Market-Based Liquidity Transformation

Kathy Yuan

May 20, 2026

Abstract

We develop a theory of liquidity transformation in which the secondary-market price of a tradable wrapper (corporate bond ETFs, CLO ETFs, and stablecoins) plays a dual role: it clears the wrapper market against dealer-authorised- participant (AP) arbitrage, and serves as the public signal on which long-term holders condition their hold-or-sell decision and the AP conditions its balance-sheet commitment. Uniqueness of equilibrium depends on wrapper-side absorption capacity as well as on private-signal precision. Adverse shocks generate a *mild* regime with a sustained wrapper discount or a *severe* regime in which the sponsor's AUM rule triggers basket liquidation at fire-sale prices. The tradable wrapper produces two pecuniary externalities, addressable by a Pigouvian redemption tax and an AP balance-sheet subsidy. Fragility runs through dealer balance sheets, so regulation should target balance-sheet capacity rather than reserve adequacy.

1 Introduction

The financial system is transitioning from bank-based to market-based intermediation. 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. 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 (exchange-traded funds, collateralised loan obligations, dealer portfolio trades, stablecoins) 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 insurance 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 the first model of liquidity transformation in which the secondary-market price does what no bank deposit can: it aggregates dispersed private information from investors and from the dealer-AP, and serves as the public signal that long-term holders use to decide whether to hold or to sell, and that the dealer-AP uses to decide how much balance-sheet capacity to commit to wrapper arbitrage. Bank-deposit architectures have no market price. Existing market-based architectures give the price a clearing role but not an informational one. We let the same price perform both at once.

These wrappers share a common institutional architecture. Each is created and operated by a *sponsor*: a fund manager such as BlackRock, State Street, or Invesco in the ETF context, the arranger for a collateralised loan obligation, the issuer for a stablecoin. The sponsor defines the underlying basket, contracts with one or more APs to handle primary-market creation and redemption, collects management fees, and retains the right to close the wrapper when assets-under-management fall below a threshold. The wrapper itself is market-priced rather than redeemable at par; redemption is in-kind through APs rather than first-come-first-served sequential service; the basket of underlying securities is sold into the OTC market when the sponsor invokes the closure rule, and depressed OTC prices feed back into the wrapper's NAV. The most familiar examples are ETFs wrapped around illiquid securities: corporate bond ETFs (LQD 2002, HYG and JNK 2007); gold ETFs (GLD 2004); broad commodity-index ETFs (DBC, GSG); bank-loan ETFs (the Invesco Senior Loan ETF, ticker BKLN, launched 2011); cryptocurrency spot ETFs (IBIT, FBTC, 2024); collateralised-loan-obligation ETFs (JAAA, CLOZ, 2020–2024); and a growing pipeline of private-credit ETFs. Outside the ETF wrapper the same architecture appears in dealer portfolio trading of corporate bonds (Li et al., 2025), in collateralised loan obligations (Diamond et al., 2025), and in stablecoins (Ma et al., 2024). In each case an authorised-participant or arranger runs an inventory book linking the wrapper secondary market to the underlying primary market, absorbing flow imbalances on its own balance sheet.

Three episodes motivate the wrapper-failure structure we model. Petajisto (2017) and Pan and Zeng (2019) document that ETFs trade at premiums and discounts to NAV that are larger and more persistent than the AP-arbitrage story alone predicts. In March 2020, LQD traded at a 5% discount and HYG at 7–8% as APs stopped fully arbitraging because of widened bid-ask spreads on individual bonds, strained dealer balance sheets, and de facto more expensive creation/redemption; the Fed's Secondary Market Corporate Credit Facility (SMCCF) closed the discounts within weeks (Falato et al., 2021; Haddad et al., 2021; O'Hara and Zhou, 2021). The Grayscale Bitcoin Trust (GBTC) sustained-discount episode of 2021–2023 illustrates the limiting case: GBTC traded at 20–50% discounts because it was a closed-end trust without AP creation/redemption ($\chi = 0$ in the model's notation), and the discount only closed upon spot-ETF conversion in January 2024, when in-kind creation/redemption was enabled. The same closure-and-fire-sale pattern appears outside the ETF wrapper in the May 2022 collapse of TerraUSD, whose collateral was liquidated at depressed prices. These episodes indicate two regimes of wrapper failure: a *mild* regime with sustained discount but basket intact (GBTC), and a *severe* regime with sponsor liquidation and basket fire-sale (March 2020-style stress and TerraUSD).

The severe regime mechanism is sponsor-driven liquidation. Each wrapper has a finite economic lifespan: the sponsor maintains an internal assets-under-management cutoff and closes the wrapper when assets persist below the threshold. Closure stops creations, the basket is sold through the AP (or the arranger or issuer for non-ETF wrappers), and proceeds are distributed pro-rata. Closure activity is empirically concentrated in small and specialised wrappers that sit close to the cutoff: US ETF closures run at roughly 100–300 per year against a universe of about 3,500 funds (a 3–8% annual closure rate, much higher than for mutual funds), with sponsor AUM thresholds typically in the \$25M–\$100M range 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 basket fire-sale loss.

Based on these empirical observations, we propose a model of liquidity wrappers and study their sources of 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 a liquid basket. By the law of large numbers the basket 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 long-term 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 dealer-AP arbitrages the wrapper-vs-NAV discount by buying wrapper shares and unwinding the basket in the OTC market for the underlying securities.[1] It performs liquidity transformation by exploiting its market-making position in both the illiquid underlying securities and the liquid wrapper. However, AP supply of liquidity to the wrapper is constrained by its finite balance-sheet capacity, which is shared across arbitrage and OTC inventory-making. Two regimes of wrapper failure emerge endogenously. In the *mild* regime, the wrapper trades at a sustained discount because the AP withdraws from arbitrage when the price under-reacts to fundamentals; the basket is intact, so $T = 1$ sellers absorb the loss through the discount and patient holders receive NAV at $T = 2$. The GBTC episode is the canonical case. In the *severe* regime, holder selling depletes wrapper AUM below the sponsor's closure threshold; the wrapper is liquidated, the residual basket is dumped into the OTC bond market through APs, and combined OTC flow exceeds dealer capacity. The resulting fire-sale impairs the basket by an amount D that patient holders bear. Bond-ETF closures and the March 2020 episode are the empirical analogue. The two regimes share a common equilibrium structure but differ in who bears the loss: $T = 1$ sellers in mild, patient holders in severe. 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. The same price is also a public signal: long-term holders observe it before they choose between holding to maturity and selling at $T = 1$, and the dealer-AP observes it before sizing the balance-sheet capacity it commits to wrapper arbitrage. The two roles can pull in opposite directions. When wrapper-side absorption is small, each unit of strategic redemption has a large price impact, so a holder is penalised for joining the selling. Substitution is the dominant force and the equilibrium is unique. When wrapper-side absorption is large, the price flattens against selling and the substitution penalty is weak; the price then moves predominantly with fundamentals, and a low observed price signals weak fundamentals, inducing long-term holders to sell and the dealer-AP to pull back its arbitrage commitment. The information channel can dominate and wrapper runs become self-fulfilling. The same equilibrium object governs both effects, and the balance between them determines whether the equilibrium is unique. The uniqueness condition depends on wrapper-side absorption capacity including market-maker depth and aggregate AP balance-sheet capacity, in addition to the precision of private information; the absorption-capacity margin is what is new relative to standard global-games models.

The dual role of the secondary-market price is the structural feature that distinguishes market-based liquidity transformation from its bank-based predecessor. In Diamond and Dybvig (1983) and Goldstein and Pauzner (2005), no market price exists: bank depositors coordinate through the sequential-service rule, and the resource that runs deplete is the intermediary's liquid pool. In Pan and Zeng (2019), the wrapper price functions only as a clearing object that AP arbitrage discipline pins to a band; long-term holders are not modelled and do not extract information from the price. Ozdenoren and Yuan (2008) formalise the dual role of price in a speculative-attack setting but the wrapper-arbitrage architecture is absent. We integrate the two by letting the same wrapper price clear AP arbitrage demand against mechanical and strategic selling, and act as the public signal on which both long-term holders condition their hold-or-sell decision and the dealer-

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

AP conditions its balance-sheet commitment. The contribution is the coupling: the substitution-versus-information uniqueness condition is a function of AP balance-sheet capacity, which neither existing literature can deliver. Within the unique-equilibrium region, AP warehousing capacity is unambiguously stabilising: it narrows the level of the wrapper-vs-NAV spread in the normal and mild regimes and shrinks the severe regime through the equilibrium run threshold. The substitution-versus-information condition determines the boundary of this region. The fragility margin in the severe regime is the sponsor's AUM-closure rule, not OTC dealer capacity *per se*: closure-driven basket dumps overwhelm OTC capacity once the closure condition is met. The framework applies to any non-bank intermediation that operates through a continuously-priced wrapper backed by illiquid assets: bond ETFs, gold and broad-commodity ETFs, bank-loan ETFs, crypto spot ETFs, closed-end collateralised loan obligations, and stablecoins. Bond ETFs are the cleanest empirical application because the data have the richest cross-section, but the dual-role mechanism is general.

Our comparative statics deliver several empirically testable hypotheses, including how bond-level shorting and ETF shorting affect bond prices and ETF discounts. We also characterise welfare and the regulatory implications. The wrapper-transformation gain decomposes into liquidity-transformation and fire-sale components. The welfare analysis identifies two pecuniary externalities, both new to the market-based architecture and both delivered through market-clearing rather than through depletion of an intermediary's reserves. The first is a run externality at the long-term-holder SELL cutoff. When a holder switches from HOLD to SELL, she ignores three losses she imposes on others: a higher probability of severe failure that lowers the payoff to patient holders who stay, a deeper wrapper-vs-NAV spread that lowers the liquidation proceeds received by every other shocked wrapper seller, and the basket fire-sale impairment imposed on shocked direct sellers when the closure flow exhausts dealer capacity. The second is an under-provision of dealer-AP warehousing capacity. The dealer-AP equates the private return on its own arbitrage trades to the balance-sheet cost of holding inventory, but does not count the gain that a tighter spread confers on every other shocked seller, nor the reduction in severe-failure probability that benefits patient holders and direct-holder shocked sellers. Both wedges share a common reason: each agent sees only the spread on its own trade, and neither sees the spillover on every other inframarginal trader. Neither wedge is the Diamond and Dybvig (1983) reserve-depletion externality, where the bank's liquid pool runs dry on a first-come first-served basis. Both are clearing-market externalities in the spirit of Lorenzoni (2008) and Stein (2012), here delivered through wrapper secondary-market clearing and through the OTC fire-sale link rather than through an explicit collateral constraint.

Each wedge maps to a policy tool. A balance-sheet transfer to dealer-APs closes the AP under-provision of warehousing capacity. A facility on the general design of the Federal Reserve's March 2020 Secondary Market Corporate Credit Facility is the closest institutional analogue, although the SMCCF operated through secondary-market purchases of bonds and bond ETFs that relieved dealer-AP inventory pressure indirectly, rather than through direct capital infusion. A Pigouvian charge on strategic redemption internalises the run externality directly. The two instruments are not redundant: each targets a different side of the wrapper secondary market, and both are needed to implement the planner's first-best. Bank-era tools, deposit insurance and reserve requirements, target an intermediary's reserve position and so address neither margin. As liquidity transformation migrates from banks to markets, the regulatory perimeter 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, noise traders, and dealer-APs. Section 5 specifies wrapper secondary-market clearing and the OTC bond market. Section 6 defines the linear-threshold rational-expectations equilibrium and proves existence and uniqueness. Section 7 conducts the equilibrium analysis. Section 8 discusses the closed-end CLO as a related institutional form. Section 9 develops the welfare analysis and the policy implications. Section 10 concludes. The appendix collects all proofs.

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. 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 at the intersection of these literatures is to put the dual role of the secondary-market price at the centre of non-bank liquidity transformation. The wrapper price clears AP arbitrage against mechanical and strategic selling, and the same price is the public signal on which long-term holders condition their hold-or-sell decision and the dealer-AP conditions its balance-sheet commitment. The substitution-versus-information uniqueness condition that follows depends on wrapper-side absorption capacity, the dealer-intermediation primitive of the architecture, in addition to the precision of private information. Wrapper-side absorption is the sum of market-maker depth, supplied by dealers, and balance-sheet warehousing, supplied by APs. When this absorption is small, each unit of selling moves the price by more, strengthening the substitution penalty and delivering uniqueness; when absorption is large, the price-impact channel flattens and the information role of the price dominates. 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 dealer-APs whose private signal of the fundamental is the source of their information advantage.

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: long-term 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 long-term holders observe in their hold-or-sell decision and the dealer-AP observes in sizing its balance-sheet commitment to arbitrage. This coupling delivers two new objects that are absent in Pan-Zeng: a substitution-versus-information uniqueness condition that depends on AP balance-sheet capacity, and a *severe* regime in which the residual basket is dumped into the OTC market through an AUM-triggered closure 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), Haddad et al. (2021), and O'Hara and Zhou (2021) document the March 2020 bond-ETF dislocation and the role of dealer balance-sheet constraints. 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 8.

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 $\text{NAV}_2 = \mu + \theta$. The $T = 1$ NAV is $\text{NAV}_1 = \mu + E_1[\theta]$.

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 $P_{i,1}(1 - d_H) - D\,\mathbf{1}\{\text{severe}\}$, where d_H is the baseline OTC discount and D is the additional severe-state basket fire-sale loss endogenised in Section 5.2. The wrapper trades on a continuous exchange at price P_E^t. Risk-free rate $r_f = 0$. We normalise to $\bar{P} \equiv E_0[\text{NAV}_1] = \mu$, 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 a per-AP warehousing capacity χ; they do not force the sponsor to liquidate basket assets to pay cash. Patient long-term 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 maintains an AUM threshold $\bar{A} > 0$: if remaining wrapper assets at $T = 1$ fall below $\bar{A}$, the sponsor closes the wrapper, sells the residual basket into the OTC market through APs, and distributes the cash proceeds pro-rata to remaining holders. 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. AUM at $T = 1$ is the mass of remaining long-term holders times the basket price. We model the rule using a fixed share-count threshold: closure occurs when the remaining mass of long-term holders falls below $\bar{A}/\mu$, equivalently when AUM valued at the prior basket price $\bar{P} = \mu$ falls below $\bar{A}$. Using μ rather than the realised $T = 1$ basket value $\mu + E_1[\theta]$ to value the remaining mass is a deliberate simplification: it makes the closure trigger a sharp condition on aggregate holder selling alone, removing the noise-shock contamination that would otherwise enter through the basket price. Empirically, sponsor closure decisions are based on trailing AUM relative to a fixed dollar floor (typically \$25M–\$100M), so a par-value-anchored share-count cutoff captures the institutional rule reasonably well. We make the liquidation condition explicit in Section 4.2.

3.4 Information

Each long-term 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.

Dealer-APs receive a common signal

$$\tilde{s}_A \;=\; \theta + \sigma_a \tilde{\varepsilon}_A, \qquad \tilde{\varepsilon}_A \sim N(0,1), \qquad \tau_a = 1/\sigma_a^2. \tag{1}$$

The AP signal is observed only by the AP at $T = 1$. The shock $\tilde{\varepsilon}_A$ is independent of θ, $\tilde{\xi}$, and the holder shocks.

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. Long-term 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 dealer-AP solves its arbitrage problem using $(\tilde{s}_A, P_E^1, \chi)$. Wrapper and OTC markets clear.
- **$T = 2$.** Idiosyncratic shocks ε_i realise. Wrapper holders receive $\text{NAV}_2 - D \cdot \mathbf{1}\{\text{severe failure at } T = 1\}$. Direct holders receive v_i.

4 Agents

4.1 Investors at $T = 0$

A unit mass of investors, each endowed with one unit of capital at $T = 0$, is risk-neutral over $T = 2$ wealth and evaluates choices by expected value. We adopt risk-neutrality at $T = 0$ to keep the patient-holder fire-sale loss tractable as a deterministic deduction conditional on the severe state, rather than requiring a binary-lottery risk premium. The wrapper's economic role in this paper is liquidity insurance (smoothing the shocked-investor sale price) and fire-sale exposure (the cost the patient holder bears in the severe regime), and both are first-order in expected value.

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 $T = 2$.[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 an AUM 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 impairment 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. The wrapper advantage is

$$\Delta(\psi_h) \;=\; \psi_h(d_H \bar{P} + D\pi_s - \Delta_0) \;-\; (1 - \psi_h)(d_H \bar{P} + D)\pi_s \;-\; \phi. \tag{4}$$

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 impairment $D\pi_s$ 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. The third term is the wrapper fee.

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

$$\psi^* \;=\; \frac{\phi + (d_H \bar{P} + D)\pi_s}{d_H \bar{P} + D\pi_s - \Delta_0 + (d_H \bar{P} + D)\pi_s}. \tag{5}$$

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

[2] The ψ-sorting describes the flexible investor base that freely chooses between Direct and Wrapper. Buy-and-hold institutional holders (insurance companies, pension target-date funds, BDC sleeves) hold ETFs for reasons exogenous to the marginal-type calculation (index mandates, regulatory accounting, diversification under risk-management constraints) and so sit outside this sorting. They are not the source of fragility we study.

$T = 0$ **aggregates.** Wrapper-ownership share $\Omega_W = 1 - F(\psi^*)$. Long-term wrapper-holder mass and mechanical wrapper-selling mass are

$$\kappa \;=\; \int_{\psi^*}^{1} (1-\psi)\, dF(\psi), \qquad \lambda \;=\; \int_{\psi^*}^{1} \psi\, dF(\psi), \tag{6}$$

respectively, and direct-holder shocked OTC selling is $\lambda_b = \int_0^{\psi^*} \psi\, dF(\psi)$.

4.2 Long-term wrapper holders at $T = 1$

Long-term 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:

$$\text{Payoff} = \begin{cases} \mu + \theta - (d_H \bar{P} + D) \cdot \mathbf{1}\{A > A^{\text{liq}}\} & \text{if } a_h^1 = \text{HOLD} \\ P_E^1 & \text{if } a_h^1 = \text{SELL} \end{cases} \tag{7}$$

where $A > A^{\text{liq}}$ is the sponsor-closure event defined below in (10), which 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.

A monotone strategy is a threshold on the holder's posterior on θ:

$$z(\tilde{s}_h, P_E^1) \;=\; \frac{\tau_s \tilde{s}_h + \tau_p y(P_E^1)}{\hat{\tau}}, \qquad \hat{\tau} = \tau_\theta + \tau_s + \tau_p,$$

where $y(P_E^1)$ is the sufficient statistic of the endogenous public signal P_E^1 for θ. A long-term holder sells iff $z < z^*$; equivalently, iff

$$\tilde{s}_h + k\, y(P_E^1) \;\le\; c^*, \qquad k = \tau_p/\tau_s, \qquad c^* = z^* \hat{\tau}/\tau_s. \tag{8}$$

Each long-term 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 (8), holder i sells iff

$$\theta + \sigma_s \tilde{\varepsilon}_h^i + k\, y(P_E^1) \;\le\; c^* \quad\Longleftrightarrow\quad \tilde{\varepsilon}_h^i \;\le\; \frac{c^* - k\, y(P_E^1) - \theta}{\sigma_s}.$$

Conditional on (θ, y), the fraction of long-term holders who sell is therefore

$$A(\theta, y) \;\equiv\; \Phi\!\left(\frac{c^* - k\, y(P_E^1) - \theta}{\sigma_s}\right), \tag{9}$$

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 $\kappa A(\theta, y)$. Two properties are immediate: A is decreasing in θ (better fundamentals lower the fraction selling) and decreasing in y (a higher price-extracted public signal lowers the fraction selling through $k > 0$); the slope per unit of θ or ky is $-\Phi'(\cdot)/\sigma_s$, scaling inversely with private-signal noise.

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 AUM at $T = 1$

falls below $\bar{A}$, equivalently when the mass of long-term holders that do not sell is below $\bar{A}/\mu$. Using the aggregate selling expression $\kappa A(\theta, y)$ from (9), the remaining mass is $\kappa(1 - A(\theta, y))$ and closure occurs iff

$$A(\theta, y) \; > \; A^{\text{liq}} \; \equiv \; 1 - \frac{\bar{A}}{\kappa\mu}. \tag{10}$$

The trigger (10) is a threshold on aggregate selling, hence on θ given y. On closure, the residual basket $\kappa(1 - A)$ is dumped into the OTC market through APs. The sponsor's mandate is to liquidate the entire residual at once, not in small slices, so the closure-induced basket dump exhausts dealer capacity, and we model the fire-sale impairment D as triggered directly by the closure event $A > A^{\text{liq}}$. The indicator $\mathbf{1}\{A > A^{\text{liq}}\}$ in (7) therefore captures the joint event of closure plus fire-sale-priced basket sale. Patient holders receive the OTC sale proceeds $\mu + \theta - d_H\bar{P} - D$ rather than the $T = 2$ NAV $\mu + \theta$, with the shortfall split between the baseline discount $d_H\bar{P}$ and the fire-sale lump-sum D. This is the in-kind analogue of cash-redemption dilution: in both cases the patient holder bears the basket fire-sale loss when the wrapper is forced to liquidate assets at depressed prices, but the trigger is sponsor closure under the AUM rule rather than cash-redemption pressure on liquid reserves.

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

$$D_m(P_E^1) \; = \; \kappa\,\Phi\Big(-\alpha(P_E^1 - \mu) + \sigma_\xi\tilde{\xi} + n\tilde{W}\Big) \; + \; \lambda, \tag{11}$$

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 $n\tilde{W}$ is the aggregate AP order flow received by the MM. 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 $\kappa\Phi((c^* - ky - \theta)/\sigma_s)$, and the additive λ matches the mechanical-selling baseline. Because MM demand does not depend on θ, market makers contribute no fundamentals information to the price; the price aggregates information only through $n\tilde{W}$ in the Φ argument. Clearing then reduces to an exactly linear equation in P_E^1 and the price-extracted public signal $y(P_E^1)$ is also linear in P_E^1, so the Bayesian posterior remains closed-form under Gaussian noise; the only fundamentals-bearing private signals in the economy are the holders' $\tilde{s}_h$ and the AP's $\tilde{s}_A$.

The $\kappa\Phi(\cdot)$ piece is a reduced-form modelling choice that buys 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 run-feedback channel correspondingly amplified. Reading $\kappa\Phi(\cdot)$ as a shorthand for dealer balance sheets, internalisation, and dynamic inventory management lets us capture these stabilising forces without modelling them explicitly.

4.4 Dealer-Authorised Participants

A continuum of dealer-APs of mass n. AP capacity $\chi > 0$ is a deterministic constant. Each AP receives the common private signal $\tilde{s}_A = \theta + \sigma_a \tilde{\varepsilon}_A$ at $T = 1$. The AP's information at $T = 1$ is $\mathcal{I}_{AP} = \{P_E^1, \tilde{s}_A\}$. The price P_E^1 is public, but the AP has an information advantage over investors because it knows its own signal $\tilde{s}_A$. The linear-equilibrium price equation is a linear function of the fundamental and the noise shocks:

$$p \;=\; \mathcal{A} + \mathcal{B}\theta + \mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A. \tag{12}$$

Rational expectations require that the conjectured coefficients used by agents to invert P_E^1 for the public signal coincide with the coefficients of the realised market-clearing price function. This is a fixed-point condition on $(\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G})$.

Given that the conjecture is unique (which we verify later), the price-extracted public signal is then contaminated by two noise components:

$$y(P_E^1) \;=\; \theta + \tilde{u}, \qquad \tilde{u} \;=\; \frac{\mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A}{\mathcal{B}}, \tag{13}$$

with public-signal precision $\tau_p \equiv \mathcal{B}^2/(\mathcal{D}^2 + \mathcal{G}^2)$. Investors see only y, so they use it at precision τ_p. The AP, knowing $\tilde{s}_A$, can subtract the $\tilde{\varepsilon}_A$ component out of the price-extracted signal and form a filtered signal

$$y^{AP} \;\equiv\; \frac{y - r\,\tilde{s}_A}{1 - r} \;=\; \theta + \frac{\mathcal{D}\tilde{\xi}}{\mathcal{B}(1 - r)}, \qquad r \;\equiv\; \frac{\mathcal{G}}{\mathcal{B}\sigma_a}, \tag{14}$$

which is conditionally independent of $\tilde{s}_A$ given θ and has precision

$$\tau_p^{AP} \;=\; \frac{\mathcal{B}^2(1 - r)^2}{\mathcal{D}^2}. \tag{15}$$

The AP-signal noise channel is removed by the AP filter. The AP's information advantage $\tau_p^{AP} - \tau_p > 0$ widens with σ_a (more $\tilde{\varepsilon}_A$ contamination of the price, all of which only the AP can filter).

The AP's posterior on θ uses the filtered signal y^{AP} and the private signal $\tilde{s}_A$, which are conditionally independent given θ:

$$E[\theta \mid \mathcal{I}_{AP}] \;=\; \beta_p^{AP}\, y^{AP} \;+\; \beta_a^{AP}\, \tilde{s}_A, \tag{16}$$

with

$$\beta_p^{AP} \;=\; \frac{\tau_p^{AP}}{\tau_\theta + \tau_p^{AP} + \tau_a}, \qquad \beta_a^{AP} \;=\; \frac{\tau_a}{\tau_\theta + \tau_p^{AP} + \tau_a}. \tag{17}$$

The AP runs an intra-period arbitrage book between the wrapper and the underlying basket. The position W can be positive (long wrapper, short basket; profitable when the wrapper trades at a discount to NAV) or negative (short wrapper, long basket; profitable when the wrapper trades at a premium). The per-share arbitrage profit is the wrapper-vs-NAV spread net of OTC unwinding cost,

$$\pi^{AP}_{\text{per share}} \;=\; (\mu + E[\theta \mid \mathcal{I}_{AP}]) - d_h/\chi - P_E^1,$$

with d_h/χ the per-share OTC discount faced by an AP of capacity χ.[3] Carrying W shares incurs the convex inventory charge $W^2/(2\chi)$. The AP solves

$$\max_W \; W \cdot [\mu + \beta_p^{AP} y^{AP} + \beta_a^{AP} \tilde{s}_A - d_h/\chi - P_E^1] \;-\; \frac{W^2}{2\chi}. \tag{18}$$

[3] Modelling the per-share OTC cost as d_h/χ rather than d_h assumes scale economies in OTC trading: a higher-capacity AP faces a lower per-share unwinding cost, reflecting better dealer relationships, volume discounts, and more axes to find the other side of the basket.

Differentiating in W and setting to zero gives

$$\tilde{W}^*(P_E^1; \tilde{s}_A) \;=\; \chi \cdot \Delta_A(P_E^1, \tilde{s}_A) \;-\; d_h, \tag{19}$$

where the AP's expected per-share spread is

$$\Delta_A(P_E^1, \tilde{s}_A) \;\equiv\; \mu + \beta_p^{AP} y^{AP} + \beta_a^{AP} \tilde{s}_A - P_E^1. \tag{20}$$

The rest of the analysis focuses on the wrapper-discount region ($\Delta_A > d_h/\chi$, hence $\tilde{W}^* > 0$). The premium region ($\Delta_A < -d_h/\chi$, $\tilde{W}^* < 0$) is symmetric. Aggregate AP demand is $n\tilde{W}^*$.

The parameter χ governs three things: (i) the inventory cost the AP pays for warehousing, (ii) the OTC unwinding cost the AP faces when liquidating the basket, and (iii) the price elasticity of AP demand on the wrapper secondary market.

The AP forms its posterior on θ partly from the wrapper price (through y^{AP}), so heavier wrapper selling pulls AP demand in the same direction. Long-term holders sell when their private signals are bad. Their selling pushes P_E^1 down through clearing. The AP, observing the price, infers fundamentals are weak. Its posterior $\beta_p^{AP} y^{AP} + \beta_a^{AP} \tilde{s}_A$ falls (on the y^{AP} piece). Whether the AP then buys more or less depends on whether the price falls faster or slower than its posterior:

$$\frac{\partial \Delta_A}{\partial P_E^1} \;=\; \frac{\beta_p^{AP}}{\mathcal{B}(1-r)} - 1. \tag{21}$$

If the price overreacts to fundamentals ($\mathcal{B}(1-r) > \beta_p^{AP}$), the spread widens when the price falls and the AP arbitrages back. Selling pressure is dampened. If the price underreacts ($\mathcal{B}(1-r) < \beta_p^{AP}$), the AP cuts its expected NAV faster than the price falls, the spread narrows, and the AP pulls back. Selling feeds on itself through AP-learning. The destabilising condition holds when the AP relies heavily on the price relative to its prior and its private signal.

5 Markets and clearing

5.1 Wrapper secondary-market clearing

Total wrapper sales at $T = 1$ are mechanical λ plus strategic $\kappa A(\theta, y)$, with $A(\theta, y)$ from (9). The MM is the sole counterparty on the buy side (Section 4.3); the AP order flow $n\tilde{W}$ enters through the MM's quote in (11) 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 ($\chi\Delta_A > d_h$) is

$$\kappa\, \Phi\Big(-\alpha(P_E^1 - \mu) + \sigma_\xi \tilde{\xi} + n\tilde{W}\Big) + \lambda \;=\; \lambda + \kappa\, \Phi\left(\frac{c^* - ky - \theta}{\sigma_s}\right). \tag{22}$$

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

$$-\,\alpha(P_E^1 - \mu) + \sigma_\xi \tilde{\xi} + n\tilde{W} \;=\; \frac{c^* - ky - \theta}{\sigma_s}, \tag{23}$$

with no Taylor approximation. Substituting $n\tilde{W} = n\chi(\mu + \beta_p^{AP} y^{AP} + \beta_a^{AP} \tilde{s}_A - P_E^1) - nd_h$ from (19) yields the price-coefficient matching equations of Section 6.2. Because the AP filters its own signal noise out of the price (replacing y with y^{AP}), the AP-side $\tilde{\varepsilon}_A$ loading in clearing is exactly $n\chi\beta_a^{AP}\sigma_a$ (through $\tilde{s}_A$ only); the indirect $\tilde{\varepsilon}_A$ contribution that would come through y is removed by the filter.

5.2 OTC market and the fire-sale link

Three components of basket-selling flow reach the OTC market at $T = 1$: discretionary AP arbitrage $nW^* = n[\chi\Delta - d_h]_+$ (the short leg of the AP's long-wrapper book, scaling up as the wrapper-vs-NAV spread widens); direct-holder shocked sales λ_b; and the sponsor's closure flow $\mathcal{L} = \kappa(1 - A(\theta, y))$, which hits the market when the AUM-closure condition (10) is met. The closure flow is a forced-sale of the residual basket in a single window, so it dominates the discretionary flows. The discretionary AP-arbitrage flow nW^* is small in the severe regime for two reasons. First, the wrapper-vs-NAV spread is wide enough that AP arbitrage has reached the limit of available capacity χ at the prevailing inventory cost. Second, AP price-learning becomes self-reinforcing in the destabilising region (as shown in Section 4.4), which pulls the AP back from absorbing further flow. The severe-failure event is therefore empirically a closure-driven fire-sale, consistent with the bond-ETF closure pattern documented in 2020–2022.

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

$$P_{i,1}^{\text{sale}} \;=\; \mu + \theta \;-\; d_H \bar{P} \;-\; D \cdot \mathbf{1}\{A > A^{\text{liq}}\}, \tag{24}$$

$$\text{basket fire-sale loss} \;=\; D \cdot \mathbf{1}\{A > A^{\text{liq}}\} \;=\; \begin{cases} 0 & \text{if } A \le A^{\text{liq}}, \\ D & \text{if } A > A^{\text{liq}}. \end{cases} \tag{25}$$

Throughout, $d_H \in (0, 1)$ is the baseline OTC discount expressed as a fraction of $\bar{P}$, $D \ge 0$ is the fire-sale impairment 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 dumps the residual basket, in which case sellers face a step-function impairment.

5.3 Failure regime triggers

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

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

$$\Delta(\theta, \tilde{\epsilon}) \;>\; d_H \bar{P}, \tag{26}$$

where $\Delta = \mu - P_E^1$. 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.

Severe trigger. The mild trigger fires *and* the sponsor's AUM-closure condition fires:

$$\Delta(\theta, \tilde{\epsilon}) > d_H \bar{P} \quad \text{and} \quad A(\theta, y) > A^{\text{liq}}, \tag{27}$$

with A^{liq} from (10). Both conditions are needed: when only the AUM condition fires, AP arbitrage is still absorbing strategic selling and the closure flow at the wrapper-vs-NAV margin, so the basket clears at the baseline OTC discount; when only the mild condition fires, the sponsor does not close and the residual basket is not dumped. When both fire, AP arbitrage is exhausted, the sponsor closes, the residual basket $\mathcal{L} = \kappa(1 - A)$ is dumped wholesale, and the dealer-capacity breach lump-sum D is incurred. The severe-failure event is therefore a holder-driven coordination event mediated by sponsor closure, with APs withdrawn in the relevant range.

Both triggers are state events, but evaluating their probabilities and the implied θ-thresholds requires the equilibrium price coefficients $(\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G})$.

6 Equilibrium

6.1 Definition

Definition 1 (Linear-threshold rational-expectations equilibrium)**.** A linear-threshold rational-expectations equilibrium is a tuple

$$(\psi^*,\ k,\ c^*,\ (\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G}),\ \tilde{W}^*(\cdot,\cdot),\ \kappa, \lambda, \lambda_b)$$

such that:

(E.1) *Marginal investor at $T = 0$.* $\Delta(\psi^*) = 0$ from (4), and aggregate masses $\kappa, \lambda, \lambda_b$ are the integrals over the resulting partition (6).

(E.2) *Long-term holders at $T = 1$.* Holders follow the threshold strategy (8) with $k = \tau_p/\tau_s$ and c^* from the indifference condition

$$E[P_E^1 \mid z^*] \ = \ \mu + z^* \ - \ (d_H \bar{P} + D) \cdot \Pr(\text{severe} \mid z^*). \tag{28}$$

(E.3) *APs at $T = 1$.* $\tilde{W}^*(P_E^1; \tilde{s}_A)$ solves (18).

(E.4) *Wrapper clearing.* (22) holds pointwise in $(\theta, \tilde{\xi}, \tilde{\varepsilon}_A)$.

(E.5) *OTC market.* Prices and basket fire-sale loss follow (24) and (25).

(E.6) *Rational expectations.* Conjectured price coefficients used in (E.2)–(E.3) equal the realised coefficients from (E.4).

6.2 Existence and uniqueness

The equilibrium price coefficients are obtained by substituting the AP posterior into clearing and matching coefficients with the linear-price conjecture (12) (Appendix A.1, Lemma A.1). A useful structural feature emerges in this construction: τ_p^{AP}, and through it the AP posterior weights $\beta_p^{AP}, \beta_a^{AP}$ in (17), depends only on (σ_s, σ_ξ), not on κ, k, or c^*. The AP posterior weights are therefore parameter-determined rather than equilibrium objects, which collapses the fixed point in k to a single equation rather than a joint system.

Proposition 1 (Linear-threshold equilibrium)**.** *Suppose:*

1. *Market-maker depth α lies in a non-empty range $[\underline{\alpha}, \bar{\alpha})$ defined by the deep parameters $(d_H, D, \sigma_s, \sigma_\xi, \sigma_a, n\chi, \tau_\theta)$ via Lemma A.1. The lower bound $\underline{\alpha}$ ensures the noisy-trader-region property: the equilibrium price loads on the fundamental with weight $\mathcal{B} \le 1$. The upper bound $\bar{\alpha}$ ensures the substitution-versus- information uniqueness condition*

$$\frac{1}{\Lambda} \ > \ \frac{d_H \bar{P} + D}{\sqrt{2\pi}\, \hat{\Sigma}(\sigma_s)},$$

where $\Lambda = \alpha + n\chi$ is wrapper-side absorption capacity and $\hat{\Sigma}(\sigma_s) \in (0, 1]$ is the marginal holder's posterior standard deviation of $\theta + ky$ normalised by private-signal noise; see (46) in Appendix A.1 for the closed form.

2. *The AP's private signal is not perfectly informative ($\sigma_a > 0$).*
3. *The sorting condition of Section 4.1 holds.*
4. *The $T = 0$ feedback satisfies the contraction condition* (50) *in Appendix A.1, which bounds the per-round gain to the strategic-complementarity loop $\kappa \to \pi_s \to \psi^* \to \kappa$ strictly below one.*

Then there exists a unique linear-threshold rational-expectations equilibrium, with the following closed-form characterisation:

- *Price coefficients $(\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G})$ given by* (40) *in Appendix A.1.*
- *AP-filtered precision*

$$\tau_p^{AP} \;=\; \frac{1}{\sigma_s^2\,\sigma_\xi^2}, \tag{29}$$

  depending only on the holder-signal noise σ_s and the wrapper noise-trader scale σ_ξ, feeding into the AP posterior weights $\beta_p^{AP}, \beta_a^{AP}$ in (17)*.*
- *Expected wrapper-vs-NAV spread*

$$\Delta_0 \;\equiv\; E_0[\Delta] \;=\; -\mathcal{A} \;=\; \frac{c^*/\sigma_s + nd_h}{\Lambda}, \tag{30}$$

  which compensates the AP for residual strategic-selling pressure c^*/σ_s and the round-trip OTC cost nd_h. The mass parameters λ and κ drop out of Δ_0.
- *(k^*, c^*) jointly solve the Bayesian-weighting condition $k = \tau_p/\tau_s$ from (E.2) and the indifference condition* (28)*; the marginal investor ψ^* is given by* (5)*.*

Proof. See Appendix A.1. □

Holders' selling decisions are strategic complements through the AUM-closure trigger: heavier redemption raises the probability that the sponsor closes the wrapper, raises the severe-state loss for those who continue to hold, and tips 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 $(c^* - ky - \theta)/\sigma_s$ 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 dealer-AP reads the same price signal and also pulls back its arbitrage as the price falls, reinforcing the drop. Through this channel the endogenous price amplifies the complementarity. Noisier private signals (larger σ_s) weaken it by lowering the loading 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 redemption 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 $\Lambda = \alpha + n\chi$: small Λ means each unit of strategic redemption induces a sharp price drop, so substitution is strong; large Λ flattens the price response and weakens the penalty on sellers. The substitution role is the uniqueness force. The condition $1/\Lambda > (d_H\bar{P} + D)/(\sqrt{2\pi}\,\hat{\Sigma})$ in Proposition 1 requires substitution to dominate the information role at the equilibrium threshold: $1/\Lambda$ measures substitution strength,

$(d_H\bar{P} + D)/\hat{\Sigma}$ measures the run-feedback payoff of coordinating, and $\hat{\Sigma}$ scales with the noise in the price-extracted signal. Noisy private signals (large σ_s) are necessary for uniqueness but not sufficient; the absorption-versus-penalty condition must hold as well.

7 Equilibrium Analysis

7.1 Three regimes and basket fire-sale loss

The equilibrium delivers three regimes indexed jointly by the wrapper spread Δ_0 and the sponsor-closure indicator $\mathbf{1}\{A > A^{\text{liq}}\}$. Figure 1 illustrates how the equilibrium prices and payoffs move across the three regimes as the fundamental θ varies.

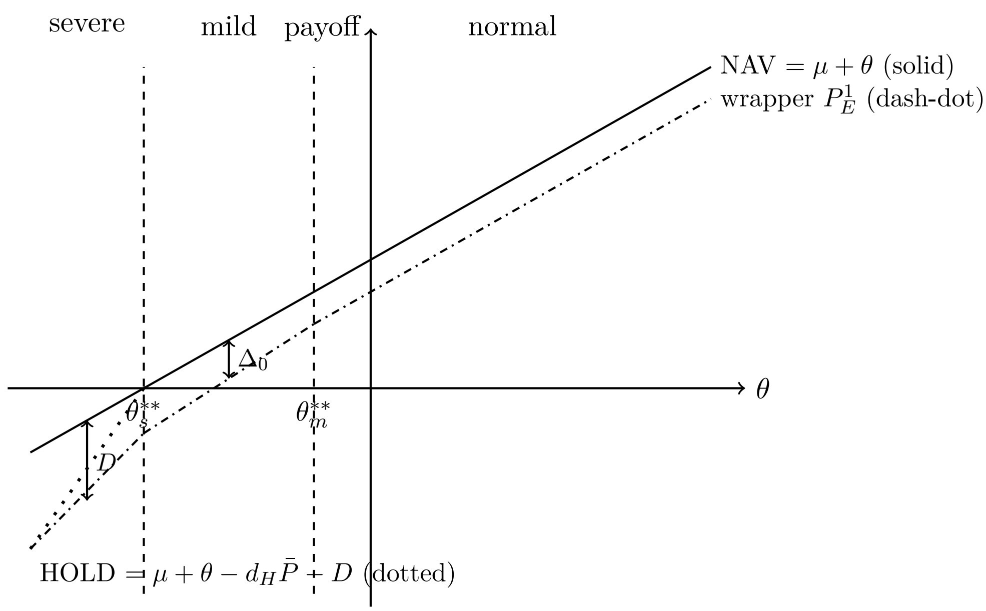

Figure 1: Equilibrium prices and payoffs across the three regimes as θ varies. Solid: NAV $\mu + \theta$. Dash-dot: wrapper price P_E^1. Dotted: HOLD payoff in severe failure, $\mu + \theta - d_H\bar{P} - D$ (in the normal and mild regimes HOLD coincides with NAV). Thresholds θ_m^{**} and θ_s^{**} are given by (32) and (34).

In the *normal regime* ($\Delta \le d_H\bar{P}$, $A < A^{\text{liq}}$), AP arbitrage holds the wrapper near NAV and strategic redemption is well below the sponsor's closure floor. SELL pays the linear-REE wrapper price P_E^1; HOLD pays $\mu + \theta$; the basket is intact. In the *mild-failure regime* ($\Delta > d_H\bar{P}$, $A < A^{\text{liq}}$), the wrapper price has decoupled from NAV but strategic redemption is still below the closure floor. Sellers absorb the wrapper discount; holders still receive $\mu + \theta$. In the *severe-failure regime* ($\Delta > d_H\bar{P}$ *and* $A > A^{\text{liq}}$), the wrapper has decoupled *and* the sponsor's AUM-closure condition is met; the residual basket $\mathcal{L} = \kappa(1 - A)$ is dumped 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. Substituting the linear price $p = \mathcal{A} + \mathcal{B}\theta + \mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A$ from (12) into the mild trigger (26) gives a state-dependent condition. Conditional on θ, $(\tilde{\xi}, \tilde{\varepsilon}_A)$ is standard bivariate normal, so the failure probability is

$$\Pr(\text{mild} \mid \theta) \;=\; \Phi\!\left(\frac{\theta_m^{**} - \theta}{\sigma^{\text{mild}}}\right), \tag{31}$$

where $\sigma^{\text{mild}} = \sqrt{\mathcal{D}^2 + \mathcal{G}^2}/\mathcal{B}$ and the threshold on θ at zero noise is

$$\theta^{**}_m \;=\; \frac{c^*/\sigma_s - (\alpha + n\chi)d_H\bar{P} + nd_h}{n\chi(\beta^{AP}_p + \beta^{AP}_a) + (1+k)/\sigma_s}. \tag{32}$$

We next characterise the threshold for the severe failure regime. The severe trigger (27) combines the mild condition with the AUM-closure condition $A(\theta, y) > A^{\text{liq}}$. Plugging (9) into the AUM condition gives $\theta + ky < c^* - \sigma_s\Phi^{-1}(A^{\text{liq}})$, so conditional on θ,

$$\Pr(\text{AUM} \mid \theta) \;=\; \Phi\left(\frac{\theta^{**}_{\text{AUM}} - \theta}{\sigma^{\text{sev}}}\right), \tag{33}$$

with $\theta^{**}_{\text{AUM}} = (c^* - \sigma_s\Phi^{-1}(A^{\text{liq}}))/(1+k)$ and $\sigma^{\text{sev}} = (k/\mathcal{B})\sqrt{\mathcal{D}^2 + \mathcal{G}^2}/(1+k)$. The severe threshold at zero noise is the binding of the two triggers,

$$\theta^{**}_s \;=\; \min\{\theta^{**}_{\text{AUM}}, \theta^{**}_m\}. \tag{34}$$

For severe to be a meaningful event distinct from mild we impose the regime-defining condition $\theta^{**}_{\text{AUM}} < \theta^{**}_m$: the AUM threshold lies below the mild threshold, so AUM-closure is the binding trigger and $\theta^{**}_s = \theta^{**}_{\text{AUM}}$. We maintain this condition throughout. Severe failure is then strictly nested inside mild, and the regime structure on θ is

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

In the normal and mild regimes ($D = 0$), the indifference condition reduces to $E[P^1_E \mid z^*] = \mu + z^*$, a pure rational-expectations consistency: coordination is purely informational. In the severe regime, $D \cdot \Pr(\text{severe} \mid z^*)$ delivers Diamond-Dybvig strategic complementarity.

Larger AP capacity stabilises the wrapper-vs-NAV spread: higher χ raises AP absorption $n\chi$, narrowing Δ_0 through (30) (because $\Lambda = \alpha + n\chi$), and lowering the mild-failure probability. The severe-failure threshold θ^{**}_s does not depend on χ directly because the AUM trigger is a holder-side coordination event: the closure-induced basket dump $\mathcal{L}$ dominates regardless of AP capacity, and the discretionary AP-arbitrage flow nW^* is small in stress because APs withdraw. χ enters severe failure only through the equilibrium feedback in c^*, which lowers c^* when χ is large and shrinks the severe region. AP capacity is therefore unambiguously stabilising within the unique-equilibrium region of Proposition 1.

7.2 Comparative statics

All comparative-statics and welfare results in this section and in Section 9 apply within the unique-equilibrium region of Proposition 1, that is, on the parameter set where the substitution-versus-information condition $1/\Lambda > (d_H\bar{P} + D)/(\sqrt{2\pi}\,\hat{\Sigma})$ holds. Outside that region the equilibrium is not unique and the derivatives below are not well-defined as parameter perturbations.

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 $(\kappa, c^*, \psi^*, \lambda, \lambda_b)$ fixed at their equilibrium values. We do *not* hold the AP-side endogenous objects $(k, \beta^{AP}_p, \beta^{AP}_a, \tau_p)$ fixed: by Lemma A.1 these are exact closed-form functions of the primitive parameters, so any parameter perturbation moves them mechanically. The full REE total derivative therefore decomposes as

$$\frac{dX}{d\theta_j} \;=\; \underbrace{\frac{\partial X}{\partial \theta_j}}_{\text{direct effect (this section)}} + \underbrace{\frac{\partial X}{\partial k}\frac{dk}{d\theta_j} + \frac{\partial X}{\partial \beta^{AP}_p}\frac{d\beta^{AP}_p}{d\theta_j} + \frac{\partial X}{\partial \beta^{AP}_a}\frac{d\beta^{AP}_a}{d\theta_j}}_{\text{informational feedback}} + \underbrace{\frac{\partial X}{\partial c^*}\frac{dc^*}{d\theta_j}}_{\text{run-feedback through } c^*}$$

for $X \in \{\theta_m^{**}, \theta_s^{**}, \Delta_0\}$ and θ_j a primitive parameter. The direct effect is the sign of the parameter's first-order impact on the threshold formulas (32)–(34) or on the expected spread (30); the informational feedback runs through the equilibrium signal-precision mapping in Lemma A.1; the run-feedback runs through the indifference condition. We isolate the direct effect because it has a clean structural interpretation; we discuss the informational and run-feedback channels qualitatively after each corollary, and note explicitly when they could overturn the direct sign. Proofs are in Appendix A.2.

Corollary 1 (Equilibrium discount and wrapper-transformation gain)**.** *The equilibrium wrapper-vs-NAV spread* (30) *is*

$$\Delta_0 \;=\; \frac{c^*/\sigma_s + nd_h}{\Lambda}, \qquad \Lambda \;=\; \alpha + n\chi.$$

The wrapper trades in the normal regime iff $\Delta_0 \leq d_H \bar{P}$. In that 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

$$\frac{\partial\Delta_0}{\partial\chi} \;<\; 0, \qquad \frac{\partial\Delta_0}{\partial\alpha} \;<\; 0, \qquad \frac{\partial\Delta_0}{\partial c^*} \;>\; 0, \qquad \frac{\partial\Delta_0}{\partial d_h} \;>\; 0.$$

Note that the mechanical-mass parameter λ does not affect Δ_0 because the additive λ on MM's quote in (11) absorbs mechanical sales pointwise; the strategic-mass parameter κ does not affect Δ_0 because the $\kappa\Phi(\cdot)$ structures cancel exactly in clearing.

The equilibrium spread is the price the AP charges to clear the strategic-selling argument (c^*/σ_s, the slope at which strategic selling enters the exact clearing equation (23)) plus the per-share OTC unwinding cost (nd_h), divided by the combined wrapper-side absorption capacity ($\Lambda = \alpha + n\chi$). When AP capacity (χ) or market-maker depth (α) is large relative to residual flow, the spread is tight and the wrapper closely tracks NAV. A more pessimistic strategic threshold (c^*) widens the spread; a higher OTC unwinding cost (d_h) is passed through to the spread. Mechanical sales λ are absorbed point-wise by MM and do not feed into the spread.

The wrapper transformation provides liquidity insurance to shocked investors: they sell at $\mu - \Delta_0$ in expectation rather than facing the direct-OTC sale price $\bar{P}(1 - d_H) = \mu - d_H\bar{P}$. The gain per shocked share is $g = d_H\bar{P} - \Delta_0$. The wrapper helps exactly when $\Delta_0 < d_H\bar{P}$, and the gain rises with AP capacity and market-maker depth. 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)**.** *The failure thresholds* (32)–(34) *satisfy*

$$\frac{\partial\theta_m^{**}}{\partial(n\chi)} < 0, \qquad \frac{\partial\theta_s^{**}}{\partial\bar{A}} > 0, \qquad \frac{\partial\theta_s^{**}}{\partial\kappa} < 0, \qquad \frac{\partial\theta_r^{**}}{\partial c^*} > 0 \textit{ for } r \in \{m, s\}.$$

Note also that the mechanical-mass parameter λ has no direct effect on either threshold (it cancels in clearing). The direct-holder OTC-flow parameter λ_b has no direct effect on θ_s^{**} either, because the severe-failure trigger is the sponsor's AUM closure.

More wrapper-side absorption capacity ($n\chi$) tightens the AP-arbitrage spread Δ_0, lowering the mild-failure threshold. A higher sponsor AUM floor ($\bar{A}$) requires less holder selling to trigger closure, expanding the severe region. A larger strategic-holder mass (κ) provides a larger cushion of holders that need to defect before the closure condition is met, shrinking the severe region. A more pessimistic strategic threshold (c^*) widens the strategic-selling argument that drives both triggers, pushing both thresholds up. Mechanical wrapper-side sales (λ) are absorbed pointwise by MM and have no effect.

The informational feedback in $\partial\theta_m^{**}/\partial(n\chi)$ is worth tracing. The direct effect above holds the equilibrium signal weight $k = \tau_p/\tau_s$ fixed. In equilibrium τ_p also depends on $n\chi$: AP order flow injects both information (about θ, via $\tilde{s}_A$) and noise (the AP signal noise $\tilde{\varepsilon}_A$) into the price. The net effect $\partial\tau_p/\partial(n\chi)$ is parameter-dependent but typically positive: more AP demand makes the price more informative. Strategic holders then place a higher weight k on the price, which raises the denominator $n\chi(\beta_p^{AP} + \beta_a^{AP}) + (1+k)/\sigma_s$ of (32) and reinforces the direct sign $\partial\theta_m^{**}/\partial(n\chi) < 0$. Run-feedback through c^* works in the same direction: more AP capacity narrows the spread, makes the marginal holder less inclined to sell, lowers c^*, and through $\partial\theta_r^{**}/\partial c^* > 0$ (for $r \in \{m, s\}$) lowers both thresholds further.

Corollary 3 (Informational comparative statics)**.** *The AP's filtered-signal precision* $\tau_p^{AP} = 1/(\sigma_\xi^2\sigma_s^2)$ *satisfies*

$$\frac{\partial\tau_p^{AP}}{\partial\sigma_\xi} < 0, \qquad \frac{\partial\tau_p^{AP}}{\partial\sigma_s} < 0,$$

and through τ_p^{AP}*, the AP posterior weight* β_p^{AP} *is strictly increasing in* τ_p^{AP}*. The mild-failure threshold satisfies*

$$\frac{\partial\theta_m^{**}}{\partial\beta_p^{AP}} < 0,$$

while θ_s^{**} *does not depend on* β_p^{AP} *directly.*

More noise-trader noise (σ_ξ) raises the noise floor in the price, lowering the AP's filtered-signal precision and the AP's reliance on the price. Noisier private holder signals (larger σ_s) attenuate the strategic-selling argument that drives the price, also lowering precision. A stronger AP price-learning channel (higher β_p^{AP}) tightens AP wrapper-side absorption, narrowing the wrapper discount and lowering the mild-failure threshold. The severe-failure threshold is unaffected *directly* because the AUM trigger is mechanical: it depends on the fraction of strategic holders who redeem, not on AP arbitrage activity. *Indirectly*, however, AP price-learning still moves the severe threshold through equilibrium c^*: a narrower wrapper discount raises the marginal holder's expected SELL payoff, lowers c^*, and through $\partial\theta_s^{**}/\partial c^* > 0$ shrinks the severe region as well.

Corollary 4 (AP-signal precision σ_a)**.** *The non-AP precision* τ_p *from* (41) *is affected by* σ_a *through two channels:*

1. Noise-floor channel*: holding* β_a^{AP} *fixed,* $\partial\tau_p/\partial\sigma_a < 0$ *via the* $(n\chi\beta_a^{AP})^2\sigma_a^2$ *term in the denominator.*
2. AP-information-dilution channel*: through* $\tau_a = 1/\sigma_a^2$ *and* (17)*,* $\partial\beta_a^{AP}/\partial\sigma_a < 0$*, which shrinks the AP-noise contribution.*

The net sign of $\partial\tau_p/\partial\sigma_a$ *is ambiguous and depends on the relative magnitudes of* τ_a *and* $\tau_\theta + \tau_p^{AP}$*.*

A noisier AP signal both injects more $\tilde{\varepsilon}_A$ into the price (lowering precision for non-AP agents) and reduces the AP's own reliance on $\tilde{s}_A$ (pulling AP demand back toward the price-only benchmark, dampening the $\tilde{\varepsilon}_A$ transmission). When the AP signal is already noisy (τ_a small), the first channel dominates: more noise mostly hurts. When the AP signal is precise (τ_a large), the second channel dominates: marginal noise mostly dilutes AP-side signalling.

Corollary 5 (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 $\pi_s = \Pr(\theta < \theta_s^{**})$ from* (33)*. The loss is strictly increasing in the lump-sum penalty D and in the sponsor-closure threshold $\bar{A}$ (which raises θ_s^{**} and hence π_s), and strictly decreasing in the strategic-holder mass κ (which lowers A^{liq} and shrinks the severe region).*

Severe failure is driven by the sponsor's AUM-closure rule: the wrapper closes when too many strategic holders redeem, and the residual basket dump overwhelms OTC dealer capacity, regardless of χ in partial equilibrium. The expected severe-state loss therefore scales with the size of the impairment (D) and the size of the closure region. A higher AUM floor $\bar{A}$ makes the sponsor more conservative, expanding the closure region, raising π_s, and raising the expected loss. A larger strategic-holder mass κ raises the denominator of the closure ratio $\bar{A}/(\kappa\mu)$, shrinking A^{liq} and the severe region; equivalently, more strategic holders provide a thicker cushion before AUM falls below the floor.

Under the AUM trigger, AP capacity χ does not enter θ_s^{**} directly: severe failure is a holder-side coordination event, not an AP-flow event. Higher χ tightens the wrapper-vs-NAV discount Δ_0 (Corollary 1) and through the equilibrium feedback in the indifference condition lowers c^*, which then shrinks the severe region. AP capacity is therefore unambiguously stabilising in this model.

8 Discussion: closed-end CLO architecture

The model treats the open-end ETF as the canonical wrapper because the dealer-AP arbitrage channel is the central mechanism. Closed-end vehicles for illiquid securities exist, however, and the leveraged-loan market provides a clean institutional contrast. A collateralised loan obligation (CLO) holds a portfolio of leveraged loans and issues tranched closed-end claims that do not allow $T = 1$ redemption. CLOs cannot run, so the wrapper-coordination story does not apply to them. Two other economic roles distinguish CLOs from direct holding of the underlying loans.

The first is diversification. A CLO holds many loans, so each tranche bears diversified exposure to the underlying basket rather than the idiosyncratic-loan exposure that a direct lender would carry. The risk-neutral framework in this paper is silent on this channel because $E[\varepsilon_i] = 0$ at the security level, but it is empirically central for institutional investors who face binding solvency or leverage constraints on idiosyncratic exposure.

The second is liquidity transformation in the secondary market for CLO tranches. The underlying leveraged loans are illiquid, with high OTC search costs and large bid-ask spreads. The tranched CLO claims, by contrast, are standardised securities. The senior tranches (AAA, AA) trade in dealer-intermediated markets with reliably more depth than the underlying loans. Even the equity tranche, often held to maturity by long-horizon investors, can in stress periods be sold at a secondary-market discount that is bounded by the underlying loan-market discount, providing a partial liquidity improvement on the illiquid underlying for the marginal seller.

Diamond et al. (2025) document that in the post-2008 leveraged-loan market CLOs absorbed the bulk of new loan issuance and were not subject to runs in either March 2020 or in the 2022 leveraged-loan correction, while the open-end loan ETF (the Invesco Senior Loan ETF, ticker BKLN) traded at sharp discounts as APs faced strained dealer balance sheets. The persistent coexistence of CLOs and BKLN in the same underlying (leveraged loans), held by different investor clienteles (institutional buy-and-hold for CLOs, retail or short-horizon for BKLN), is consistent with the two architectures filling complementary roles. Our model captures the open-end side rigorously; the CLO side enters as a substitute architecture that institutional investors choose when the costs of being subject to wrapper-run coordination outweigh the costs of bearing a lockup.

9 Welfare

9.1 Welfare criterion

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

$$\mathscr{W} \;=\; \int_0^{\psi^*} U^D(\psi)\, dF(\psi) \;+\; \int_{\psi^*}^1 U^W(\psi)\, dF(\psi), \tag{35}$$

with cutoff ψ^* from (5) and expected payoffs

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

The wrapper-holder payoff uses the linear-REE 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.

9.2 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 AP exhaustion event):

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

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 dumps 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 $\lambda = \int_{\psi^*}^1 \psi\, dF$, $\kappa = \int_{\psi^*}^1 (1-\psi)\, dF$, and $\lambda_b = \int_0^{\psi^*} \psi\, dF$:

$$\Delta\mathscr{W}^W \;=\; \mathcal{G}_{\mathrm{liq}} \;-\; \mathcal{G}_{\mathrm{fire}} \;-\; \phi\,(1 - F(\psi^*)), \tag{36}$$

where the two gain channels are

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

Liquidity transformation $\mathcal{G}_{\mathrm{liq}}$ 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 linear-REE expectation $E_0[P_E^1] = \mu - \Delta_0$ already integrates across the normal and mild-failure regimes. Fire-sale loss $\mathcal{G}_{\mathrm{fire}}$ has two components, both loaded on $\pi_s = \Pr(\text{severe})$: the total severe-state basket fire-sale loss $d_H \bar{P} + D$ paid by patient long-term 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 wrapper raises welfare iff $\mathcal{G}_{\text{liq}} > \mathcal{G}_{\text{fire}} + \phi(1 - F(\psi^*))$. In the normal regime ($\pi_s = 0$) the fire-sale channel vanishes, so the wrapper strictly raises welfare whenever the liquidity gain $g\lambda$ exceeds the aggregate expense $\phi(1 - F(\psi^*))$.

9.3 The externality wedge

The equilibrium threshold c^* is set by marginal-type indifference (28). A planner who picks c^* to maximise $\mathscr{W}$ internalises a channel that the marginal type does not. The wedge (37) below is the *investor-side* welfare gradient, $d\mathscr{W}^{\text{inv}}/dc^*$; the corresponding total-surplus gradient $d\mathscr{W}^{\text{total}}/dc^* = d\mathscr{W}^{\text{inv}}/dc^* + n\,\partial\Pi^{AP}/\partial c^*$ adds an aggregate AP-profit term (mass n of identical APs) that partially offsets the wrapper-spread channel and is treated in Appendix A.4. Differentiating $\mathscr{W}$ with respect to c^*:

$$\frac{d\mathscr{W}}{dc^*} = \underbrace{(\text{effect via marginal type})}_{=0 \text{ at } c^*_{\text{eq}} \text{ by envelope}} + \underbrace{\left[-\,\kappa\,\frac{\partial\big((d_H\bar{P} + D)\,\Pr(\text{severe} \mid c^*)\big)}{\partial c^*}\right]}_{\text{externality on inframarginal patient holders}} + \underbrace{\left[-\,\lambda\,\frac{\partial\Delta_0}{\partial c^*}\right]}_{\text{externality on inframarginal shocked-wrapper sellers}} + \underbrace{\left[-\,\lambda_b\,D\,\frac{\partial\pi_s}{\partial c^*}\right]}_{\text{externality on inframarginal shocked-direct sellers}}. \tag{37}$$

($\mathscr{W}$ is the $T = 0$ ex-ante expectation (35), so the shocked-seller externalities run through the realised OTC sale prices: for shocked wrapper sellers through $E[P_E^1] = \mu - \Delta_0$, and for shocked direct sellers through the OTC sale price $\mu - d_H\bar{P} - D\pi_s$.) At the equilibrium c^*_{eq} the first term vanishes (the marginal type is indifferent), but the second, third, and fourth terms are all strictly negative. Raising c^* pushes more long-term holders into the SELL region, which raises wrapper-side flow and brings A closer to A^{liq}. 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 (37)). 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 (37); 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 (37)). All three groups bear pecuniary losses that the marginal long-term holder ignores when she crosses the indifference threshold, so the channels reinforce: the equilibrium c^* 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 loads on c^* in the mild regime, i.e., as long as wrapper-side flow responds to long-term-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.

9.4 AP capacity and the under-provision of liquidity

The externality wedge in Section 9.3 treats AP capacity χ as fixed and asks how the planner would shift the long-term-holder cut-off c^*. 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 (35) depends on χ through two equilibrium objects: the wrapper-vs-NAV spread Δ_0 and the severe-failure probability π_s. Higher χ lowers Δ_0 directly (Corollary 1, $\partial\Delta_0/\partial\chi < 0$), and it lowers π_s through the equilibrium feedback in (28), because a tighter Δ_0 raises expected P_E^1, lowers c^*, and shrinks the SELL region of long-term holders. (Higher χ also lowers the mild-failure probability π_m, but π_m does not appear in welfare derivatives because the linear-REE expectation $E_0[P_E^1] = \mu - \Delta_0$ already integrates across the normal and mild-failure regimes.) Substituting:

$$\frac{d\mathscr{W}}{d\chi} \;=\; \underbrace{-\,\lambda\,\frac{\partial\Delta_0}{\partial\chi}}_{\text{tighter wrapper-vs-NAV spread}} \;+\; \underbrace{-\,\big[(d_H\bar{P} + D)\,\kappa + D\,\lambda_b\big]\,\frac{\partial\pi_s}{\partial\chi}}_{\text{lower severe-failure probability}}\,. \tag{38}$$

Both partials are negative, so both terms are positive: within the unique-equilibrium region (Proposition 1), larger AP capacity is welfare-improving on every margin. The substitution- versus-information condition $1/\Lambda > (d_H\bar{P} + D)/(\sqrt{2\pi}\,\hat{\Sigma})$ defines the boundary of this region; raising χ indefinitely eventually moves Λ outside it, beyond which the welfare gradient (38) ceases to be the relevant object because the equilibrium itself is no longer unique.[4]

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 dealer-APs choose balance-sheet capacity $\chi \geq 0$ at constant marginal cost $\rho_\chi > 0$. A per-unit regulatory subsidy $s_\chi \geq 0$ to AP balance sheet lowers the effective marginal cost to $\rho_\chi - s_\chi$. At the $T = 1$ optimum $\tilde{W}^*(\chi) = \chi\Delta_A - d_h$ from (19), the per-realisation arbitrage profit is $\pi_1^{AP}(\chi) = (\chi/2)(\Delta_A - d_h/\chi)^2$, and the $T = -1$ expected profit net of capacity cost is

$$\Pi^{AP}(\chi; s_\chi) \;=\; \tfrac{1}{2}\,E_{-1}[\chi(\Delta_A - d_h/\chi)^2] \;-\; (\rho_\chi - s_\chi)\,\chi. \tag{39}$$

Lemma 1 (AP entry optimum)**.** *The AP's first-order condition $\Pi^{AP\prime}(\chi; s_\chi) = 0$ has a unique interior solution $\chi_{eq}(s_\chi)$, strictly increasing in s_χ.*

Proof. See Appendix A.3. □

The AP internalises its own profit gradient $\partial\Pi^{AP}/\partial\chi$ but not the investor-welfare gradient $d\mathscr{W}/d\chi$ from (38), where $\mathscr{W}$ is the investor-side aggregate of (35). There is a continuum of identical APs of mass n, so aggregate AP profit is $n\Pi^{AP}$ and total surplus is $\mathscr{W}^{\text{inv}} + n\Pi^{AP}$ (App A.4); the planner condition is $\partial\mathscr{W}^{\text{inv}}/\partial\chi + n\,\partial\Pi^{AP}/\partial\chi = 0$ while each AP's private condition is $\partial\Pi^{AP}/\partial\chi = 0$, so the per-AP wedge is $\partial\mathscr{W}^{\text{inv}}/\partial\chi$ divided by n. The AP's gross marginal revenue from one more unit of capacity is the expected spread Δ_0 net of the per-share OTC unwinding cost, integrated over its own arbitrage volume; as χ rises Δ_0 falls, so gross marginal revenue declines along the schedule and the AP equates it with the marginal capacity cost $\rho_\chi - s_\chi$. Three components of $d\mathscr{W}/d\chi$ 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 for the κ-mass of patient holders by reducing the expected basket

[4] (38) is the total derivative of (35) 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-type definition (5). Appendix A.5 gives the full derivation: the total derivative reduces exactly to the partial derivative with respect to χ holding ψ^* fixed, and that partial equals (38) term-by-term.

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 capacity.

The wedge admits a clean policy mapping. A balance-sheet transfer to dealer-APs raises effective χ and closes part of the gap. 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 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.

9.5 Non-redundancy of the two corrective instruments

Sections 9.3 and 9.4 identify two pecuniary externalities: a run externality at the long-term-holder threshold c^*, and an under-provision of dealer-AP warehousing capacity χ. 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 first-best, but two distinct instruments can.

Proposition 2 (Non-redundancy of corrective instruments)**.** *Let $\tau_R \geq 0$ denote a Pigouvian charge on strategic redemption, paid by long-term holders who SELL at $T = 1$, and let $s_\chi \geq 0$ denote a per-unit subsidy to dealer-AP balance-sheet capacity. Let $(c^{*,p}, \chi^p)$ denote the planner's first-best pair, defined by the joint first-order conditions $\partial \mathscr{W}^{total}/\partial c^* = 0$ and $\partial \mathscr{W}^{\,total}/\partial \chi = 0$ at the implied equilibrium. Then:*

- *(i) For any $\tau_R \geq 0$ with $s_\chi = 0$, the implemented χ remains strictly below χ^p whenever the AP under-provision wedge in* (38) *is non-zero.*
- *(ii) For any $s_\chi \geq 0$ with $\tau_R = 0$, the implemented c^* remains strictly above $c^{*,p}$ whenever the externality wedge in Section 9.3 is non-zero.*
- *(iii) There exists a unique pair $(\tau_R^*, s_\chi^*) \geq 0$ that jointly implements $(c^{*,p}, \chi^p)$.*

Proof. See Appendix A.4. □

The economic content of Proposition 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 long-term holder, lowering c^* 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.

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 serves as the public signal on which long-term holders condition their hold-or-sell decision and the dealer-AP conditions its balance-sheet commitment. The dual role is the structural feature that distinguishes market-based intermediation from its bank-based predecessor. In Diamond-Dybvig the intermediary has no market price and coordination runs through the sequential-service rule on the bank's liquid reserves. Here the same price object governs both substitution and information, and the balance between the two determines whether the architecture is fragile. Three structural results follow.

First, fragility hinges on market microstructure rather than on the sequential-service rule. Whether multiple equilibria are possible depends on whether the secondary-market price acts more as a coordination signal for runs or as a penalty on coordinated selling. Smaller wrapper-side absorption makes the price sharply responsive to selling: each unit of strategic redemption pushes the spread wider, dampening selling incentives through the substitution channel. Larger absorption flattens the price response and weakens the penalty; the price then moves predominantly with fundamentals, and a low observed price coordinates runs through the information channel. The substitution-versus-information condition, in the spirit of Ozdenoren and Yuan (2008), makes this precise. The condition depends jointly on wrapper-side absorption capacity Λ and on the precision of private information. Noisy private signals are necessary but not sufficient for uniqueness; the structural absorption-versus-penalty condition must hold as well. The absorption-capacity margin is what is new relative to standard global-games models of coordination.

Second, the market intermediaries that operate the wrapper, namely dealer-APs and market makers, are not just arbitrageurs of NAV deviations. They are the channel through which the substitution role of prices operates. Wrapper-side absorption capacity has two parts: market-maker depth and aggregate AP balance-sheet capacity. AP balance-sheet capacity is the new structural primitive of this architecture, analogous to bank reserves in Diamond-Dybvig but operating through inventory rather than through par convertibility. Within the unique-equilibrium region, higher AP capacity tightens the level of the equilibrium wrapper-vs-NAV spread, lowers the mild-failure probability, and through equilibrium feedback shrinks the severe region; aggregate investor welfare is unambiguously increasing in AP capacity. The substitution-versus-information condition determines the boundary of the region.

Third, market-based liquidity transformation generates two distinct pecuniary externalities that map to two distinct corrective instruments. A run externality arises at the long-term-holder SELL cutoff: a holder who switches from HOLD to SELL ignores three inframarginal losses she imposes on others, on patient holders (through a higher severe-failure probability), on shocked wrapper sellers (through a deeper wrapper-vs-NAV spread), and on shocked direct sellers (through the basket fire-sale impairment). A Pigouvian charge on strategic redemption addresses this wedge. A second externality is the under-provision of AP warehousing: the AP equates the private return from one more unit of arbitrage (the spread on its own volume) to the balance-sheet cost of holding inventory, but does not count the inframarginal-trader gains from a tighter spread and a lower severe-failure probability. A subsidy to AP balance-sheet capacity addresses this wedge. The two wedges share a common reason: the AP and the marginal selling holder each see only the spread on their own trade, and neither sees the spillover on every other inframarginal trader. A single instrument cannot implement the planner's first-best; both are needed. The closest institutional analogue to the AP balance-sheet subsidy is the Federal Reserve's March 2020 Secondary Market Corporate Credit Facility, although the SMCCF operated through secondary-market purchases that relieved dealer-AP inventory pressure indirectly rather than through direct capital infusion. Bank-era policy tools, deposit insurance and reserve requirements, address neither margin: a market-priced wrapper run depletes dealer-AP balance-sheet capacity and the price's substitution role, not the intermediary's liquid reserves.

The model predicts that wrappers backed by deeper underlying-securities markets, with more APs and stronger market-maker depth, should be less prone to discount runs. Mapping this to the cross-section of corporate bond, bank loan, CLO, and crypto ETFs is a natural empirical exercise.

More broadly, a central role of the market price in any modern financial market is to aggregate dispersed information held by heterogeneous participants and to broadcast it as a public signal that those participants then condition on. As liquidity transformation migrates from banks to markets, this role becomes essential: the same price that clears the wrapper market also informs the decisions of every participant who observes it. A model of market-based financial intermediation that suppresses the price's informational role abstracts away from what makes markets different from banks. The dual-role framework introduced here is one way to bring this role inside the model; future work on bond ETFs, CLO ETFs, stablecoins, and other non-bank wrappers should incorporate it.

References

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

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.

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.

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.

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.

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.

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.

A Proofs

A.1 Proof of Proposition 1

Define the shorthands

$$X \equiv n\chi\beta_p^{AP} + 1/\sigma_s, \qquad Y \equiv n\chi(\beta_p^{AP} + \beta_a^{AP}) + 1/\sigma_s, \qquad W \equiv 1/\sigma_s, \qquad \Lambda \equiv \alpha + n\chi.$$

We first establish a lemma giving the matching coefficients of the price function and the self-consistency for k. The proof of Proposition 1 follows.

Lemma A.1 (Matching coefficients, AP precision, and self-consistency for k)**.** *In the wrapper-discount region of the linear-threshold equilibrium, the coefficients of the price function $p = \mathcal{A} + \mathcal{B}\theta + \mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A$ satisfy*

$$\begin{aligned} \Lambda\mathcal{A} &= -(c^*/\sigma_s + nd_h), \\ \Lambda\mathcal{B} &= Y + k/\sigma_s, \\ \mathcal{D} &= \sigma_\xi\,\mathcal{B}\,X/(YW), \\ \mathcal{G} &= n\chi\beta_a^{AP}\sigma_a\,\mathcal{B}/Y. \end{aligned} \tag{40}$$

The AP's filtered-signal precision simplifies to

$$\tau_p^{AP} = \frac{W^2}{\sigma_\xi^2} = \frac{(1/\sigma_s)^2}{\sigma_\xi^2},$$

which depends only on σ_s and σ_ξ. The non-AP public-signal precision is

$$\tau_p = \frac{Y^2\,\tau_p^{AP}}{X^2 + \tau_p^{AP}\,(n\chi\beta_a^{AP}\sigma_a)^2}, \tag{41}$$

and the threshold-weight self-consistency $k = \tau_p/\tau_s$ from (E.2) is equivalent to

$$k\,\tau_s\Big[X^2 + \tau_p^{AP}\,(n\chi\beta_a^{AP}\sigma_a)^2\Big] = Y^2\,\tau_p^{AP}. \tag{42}$$

The strategic-mass parameter κ does not appear in any of these expressions because it cancels in clearing (22)*; the mechanical-mass parameter λ does not appear because the additive λ on each side of clearing cancels.*

Proof of Lemma A.1. We derive the matching equations from the exact clearing equation (23) sequentially.

First, the Φ structures cancel. The MM-demand specification (11) has $\kappa\Phi(\cdot)$ on demand mirroring the $\kappa\Phi(\cdot)$ strategic-selling mass on supply, and the additive λ on demand mirroring the additive λ mechanical-selling baseline on supply. Both pairs cancel in (22), reducing it to the argument-equating form (23):

$$-\alpha(P_E^1 - \mu) + \sigma_\xi\tilde{\xi} + n\tilde{W} = (c^* - ky - \theta)/\sigma_s.$$

Second, substitute $n\tilde{W}$, $\tilde{s}_A$, and y^{AP}. Use $n\tilde{W} = n\chi(\beta_p^{AP}y^{AP} + \beta_a^{AP}\tilde{s}_A - p) - nd_h$ from (19), $\tilde{s}_A = \theta + \sigma_a\tilde{\varepsilon}_A$, and the unit-coefficient form

$$y^{AP} = \theta + \frac{\mathcal{D}\tilde{\xi}}{\mathcal{B}(1-r)}, \qquad r \equiv \mathcal{G}/(\mathcal{B}\sigma_a).$$

The filter zeroes out the $\tilde{\varepsilon}_A$ component of the price; the AP-side $\tilde{\varepsilon}_A$ loading in clearing is exactly $n\chi\beta_a^{AP}\sigma_a$.

Third, substitute the public signal y in the holder term using $y = \theta + (\mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A)/\mathcal{B}$.

Fourth, collect coefficients. Combining the first three substitutions, with $p = P_E^1 - \mu$:

$$\begin{aligned}\Lambda\, p \;&=\; \big[n\chi(\beta_p^{AP} + \beta_a^{AP}) + (1+k)/\sigma_s\big]\theta \\ &\quad+ \Big[\sigma_\xi + \frac{n\chi\beta_p^{AP}}{1-r}\,\frac{\mathcal{D}}{\mathcal{B}} + \frac{k}{\sigma_s}\,\frac{\mathcal{D}}{\mathcal{B}}\Big]\tilde{\xi} \\ &\quad+ \Big[n\chi\beta_a^{AP}\sigma_a + \frac{k}{\sigma_s}\,\frac{\mathcal{G}}{\mathcal{B}}\Big]\tilde{\varepsilon}_A \;-\; c^*/\sigma_s - nd_h.\end{aligned}$$

Match to $p = \mathcal{A} + \mathcal{B}\theta + \mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A$.

Fifth, match the constant and θ:

$$\Lambda\mathcal{A} \;=\; -(c^*/\sigma_s + nd_h), \qquad \Lambda\mathcal{B} \;=\; n\chi(\beta_p^{AP} + \beta_a^{AP}) + (1+k)/\sigma_s \;=\; Y + k/\sigma_s,$$

so $\Lambda\mathcal{B} - k/\sigma_s = Y$.

Sixth, match $\tilde{\varepsilon}_A$ and compute r and $1-r$:

$$\Lambda\mathcal{G} \;=\; n\chi\beta_a^{AP}\sigma_a + \frac{k}{\sigma_s}\,\frac{\mathcal{G}}{\mathcal{B}} \;\Rightarrow\; \mathcal{G}(\Lambda\mathcal{B} - k/\sigma_s)/\mathcal{B} \;=\; n\chi\beta_a^{AP}\sigma_a \;\Rightarrow\; \mathcal{G} \;=\; \frac{n\chi\beta_a^{AP}\sigma_a\,\mathcal{B}}{Y}.$$

Then

$$r \;=\; \frac{\mathcal{G}}{\mathcal{B}\sigma_a} \;=\; \frac{n\chi\beta_a^{AP}}{Y}, \qquad 1-r \;=\; \frac{Y - n\chi\beta_a^{AP}}{Y} \;=\; \frac{X}{Y}, \qquad \frac{1}{1-r} \;=\; \frac{Y}{X}.$$

Seventh, match $\tilde{\xi}$ and compute precisions. The $\tilde{\xi}$ matching gives

$$\mathcal{D}\left(\Lambda\mathcal{B} - \frac{n\chi\beta_p^{AP}}{1-r} - \frac{k}{\sigma_s}\right) \;=\; \sigma_\xi\,\mathcal{B}.$$

Using the constant-and-θ matching and the $\tilde{\varepsilon}_A$ matching above,

$$\Lambda\mathcal{B} - \frac{n\chi\beta_p^{AP}}{1-r} - \frac{k}{\sigma_s} \;=\; Y - \frac{n\chi\beta_p^{AP}\,Y}{X} \;=\; Y\,\frac{X - n\chi\beta_p^{AP}}{X} \;=\; \frac{YW}{X},$$

so $\mathcal{D} = \sigma_\xi\,\mathcal{B}\,X/(YW)$. By construction $\tau_p^{AP} = \mathcal{B}^2(1-r)^2/\mathcal{D}^2$. Substituting:

$$\tau_p^{AP} \;=\; \frac{\mathcal{B}^2X^2/Y^2}{\sigma_\xi^2\mathcal{B}^2X^2/(YW)^2} \;=\; \frac{W^2}{\sigma_\xi^2} \;=\; \frac{1/\sigma_s^2}{\sigma_\xi^2}.$$

The non-AP precision is $\tau_p = \mathcal{B}^2/(\mathcal{D}^2 + \mathcal{G}^2)$. Substituting,

$$\mathcal{D}^2 + \mathcal{G}^2 \;=\; \frac{\mathcal{B}^2}{Y^2}\left[\frac{X^2\sigma_\xi^2}{W^2} + (n\chi\beta_a^{AP}\sigma_a)^2\right] \;=\; \frac{\mathcal{B}^2}{Y^2}\left[\frac{X^2}{\tau_p^{AP}} + (n\chi\beta_a^{AP}\sigma_a)^2\right],$$

so

$$\tau_p \;=\; \frac{Y^2\,\tau_p^{AP}}{X^2 + \tau_p^{AP}\,(n\chi\beta_a^{AP}\sigma_a)^2}.$$

Multiplying $k = \tau_p/\tau_s$ by τ_s and clearing the denominator gives (42). □

We now prove Proposition 1. The equilibrium is a tuple $(\psi^*, k, c^*, (\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G}), \tilde{W}^*, \Delta_0, \theta_m^{**}, \theta_s^{**}, \kappa, \lambda, \lambda_b)$ satisfying (E.1)–(E.6) of Definition 1. Throughout we hold $(\kappa, \lambda, \lambda_b)$ at their equilibrium values from (E.1) and treat ψ^* as the residual that closes (E.1).

First, τ_p^{AP} and the AP posterior weights are pinned down by parameters. By Lemma A.1, $\tau_p^{AP} = (1/\sigma_s)^2/\sigma_\xi^2$ depends only on σ_s and σ_ξ. The AP posterior weights $\beta_p^{AP} = \tau_p^{AP}/(\tau_\theta + \tau_p^{AP} + \tau_a)$ and $\beta_a^{AP} = \tau_a/(\tau_\theta + \tau_p^{AP} + \tau_a)$ are therefore pure parameter functions; they do not depend on κ, k, or c^*. The shorthands X and Y are similarly pinned down by parameters alone.

Second, k^* exists and is unique. By Lemma A.1, the self-consistency equation (42) pins down k. Since $X^2 + \tau_p^{AP}(n\chi\beta_a^{AP}\sigma_a)^2 > 0$ and $Y^2\tau_p^{AP} > 0$ under $\sigma_a > 0$, (42) has the unique positive solution

$$k^* = \frac{Y^2\,\tau_p^{AP}}{\tau_s\Big[X^2 + \tau_p^{AP}(n\chi\beta_a^{AP}\sigma_a)^2\Big]}. \tag{43}$$

By the parameter-pinning above, k^* is a pure parameter function (independent of κ and c^*).

Third, c^* exists and is unique given k^*. The marginal-type indifference condition (28) is

$$E[P_E^1 \mid z^*] - \mu - z^* + (d_H\bar{P} + D)\cdot \Pr(\text{severe} \mid z^*) = 0, \tag{44}$$

where $z^* = \tau_s c^*/\hat{\tau}$ and $\hat{\tau} = \tau_\theta + \tau_s + \tau_p$ with $\tau_p = k^*\tau_s$. In a symmetric equilibrium the individual cutoff equals the aggregate cutoff, both equal to c^*. Both objects in (44) therefore depend on c^* through three channels: the marginal type's posterior $z^*(c^*)$, the equilibrium price intercept $\mathcal{A}(c^*)$, and the conditional severe-failure probability $\Pr(\text{severe} \mid z^*; c^*)$. Under the severe-regime definition in Section 7.1, the severe event is the joint event that the mild-failure condition and the AUM-closure condition are both met. Under the regime-defining condition $\theta_{\text{AUM}}^{**} < \theta_m^{**}$ maintained throughout the paper (Section 7.1), the AUM-closure condition is the binding trigger, so $\Pr(\text{severe}) = \Pr(\theta + ky < c^* - \sigma_s\Phi^{-1}(A^{\text{liq}}))$ from (33).

Define

$$H(c^*) \equiv E[P_E^1 \mid z^*(c^*); c^*] - \mu - z^*(c^*) + (d_H\bar{P} + D)\,\Pr(\text{severe} \mid z^*(c^*);\, c^*),$$

where the second argument of $E[P_E^1 \mid z^*; c^*]$ and $\Pr(\text{severe} \mid z^*; c^*)$ flags the aggregate dependence on c^*. Existence follows from continuity and boundary behaviour: H is continuous, $-z^*(c^*) = -\tau_s c^*/\hat{\tau}$ dominates at extremes, so $H(c^*) \to +\infty$ as $c^* \to -\infty$ and $H(c^*) \to -\infty$ as $c^* \to +\infty$. The intermediate-value theorem gives at least one solution.

For uniqueness, we take the *total* derivative of H with respect to c^*. Using $S = \tilde{s}_h + ky$ as the marginal type's sufficient statistic (with $S = c^*$ at indifference), joint normality gives $E[\theta \mid S = c^*] = \tau_s c^*/\hat{\tau}$ and $E[y \mid S = c^*] = c^*/(1+k)$, where $\hat{\tau} = \tau_\theta + \tau_s(1+k)$. Since the wrapper price reduces to $P_E^1 = \mu + \mathcal{A} + \mathcal{B}y$ once the noise is absorbed into y, the conditional expected price has slope $\partial E[P_E^1 \mid S = c^*]/\partial c^* = \mathcal{B}/(1+k)$. The derivative decomposes into three channels:

$$H'(c^*) = \underbrace{\Big(\frac{\mathcal{B}}{1+k} - \frac{\tau_s}{\hat{\tau}}\Big) + (d_H\bar{P} + D)\,\frac{\partial \Pr(\text{severe} \mid z^*)}{\partial z^*}\,\frac{\tau_s}{\hat{\tau}}}_{\text{(a) individual / posterior channel}} + \underbrace{\frac{\partial \mathcal{A}}{\partial c^*}}_{\text{(b) price-intercept channel}} + \underbrace{(d_H\bar{P} + D)\,\frac{\partial \Pr(\text{severe} \mid z^*; c^*)}{\partial c^*}\bigg|_{\text{aggregate}}}_{\text{(c) strategic-complementarity channel}}. \tag{45}$$

Channel (a) is the direct effect through the marginal type's signals. The first piece $\mathcal{B}/(1+k) - \tau_s/\hat{\tau}$ compares the slope of the expected price in c^* with the slope of the expected NAV; because $\tau_s/\hat{\tau} <$

$1/(1+k)$ (the posterior puts non-zero weight on the prior τ_θ), the first piece can be strictly positive even when $\mathcal{B} \leq 1$. The second piece is the severe-probability slope: $\partial \Pr(\text{severe} \mid z^*)/\partial z^* < 0$ (higher posterior on θ lowers the conditional severe probability), so the second piece is non-positive.

Channel (b) is the price-intercept channel. From (30), $\mathcal{A} = -(c^*/\sigma_s + nd_h)/\Lambda$, so

$$\frac{\partial \mathcal{A}}{\partial c^*} \;=\; -\frac{1}{\sigma_s \Lambda} \;<\; 0.$$

Higher c^* raises the strategic-selling argument $(c^* - ky - \theta)/\sigma_s$ in the exact clearing equation (23) by $1/\sigma_s$ per unit; the MM absorbs this through a wider discount, lowering $E[P_E^1 \mid z^*]$. The Φ-cancellation in (22) means the strategic-mass parameter κ does not appear in this channel.

Channel (c) is the run feedback. Under the AUM-trigger formulation of severe failure, the severe event boundary $\theta + ky = c^* - \sigma_s \Phi^{-1}(A^{\text{liq}})$ shifts at slope 1 in c^*. The direct partial derivative is

$$\frac{\partial \Pr(\text{severe} \mid z^*; c^*)}{\partial c^*} \;=\; f_{\theta+ky|z^*}(c^* - \sigma_s \Phi^{-1}(A^{\text{liq}})),$$

where $f_{\theta+ky|z^*}$ is the conditional density of the random variable $\theta + ky$ given the marginal type's posterior z^*. Substituting $y = \theta + (\mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A)/\mathcal{B}$ gives $\theta + ky = (1+k)\theta + (k/\mathcal{B})(\mathcal{D}\tilde{\xi} + \mathcal{G}\tilde{\varepsilon}_A)$, which has conditional standard deviation $\Sigma(\sigma_s)$ combining the prior standard deviation of θ and the noise contributions (an exact expression is derived below). Channel (c) is therefore bounded by

$$\big|\text{Channel (c)}\big| \;\leq\; \frac{d_H \bar{P} + D}{\sqrt{2\pi}\,\Sigma(\sigma_s)},$$

the supremum of a Gaussian density times the slope-1 boundary shift. The sign is strictly positive whenever the severe regime has positive measure.

Channel (b) is strictly negative and Channel (c) is strictly positive. Channel (a)'s second piece is non-positive. Its first piece $\mathcal{B}/(1+k) - \tau_s/\hat{\tau}$ can be of either sign: under hypothesis (1) of Proposition 1, $\mathcal{B} \leq 1$ bounds it above by $1/(1+k) - \tau_s/\hat{\tau}$, but this bound does not sign it. Channel (c) is bounded above in absolute value by $(d_H \bar{P} + D)/(\sqrt{2\pi}\,\Sigma(\sigma_s))$, where $\Sigma(\sigma_s)$ is the conditional standard deviation of $X \equiv \theta + ky$ given the marginal type's information. Treat X as the quantity of interest and observe that the marginal type's noisy information $(\tilde{s}_h = c^*, y)$ is equivalent to a single noisy signal $S = \tilde{s}_h + ky = X + \sigma_s \tilde{\varepsilon}_h$ about X with noise variance σ_s^2 (since y is observed, ky is known, and the only residual noise in the signal about X is $\sigma_s \tilde{\varepsilon}_h$). Standard normal signal extraction gives

$$\Sigma^2(\sigma_s) \;=\; \frac{\mathrm{Var}(X)\,\sigma_s^2}{\mathrm{Var}(X) + \sigma_s^2}, \qquad \mathrm{Var}(X) \;=\; \frac{(1+k)^2}{\tau_\theta} \;+\; k^2\,\frac{\mathcal{D}^2 + \mathcal{G}^2}{\mathcal{B}^2}, \tag{46}$$

evaluated at the equilibrium k and price coefficients. Equivalently, in precision form, $1/\Sigma^2 = 1/\mathrm{Var}(X) + 1/\sigma_s^2$. Note $\Sigma^2 \leq \sigma_s^2$ always: more information can only reduce conditional variance.

A sufficient condition for $H'(c^*) < 0$ is that Channel (b) dominates the sum of Channel (c)'s upper bound and Channel (a)'s positive part:

$$\frac{1}{\sigma_s \Lambda} \;>\; \frac{d_H \bar{P} + D}{\sqrt{2\pi}\,\Sigma(\sigma_s)} \;+\; \max\!\left\{0,\; \frac{\mathcal{B}}{1+k} - \frac{\tau_s}{\hat{\tau}}\right\}. \tag{47}$$

Both terms on the right-hand side are deterministic functions of model primitives evaluated at the equilibrium $(k, \mathcal{B}, \tau_s, \hat{\tau})$. Multiplying both sides by σ_s and defining the normalised conditional standard deviation $\hat{\Sigma}(\sigma_s) \equiv \Sigma(\sigma_s)/\sigma_s$, the condition takes the equivalent form

$$\frac{1}{\Lambda} \;>\; \frac{d_H \bar{P} + D}{\sqrt{2\pi}\,\hat{\Sigma}(\sigma_s)} \;+\; \sigma_s\,\max\!\left\{0,\; \frac{\mathcal{B}}{1+k} - \frac{\tau_s}{\hat{\tau}}\right\}, \tag{48}$$

which we call the substitution-versus-run-feedback condition. The left-hand side is the inverse of the wrapper-side absorption capacity and measures the strength of the price's substitution role: a small Λ means each unit of strategic redemption induces a sharp price drop and a strong penalty on sellers. The right-hand side measures the run-feedback payoff, with $\hat{\Sigma}$ controlling the sharpness of the AUM-closure boundary in c^*-space. Both sides are explicit continuous functions of model primitives at given σ_s; (48) is checked directly, not as a limit. Under (48), $H'(c^*) < 0$ throughout, so the solution to $H(c^*) = 0$ is unique. Call it $c^*(\kappa)$.

Fourth, ψ^* exists and is unique, with aggregate consistency. From the AP-precision pinning (E.2 and E.6), the price coefficients $(\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G})$ are pinned down by c^* (they do not depend on κ, which cancels in clearing). The pair $(k^*, c^*(\kappa))$ behaves as follows: k^* is a pure parameter function (from the parameter-pinning that fixes τ_p^{AP} in Lemma A.1), while c^* varies with κ through the run-feedback channel (45) in the closure flow $\mathcal{L} = \kappa(1 - A)$. Substituting back, the failure probabilities π_m, π_s, the wrapper-vs-NAV spread Δ_0, and the basket fire-sale loss D at equilibrium are all functions of κ. The marginal-investor condition (E.1) becomes

$$\Delta(\psi^*; \kappa) \;=\; 0, \tag{49}$$

with Δ from (4). Under the sorting condition stated in Section 4.1, Δ is strictly increasing in ψ, so for each κ there exists a unique $\psi^*(\kappa)$ solving (49).

Aggregate consistency (E.1) requires $\kappa = \int_{\psi^*(\kappa)}^{1} (1 - \psi)\, dF(\psi)$ and similarly for λ, λ_b. Define $T(\kappa) \equiv \int_{\psi^*(\kappa)}^{1} (1 - \psi)\, dF(\psi)$. $T(\kappa)$ is continuous in κ (since ψ^* is continuous in κ), bounded in $[0, E_F[1 - \psi]]$, so $T : [0, \bar{\kappa}] \to [0, \bar{\kappa}]$ for $\bar{\kappa} = E_F[1 - \psi]$. By Brouwer's fixed-point theorem, T has at least one fixed point.

For uniqueness of the fixed point we derive a contraction condition on $T(\kappa)$ explicitly. Differentiate (49) implicitly:

$$\frac{d\psi^*}{d\kappa} \;=\; -\left.\frac{\partial\Delta/\partial\kappa}{\partial\Delta/\partial\psi}\right|_{(\psi^*, \kappa)}.$$

By the sorting condition, $\partial\Delta/\partial\psi > 0$ uniformly on $[0, 1]$. The κ-channel enters Δ through $\pi_s(\kappa)$ via the AUM trigger $A^{\text{liq}} = 1 - \bar{A}/(\kappa\mu)$. By Corollary 2, $\partial\theta_s^{**}/\partial\kappa < 0$: a larger long-term-holder mass raises the closure threshold A^{liq} and so requires more selling to trigger closure, shrinking the severe region and lowering π_s. The π_s-channel contributes

$$\left[(1 - \psi)(d_H \bar{P} + D) - \psi D\right] \cdot |\partial\pi_s/\partial\kappa|$$

to $\partial\Delta/\partial\kappa$. In the relevant range $\psi < (d_H \bar{P} + D)/(d_H \bar{P} + 2D)$ the bracket is strictly positive (patient-holder loss dominates), so $\partial\Delta/\partial\kappa > 0$: more long-term holders make the wrapper safer for the marginal type. This is strategic complementarity. Consequently $d\psi^*/d\kappa < 0$ and

$$T'(\kappa) \;=\; -(1 - \psi^*(\kappa))\, f(\psi^*(\kappa))\, \frac{d\psi^*}{d\kappa} \;>\; 0,$$

so T is strictly increasing in κ.

We impose the explicit contraction condition

$$(1 - \psi^*(\kappa))\, f(\psi^*(\kappa))\, \left.\frac{\partial\Delta/\partial\kappa}{\partial\Delta/\partial\psi}\right|_{(\psi^*(\kappa), \kappa)} \;<\; 1 \quad \text{for all } \kappa \in [0, \bar{\kappa}]. \tag{50}$$

The left-hand side is the per-round gain to the κ-feedback: a unit increase in κ lowers π_s, making the wrapper safer; the marginal cutoff ψ^* then falls, and the long-term-holder mass $(1 - \psi^*) f(\psi^*)$

enters the wrapper at rate $|d\psi^*/d\kappa|$. Condition (50) says this per-round gain is strictly below one, so the feedback contracts. The condition is a direct algebraic restriction on primitives via the implicit-function expressions $\partial\Delta/\partial\kappa = [(1-\psi)(d_H\bar{P}+D)-\psi D]\cdot|\partial\pi_s/\partial\kappa|$ and $\partial\Delta/\partial\psi = (d_H\bar{P} + D\pi_s - \Delta_0) + (d_H\bar{P}+D)\pi_s$, together with $\partial\pi_s/\partial\kappa$ evaluated through Corollary 2: $\partial\pi_s/\partial\kappa$ scales with $\bar{A}/(\kappa^2\mu)$, so the feedback is bounded by primitives $(\bar{A}, \mu, \sigma_s, d_H, D, f_{\max})$ at the equilibrium.

Under (50), $T'(\kappa) < 1$ everywhere. Define $G(\kappa) \equiv T(\kappa) - \kappa$; then $G'(\kappa) = T'(\kappa) - 1 < 0$, so G is strictly decreasing. G is continuous, $G(0) = T(0) > 0$ (since $\psi^*(0) < 1$ generically), and $G(\bar{\kappa}) = T(\bar{\kappa}) - \bar{\kappa} \leq 0$. By the intermediate value theorem, G has exactly one zero in $[0, \bar{\kappa}]$, hence T has a unique fixed point κ^*, and $\psi^* = \psi^*(\kappa^*)$ is uniquely determined.

Fifth, failure events are smooth in θ. $\sigma_a > 0$ ensures that $\Pr(\text{mild} \mid \theta)$ and $\Pr(\text{severe} \mid \theta)$ from (31) and the analogous expression for severe failure are continuous Φ-functions rather than step functions. This prevents the kink at $\theta = \theta_r^{**}$ that would otherwise generate multiple equilibria via discontinuous $D\cdot\Pr(\text{severe})$ in (44). Combined with the strict monotonicity of H established above, this delivers the smooth monotone dependence of π_s, π_m on κ that the implicit-function derivation of $\psi^*(\kappa)$ requires.

The five steps above establish existence and uniqueness of $(\kappa^*, \psi^*, k^*, c^*)$. The remaining objects, the price coefficients $(\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G})$, the AP warehouse $\tilde{W}^*$, the equilibrium spread Δ_0, and the failure thresholds $\theta_m^{**}, \theta_s^{**}$, are determined as explicit functions of c^* via (40), (19), (30), and (32)–(34). The aggregates λ, λ_b are determined by ψ^* via (6) and the analogous expression for λ_b. □

A.2 Proofs of corollaries

Throughout, write $N_m = c^*/\sigma_s + nd_h - (\alpha + n\chi)d_H$ for the numerator of (32) and $D_m^{(m)} = n\chi(\beta_p^{AP} + \beta_a^{AP}) + (1+k)/\sigma_s$ for its denominator (mild trigger uses the realised wrapper discount Δ_0). The severe-failure threshold from (34) is $\theta_s^{**} = (c^* - \sigma_s\Phi^{-1}(A^{\text{liq}}))/(1+k)$ with $A^{\text{liq}} = 1 - \bar{A}/(\kappa\mu)$. We take $N_m > 0$, $D_m^{(m)} > 0$, and $A^{\text{liq}} \in (0,1)$ in the parameter region of interest. The mechanical-mass parameter λ does not appear in either threshold (it cancels in clearing).

Proof of Corollary 1. Write $N \equiv c^*/\sigma_s + nd_h$ and $\Lambda = \alpha + n\chi$, so $\Delta_0 = N/\Lambda$. Since $N > 0$ and $\Lambda > 0$,

$$\frac{\partial\Delta_0}{\partial\chi} = -\frac{nN}{\Lambda^2} < 0, \qquad \frac{\partial\Delta_0}{\partial\alpha} = -\frac{N}{\Lambda^2} < 0,$$
$$\frac{\partial\Delta_0}{\partial c^*} = \frac{1}{\sigma_s\Lambda} > 0, \qquad \frac{\partial\Delta_0}{\partial d_h} = \frac{n}{\Lambda} > 0.$$

$\partial\Delta_0/\partial\lambda = 0$ and $\partial\Delta_0/\partial\kappa = 0$ because neither parameter appears in N or Λ. The normal regime condition $\Delta_0 \leq d_H\bar{P}$ is the parameter condition that makes mild failure non-binding (Section 7.1). In that region, a shocked investor selling the wrapper at $T = 1$ obtains the wrapper price $E_0[P_E^1] = \mu + \mathcal{A} = \mu - \Delta_0$, while a shocked direct holder is paid $\bar{P}(1-d_H) = \mu - d_H\bar{P}$ (using $\bar{P} = \mu$ at $\bar{\theta} = 0$). Subtracting gives the wrapper-transformation gain $g = d_H\bar{P} - \Delta_0$, both terms in dollars. □

Proof of Corollary 2. For $\partial\theta_m^{**}/\partial(n\chi)$, write $\theta_m^{**} = N_m/D_m^{(m)}$ and apply the quotient rule: $\partial N_m/\partial(n\chi) = -d_H < 0$ and $\partial D_m^{(m)}/\partial(n\chi) = \beta_p^{AP} + \beta_a^{AP} > 0$, so the numerator is $D_m^{(m)}(-d_H) - N_m(\beta_p^{AP} + \beta_a^{AP}) < 0$, giving $\partial\theta_m^{**}/\partial(n\chi) < 0$. For $\partial\theta_s^{**}/\partial\bar{A}$, $\partial A^{\text{liq}}/\partial\bar{A} = -1/(\kappa\mu) < 0$ and $\partial\Phi^{-1}(A^{\text{liq}})/\partial A^{\text{liq}} = 1/\phi(\Phi^{-1}(A^{\text{liq}})) > 0$, so $\partial\Phi^{-1}/\partial\bar{A} < 0$ and $\partial\theta_s^{**}/\partial\bar{A} = -\sigma_s/(1+k)\cdot\partial\Phi^{-1}/\partial\bar{A} >$

0. For $\partial\theta_s^{**}/\partial\kappa$, $\partial A^{\text{liq}}/\partial\kappa = \bar{A}/(\kappa^2\mu) > 0$, so $\partial\Phi^{-1}/\partial\kappa > 0$ and $\partial\theta_s^{**}/\partial\kappa = -\sigma_s/(1+k)\cdot\partial\Phi^{-1}/\partial\kappa < 0$. For $\partial\theta_r^{**}/\partial c^*$, $\partial N_m/\partial c^* = 1/\sigma_s > 0$ with $\partial D_m^{(m)}/\partial c^* = 0$, giving $\partial\theta_m^{**}/\partial c^* > 0$; and $\partial\theta_s^{**}/\partial c^* = 1/(1+k) > 0$ directly. Finally, λ and λ_b do not appear in θ_s^{**} from (34), so the corresponding partials are zero. □

Proof of Corollary 3. The closed form $\tau_p^{AP} = (1/\sigma_s)^2/\sigma_\xi^2 = 1/(\sigma_s^2\sigma_\xi^2)$ follows from substituting (40) into (15) (derived in the proof of Proposition 1). Differentiating:

$$\frac{\partial\tau_p^{AP}}{\partial\sigma_\xi} = -\frac{2}{\sigma_s^2\sigma_\xi^3} < 0, \qquad \frac{\partial\tau_p^{AP}}{\partial\sigma_s} = -\frac{2}{\sigma_s^3\sigma_\xi^2} < 0.$$

τ_p^{AP} does not depend on κ: the Φ-cancellation removes κ from the price coefficients. $\beta_p^{AP} = \tau_p^{AP}/(\tau_\theta + \tau_p^{AP} + \tau_a)$ satisfies $\partial\beta_p^{AP}/\partial\tau_p^{AP} = (\tau_\theta+\tau_a)/(\tau_\theta+\tau_p^{AP}+\tau_a)^2 > 0$. For the mild threshold: $\partial N_m/\partial\beta_p^{AP} = 0$ and $\partial D_m^{(m)}/\partial\beta_p^{AP} = n\chi > 0$, so $\partial\theta_m^{**}/\partial\beta_p^{AP} = -N_m\cdot n\chi/(D_m^{(m)})^2 < 0$. The severe threshold $\theta_s^{**} = (c^* - \sigma_s\Phi^{-1}(A^{\text{liq}}))/(1+k)$ does not contain β_p^{AP}, so $\partial\theta_s^{**}/\partial\beta_p^{AP} = 0$ in partial equilibrium: the AUM trigger is on holder selling, which does not depend on AP demand. □

Proof of Corollary 4. From (41),

$$\tau_p = \frac{Y^2\tau_p^{AP}}{X^2 + \tau_p^{AP}(n\chi\beta_a^{AP}\sigma_a)^2}.$$

First, we fix β_a^{AP} at its equilibrium value; then since τ_p^{AP}, X, Y do not depend on σ_a, $\partial\tau_p/\partial\sigma_a = -2Y^2\tau_p^{AP}\cdot\tau_p^{AP}(n\chi\beta_a^{AP})^2\sigma_a/(\text{denom})^2 < 0$.

Next, we show the effect on β_a^{AP} (AP-information dilution). $\tau_a = 1/\sigma_a^2$, so $\partial\tau_a/\partial\sigma_a = -2/\sigma_a^3 < 0$, and $\beta_a^{AP} = \tau_a/(\tau_\theta + \tau_p^{AP} + \tau_a)$ satisfies $\partial\beta_a^{AP}/\partial\tau_a = (\tau_\theta + \tau_p^{AP})/(\tau_\theta + \tau_p^{AP} + \tau_a)^2 > 0$. So $\partial\beta_a^{AP}/\partial\sigma_a < 0$. The product $\beta_a^{AP}\sigma_a = \tau_a\sigma_a/(\tau_\theta + \tau_p^{AP} + \tau_a) = 1/[\sigma_a(\tau_\theta + \tau_p^{AP} + \tau_a)]$, so

$$\frac{\partial(\beta_a^{AP}\sigma_a)}{\partial\sigma_a} = -\frac{1}{\sigma_a^2(\tau_\theta + \tau_p^{AP} + \tau_a)} - \frac{1}{\sigma_a(\tau_\theta + \tau_p^{AP} + \tau_a)^2}\cdot\frac{\partial\tau_a}{\partial\sigma_a}.$$

Substituting $\partial\tau_a/\partial\sigma_a = -2/\sigma_a^3$:

$$\frac{\partial(\beta_a^{AP}\sigma_a)}{\partial\sigma_a} = \frac{1}{\sigma_a^2(\tau_\theta + \tau_p^{AP} + \tau_a)}\left[-1 + \frac{2\tau_a}{\tau_\theta + \tau_p^{AP} + \tau_a}\right].$$

The sign of the bracket is positive iff $2\tau_a > \tau_\theta + \tau_p^{AP} + \tau_a$, i.e., $\tau_a > \tau_\theta + \tau_p^{AP}$ (precise AP signal). The sign is negative when $\tau_a < \tau_\theta + \tau_p^{AP}$ (noisy AP signal). The full $\partial\tau_p/\partial\sigma_a$ inherits this ambiguity: when the bracket is negative, the direct noise-floor channel is reinforced ($\beta_a^{AP}\sigma_a$ shrinks in σ_a, but slower than σ_a grows, so the product $(n\chi\beta_a^{AP}\sigma_a)^2$ may rise or fall depending on magnitudes). □

Proof of Corollary 5. The expected basket fire-sale loss is $E[D\cdot\mathbf{1}\{\text{severe}\}] = D\cdot\Pr(\theta < \theta_s^{**})$ by definition. Strict monotonicity in D is immediate. Strict monotonicity in $\bar{A}$ follows from $\partial\theta_s^{**}/\partial\bar{A} > 0$ (Corollary 2), which raises $\Pr(\theta < \theta_s^{**})$ for any non-degenerate prior on θ. Strict monotonicity in κ (decreasing) follows analogously from $\partial\theta_s^{**}/\partial\kappa < 0$. AP capacity χ does not enter θ_s^{**} directly because the AUM trigger (10) is on aggregate holder selling, not on AP-flow capacity. □

A.3 Proof of Lemma 1

The gross-revenue term in (39) is strictly concave in χ at the AP's interior optimum. Expanding expectation gives

$$\Pi^{AP}_{\text{gross}}(\chi) \;=\; \tfrac{\chi}{2}\big[\Delta_0(\chi)^2 + V(\chi)\big] \;-\; d_h\,\Delta_0(\chi) \;+\; \frac{d_h^2}{2\chi},$$

where $\Delta_0(\chi) = \mathbb{E}_{-1}[\Delta_A]$ and $V(\chi) \equiv \mathrm{Var}_{-1}(\Delta_A)$. Write $N \equiv c^*/\sigma_s + nd_h$ for the spread numerator (Corollary 1); from Lemma A.1, $\Delta_0 = N/\Lambda$ with $\Lambda = \alpha + n\chi$, and the noise loadings entering V inherit the same $1/\Lambda$ scaling through $(\mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{G})$, so $\chi V(\chi) = O(1/\chi)$ at large χ and $\chi\Delta_0(\chi)^2 = \chi N^2/\Lambda^2$ is hump-shaped with peak at $\chi = \alpha/n$. The AP's interior optimum lies past this peak under the maintained parametric range, where direct calculation gives $\Pi^{AP\prime\prime}_{\text{gross}}(\chi) < 0$: the dominant concave contribution $-d_h\Delta_0''(\chi) = -2Nd_hn^2/\Lambda^3$ outweighs the convex residual d_h^2/χ^3 whenever $\chi^3 > d_h\Lambda^3/(2Nn^2)$, a mild lower bound on χ within the maintained range. So the first-order condition $\Pi^{AP\prime}(\chi; s_\chi) = 0$ has a unique solution $\chi_{\text{eq}}(s_\chi)$. By the implicit-function theorem,

$$\frac{d\chi_{\text{eq}}}{ds_\chi} \;=\; \frac{1}{|\Pi^{AP\prime\prime}(\chi_{\text{eq}}; s_\chi)|} \;>\; 0,$$

so χ_{eq} is strictly increasing in s_χ and the map $s_\chi \mapsto \chi_{\text{eq}}(s_\chi)$ is a continuous bijection on its range. □

A.4 Proof of Proposition 2

We first show how redemption tax affects holder's indifference condition. A per-unit Pigouvian charge $\tau_R \geq 0$ is paid by long-term holders who SELL at $T = 1$; the SELL payoff becomes $P_E^1 - \tau_R$. The marginal-type indifference condition (28) becomes

$$\tilde{H}(c^*; \tau_R, \chi) \;\equiv\; H(c^*; \chi) - \tau_R \;=\; 0,$$

where H is the indifference deficit from (45). Under the uniqueness condition of Proposition 1, $\partial\tilde{H}/\partial c^* = H'(c^*) < 0$, so the equilibrium $c^*(\tau_R, \chi)$ is uniquely determined. Implicit differentiation of $\tilde{H} = H(c^*; \chi) - \tau_R = 0$ gives

$$\frac{\partial c^*}{\partial \tau_R} \;=\; \frac{1}{H'(c^*)} \;<\; 0, \qquad \frac{\partial c^*}{\partial \chi} \;=\; -\frac{\partial H/\partial\chi}{H'(c^*)},$$

so a higher redemption tax strictly lowers the selling threshold, as expected economically. The sign of $\partial c^*/\partial\chi$ depends on $\partial H/\partial\chi$; what matters for the proof is strict monotonicity of c^* in τ_R at any fixed χ.

We then compute planner's first-order conditions. There is a continuum of identical APs of mass n, each with profit $\Pi^{AP}(\chi; c^*) = \frac{1}{2}E_{-1}[\chi(\Delta_A(c^*, \chi) - d_h/\chi)^2] - \rho_\chi\chi$ from (39), evaluated at $s_\chi = 0$ since the regulatory subsidy is an internal transfer and does not affect total surplus. Aggregate AP profit is $n\Pi^{AP}$. The planner maximises total surplus

$$\mathscr{W}^{\text{total}}(c^*, \chi) \;\equiv\; \mathscr{W}^{\text{inv}}(c^*, \chi) \;+\; n\,\Pi^{AP}(\chi; c^*),$$

where $\mathscr{W}^{\text{inv}}$ is the investor-side aggregate from (35). The first-best pair $(c^{*,\text{P}}, \chi^{\text{P}})$ satisfies

$$\left.\frac{\partial \mathscr{W}^{\text{total}}}{\partial c^*}\right|_{(c^{*,\text{P}},\chi^{\text{P}})} \;=\; 0, \tag{51}$$

$$\left.\frac{\partial \mathscr{W}^{\text{total}}}{\partial \chi}\right|_{(c^{*,\text{P}},\chi^{\text{P}})} \;=\; \frac{\partial \mathscr{W}^{\text{inv}}}{\partial\chi} + n\,\frac{\partial \Pi^{AP}}{\partial\chi} \;=\; 0. \tag{52}$$

The investor-welfare gradient $\partial\mathscr{W}^{\text{inv}}/\partial\chi$ is strictly positive at every χ in the relevant range ((38): both channels are positive). The aggregate AP-profit gradient $n\,\partial\Pi^{AP}/\partial\chi$ is decreasing in χ (concavity), positive at low χ and negative at high χ; the planner's FOC (52) is the equality between the two, pinning down χ^{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 $(c^*(\tau_R,\chi),\chi)$:

$$\text{MR}^{AP}(\chi;\tau_R,s_\chi)\ \equiv\ \frac{\partial\Pi^{AP}_{\text{gross}}}{\partial\chi}\bigg|_{(c^*(\tau_R,\chi),\chi)}\ -\ (\rho_\chi-s_\chi),$$

where $\Pi^{AP}_{\text{gross}}(\chi;c^*)=\frac{1}{2}E_{-1}[\chi(\Delta_A-d_h/\chi)^2]$ is the AP's gross arbitrage revenue. The AP's private FOC is $\text{MR}^{AP}=0$. Define the wedge at the χ margin as the investor-welfare gradient the AP does not internalise:

$$W_\chi(\chi;\tau_R)\ \equiv\ \frac{\partial\mathscr{W}^{\text{inv}}}{\partial\chi}\bigg|_{(c^*(\tau_R,\chi),\chi)},$$

which is exactly the inframarginal-trader spillovers in (38): 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 crucial property is local: $W_\chi(\chi^{\text{p}};\tau_R)>0$ at the planner's χ^{p} for every τ_R in the relevant range $[0,\bar{\tau}_R]$. This holds because at any finite χ both terms of (38) are strictly positive (the spread channel is non-zero whenever $\Delta_0>0$, and the severe-probability channel is non-zero whenever $\pi_s>0$). We make no uniform claim that $W_\chi>0$ for arbitrarily large χ, where both terms vanish; we use $W_\chi>0$ only at the planner's optimum. Using (52), the planner's FOC for χ can be written as $n\,\text{MR}^{AP}(\chi^{\text{p}};\tau_R^*,0)+W_\chi(\chi^{\text{p}};\tau_R^*)=0$.

Define analogously at the c^* margin the part of total-surplus gradient that comes from inframarginal-trader spillovers, isolated from the boundary contribution at the marginal holder:

$$W_{c^*}(c^*;\tau_R,s_\chi)\ \equiv\ \frac{\partial\mathscr{W}^{\text{total}}}{\partial c^*}\bigg|_{(c^*,\chi(s_\chi))}\ -\ \underbrace{[H(c^*;\chi(s_\chi))-\tau_R]\,m(c^*)}_{\text{marginal-holder boundary term}},$$

where $H(c^*)-\tau_R$ is the marginal-holder IC residual (zero on the decentralised IC) and $m(c^*)$ is the density of the marginal type. The boundary term vanishes wherever the marginal holder is indifferent (the envelope condition), so on the decentralised IC $\partial\mathscr{W}^{\text{total}}/\partial c^*=W_{c^*}$. Expanding $\mathscr{W}^{\text{total}}=\mathscr{W}^{\text{inv}}+n\,\Pi^{AP}$, the wedge decomposes as

$$W_{c^*}\ =\ \underbrace{-\kappa\,\frac{\partial[(d_H\bar{P}+D)\,\Pr(\text{severe}\mid c^*)]}{\partial c^*}}_{\text{basket fire-sale loss on patient holders, strictly negative}}\ +\ \underbrace{-\lambda_b\,D\,\frac{\partial\pi_s}{\partial c^*}}_{\text{basket fire-sale loss on shocked direct sellers, strictly negative}}$$
$$+\ \underbrace{-\lambda\,\frac{\partial\Delta_0}{\partial c^*}\ +\ n\,\frac{\partial\Pi^{AP}}{\partial c^*}}_{\text{spread channel, near-cancellation}}\ .$$

(Using $E[P^1_E]=\mu-\Delta_0$ in the $T=0$ aggregate, the marginal-seller term collapses to $-\lambda\,\partial\Delta_0/\partial c^*$, with $\partial\Delta_0/\partial c^*>0$.) The first two terms are strictly negative when the severe regime has positive measure: a higher c^* raises the severe-failure probability π_s, so patient holders bear a deeper expected loss $(d_H\bar{P}+D)\pi_s$ and shocked direct sellers (mass λ_b) bear a deeper expected lump-sum $D\pi_s$, both with D a real deadweight loss to total surplus. The third term is the net spread effect.

By the envelope theorem on each AP's optimisation, $\partial\Pi^{AP}/\partial c^* = \tilde{W}^*\,\partial\Delta_A/\partial c^*$ where $\tilde{W}^*$ is the per-AP arbitrage volume. The threshold c^* enters Δ_A only through the deterministic intercept $\mathcal{A}$ of the linear price function, so $\partial\Delta_A/\partial c^* = -\partial\mathcal{A}/\partial c^* = \partial\Delta_0/\partial c^*$ holds state-by-state. Aggregate AP-side spread sensitivity is therefore $n\,\partial\Pi^{AP}/\partial c^* = n\tilde{W}^*\,\partial\Delta_0/\partial c^*$. The spread channel collapses to

$$-\lambda\,\frac{\partial\Delta_0}{\partial c^*} + n\,\frac{\partial\Pi^{AP}}{\partial c^*} \;=\; -\,(\lambda - n\tilde{W}^*)\,\frac{\partial\Delta_0}{\partial c^*},$$

with $\partial\Delta_0/\partial c^* > 0$. The sign of the residual depends on whether $\lambda \gtrless n\tilde{W}^*$. With the AP at or near capacity in severe states, the natural ordering is $\lambda > n\tilde{W}^*$ (mechanical-seller volume exceeds AP-absorbed volume), so the spread channel is itself non-positive and reinforces the basket fire-sale term. The crucial property is local: $W_{c^*}(c^{*,\mathrm{p}};\tau_R,s_\chi) < 0$ at the planner's $c^{*,\mathrm{p}}$ for every (τ_R,s_χ) in the relevant range $[0,\bar{\tau}_R]\times[0,\bar{s}_\chi]$. A sufficient condition is

$$\Big[\kappa\,\frac{\partial[(d_H\bar{P}+D)\pi_s]}{\partial c^*} \;+\; \lambda_b\,D\,\frac{\partial\pi_s}{\partial c^*}\Big]\Big|_{c^{*,\mathrm{p}}} \;>\; \max\Big\{0,\;(n\tilde{W}^* - \lambda)\,\frac{\partial\Delta_0}{\partial c^*}\Big\}\Big|_{c^{*,\mathrm{p}}}, \tag{53}$$

where the maximum keeps the bound informative both when $n\tilde{W}^* > \lambda$ (spread residual positive, basket fire-sale must dominate) and when $n\tilde{W}^* \le \lambda$ (spread channel already non-positive, the basket fire-sale term alone delivers $W_{c^*} < 0$). We make no uniform claim that $W_{c^*} < 0$ for arbitrarily large τ_R, where c^* falls, $\pi_s \to 0$, and $W_{c^*} \to 0$; we use $W_{c^*} < 0$ only at the planner's optimum.

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

For the direction of the inequality, evaluate MR^{AP} at χ^{p}. From the planner's FOC at $(c^{*,\mathrm{p}},\chi^{\mathrm{p}})$, $\mathrm{MR}^{AP}(\chi^{\mathrm{p}};\tau_R^*,0) = -W_\chi(\chi^{\mathrm{p}};\tau_R^*)/n < 0$. At the implementing tax τ_R^* (see part iii), the AP would earn negative marginal revenue at χ^{p}: each additional unit of capacity at χ^{p} costs the AP more than its private return, because the social return to each AP exceeds the private return by exactly W_χ/n. By strict concavity of Π^{AP} in χ, the AP's privately optimal χ satisfies $\chi_{\mathrm{eq}}(\tau_R^*,0) < \chi^{\mathrm{p}}$. The same conclusion holds for any $\tau_R \in [0,\bar{\tau}_R]$ in the maintained range by continuity of χ_{eq} in τ_R together with the local strict-positivity $W_\chi(\chi^{\mathrm{p}};\tau_R) > 0$: at each τ_R in this range, the AP's marginal revenue evaluated at χ^{p} is strictly negative, so the AP's optimum lies below χ^{p}. □

Proof of part (ii). Set $\tau_R = 0$ and fix any $s_\chi \ge 0$. The holder's indifference condition reduces to $H(c^*;\chi(s_\chi)) = 0$, which pins down $c^* = c^*_{\mathrm{eq}}(0,\chi(s_\chi))$ uniquely under Proposition 1. Suppose for contradiction $c^*_{\mathrm{eq}}(0,\chi(s_\chi)) = c^{*,\mathrm{p}}$. Two equations must hold at $c^{*,\mathrm{p}}$ under this hypothesis. First, the decentralised IC at $\tau_R = 0$: $H(c^{*,\mathrm{p}};\chi(s_\chi)) = 0$, so the marginal-holder IC residual at $c^{*,\mathrm{p}}$ is exactly zero. Second, the planner's FOC (51): $\partial\mathscr{W}^{\mathrm{total}}/\partial c^*|_{c^{*,\mathrm{p}}} = 0$. By the envelope condition, when the IC residual is zero at $c^{*,\mathrm{p}}$ the boundary contribution to the total-surplus gradient vanishes and $\partial\mathscr{W}^{\mathrm{total}}/\partial c^* = W_{c^*}(c^{*,\mathrm{p}};0,s_\chi)$ at this point. Combining: $W_{c^*}(c^{*,\mathrm{p}};0,s_\chi) = 0$, contradicting $W_{c^*} < 0$ at $c^{*,\mathrm{p}}$. So $c^*_{\mathrm{eq}}(0,\chi(s_\chi)) \neq c^{*,\mathrm{p}}$.

For the direction, note that the marginal holder's IC at $c^* = c^{*,\mathrm{p}}$ gives $H(c^{*,\mathrm{p}};\chi(s_\chi)) > 0$ because H is strictly decreasing in c^* (Proposition 1) and the planner's $c^{*,\mathrm{p}}$ is strictly below the marginal-holder c^*_{eq} at $\chi(s_\chi)$ for which $H = 0$. Since the marginal holder requires $H \le 0$ to weakly prefer HOLD, the decentralised $c^*_{\mathrm{eq}}(0,\chi(s_\chi))$ exceeds $c^{*,\mathrm{p}}$ for any $s_\chi \ge 0$. □

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

$$H(c^*;\chi)\ -\ \tau_R\ =\ 0, \tag{54}$$

$$\mathrm{MR}^{AP}(\chi;\tau_R,s_\chi)\ =\ 0, \tag{55}$$

with target $(c^*,\chi)=(c^{*,\mathrm{p}},\chi^{\mathrm{p}})$. (54) at the target gives

$$\tau_R^*\ \equiv\ H(c^{*,\mathrm{p}};\chi^{\mathrm{p}}),$$

which is strictly positive by the argument in part (ii) (H is strictly decreasing in c^* and $c^{*,\mathrm{p}} < c^*_{\mathrm{eq}}(0,\chi^{\mathrm{p}})$). Substituting τ_R^* into (55) at $\chi=\chi^{\mathrm{p}}$:

$$s_\chi^*\ \equiv\ \rho_\chi\ -\ \frac{\partial\Pi^{AP}_{\mathrm{gross}}}{\partial\chi}\bigg|_{(c^{*,\mathrm{p}},\chi^{\mathrm{p}})}\ =\ \frac{W_\chi(\chi^{\mathrm{p}};\tau_R^*)}{n}\ >\ 0,$$

where the second equality uses the planner's FOC (52): $\partial\Pi^{AP}_{\mathrm{gross}}/\partial\chi=\rho_\chi-W_\chi/n$ at the planner's optimum. The per-AP Pigouvian subsidy equals the per-AP share of the marginal inframarginal-trader spillover at the planner's optimum, the textbook Pigouvian formula scaled by the mass n of identical APs. Positivity follows from $W_\chi>0$.

Uniqueness: τ_R^* is the unique solution of (54) at $(c^{*,\mathrm{p}},\chi^{\mathrm{p}})$ by strict monotonicity of H in c^*. Given τ_R^*, s_χ^* is the unique value satisfying (55) at χ^{p} by strict concavity of Π^{AP} (the AP's gross MR is a strictly decreasing function of χ). Both values are strictly positive.

Implementation verification: at (τ_R^*,s_χ^*), the AP's FOC (55) is satisfied at χ^{p}, so the AP chooses χ^{p}. Given χ^{p}, the holder's IC (54) is satisfied at $c^{*,\mathrm{p}}$, so the equilibrium c^* is $c^{*,\mathrm{p}}$. Both planner FOCs are satisfied by definition of $(c^{*,\mathrm{p}},\chi^{\mathrm{p}})$. □

A.5 Derivation of $d\mathscr{W}/d\chi$: sorting-term cancellation

This appendix derives (38) from the investor-welfare aggregate (35), showing that the $T=0$ sorting term vanishes exactly and that the total derivative reduces to the partial holding the $T=0$ partition ψ^* fixed.

We first compute the total derivative. Starting from (35),

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

with $U^D(\psi)=\mu-\psi(d_H\bar{P}+D\pi_s)$ depending on χ through π_s, and $U^W(\psi)=\mu-(1-\psi)(d_H\bar{P}+D)\pi_s-\psi\Delta_0-\phi$ depending on χ through π_s and Δ_0. The total derivative with respect to χ contains four pieces:

$$\frac{d\mathscr{W}}{d\chi}\ =\ \underbrace{-U^D(\psi^*)\,f(\psi^*)\frac{d\psi^*}{d\chi}}_{\text{lower limit moves}}\ +\ \underbrace{U^W(\psi^*)\,f(\psi^*)\frac{d\psi^*}{d\chi}}_{\text{upper limit moves}}\ +\ \underbrace{\int_0^{\psi^*}\frac{\partial U^D}{\partial\chi}\,dF(\psi)}_{\text{Direct integrand}}\ +\ \underbrace{\int_{\psi^*}^1\frac{\partial U^W}{\partial\chi}\,dF(\psi)}_{\text{Wrapper integrand}}. \tag{56}$$

The first two pieces combine to

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

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 (5). 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 c^* to re-equilibrate via (28). 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

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

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

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

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 terms of (38). The χ-derivatives $\partial \Delta_0/\partial \chi$ and $\partial \pi_s/\partial \chi$ here are total derivatives in χ that allow c^* to re-equilibrate via (28) 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. □