clienteles in the CLOs and short-horizon investors in BKLN, is the coexistence the proposition predicts.[19] And when the fee base at either architecture falls short of the operating cost, the vehicle is not launched or is retired, which is the closure economics of Appendix D operating at the launch decision.

The choice of architecture also prices the run ex ante. Closure risk does not only destroy value when it strikes; anticipated, it pushes sponsors toward the closed-end form, and the closed-end form provides no anchored liquidity transformation: a shocked seller faces the untethered price, and the anchored sale price that is the wrapper's liquidity service is lost. A regime with high panic-driven closure risk therefore loses liquidity transformation through two channels, the failures that occur and the wrappers that are never launched. Two caveats bound the comparison. Risk neutrality makes the closed-end price dispersion costless to shocked sellers, and the model's secondary price does not capitalise fees, both of which flatter the closed-end fund; correcting either would push the

[19] Two roles of the closed-end CLO sit outside the model. Tranching diversifies the idiosyncratic loan exposure that a direct lender would hold, a channel on which the risk-neutral framework is silent, and the standardised senior tranches trade in dealer-intermediated markets with more depth than the underlying loans, a partial liquidity improvement even without redemption.

threshold further in the wrapper's favour and, as the clientele discussion above notes, restore the observed clientele split, so $\bar{\pi}$ is conservative.

The attachment assumption does part of the work in the comparison: the open-end form wins in part because the mandate base is assigned to it. The economic basis of the assignment is the mandate contracts stated above. The conclusion is also not knife-edge in the assignment. Suppose a fraction $\eta \in [0,1]$ of the base were willing to hold the closed-end fund, so that its fee base becomes $\eta\,\Omega_B + A_C$. The threshold becomes $\bar{\pi} = 1 - [(\eta\,\Omega_B + A_C)(1-\pi_s^C) + c_Q\,E_0[Q]/\phi]/(\Omega_B + A_W)$, and at the base parameter values the sponsor still chooses the wrapper for every η up to 0.94: the open-end form survives unless nearly all of the mandate base would migrate to an unanchored price. The content of Proposition OA.1 is the trade-off between collecting fees on the mandate base and adding the redemption route to closure, and its comparative statics hold for any attachment fraction.

OA.2 Welfare analysis

This appendix develops the normative counterpart of the positive policy analysis in Section 8: the welfare criterion, the externality wedges, and the two-instrument implementation. The treatment of the criterion, including whether market-maker and OTC-dealer surplus enter aggregate welfare, is discussed at the end of Section 8.

OA.2.1 Welfare criterion

Aggregate investor welfare under the equilibrium asset partition is the utilitarian sum of expected payoffs, weighted by the type density:

$$\mathscr{W} \;=\; m\int_0^{\psi^*} U^D(\psi)\,dF(\psi) \;+\; m\int_{\psi^*}^1 U^W(\psi)\,dF(\psi), \tag{52}$$

with cutoff ψ^* from (6) and expected payoffs

$$\begin{aligned} U^D(\psi) \;&=\; \mu - \psi\,(d_H\bar{P} + D\pi_s), \\ U^W(\psi) \;&=\; \mu - (1-\psi)\big[(d_H\bar{P} + D)\pi_s - \mathrm{OV}\big] - \psi\,\Delta_0 - \phi. \end{aligned}$$

The wrapper-holder payoff uses the equilibrium expectation $E_0[P_E^1] = \mu - \Delta_0$ (Corollary 1) for the shocked wrapper-seller mass ψ, so that the shocked-seller loss Δ_0 is the same object that governs the AP price-impact problem. Investors are risk-neutral over $T = 2$ wealth, so $U^a(\psi)$ is the expected payoff under asset a; $\mathscr{W}$ is the unweighted utilitarian aggregate and is Pareto-comparable across the type distribution. The option value OV from (4) is the value of the patient holder's $T = 1$ sell decision. Because that decision is optimal state by state, policy derivatives of OV at the $T = 1$ action margin vanish by the envelope theorem; this is what the derivative calculations below use when they discard cutoff-variation terms. The direct dependence of OV on a policy parameter through the price, the posterior, and the closure probability in the selling region remains: for each parameter it adds the term $\kappa\, E_0[\partial(\text{SELL}-\text{HOLD})/\partial\text{parameter}\cdot \mathbf{1}\{\text{sell}\}]$ to the corresponding welfare derivative. We do not sign these terms analytically; the derivative decompositions below are stated for the always-hold component of U^W, and we flag the option term where its omission could matter.

The criterion counts the two sides of the liquidity transformation: the investors here, and the AP when we turn to the capacity margin in Section OA.2.4. These are the two sides of the liquidity transformation the wrapper performs: investors demand liquidity when shocks arrive, and the AP supplies it. Noise traders, wrapper market makers, OTC dealers, and the buy-and-hold institutional base Ω_B do not enter. These agents are represented by parametric schedules, the noise-demand shock, the market maker's quoted demand (12), the OTC discount, and the mandate that holds Ω_B: they take no optimising decision and have no utility function inside the model, and a welfare statement requires preferences, so any surplus assigned to them would be an assumption imported from outside the model rather than an object the model delivers. The institutional base is the exclusion a reader may most want reversed, because the base bears the closure loss. That reversal is a known quantity and moves in one direction: the base's loss is proportional to the severe-state probability the criterion already prices, so counting it scales the severe-state externality by $(\kappa+\Omega_B)/\kappa$, moves the planner's first-order conditions, and raises the optimal corrective charge without changing any sign. The exclusion of the OTC dealers also has a substantive reading. The fire-sale discount is a price below the assets' $T=2$ value, so the sellers' loss is in part the dealers' gain, and a criterion that counted dealer surplus would treat that part as a transfer rather than a loss. This is a difference from Diamond and Dybvig (1983), where early liquidation lowers the real return on the project. Our criterion measures the surplus of the agents who demand and supply liquidity transformation, so the discount enters as the price investors pay for immediacy, whichever agent receives it. The sponsor is an optimising agent, but its surplus is fee income rather than liquidity transformation, so it too stays outside the criterion. Its inclusion would work in the same direction as the institutional base's: closure destroys the fee stream net of the operating cost, so counting the sponsor adds a loss term proportional to π_s times the fee base $\phi\,(\Omega_B+A_W)$, and because both instruments lower π_s, the term raises the optimal tax and subsidy without changing any sign. The maintained criterion is therefore the conservative one: the omissions work in the same direction, so every corrective instrument the analysis recommends would be recommended more strongly under a criterion that counted the base or the sponsor, and none would be overturned. Extending the criterion to intermediary surplus would require modelling their objectives explicitly, and we leave this open.

OA.2.2 Aggregate decomposition of the wrapper-transformation gain

The welfare gain from making the tradable wrapper available is measured against the all-Direct counterfactual in which every investor holds direct and the severe regime never arises ($\pi_s=0$, no wrapper, no sponsor closure):

$$\Delta\mathscr{W}^W \;\equiv\; \mathscr{W} \;-\; \mathscr{W}^{\rm cf}, \qquad \mathscr{W}^{\rm cf} \;=\; m\int_0^1 \left[\mu - \psi\, d_H \bar{P}\right] dF(\psi).$$

The all-Direct counterfactual has $\pi_s = 0$ as a structural consequence, not an arbitrary assumption. Each direct holder owns one security i from the continuum $[0, 1]$ and sells i alone when shocked, so per-security OTC selling is dispersed and no dealer faces concentrated flow. The wrapper architecture is what creates the possibility of basket-wide coordination: the sponsor's AUM-closure rule sells the basket across all i at once, and every OTC dealer faces concentrated flow at the same moment. The basket fire-sale loss D exists only under the wrapper architecture, because only the wrapper has a mechanism that coordinates basket-wide selling. Substituting the $T = 0$ aggregates $\lambda = m\int_{\psi^*}^1 \psi\, dF$, $\kappa = m\int_{\psi^*}^1 (1-\psi)\, dF$, and $\lambda_b = m\int_0^{\psi^*} \psi\, dF$:

$$\Delta\mathscr{W}^W \;=\; \mathcal{G}_{\rm liq} \;+\; \mathcal{G}_{\rm opt} \;-\; \mathcal{G}_{\rm fire} \;-\; m\,\phi\,(1 - F(\psi^*)), \tag{53}$$

where the channels are

$$\begin{aligned}
\mathcal{G}_{\rm liq} \;&=\; (d_H \bar{P} - \Delta_0)\,\lambda, \\
\mathcal{G}_{\rm opt} \;&=\; \kappa\,\mathrm{OV}, \\
\mathcal{G}_{\rm fire} \;&=\; \pi_s\left[(d_H \bar{P} + D)\,\kappa \;+\; D\,\lambda_b\right].
\end{aligned}$$

Liquidity transformation $\mathcal{G}_{\rm liq}$ is the wrapper-transformation gain $g = d_H \bar{P} - \Delta_0$ from Corollary 1 multiplied by the wrapper-side mechanical-flow mass λ. The gain is unconditional in π_m because the equilibrium expectation $E_0[P_E^1] = \mu - \Delta_0$ already integrates across the normal and mild-failure regimes. Fire-sale loss $\mathcal{G}_{\rm fire}$ has two components, both scaled by $\pi_s = \Pr(\text{severe})$: the total severe-state basket fire-sale loss $d_H \bar{P} + D$ paid by patient holders (mass κ), and the basket fire-sale loss D paid by shocked direct sellers (mass λ_b) when their OTC sales clear at the fire-sale price. The wrapper-side fire-sale discount on shocked wrapper sellers is already absorbed into Δ_0 and so does not appear as a separate term.

The option gain $\mathcal{G}_{\rm opt} = \kappa\,\mathrm{OV}$ is the value of the patient wrapper holders' sell option, non-negative because OV is an increment over the always-hold payoff. The wrapper raises welfare iff $\mathcal{G}_{\rm liq} + \mathcal{G}_{\rm opt} > \mathcal{G}_{\rm fire} + m\phi(1 - F(\psi^*))$. In the normal regime ($\pi_s = 0$) the fire-sale channel vanishes, so the wrapper strictly raises welfare whenever the liquidity gain $g\lambda$ alone exceeds the aggregate expense $m\phi(1 - F(\psi^*))$; the option gain only strengthens this.

OA.2.3 The externality wedge

The equilibrium cutoff function $s^*(\cdot)$ is set by pointwise marginal-type indifference (34). A planner who could shift the cutoff to maximise $\mathscr{W}$ internalises a channel that the marginal type does not. We measure the wedge along a uniform shift of the cutoff function, $s^*(\cdot) + \varepsilon$, which is the direction a per-unit redemption charge moves the equilibrium (Section OA.2.5); the derivative is evaluated at $\varepsilon = 0$. The wedge (54) below is the *investor-side* welfare gradient, $d\mathscr{W}^{\text{inv}}/d\varepsilon$; the corresponding total-surplus gradient $d\mathscr{W}^{\text{total}}/d\varepsilon = d\mathscr{W}^{\text{inv}}/d\varepsilon + \partial\Pi^{AP}/\partial\varepsilon$ adds an aggregate AP-profit term that partially offsets the wrapper-spread channel and is treated in Appendix OA.3.2. Differentiating $\mathscr{W}$ with respect to ε:

$$\begin{aligned}\frac{d\mathscr{W}}{d\varepsilon} = \quad & \underbrace{\text{(effect via marginal type)}}_{=0 \text{ pointwise at every } y \text{ by envelope}} \quad + \quad \underbrace{\left[-\,\kappa\,\frac{\partial((d_H\bar{P}+D)\ \Pr(\text{severe}))}{\partial\varepsilon}\right]}_{\text{externality on inframarginal patient holders}} \\ & + \quad \underbrace{\left[-\,\lambda\,\frac{\partial\Delta_0}{\partial\varepsilon}\right]}_{\text{externality on inframarginal shocked-wrapper sellers}} \quad + \quad \underbrace{\left[-\,\lambda_b\,D\,\frac{\partial\pi_s}{\partial\varepsilon}\right]}_{\text{externality on inframarginal shocked-direct sellers}} \quad . \end{aligned} \tag{54}$$

($\mathscr{W}$ is the $T = 0$ ex-ante expectation (52), so the shocked-seller externalities run through the realised OTC sale prices: for shocked wrapper sellers through $E[P^1_E] = \mu - \Delta_0$, and for shocked direct sellers through the OTC sale price $\mu - d_H\bar{P} - D\pi_s$.) The first term vanishes because on the equilibrium cutoff the marginal holder is indifferent at every realised y: the pointwise condition (34) makes the boundary contribution zero state by state, not only on average. The second, third, and fourth terms are all strictly negative. Raising the cutoff pushes more patient holders into the SELL region, which lowers the price at every state, widens the AP's wedge, and brings the redemption wave closer to the sponsor's floor. This has three effects. First, it raises $(d_H\bar{P} + D)\,\Pr(\text{ severe})$, which lowers the HOLD payoff for the κ-mass of patient holders who continue to hold (the second term of (54)). Second, it raises the AP unwinding flow and so widens the wrapper-vs-NAV spread Δ_0, lowering the expected liquidation proceeds received by the λ-mass of inframarginal shocked wrapper sellers (the third term of (54); these are the T=0 wrapper buyers who are hit by a liquidity shock at T=1). Third, the higher volume of strategic selling raises the severe-state probability π_s, which imposes the lump-sum fire-sale loss D on the λ_b-mass of shocked direct sellers who unwind through OTC dealers in the severe state (the fourth term of (54)). All three groups bear pecuniary losses that the marginal patient holder ignores when she crosses the indifference threshold, so the channels reinforce: the equilibrium cutoff function is too high relative to the planner's optimum, and the decentralised equilibrium has too much SELL-ing.

The wedge is the standard pecuniary externality of fire-sale models (Lorenzoni, 2008; Stein, 2012), here delivered through clearing in both the wrapper market (the second channel) and the OTC market (the third and fourth channels, via the basket fire-sale event) rather than through an explicit collateral constraint. It is non-zero whenever the severe regime has positive measure in equilibrium or whenever AP unwinding depends on the cutoff function in the mild regime, i.e., as long as wrapper-side flow responds to patient-holder selling. The AP-vs-direct flow split that determines the relative weights of the third and fourth channels is set by balance-sheet parameters, not social cost.

OA.2.4 AP capacity and the under-provision of liquidity

The externality wedge in Section OA.2.3 treats AP capacity χ as fixed and asks how the planner would shift the patient-holder cutoff function. A second welfare question is how $\mathscr{W}$ responds to χ itself, and whether the AP's private choice of χ matches the social optimum.

Aggregate investor welfare (52) depends on χ through three equilibrium objects: the wrapper-vs-NAV spread Δ_0, the severe-failure probability π_s, and the sell option OV. The decomposition below signs the first two; the option term follows the convention stated with the criterion and is not signed. Higher χ deepens absorption and, on the maintained region where absorption dominates the closure charge and the warehousing cost, lowers Δ_0: $\partial\Delta_0/\partial\chi < 0$ (Corollary 1 states the two-sided structure). Its effect on π_s combines the two channels of Corollary 2: the tighter wedge and the equilibrium feedback through the pointwise condition (34) lower π_s, while the faster redemption per unit of wedge raises the closure boundary $\theta_c(Q)$ and raises it. At the base parameter values the redemption side dominates: raising χ raises π_s (Section 7.4), so the base values lie outside the region where the wedge-narrowing effect dominates, and the net welfare effect of trading-book capacity is ambiguous there. The unambiguous capacity margin is the warehouse: raising χ_w lowers the warehousing cost, narrowing Δ_0, and lowers π_s with no offsetting depletion channel; both derivatives are negative at the base values in our computations. (Higher χ also lowers the mild-failure probability π_m, but π_m does not appear in welfare derivatives because the equilibrium expectation $E_0[P_E^1] = \mu - \Delta_0$ already integrates across the normal and mild-failure regimes.) Substituting:

$$\frac{d\mathscr{W}}{d\chi} = \underbrace{-\lambda\,\frac{\partial\Delta_0}{\partial\chi}}_{\text{tighter wrapper-vs-NAV spread}} + \underbrace{-\left[(d_H\bar{P}+D)\,\kappa + D\,\lambda_b\right]\frac{\partial\pi_s}{\partial\chi}}_{\text{lower severe-failure probability}}. \tag{55}$$

The full derivative adds the option term $\kappa\, d\mathrm{OV}/d\chi$, the direct dependence of the sell option on capacity in the selling region, which we do not sign analytically; in our computations at the base values the option term is positive in both capacity directions and equal to 0.000006 for the warehouse, under two percent of the two displayed components' sum, so the signs below are unchanged by it. Where the wedge-narrowing effect dominates, both partials are negative and both terms are positive; at the base values the severe-probability term carries the opposite sign for the trading book. The same decomposition applied to χ_w has two positive terms at the base values: the warehousing cost falls, narrowing Δ_0, and the severe probability falls with no offsetting depletion channel, so both signed components of $d\mathscr{W}/d\chi_w$ are positive. Trading capacity also enters the uniqueness condition (37), through the redemption response $\bar{S}_Q$ on the destabilising side rather than through learning: the valuation dispersion is invariant to every balance-sheet parameter, but a larger trading book raises the redemption response, so large capacity support can tighten the condition as well as moving the location and likelihood of the severe regime.[20]

[20](55) is the total derivative of (52) with the equilibrium $T=0$ partition $\psi^*(\chi)$ adjusting at each χ. Differentiating produces a sorting term $\Delta(\psi^*)\,f(\psi^*)\,d\psi^*/d\chi$ that vanishes exactly because $\Delta(\psi^*)=0$ at every χ by the marginal-

To compare the AP's private choice with this investor-welfare gradient, we augment the model with an entry stage at $T = -1$ in which each dealer-AP chooses its warehouse capacity $\chi_w/n \geq 0$ at constant marginal cost $\rho_\chi > 0$; a per-unit regulatory subsidy $s_\chi \geq 0$ lowers the effective marginal cost to $\rho_\chi - s_\chi$. We state the entry stage for the warehouse because it is the capacity margin whose welfare gradient is unambiguous at the base values; the identical envelope argument applies to the trading book. At the capacity stage each AP takes the AP sector's play and the price system as given: its own capacity affects only its own warehouse charge in (17). (The common-pool internalisation of Proposition 5 operates within the trading stage, not at the entry margin.) By the envelope theorem on the AP's $T = 1$ problem, the marginal value of own capacity is the shadow value of the warehouse charge, half the squared warehoused position per unit of capacity, and at the symmetric entry equilibrium the private first-order condition reads

$$\Gamma(\chi_w) \;\equiv\; \frac{E_{-1}\big[w^*(y;\chi_w)^2\big]}{2\chi_w^2} \;=\; \rho_\chi - s_\chi, \tag{56}$$

where $w^*(y;\chi_w)$ is the AP sector's equilibrium warehousing from Section 4.4. Γ is strictly decreasing on the maintained region: the warehoused position per unit of capacity is the ratio $\chi/(\chi + \chi_w)$ times the wedge net of the closure charge and the redemption crowd-out: the ratio falls strictly in χ_w, and on the maintained region the remaining factor does not rise, with $\Gamma(0^+) > \rho_\chi$ maintained and $\Gamma \to 0$ at large χ_w.

Lemma OA.1 (AP entry optimum)**.** *On the maintained region where* Γ *is strictly decreasing with* $\Gamma(0^+) > \rho_\chi$*, the symmetric entry equilibrium* $\chi_{eq}(s_\chi)$ *solving* (56) *exists, is unique, and is strictly increasing in* s_χ*.*

Proof. See Appendix OA.3.1. □

The AP internalises its own profit gradient but not the investor-welfare gradient $d\mathscr{W}/d\chi_w$, the χ_w version of (55), where $\mathscr{W}$ is the investor-side aggregate of (52). Writing Π^{AP} for aggregate AP entry profit, total surplus is $\mathscr{W}^{\text{inv}} + \Pi^{AP}$ (Appendix OA.3.2). Because one unit of own capacity is one unit of AP sector capacity, the aggregate profit gradient at the subsidy-free point equals the private marginal value, $\partial\Pi^{AP}/\partial\chi_w = \Gamma(\chi_w) - \rho_\chi$, so the planner's condition is $\partial\mathscr{W}^{\text{inv}}/\partial\chi_w + \Gamma(\chi_w) - \rho_\chi = 0$ while the private condition is $\Gamma(\chi_w) = \rho_\chi$: the wedge between the two is the investor-welfare gradient itself. The AP's marginal value of capacity is the shadow value of storing the arbitrage position; as χ_w rises the warehousing cost falls and the warehoused position per unit of capacity falls, so the marginal value declines along the schedule and the AP equates it with the marginal capacity cost $\rho_\chi - s_\chi$. Three components of $d\mathscr{W}/d\chi_w$ never enter this calculation. First, a tighter Δ_0 raises liquidation proceeds for every inframarginal shocked seller (the λ-mass) at the original spread, a pecuniary spillover the AP does not collect. Second, a lower π_s raises the HOLD payoff

type definition (6). Appendix OA.3.3 gives the full derivation: the sorting term vanishes, so the total derivative equals the expression (55) evaluated at the current partition, with the derivatives inside read as full equilibrium derivatives that include the induced shift in ψ^*; it does not reduce to a partial holding ψ^* fixed, because π_s and Δ_0 depend on the partition through $(\kappa, \lambda, \Omega_W)$. The same cancellation applies verbatim with χ_w in place of χ.

for the κ-mass of patient holders by reducing the expected basket fire-sale loss $(d_H \bar{P} + D)\pi_s$ they bear. Third, a lower π_s also lowers the basket fire-sale loss on direct-holder shocked sellers, who unwind through the OTC market rather than the wrapper. The AP captures the spread on its own trades; the planner counts the spread plus the regime-probability gains across both investor groups. The decentralised AP under-supplies warehouse capacity: both signed components of the gradient are positive at the base values, and in our computations the option term is positive as well, so the conclusion does not depend on leaving it unsigned.

The wedge admits a clean policy mapping. A balance-sheet transfer to dealer-APs raises effective χ_w and closes part of the gap, which is also the strain section's reading of the SMCCF, a transfer of inventory off dealer warehouses. The Federal Reserve's March 2020 Secondary Market Corporate Credit Facility is a close institutional analogue, although the SMCCF operated primarily through secondary-market purchases of bonds and bond ETFs that relieved dealer-AP inventory pressure indirectly, rather than through direct capital infusion to APs. A facility on this general design, targeting AP balance sheet more directly, is the natural reading of the model's prescription. Conversely, an AP-capital charge that raises the funding cost of χ_w moves welfare in the opposite direction, since constrained APs respond by reducing inventory. In this model the binding regulatory variable is the balance-sheet capacity available to dealer-APs, not the level of liquid reserves held inside the wrapper.

OA.2.5 Non-redundancy of the two corrective instruments

Sections OA.2.3 and OA.2.4 identify two pecuniary externalities: a run externality at the patient-holder cutoff, and an under-provision of dealer-AP warehouse capacity χ_w. The two wedges run through overlapping welfare channels. A tighter Δ_0 raises liquidation proceeds for inframarginal shocked sellers (both wedges), and a lower π_s reduces the basket fire-sale loss on patient holders and on direct-holder shocked sellers (both wedges). One might therefore ask whether the two are really independent externalities, or whether they are a single underlying spillover (the marginal trader ignores inframarginal-trader losses) showing up on two margins.

They are independent in the policy-relevant sense: a single corrective instrument cannot implement the planner's chosen allocation, but two distinct instruments can.

Proposition OA.2 (Non-redundancy of corrective instruments)**.** *Let $\tau_R \geq 0$ denote a Pigouvian charge on strategic selling, paid by patient holders who sell at $T = 1$ (distinct from the AP redemption fees of Proposition 6(ii), whose incidence is on the primary market), and let $s_\chi \geq 0$ denote a per-unit subsidy to dealer-AP warehouse capacity χ_w, the entry decision of Section OA.2.4, maintained on $0 \leq s_\chi < \rho_\chi$. The charge lowers the payoff from selling to $P_E^1 - \tau_R$, so the equilibrium cutoff function $s^*(\cdot\,; \tau_R, \chi_w)$ is strictly decreasing in τ_R at every y (pointwise, by the implicit function theorem on* (34) *under* (37)*). Let (τ_R^p, χ_w^p) denote the planner's allocation, the maximiser of $\mathscr{W}^{total}$ over the maintained instrument ranges, characterised by the joint first-order conditions $\partial \mathscr{W}^{total} / \partial \tau_R = 0$ and $\partial \mathscr{W}^{total} / \partial \chi = 0$ at an interior optimum and by the corresponding Kuhn–Tucker conditions at the tax bound $\bar{\tau}_R$, the case our computations indicate; the associated cutoff* *function is $s^p(\cdot) = s^*(\cdot\,; \tau_R^p, \chi_w^p)$. The statements below are local along the instrument-indexed equilibrium family, with the equilibrium-feedback signs verified computationally under the maintained conditions. Then:*

(i) At the planner's tax τ_R^p with $s_\chi = 0$, the implemented χ_w lies strictly below χ_w^p whenever the AP under-provision wedge in (55) *is non-zero; at any other τ_R in the maintained range at which the wedge $W_{\chi_w}(\chi_w^p; \tau_R)$ is strictly positive, the implemented χ_w differs from χ_w^p.*

(ii) For any s_χ in the maintained range with $\tau_R = 0$ at which the externality wedge of Section OA.2.3 is strictly positive, total surplus strictly rises in τ_R at the implemented equilibrium, and at the planner's capacity the implemented cutoff function lies strictly above $s^p(\cdot)$ at every y.

(iii) There exists a unique pair $(\tau_R^*, s_\chi^*) \geq 0$ that jointly implements $(s^p(\cdot), \chi_w^p)$; the implementation is of the planner's instrument-constrained target, with $\tau_R^* = \tau_R^p$ by construction and the content carried by the subsidy formula.

Proof. See Appendix OA.3.2. □

The economic content of Proposition OA.2 is that the two corrective instruments operate on opposite sides of the wrapper secondary market. A redemption tax targets the demand side: it raises the cost of selling for the marginal patient holder, lowering the cutoff function at every state and shrinking the SELL region. An AP balance-sheet subsidy targets the supply side: it expands AP absorption capacity, tightening Δ_0 and lowering π_s through the indifference-condition feedback. Neither instrument can substitute for the other because their incidence is on different equilibrium objects. Both instruments are needed at an interior optimum.

The existence of two implementing instruments is the standard two-targets-two-instruments structure: once two distinct externalities are identified, Tinbergen guarantees an implementing pair. The paper's contribution is to uncover the sources of fragility in market-based liquidity transformation. The dual role of the secondary-market price identifies which externalities arise (a spread externality and a basket fire-sale externality) and which markets they show up in (the wrapper market and the OTC market). The mapping to the SMCCF-style AP balance-sheet facility on one side and the SEC-style redemption restriction on the other interprets these externalities for policy and does not claim new instrument design.

OA.3 Proofs for the Online Appendix

OA.3.1 Proof of Lemma OA.1

Fix the environment of AP j: the price system, the others' redemptions, and the public closure probability. Its own warehouse capacity c_j ($c_j = \chi_w/n$ at symmetry) enters the objective (17) only through the warehouse charge $w_j^2/(2c_j)$. By the envelope theorem, the value V_j of the $T = 1$ problem satisfies $V_j'(c_j) = w_j^{*2}/(2c_j^2) = m_j^2/2$, with $m_j \equiv w_j^*/c_j$ the warehoused position per unit of capacity. m_j is strictly decreasing in c_j wherever $w_j^* > 0$: the w-condition gives $m_j =$

$\Delta_A - L\,\Pi_c - W_j\,n/\chi$, and, holding the closure probability and its slope at the sector values for the individual capacity perturbation (exact to first order in the AP's $1/n$ share, with the curvature terms in Π_c'' dominated on the maintained region by the same margin as (46)), the two first-order conditions form a linear system in (q_j, w_j) with strictly negative own-slopes (the dealer-bid slope λ_o and the closure-probability slope enter with negative sign), so a rise in c_j raises w_j^* less than proportionally and lowers m_j. Hence V_j is concave in own capacity on the maintained region, and the entry profit $E_{-1}[V_j(c_j)] - (\rho_\chi - s_\chi)c_j$ has a unique maximiser characterised by $E_{-1}[m_j^2]/2 = \rho_\chi - s_\chi$.

At the symmetric point $c_j = \chi_w/n$ and $w_j^* = w^*(y;\chi_w)/n$, so $m_j = w^*/\chi_w$ and the first-order condition is (56): $\Gamma(\chi_w) = \rho_\chi - s_\chi$. On the maintained region Γ is continuous and strictly decreasing with $\Gamma(0^+) > \rho_\chi - s_\chi$ and $\Gamma(\chi_w) \to 0$ as $\chi_w \to \infty$, so a solution exists and is unique on the maintained subsidy range $0 \le s_\chi < \rho_\chi$; at $s_\chi \ge \rho_\chi$ the net capacity cost is non-positive and the entry problem is unbounded without a capacity cap. By the implicit-function theorem,

$$\frac{d\chi_{w,\mathrm{eq}}}{ds_\chi} \;=\; \frac{1}{|\Gamma'(\chi_{w,\mathrm{eq}})|} \;>\; 0,$$

so $\chi_{w,\mathrm{eq}}$ is strictly increasing in s_χ. Strict concavity of each AP's entry profit in its own capacity makes the first-order condition sufficient, so the symmetric point is an equilibrium of the entry game. □

OA.3.2 Proof of Proposition OA.2

We first show how the redemption tax affects the holder's indifference condition. The derivatives in this proof are local statements along the tax-indexed equilibrium family, with the equilibrium-feedback signs verified computationally under the maintained conditions; the assembly argument of Appendix B.1 supplies the monotonicity they use. A per-unit Pigouvian charge $\tau_R \ge 0$ is paid by patient holders who sell at $T = 1$; the payoff from selling becomes $P_E^1 - \tau_R$. The pointwise indifference condition (34) becomes, at each realised y,

$$G(s^*(y;\tau_R,\chi_w);y) \;+\; \tau_R \;=\; 0,$$

where G is the hold-minus-sell difference of the pointwise condition (34), fully reduced: evaluated at the equilibrium price, redemption, and closure probability that the cutoff function induces at the realised y. Under condition (37) of Proposition 1, the reduced difference is increasing in the cutoff: the individual single-crossing (29) gives the direct term, and the assembly argument of Appendix B.1 signs the equilibrium feedback through the price and redemptions, which is what $f' < 0$ delivers. The equilibrium cutoff $s^*(y;\tau_R,\chi_w)$ is therefore uniquely determined at each y. Implicit differentiation gives

$$\frac{\partial s^*(y)}{\partial \tau_R} \;=\; -\frac{1}{\partial G/\partial s} \;<\; 0 \;\text{ at every } y, \qquad \frac{\partial s^*(y)}{\partial \chi_w} \;=\; -\frac{\partial G/\partial \chi_w}{\partial G/\partial s},$$

so a higher redemption tax strictly lowers the cutoff function pointwise. The sign of $\partial s^*/\partial\chi$ depends on $\partial G/\partial\chi$; what matters for the proof is strict pointwise monotonicity of the cutoff function in τ_R at any fixed χ.

We then compute planner's first-order conditions. There are n identical APs; aggregate AP entry profit is $\Pi^{AP}(\chi_w;\tau_R) = E_{-1}[V(\chi_w;\tau_R)] - \rho_\chi\chi$, where V is the AP sector's realised trading-stage value from (17), evaluated at the implied equilibrium cutoff $s^*(\cdot\,;\tau_R,\chi_w)$ and at $s_\chi = 0$ since the regulatory subsidy is an internal transfer and does not affect total surplus. The planner chooses the instrument pair to maximise total surplus

$$\mathscr{W}^{\text{total}}(\tau_R,\chi_w) \;\equiv\; \mathscr{W}^{\text{inv}}(\tau_R,\chi_w) \;+\; \Pi^{AP}(\chi_w;\tau_R),$$

where $\mathscr{W}^{\text{inv}}$ is the investor-side aggregate from (52), both evaluated at the implied equilibrium cutoff (the tax revenue itself is a transfer and drops out of total surplus). At an interior optimum the planner's pair $(\tau_R^{\text{p}},\chi_w^{\text{p}})$ satisfies

$$\left.\frac{\partial\mathscr{W}^{\text{total}}}{\partial\tau_R}\right|_{(\tau_R^{\text{p}},\chi_w^{\text{p}})} \;=\; 0, \tag{57}$$

$$\left.\frac{\partial\mathscr{W}^{\text{total}}}{\partial\chi_w}\right|_{(\tau_R^{\text{p}},\chi_w^{\text{p}})} \;=\; \frac{\partial\mathscr{W}^{\text{inv}}}{\partial\chi_w} + \frac{\partial\Pi^{AP}}{\partial\chi_w} \;=\; 0, \tag{58}$$

with associated cutoff function $s^{\text{p}}(\cdot) = s^*(\cdot\,;\tau_R^{\text{p}},\chi_w^{\text{p}})$. The investor-welfare gradient $\partial\mathscr{W}^{\text{inv}}/\partial\chi$ is strictly positive at every χ in the relevant range ((55): both channels are positive). The aggregate AP-profit gradient $\partial\Pi^{AP}/\partial\chi = \Gamma(\chi_w) - \rho_\chi$ (by the envelope aggregation of Appendix OA.3.1) is strictly decreasing in χ_w, positive at low χ_w and negative at high χ_w; the planner's first-order condition (58) is the equality between the two, pinning down χ_w^{p} at an interior value.

We then characterise the wedges between social and private optimisation problems. Define the AP's private marginal revenue from capacity at the implied equilibrium $(s^*(\cdot\,;\tau_R,\chi_w),\chi)$:

$$\text{MR}^{AP}(\chi;\tau_R,s_\chi) \;\equiv\; \Gamma(\chi_w;\tau_R) \;-\; (\rho_\chi - s_\chi),$$

where $\Gamma(\chi_w;\tau_R) = E_{-1}[w^*(y;\chi_w,\tau_R)^2]/(2\chi_w^2)$ is the expected squared AP sector position per unit of capacity from (56), evaluated at the tax-indexed equilibrium $(s^*(\cdot\,;\tau_R,\chi_w),\chi)$. The AP's private first-order condition is $\text{MR}^{AP} = 0$. Define the wedge for capacity as the investor-welfare gradient the AP does not internalise:

$$W_{\chi_w}(\chi;\tau_R) \;\equiv\; \left.\frac{\partial\mathscr{W}^{\text{inv}}}{\partial\chi_w}\right|_{(s^*(\cdot\,;\tau_R,\chi_w),\chi_w)},$$

which is exactly the inframarginal-trader spillovers in the χ_w version of (55): the pecuniary gain from a tighter Δ_0 accruing to inframarginal shocked sellers (λ-mass), plus the regime-probability gains accruing to patient holders and direct-holder shocked sellers (κ-mass and direct holders). The property the proof uses is local: $W_{\chi_w}(\chi_w^{\text{p}};\tau_R) > 0$ at the planner's χ_w^{p} for every τ_R in the relevant range $[0, \bar{\tau}_R]$. This holds because at the base values both terms of the χ_w version of (55) are strictly positive: the warehousing cost narrows the spread ($\partial \Delta_0 / \partial \chi_w < 0$) and the severe probability falls ($\partial \pi_s / \partial \chi_w < 0$), both verified in our computations; and in our computations the option term reinforces the wedge, $\kappa \, d\mathrm{OV}/d\tau_R = +0.015$ at the zero-tax point against the wedge 0.057, so the positivity of $W_\tau(0)$ does not depend on the omission. We make no uniform claim that $W_{\chi_w} > 0$ for arbitrarily large χ_w, where both terms vanish; we use $W_{\chi_w} > 0$ only at the planner's optimum. Using (58), the planner's first-order condition for χ_w can be written as $\mathrm{MR}^{AP}(\chi_w^{\mathrm{p}}; \tau_R^*, 0) + W_{\chi_w}(\chi_w^{\mathrm{p}}; \tau_R^*) = 0$.

Define analogously for the tax the part of the total-surplus gradient that comes from inframarginal-trader spillovers, isolated from the boundary contribution at the marginal holder:

$$W_\tau(\tau_R; s_\chi) \;\equiv\; \left.\frac{\partial \mathscr{W}^{\mathrm{total}}}{\partial \tau_R}\right|_{(\tau_R, \chi_w(s_\chi))},$$

where the boundary contribution at the marginal holder is already absent: on the equilibrium cutoff the marginal holder is indifferent at every realised y (the pointwise indifference condition $G(s^*(y); y) + \tau_R = 0$), so the boundary term vanishes state by state by the envelope condition, and the gradient consists of the inframarginal spillovers alone. Expanding $\mathscr{W}^{\mathrm{total}} = \mathscr{W}^{\mathrm{inv}} + \Pi^{AP}$ and using $\partial s^*(y)/\partial \tau_R < 0$ at every y, the wedge decomposes as

$$\begin{aligned} W_\tau \;=\; & \underbrace{-\kappa \, \frac{\partial[(d_H \bar{P} + D) \Pr(\text{severe})]}{\partial \tau_R}}_{\text{fire-sale relief for patient holders, strictly positive}} \;+\; \underbrace{-\,\lambda_b \, D \, \frac{\partial \pi_s}{\partial \tau_R}}_{\text{fire-sale relief for shocked direct sellers, strictly positive}} \\ & + \underbrace{-\,\lambda \, \frac{\partial \Delta_0}{\partial \tau_R} + \frac{\partial \Pi^{AP}}{\partial \tau_R}}_{\text{spread channel, near-cancellation}} . \end{aligned}$$

(Using $E[P_E^1] = \mu - \Delta_0$ in the $T = 0$ aggregate, the marginal-seller term collapses to $-\lambda \, \partial \Delta_0 / \partial \tau_R$, with $\partial \Delta_0 / \partial \tau_R < 0$: the tax lowers the cutoff function pointwise, hence lowers $\bar{c}^*$ and the spread.) The first two terms are strictly positive when the severe regime has positive measure: the tax lowers the cutoff, raises the price at every state, narrows the AP's wedge Δ_A, lowering redemptions and with them the closure boundary $\theta_c(Q)$, and so shrinks the closure event (26); the severe-failure probability π_s falls, patient holders bear a smaller expected loss $(d_H \bar{P} + D)\pi_s$, and shocked direct sellers (mass λ_b) bear a smaller expected lump-sum $D\pi_s$; within the criterion D is a loss, though part of it is a transfer to the excluded OTC dealers, as the criterion subsection notes. The third term is the net spread effect. By the envelope theorem on each AP's optimisation, aggregating over the AP sector, $\partial \Pi^{AP} / \partial \tau_R = E_{-1}[W^* \, \partial \Delta_A / \partial \tau_R] + R_\tau$, where W^* is the AP sector's absorption and $R_\tau \geq 0$ collects the closure-charge and dealer-bid relief terms, $-E_{-1}[(QD + wL) \, \partial \Pi_c / \partial \tau_R] - \frac{n-1}{n} \lambda_o E_{-1}[Q \, \partial Q / \partial \tau_R]$, non-negative because the tax lowers closure risk and slows redemption. The tax enters Δ_A only through the price map. Differentiating (36) along the equilibrium, the tax moves the price through three channels that share a sign at the zero-tax point: it lowers the cutoff term pointwise, and with it closure risk and warehoused inventory, so each channel raises the price and $\partial\mathscr{P}(y)/\partial\tau_R > 0$ state by state, with direct cutoff term $-(\partial s^*(y)/\partial\tau_R)/(\sigma_s\Lambda)$; the spread identity below uses only the sign. Then $\partial\Delta_A/\partial\tau_R = -\partial\mathscr{P}(y)/\partial\tau_R < 0$ and $\partial\Delta_0/\partial\tau_R = -E[\partial\mathscr{P}(y)/\partial\tau_R] < 0$. The spread channel collapses to

$$-\lambda\,\frac{\partial\Delta_0}{\partial\tau_R} + \frac{\partial\Pi^{AP}}{\partial\tau_R} \;=\; E_{-1}\Big[(\lambda - W^*)\,\frac{\partial\mathscr{P}(y)}{\partial\tau_R}\Big] \;+\; R_\tau,$$

with $\partial\mathscr{P}(y)/\partial\tau_R > 0$ state by state. The sign of the first term depends on whether $\lambda \gtreqless W^*$. With the AP sector at or near capacity in severe states, the natural ordering is $\lambda > W^*$ (mechanical-seller volume exceeds AP-absorbed volume), so the spread channel is itself non-negative and reinforces the fire-sale terms. Two properties of W_τ organise the proof. First, along the tax-indexed equilibrium family the boundary term is always absent, so the planner's first-order condition (57) at an interior optimum reads $W_\tau(\tau_R^{\mathrm{p}}; s_\chi^*) = 0$. Second, the local property the proof uses is $W_\tau(0; s_\chi) > 0$ at the zero-tax point for every s_χ in the relevant range $[0, \bar{s}_\chi]$: this is the externality wedge of Section OA.2.3 expressed in the tax direction. A sufficient condition is

$$-\Big[\kappa\,\frac{\partial[(d_H\bar{P} + D)\pi_s]}{\partial\tau_R} \;+\; \lambda_b\,D\,\frac{\partial\pi_s}{\partial\tau_R}\Big] \;>\; \max\Big\{0,\; E_{-1}\Big[(W^* - \lambda)\,\frac{\partial\mathscr{P}(y)}{\partial\tau_R}\Big]\Big\}, \tag{59}$$

evaluated at the point in question, where the maximum keeps the bound informative both when $W^* > \lambda$ on average (spread residual negative, fire-sale relief must dominate) and when $W^* \leq \lambda$ (spread channel already non-negative, the fire-sale terms alone deliver $W_\tau > 0$). We make no uniform claim that $W_\tau > 0$ for arbitrarily large τ_R, where the cutoff falls, $\pi_s \to 0$, and $W_\tau \to 0$; we use $W_\tau > 0$ only at the zero-tax point.

Numerical verification at the base parameter values. Both sign conditions are verified with the equilibrium solver at the base parameter values, by finite differences in τ_R at the zero-tax point. The tax lowers closure risk, $\partial\pi_s/\partial\tau_R = -0.044$, so the left side of (59) evaluates to $+0.0012$. The spread channel is itself non-negative: the AP sector's expected absorption is $E_{-1}[W^*] = 0.05$ against $\lambda = 0.11$, the states in which W^* exceeds λ carry 1.3 percent of the mass, and $E_{-1}[(\lambda - W^*)\,\partial\mathscr{P}/\partial\tau_R] = +0.05$, so the right side of (59) is zero and the condition holds strictly. The zero-tax wedge evaluates to $W_\tau(0) = 0.057$. For capacity, with ρ_χ set so that the zero-subsidy entry equilibrium is the base value $\chi_w = 0.75$, total surplus is strictly concave in χ_w on the examined range, with negative second differences throughout and an interior planner's capacity between 1.5 and 2 against the equilibrium value 0.75: the second-order condition for capacity holds and the under-provision result is an interior comparison. For the tax, total surplus continues to rise in τ_R over the examined range: beyond its effect on closure risk, the tax raises the equilibrium price and transfers surplus from the market maker, which is outside the criterion, to sellers inside it. The planner's tax therefore sits at the instrument bound $\bar{\tau}_R$ rather than at an interior point unless market-maker surplus enters the criterion; the proofs below use only the zero-tax wedge $W_\tau(0) > 0$ and the interior capacity comparison, both verified.

Proof of part (i). Set $s_\chi = 0$ and fix $\tau_R = \tau_R^{\mathrm{p}}$, the planner's redemption tax. The AP's first-order condition pins down χ_w via $\mathrm{MR}^{AP}(\chi; \tau_R^{\mathrm{p}}, 0) = 0$, with a unique solution $\chi_{w,\mathrm{eq}}(\tau_R^{\mathrm{p}}, 0)$ by strict monotonicity of Γ (Lemma OA.1). Suppose for contradiction $\chi_{w,\mathrm{eq}}(\tau_R^{\mathrm{p}}, 0) = \chi_w^{\mathrm{p}}$. Then $\mathrm{MR}^{AP}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}, 0) = 0$. By the planner's first-order condition at the planner's optimum, $\mathrm{MR}^{AP}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}, 0) + W_{\chi_w}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}) = 0$, so this forces $W_{\chi_w}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}) = 0$, contradicting the local strict-positivity claim at χ_w^{p}. So $\chi_{w,\mathrm{eq}}(\tau_R^{\mathrm{p}}, 0) \neq \chi_w^{\mathrm{p}}$. The argument extends to any $\tau_R \in [0, \bar{\tau}_R]$ for which $W_{\chi_w}(\chi^{\mathrm{p}}; \tau_R) > 0$ (the maintained local positivity at χ^{p}): the same contradiction goes through with τ_R^{p} replaced by τ_R.

For the direction of the inequality, evaluate MR^{AP} at χ_w^{p}. From the planner's first-order condition at $(\tau_R^{\mathrm{p}}, \chi^{\mathrm{p}})$, $\mathrm{MR}^{AP}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}, 0) = -W_{\chi_w}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}) < 0$. At the planner's tax τ_R^{p}, the AP would earn negative marginal revenue at χ_w^{p}: each additional unit of capacity at χ_w^{p} costs the AP more than its private return, because the social return to capacity exceeds the private return by exactly W_{χ_w}. Because Γ is strictly decreasing in χ, the AP's privately optimal χ satisfies $\chi_{w,\mathrm{eq}}(\tau_R^{\mathrm{p}}, 0) < \chi_w^{\mathrm{p}}$. At other values of τ_R the planner's first-order condition does not pin the sign of MR^{AP} at χ_w^{p}, so the proposition claims only $\chi_{w,\mathrm{eq}}(\tau_R, 0) \neq \chi_w^{\mathrm{p}}$ there, from the contradiction argument above; along the computed range the direction agrees with the planner's-tax case. □

Proof of part (ii). Set $\tau_R = 0$ and fix any $s_\chi \geq 0$ with implied capacity $\chi_w(s_\chi)$. The equilibrium cutoff is $s^*(\cdot\,; 0, \chi_w(s_\chi))$, unique state by state under Proposition 1. Along the tax-indexed equilibrium family the boundary term vanishes pointwise (the marginal holder is indifferent at every realised y), so

$$\left.\frac{d\mathscr{W}^{\mathrm{total}}}{d\tau_R}\right|_{\tau_R=0} \;=\; W_\tau(0; s_\chi) \;>\; 0$$

whenever the externality wedge is non-zero at the zero-tax point (the local strict positivity guaranteed by (59) there). A small positive redemption tax therefore strictly raises total surplus, so $\tau_R = 0$ is not optimal for any s_χ. In particular the planner's optimum has $\tau_R^{\mathrm{p}} > 0$, whether it is interior, where (57) requires $W_\tau(\tau_R^{\mathrm{p}}; s_\chi^*) = 0$ and this fails at $\tau_R = 0$, or at the instrument bound $\bar{\tau}_R$, the case the numerical verification above indicates at the base values. Finally, at the planner's capacity χ_w^{p}, the cutoff family is strictly decreasing in τ_R pointwise, so

$$s^*(y; 0, \chi_w^{\mathrm{p}}) \;>\; s^*(y; \tau_R^{\mathrm{p}}, \chi_w^{\mathrm{p}}) \;=\; s^{\mathrm{p}}(y) \quad \text{at every } y. \quad \square$$

Proof of part (iii). The implementation system has two conditions in two unknowns:

$$G(s^*(y); y) \;+\; \tau_R \;=\; 0 \quad \text{at every } y, \tag{60}$$

$$\mathrm{MR}^{AP}(\chi; \tau_R, s_\chi) \;=\; 0, \tag{61}$$

with target $(s^{\mathrm{p}}(\cdot), \chi_w^{\mathrm{p}})$. The pointwise indifference condition (60) is satisfied at the target by $\tau_R = \tau_R^{\mathrm{p}}$, since $s^{\mathrm{p}}(\cdot) = s^*(\cdot\,; \tau_R^{\mathrm{p}}, \chi_w^{\mathrm{p}})$ is by construction the equilibrium cutoff at the planner's instrument pair; and τ_R^{p} is the unique tax implementing $s^{\mathrm{p}}(\cdot)$ at χ_w^{p}, because the cutoff family is strictly decreasing in τ_R pointwise (two different taxes produce different cutoff functions). It is strictly positive by part (ii). Substituting τ_R^{p} into (61) at $\chi = \chi_w^{\mathrm{p}}$:

$$s_\chi^* \;\equiv\; \rho_\chi \;-\; \Gamma(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}) \;=\; W_{\chi_w}(\chi_w^{\mathrm{p}}; \tau_R^{\mathrm{p}}) \;>\; 0,$$

where the second equality uses the planner's first-order condition (58): $\Gamma = \rho_\chi - W_{\chi_w}$ at the planner's optimum. The subsidy equals the marginal inframarginal-trader spillover of capacity at the planner's optimum, the textbook Pigouvian formula; because each AP's unit of capacity is a unit of AP sector capacity, no scaling by the number of APs appears. Positivity follows from $W_{\chi_w} > 0$.

Uniqueness: $\tau_R^* = \tau_R^{\mathrm{p}}$ is the unique tax implementing the target cutoff, as above. Given τ_R^{p}, s_χ^* is the unique value satisfying (61) at χ_w^{p} because Γ is strictly decreasing in χ_w (Lemma OA.1). Both values are strictly positive.

Implementation verification: at (τ_R^*, s_χ^*), the AP's first-order condition (61) is satisfied at χ_w^{p}, so the AP chooses χ_w^{p}. Given χ_w^{p}, the pointwise indifference condition (60) delivers the cutoff $s^{\mathrm{p}}(\cdot)$. Both planner first-order conditions are satisfied by definition of $(\tau_R^{\mathrm{p}}, \chi_w^{\mathrm{p}})$. □

OA.3.3 Derivation of $d\mathscr{W}/d\chi$: sorting-term cancellation

This appendix derives (55) from the investor-welfare aggregate (52), showing that the $T = 0$ sorting term vanishes exactly and that the total derivative equals the integral expression evaluated at the current partition, with the derivatives inside read as full equilibrium derivatives.

We first compute the total derivative. Starting from (52),

$$\mathscr{W} \;=\; m \int_0^{\psi^*} U^D(\psi)\, dF(\psi) \;+\; m \int_{\psi^*}^1 U^W(\psi)\, dF(\psi),$$

with $U^D(\psi) = \mu - \psi(d_H \bar{P} + D\pi_s)$ depending on χ through π_s, and $U^W(\psi) = \mu - (1-\psi)\big[(d_H \bar{P} + D)\pi_s - \mathrm{OV}\big] - \psi\Delta_0 - \phi$, the full wrapper payoff of (52), depending on χ through π_s, Δ_0, and OV. The $T = 1$ sell decision inside OV is optimal state by state, so it contributes no derivative term by the envelope theorem; the direct dependence of OV on χ through the price, the posterior, and the closure probability in the selling region remains, and it appears below as an option term that we do not sign. The total derivative with respect to χ contains four pieces:

$$\frac{d\mathscr{W}}{d\chi} \;=\; \underbrace{-U^D(\psi^*)\, f(\psi^*) \frac{d\psi^*}{d\chi}}_{\text{lower limit moves}} \;+\; \underbrace{U^W(\psi^*)\, f(\psi^*) \frac{d\psi^*}{d\chi}}_{\text{upper limit moves}} \;+\; \underbrace{\int_0^{\psi^*} \frac{\partial U^D}{\partial \chi}\, dF(\psi)}_{\text{Direct integrand}} \;+\; \underbrace{\int_{\psi^*}^1 \frac{\partial U^W}{\partial \chi}\, dF(\psi)}_{\text{Wrapper integrand}}. \tag{62}$$

The first two pieces combine to

$$\big[U^W(\psi^*) - U^D(\psi^*)\big]\, f(\psi^*) \frac{d\psi^*}{d\chi} \;=\; \Delta(\psi^*)\, f(\psi^*) \frac{d\psi^*}{d\chi},$$

which is the $T = 0$ sorting term.

We now show sorting-term cancellation. The marginal type ψ^* is defined by the indifference $\Delta(\psi^*) = 0$ in (6). This identity is algebraic, so it holds at every χ in the relevant range, not as a local linearisation. The cancellation is exact and global:

$$\Delta(\psi^*)\, f(\psi^*) \frac{d\psi^*}{d\chi} \;=\; 0 \cdot f(\psi^*) \frac{d\psi^*}{d\chi} \;=\; 0.$$

The cancellation does not require ψ^* to be invariant in χ; $d\psi^*/d\chi$ is in general non-zero. Nor does it require (π_m, π_s, Δ_0) to be invariant to the $T = 0$ partition; through aggregate consistency $\kappa = \kappa(\psi^*)$, these equilibrium objects do depend on ψ^*, and the comparative statics of π_s, π_m, Δ_0 in χ in the integrand term below allow the cutoff function to re-equilibrate via (34). What delivers the cancellation is only that $\Delta(\psi^*) = 0$ is an algebraic identity in χ, so the boundary contribution at ψ^* to the welfare derivative carries a zero multiplier.

The Wrapper integrand uses

$$\frac{\partial U^W(\psi)}{\partial \chi} \;=\; -(d_H \bar{P} + D)(1-\psi)\frac{\partial \pi_s}{\partial \chi} \;-\; \psi\, \frac{\partial \Delta_0}{\partial \chi} \;+\; (1-\psi)\, \frac{\partial \mathrm{OV}}{\partial \chi}.$$

Integrating over $\psi \in [\psi^*, 1]$ with $\lambda = \int_{\psi^*}^1 \psi\, dF$ and $\kappa = \int_{\psi^*}^1 (1-\psi)\, dF$:

$$\int_{\psi^*}^1 \frac{\partial U^W}{\partial \chi}\, dF(\psi) \;=\; -(d_H \bar{P} + D)\kappa\, \frac{\partial \pi_s}{\partial \chi} \;-\; \lambda\, \frac{\partial \Delta_0}{\partial \chi} \;+\; \kappa\, \frac{\partial \mathrm{OV}}{\partial \chi}.$$

The Direct integrand uses

$$\frac{\partial U^D(\psi)}{\partial \chi} \;=\; -\psi D\, \frac{\partial \pi_s}{\partial \chi},$$

and integrating over $\psi \in [0, \psi^*]$ with $\lambda_b = \int_0^{\psi^*} \psi\, dF$:

$$\int_0^{\psi^*} \frac{\partial U^D}{\partial \chi}\, dF(\psi) \;=\; -\lambda_b\, D\, \frac{\partial \pi_s}{\partial \chi}.$$

Combining the Wrapper and Direct integrands and collecting the $\partial \pi_s/\partial \chi$ terms gives $(d_H \bar{P} + D)\kappa + \lambda_b D$, and the resulting components are the two signed terms of (55) plus the unsigned option term $\kappa\, \partial \mathrm{OV}/\partial \chi$. The χ-derivatives $\partial \Delta_0/\partial \chi$ and $\partial \pi_s/\partial \chi$ here are total derivatives in χ that allow the cutoff function to re-equilibrate via (34) and that allow the partition ψ^* to shift. The sorting-term argument is unaffected by either source of feedback because the boundary contribution carries the algebraic zero $\Delta(\psi^*) = 0$ as a multiplier. □

OA.4 Robustness: comparative-statics signs

This appendix recomputes the three directional claims of Section 8 away from the base parameter values: that trading-book capacity raises closure risk, that concentration among APs lowers it, and that a lower sponsor floor lowers it. At each point of two parameter planes, the loss plane $(d_H\bar{P}, D)$ and the information plane $(\sigma_\theta, \sigma_\xi)$, with all other parameters at the base values, we solve six equilibria and report the three differences

$$\Delta_\chi = \pi_s(\chi{=}2.5) - \pi_s(\chi{=}0.75), \quad \Delta_n = \pi_s(n{=}10) - \pi_s(n{=}1), \quad \Delta_{\bar{A}} = \pi_s(1.2\,\bar{A}) - \pi_s(\bar{A}),$$

where the capacity comparison holds $\chi_w = \chi/2$ as in Section 7.4. A positive Δ_χ is the capacity result, a positive Δ_n is the concentration result (fragmentation raises closure risk), and a positive $\Delta_{\bar{A}}$ is the floor result (a facility that lowers the floor lowers closure risk). Table 7 reports the maps.

No difference is negative at any computed point: all three are strictly positive away from the nearly-perfect-price limit, and the two redemption-side differences are zero to the reported precision at that limit. The clearing root is unique at every state in all six equilibria at every point. On the loss plane both dealer-side effects grow with the household discount; in the fire-sale loss they move in opposite directions, the concentration effect growing and the capacity effect shrinking as D rises, with both strictly positive throughout. On the information plane the capacity and concentration effects converge to zero from above as the price becomes nearly perfect ($\sigma_\xi = 0.5$ with $\sigma_\theta \geq 0.5$): a nearly perfect price closes the redemption margin, so the effects that operate through redemption vanish rather than reverse, which is the limit of Proposition 4. The floor effect is positive throughout, including where the redemption-side effects vanish, because the floor operates on the valuation route as well.

Table 7: Sign maps for the three directional claims of Section 8. Each row is one point of a parameter plane, with all other parameters at the base values of Table 1. π_s is the severe-failure probability at the point. $\Delta_\chi = \pi_s(\chi{=}2.5, \chi_w{=}1.25) - \pi_s(\chi{=}0.75, \chi_w{=}0.375)$; $\Delta_n = \pi_s(n{=}10) - \pi_s(n{=}1)$; $\Delta_{\bar{A}} = \pi_s(1.2\,\bar{A}) - \pi_s(\bar{A})$. The clearing root is unique at every state in all six equilibria at every row.

Point	π_s	Δ_χ	Δ_n	$\Delta_{\bar{A}}$
Loss plane $(d_H\bar{P}, D)$				
$(0.08, 0.15)$	0.036	+0.0060	+0.0036	+0.0132
$(0.08, 0.30)$	0.036	+0.0054	+0.0043	+0.0129
$(0.08, 0.45)$	0.036	+0.0048	+0.0050	+0.0126
$(0.10, 0.15)$	0.037	+0.0064	+0.0038	+0.0135
$(0.10, 0.30)$	0.036	+0.0057	+0.0045	+0.0131
$(0.10, 0.45)$	0.036	+0.0050	+0.0052	+0.0128
$(0.13, 0.15)$	0.037	+0.0068	+0.0040	+0.0139
$(0.13, 0.30)$	0.037	+0.0061	+0.0049	+0.0135
$(0.13, 0.45)$	0.037	+0.0054	+0.0055	+0.0131
$(0.16, 0.15)$	0.038	+0.0074	+0.0043	+0.0144
$(0.16, 0.30)$	0.037	+0.0066	+0.0052	+0.0139
$(0.16, 0.45)$	0.037	+0.0058	+0.0059	+0.0135
$(0.20, 0.15)$	0.039	+0.0081	+0.0047	+0.0150
$(0.20, 0.30)$	0.038	+0.0072	+0.0056	+0.0145
$(0.20, 0.45)$	0.038	+0.0064	+0.0063	+0.0140
Information plane $(\sigma_\theta, \sigma_\xi)$				
$(0.35, 0.5)$	0.006	+0.0001	+0.0001	+0.0035
$(0.35, 1.5)$	0.002	+0.0011	+0.0007	+0.0021
$(0.35, 5.0)$	0.003	+0.0013	+0.0005	+0.0064
$(0.50, 0.5)$	0.050	+0.0000	+0.0000	+0.0124
$(0.50, 1.5)$	0.037	+0.0061	+0.0049	+0.0135
$(0.50, 5.0)$	0.047	+0.0067	+0.0034	+0.0250
$(0.70, 0.5)$	0.130	+0.0000	+0.0000	+0.0166
$(0.70, 1.5)$	0.117	+0.0052	+0.0035	+0.0186
$(0.70, 5.0)$	0.141	+0.0077	+0.0044	+0.0301