Risk-Based Auto-Deleveraging*

Steven Campbell[†]
Columbia University
sc5314@columbia.edu

Natascha Hey[†]
Columbia University
nyh2112@columbia.edu

Ciamac C. Moallemi
Columbia University
ciamac@gsb.columbia.edu

Marcel Nutz
Columbia University
mnutz@columbia.edu

July 16, 2026

Abstract

Auto-deleveraging (ADL) mechanisms are a critical yet understudied component of risk management on cryptocurrency futures exchanges. When available margin and other loss-absorbing resources are insufficient to cover losses following large price moves, exchanges reduce positions and socialize losses among solvent participants via rule-based ADL protocols.

We formulate ADL as an optimization problem that minimizes the exchange's risk of loss arising from future equity shortfalls. In a single-asset, isolated-margin setting, the *minimax leverage* policy — minimizing the maximum leverage among participants — is optimal for all monotone risk measures. This policy has a transparent structure: positions are reduced first for the most highly levered accounts, and leverage is progressively equalized via a water-filling (or "leverage-draining") rule. The policy is distribution-free, wash-trade resistant, Sybil resistant, and path-independent. It provides a canonical and implementable benchmark for ADL design and clarifies the economic logic underlying queue-based mechanisms used in practice.

We further study the multi-asset, cross-margin setting, where the ADL problem becomes genuinely multi-dimensional: the exchange must allocate reductions across accounts with portfolios exposed to correlated price moves. Under the expected loss objective, asset-level shadow prices separate the problem across accounts, yielding a scalable numerical method. Naive gross leverage misleads here, ignoring within-portfolio hedging. When prices are driven by a single risk factor, the optimal policy is again water-filling, but in a factor-adjusted leverage, so better-hedged portfolios are deleveraged less.

We apply the framework to the October 10, 2025 Hyperliquid ADL event. Relative to the exchange's realized allocation, our risk-minimizing allocations achieve lower expected shortfall.

*NH and CCM are supported by the Briger Family Digital Finance Lab at Columbia University. CCM is a research advisor for Paradigm and Uniswap Labs. SC and MN are supported by a Columbia University CDFT Research Grant. SC is further supported by an NSERC Postdoctoral Fellowship (PDF-599675-2025). MN is further supported by NSF Grants DMS-2407074 and DMS-2106056.

[†]These authors contributed equally to this work.

Contents

1. Introduction

Perpetual futures exchanges in cryptocurrency markets allow for extreme leverage and continuous trading, making them vulnerable to large price shocks. When rapid price moves trigger liquidation cascades, insurance funds can be exhausted, and exchanges are forced to reduce the exposure of otherwise solvent traders through auto-deleveraging (ADL). Such events are not hypothetical: during the market-wide shock of October 10–11, 2025, multiple major venues activated ADL mechanisms to preserve solvency, as reported by Chitra [2025]. Months later, on January 30, 2026, over $2.56 billion in leveraged positions were liquidated within a single trading day [Lang, 2026].

In these regimes, when a participant has insufficient account equity, their positions must be unwound. Ideally, this is done through liquidation mechanisms that sell the positions to other market participants at mutually agreeable prices, thereby unwinding distressed positions and containing losses within defaulting accounts. However, in times of extreme duress, there may be no willing buyers in the market, or the prices that are bid may leave the exchange insolvent. In these cases, ADL mechanisms unwind the positions of a distressed participant by forcibly closing the positions of solvent participants on the other side of the market. These forced position closures occur at prices that leave the exchange solvent. In this way, ADL is a mechanism both to reduce positions and to socialize potential losses across market participants.

Similar position reduction and loss-allocation problems arise in traditional financial infrastructure. Central clearinghouses rely on default waterfalls to mutualize losses, but extreme stress can overwhelm these mechanisms. During the March 2022 nickel crisis at the London Metal Exchange (LME), rapidly rising prices threatened the exhaustion of default resources, see [Heilbron, 2025]. Rather than invoking an explicit end-of-waterfall mechanism, the exchange suspended trading and retroactively canceled executed trades — an opaque intervention followed by legal challenges and lasting damage to market confidence.

On perpetual futures exchanges, where forcible position reduction is implemented through ADL, margining may be applied on an isolated basis, where collateral is posted separately for positions in different assets, or under cross-margining, where collateral is pooled across positions in different assets and liquidation decisions become coupled. In practice, most exchanges, including BitMEX [BitMEX, 2017], Hyperliquid [Hyperliquid, 2024], and Binance [Binance, 2019] use a "queue-based" methodology, first proposed by BitMEX, that ranks accounts according to a priority metric which is the product of percentage profit-and-loss (P&L) and leverage of the account. Accounts then have their positions unwound sequentially in decreasing order of priority until the requisite quantity has been unwound. Chitra [2025] estimates that 95% of perpetual trading volume occurs on exchanges that use this queue-based ADL policy. Despite their central role in market stability, such ADL mechanisms are largely heuristic and have rarely been studied as explicit quantitative design objects.

In this paper, we formalize auto-deleveraging as an explicit risk-minimization problem faced by an exchange once the equity of an account has been exhausted. In our formulation, the exchange defines a loss function, which is the total equity shortfall aggregated across all accounts. The exchange then seeks to deleverage in a manner that minimizes the risk of future loss, where risk is quantified according to some risk measure. Risk-based objectives of this form are standard in financial risk management, where coherent risk measures provide a principled framework for evaluating tail exposure under uncertainty [Artzner et al., 1999]. Among these, conditional value-at-risk (CVaR) [Rockafellar and Uryasev, 2000] and, more generally, spectral risk measures [Acerbi, 2002], offer tractable and economically interpretable criteria that capture extreme losses while preserving convexity.

Our contributions are as follows:

- We begin in the single-asset, isolated-margining setting. Here, when the exchange is risk-neutral and minimizes expected loss (i.e., expected total future account equity shortfall), we show that the unique optimal deleveraging policy seeks to minimize the maximum leverage across all participants in the system. This *minimax leverage* policy is equivalent to a *leverage water-filling* (more accurately "leverage-draining") or threshold rule that equalizes post-intervention leverage across affected accounts. This characterization is analytical and identifies leverage as the natural variable governing optimal position reduction in this setting.

- We next consider the case where the exchange is risk-averse and minimizes the conditional value-at-risk (CVaR) of the terminal aggregate account equity shortfall. Here, we show that the minimax leverage policy remains optimal, but is no longer unique. Instead, the set of optimal policies is characterized by a cutoff in leverage: if there exist accounts with leverage above the cutoff, positions of those accounts can be reduced in an arbitrary way until all are at the cutoff, at which point the minimax leverage policy must be applied. This leverage cutoff depends on the confidence level β of the CVaR objective. We further establish that the minimax leverage policy remains optimal for the more general class of monotone risk measures.

- We show that the minimax leverage policy has a number of desirable properties: it is distribution-free, wash-trade resistant, Sybil resistant, and path-independent. In these ways, the minimax leverage policy is a natural and principled deleveraging policy.

- We then turn to the multi-asset, cross-margining setting, where the simple one-dimensional leverage ordering breaks down because default risk depends on the joint distribution of returns and on cross-asset hedges within each portfolio. We show that, under an expected-loss objective, the exchange's problem is still tractable. Introducing an asset-by-asset vector of shadow prices separates the optimization across accounts and yields an efficient numerical approach for the general case. Moreover, we identify a sharp analytic benchmark: when prices are driven by a single dominant market factor, the optimal deleveraging rule again has a water-filling structure, but in a modified factor-adjusted leverage rather than gross leverage, so that hedged portfolios are penalized less.

- We give an efficient numerical method for the general cross-margin problem. Exploiting the separable structure of the expected-loss objective, we propose a sample-average ADMM scheme [Boyd et al., 2011] that decomposes across accounts at each iteration, allowing fully parallel per-account updates. On a stylized multi-asset example, the method recovers the factor water-filling structure predicted by the theory. We then adapt the scheme to exchange scale, with refinements for the heterogeneous-asset, sparse-portfolio regime.

- We instantiate the method on the October 10, 2025 Hyperliquid ADL event, comparing the exchange's realized allocation against our risk-minimizing allocations. The model-based allocations achieve lower expected shortfall, and the difference is structural: the realized policy tends to fully close out affected accounts with little regard to leverage, whereas the risk-based allocations reduce each account only partially and in order of leverage.

At a practical level, the minimax leverage policy has important similarities to and differences from the BitMEX queue-based ADL policy that is currently used in practice. The policies are similar in that, in both cases, accounts are ranked according to a priority metric: in the minimax leverage policy, this priority is the account leverage, while in the queue-based policy, it is the product of percentage P&L and leverage. Further, in the minimax leverage policy, unwinds occur across accounts in a water-filling fashion according to the priority metric, while in the queue-based policies, unwinds occur in sequence[1] according to the priority metric. The similarities highlight that the minimax leverage policy is practically implementable, while the differences highlight suboptimality in the status quo. Indeed, beyond the fact that the BitMEX policy cannot directly be motivated by principled considerations, it satisfies none of the wash-trade resistance, Sybil resistance, and path independence properties enjoyed by the minimax leverage policy.

The choice of exchange shortfall risk as the ADL objective also deserves some discussion. In the single-asset, isolated-margin setting, one possible criticism is that this criterion may overly focus on the exchange's exposure to the side being delevered. If the market has moved down sharply, for example, and as a result short positions are delevered, one may argue that these short accounts already have significant profits and therefore pose little additional risk to the exchange. However, by definition, deleveraging is constrained to the shorts in this situation. Further, minimizing risk is one principled way of determining allocations, as opposed to the ad hoc methods currently employed in practice. Moreover, the criticism is weaker in the multi-asset, cross-margin setting, where accounts hold portfolios with offsetting exposures across assets and there is no simple notion of a uniformly "winning" side of the market. There, unrealized gains in one component of the portfolio need not imply low future shortfall risk once the joint distribution of price moves and the account's residual exposures are taken into account. More broadly, while alternative criteria are conceivable, it is not clear what objective besides risk the exchange should optimize once ADL has been triggered. For this reason, we view exchange shortfall risk as a natural and tractable starting point for the design of ADL rules. To the best of our knowledge, multi-asset ADL has not been discussed in the prior literature.

Literature review. The design of default management and loss allocation mechanisms for central clearing counterparties (CCPs) such as futures exchanges has received sustained attention in the literature, but almost exclusively in a modular fashion, with different strands focusing on distinct layers of the default waterfall. While margining, default resource sizing, and auction-based close-out procedures have been studied in detail, the final stage of the waterfall — loss allocation once prefunded resources are exhausted — has not been analyzed systematically from an optimality or mechanism-design perspective.

The broader CCP literature examines several of these upstream layers. Duffie and Zhu [2011] provide a foundational analysis of when central clearing reduces counterparty risk, emphasizing the role of multilateral netting and the trade-offs involved in concentrating exposures at a CCP. Menkveld [2017] identifies crowded positions among clearing members as an overlooked systemic risk and introduces tail-risk measures that account for correlated losses across members; Huang et al. [2021] complement this with an empirical decomposition of CCP exposure dynamics in stressed markets. From a capital-structure perspective, Huang [2019] studies how CCP capitalization interacts with for-profit incentives in the design of default resources. These strands address counterparty risk reduction, exposure measurement, and capital adequacy, but stop short of analyzing the rule by which residual losses are allocated once prefunded resources are exhausted.

Indeed, the final stage of the default waterfall has been discussed primarily in qualitative and policy-oriented work. The end-of-waterfall procedure advocated by the International Swaps and Derivatives Association is variation margin gains haircutting (VMGH), which reallocates residual losses to clearing members whose positions accrued gains during the liquidation period [International Swaps and Derivatives Association, 2013]. Surveys by Domanski et al. [2015] and Armakola and Laurent [2015] document current practices and argue that this "winner-pays" mechanism is

[1] I.e., the second-ranked account will be delevered only after the first-ranked is completely depleted, and so on.

robust and typically sufficient. Cont [2015] provides a detailed discussion of end-of-waterfall tools, including VMGH, contract tear-ups, and assessments, emphasizing incentive effects, legal feasibility, and market confidence. Duffie [2015] similarly analyzes CCP resolution from a systemic risk perspective, framing the trade-off between continuity of clearing and contagion. However, neither work proposes a formal model or optimization problem for the allocation of residual losses.

A recent exception is the work of Chitra [2025], who provides a formal analysis of ADL mechanisms in perpetual futures markets. Importantly, Chitra studies ADL under a different definition than considered here: Chitra models ADL as an ex-post loss-socialization rule that seeks to recover an exchange equity deficit by applying equity or profit haircuts to accounts. This is different from the formulation considered here, where ADL is defined as the allocation of forced position reductions to manage post-event risk.[2] Within his framework, Chitra [2025] casts haircut-based ADL as a mechanism-design problem and proves impossibility results showing that no ADL rule can simultaneously satisfy a full set of natural desiderata, including revenue, fairness, and solvency preservation. However, Chitra [2025] does not directly describe how positions should be reduced, which is the focus of this paper. On the other hand, in our framework, if the positions are transferred at a fair market price, there is no change in equity and hence no loss socialization. In this way, the two papers consider different variations of ADL and are not directly comparable. The variation we consider here is consistent with the way that ADL is currently implemented by perpetual futures exchanges in practice.

Complementing this formal literature, a recent discussion by Jia et al. [2026] provides an on-chain forensic analysis of Hyperliquid's October 10, 2025 liquidation event. They analyze the overall economics of the full liquidation pipeline, from initial liquidations to backstop takeovers and subsequent ADL unwinds. They observe that delevered short positions were ex post profitable since they were bought in at relative market lows. This challenges the notion of ADL as a loss socialization mechanism. However, individual ADL outcomes are heterogeneous because there were several waves of ADL and queue position mattered. While not a formal model of ADL, [Jia et al., 2026] is useful institutional evidence. In particular, it reinforces the importance of modeling the contract space allocation rule itself and of distinguishing forced position reduction from ex post wealth transfers, which can be hard to determine during the time of a crisis.

The water-filling rule is also related to the axiomatic literature on claims, bankruptcy, and taxation problems, surveyed by Thomson [2003, 2015]. In the base claims model, agents have claims c_i on an insufficient endowment E, and a rule selects awards a_i satisfying $0 \le a_i \le c_i$ and $\sum_i a_i = E$. Canonical rules include the proportional rule, the constrained equal awards rule, the constrained equal losses rule, and the Talmud rule, together with weighted and priority versions. This literature studies such rules through axioms including order preservation, composition, consistency, duality, and no advantageous splitting or merging. As detailed in Appendix A.4, there is an exact connection between this framework and the water-filling rule in our single-asset isolated-margin model: under a change of variables from buybacks to residual positions, the water-filling allocation corresponds to the weighted constrained equal awards rule with weights proportional to account equity. The interpretation, however, differs from the classical claims model. In the claims literature, weights are typically primitive priority or entitlement parameters, whereas in our setting they arise endogenously from risk minimization considerations. Thus our contribution in the isolated-margin case is not the abstract water-filling formula itself, but the derivation of the particular weighted claims rule from an exchange shortfall-risk minimization problem.

[2] In the recent (v3) version of [Chitra, 2025], this difference is now highlighted as follows (Section 2.4 *ibid.*, see also "corrections" on p. 9): "Production ADL is executed in *contract space*: the engine selects positions by a ranking score and forces contract-level reductions. The theoretical analysis in this paper [i.e., Chitra, 2025] is written in *wealth space* (equity haircuts and haircutable endowment)."

The multi-asset cross-margin setting is not covered by the claims literature. Although the latter contains extensions with uncertainty, multi-dimensional claims, network structure, and non-homogeneous endowments [Thomson, 2015], our cross-margin ADL problem has different primitives: signed vector positions, a scalar collateral pool for each account, vector reduction constraints, and stochastic losses driven by the joint distribution of asset prices. In particular, hedging and factor exposure are central to the cross-margin problem, whereas the base claims model is a scalar rationing problem. Thus, the single-asset water-filling rule can be identified with a weighted claims rule, but the general cross-margin ADL problem is not a special case of the claims framework.

2. Single-Asset Isolated Margining

We consider a single traded asset under isolated margining and focus on a stress scenario following a large price move. Without loss of generality, as a result of this move, a collection of long accounts becomes insolvent and is liquidated, generating an aggregate exposure of size $Q > 0$ units that must be absorbed by the rest of the system. ADL reallocates this exposure to short accounts by forcing them to reduce their positions.

We model this reallocation at a fixed execution price $p_\tau > 0$ at time τ.[3] The exchange considers all short accounts, indexed by $i = 1, \ldots, n$, and forces each of them to buy back a quantity $x_i \geq 0$, thereby reducing their short positions. Each short account i is characterized by its position size $q_i > 0$, entry price $p_i^{(e)}$, and posted margin $m_i \geq 0$. Clearly, feasibility of $x \triangleq (x_1, \ldots, x_n)$ requires

$$\sum_{i=1}^{n} x_i = Q \qquad \text{and} \qquad 0 \leq x_i \leq q_i \quad \text{for} \quad i = 1, \ldots, n,$$

and we define the feasible set as

$$\mathcal{X} \triangleq \Big\{ x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = Q, \quad 0 \leq x_i \leq q_i \text{ for all } i \Big\}.$$

Assumption 2.1 (Feasibility). $Q < \sum_{i=1}^{n} q_i$.

Note that we can extend to the case $Q = \sum_{i=1}^{n} q_i$, but in this case the problem is trivial as there is only one feasible allocation.

2.1. Equity, Losses, and Leverage

For any price level p, the post-ADL equity of account i is defined as

$$e_i(x_i, p) \;\triangleq\; q_i(p_i^{(e)} - p) \;-\; x_i(p_\tau - p) \;+\; m_i.$$

This expression equals the equity the account would have had at price p absent ADL, plus the realized P&L from buying back x_i units at p_τ rather than at p. Evaluated at $p = p_\tau$, the equity is

$$E_i \triangleq e_i(x_i, p_\tau) = q_i(p_i^{(e)} - p_\tau) + m_i. \tag{1}$$

In particular, at the time of intervention, equity is unaffected by the allocation x_i, whereas ADL reshapes the sensitivity of future equity to subsequent price movements. We will assume throughout that each candidate account for deleveraging is solvent.

[3] In practice, this price is often chosen so that the exchange remains solvent. It is not necessarily a fair market price, and as a consequence, ADL may also transfer profits or losses to the short accounts. The present work focuses on the risk exposure of the exchange, not on how losses are socialized [cf. Chitra, 2025].

Assumption 2.2 (Solvency at the ADL time). *$E_i > 0$ for all i.*

The exchange incurs losses whenever account equity becomes negative. Aggregating across all short accounts, the total exchange loss due to short accounts that would be incurred at a future time $T > \tau$ when the price is given by $p_T = p$ is

$$\mathcal{L}(x,p) \;\triangleq\; \sum_{i=1}^{n} \big(- e_i(x_i,p)\big)_+ \tag{2}$$

where $(u)_+ \triangleq \max\{u, 0\}$. We write the corresponding per-account shortfall as

$$\sigma_i(x_i,p) \triangleq \big(- e_i(x_i,p)\big)_+. \tag{3}$$

A convenient state variable at the ADL time τ is the post-ADL leverage of each account, defined as the ratio of notional exposure to equity,

$$\ell_i(x_i) \;\triangleq\; \frac{p_\tau(q_i - x_i)}{e_i(x_i,p_\tau)} = \frac{p_\tau(q_i - x_i)}{q_i(p_i^{(e)} - p_\tau) + m_i} = \frac{p_\tau(q_i - x_i)}{E_i}. \tag{4}$$

Under Assumption 2.2, ℓ_i is affine and strictly decreasing in x_i on $[0, q_i]$, with

$$\ell_i(0) = \frac{p_\tau q_i}{E_i}, \qquad \ell_i(q_i) = 0.$$

While equity determines solvency, leverage captures the sensitivity of future losses to price movements and will be the natural variable in which optimal deleveraging policies are expressed.

2.2. Risk-Based ADL

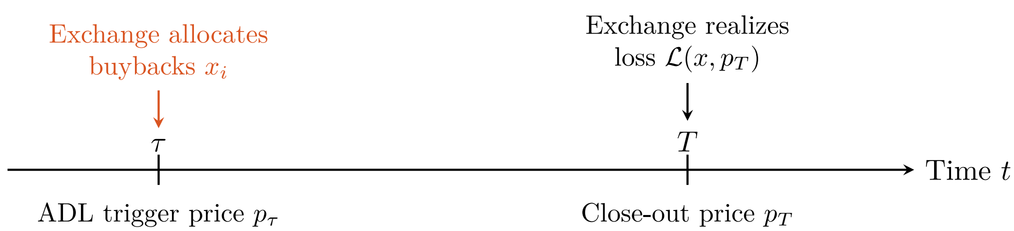

Figure 1: Timeline of the ADL allocation problem. At the trigger time τ, the exchange observes the price p_τ and chooses the buyback allocation $x = (x_1, \ldots, x_n)$. At the terminal time $T > \tau$, the close-out price p_T is realized and the exchange incurs the equity shortfall $\mathcal{L}(x, p_T)$, whose risk the allocation x is chosen to minimize.

We propose a risk-based design principle for ADL, namely to choose the buyback allocation x which minimizes a risk measure of the exchange's equity shortfall at the time horizon $T > \tau$. Given a model for the (random) price p_T at the terminal time T, the exchange evaluates the risk of the terminal loss $\mathcal{L}(x, p_T)$ through a risk measure $\rho(\cdot)$ and chooses an allocation $x \in \mathcal{X}$ that solves

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \rho\big(\mathcal{L}(x, p_T)\big) \\ \text{subject to} & x \in \mathcal{X}. \end{array} \tag{5}$$

The two-stage structure of this problem is summarized in Figure 1: at the ADL trigger time τ the exchange observes the price p_τ and commits to the buyback allocation x, while the resulting equity shortfall is only realized at the terminal time T once p_T has materialized.

We pause to discuss several modeling choices that frame the analysis.

- The formulation in (5) is static: the allocation x is chosen once at the trigger time τ and held fixed until T, with no trading by other market participants in between, no intermediate intervention by the exchange, and no re-liquidation of accounts whose equity may turn negative before T. In practice, ADL is one step in a dynamic risk-management process and may need to be repeated as market conditions evolve. We adopt the static formulation because it isolates the central allocation question — given a target quantity Q that must be unwound at τ, how should it be distributed across accounts? — and admits tractable convex analysis. We view the resulting policies as a principled benchmark against which dynamic refinements can be compared.

- Alternative objectives are conceivable: once ADL has been triggered, the exchange could in principle aim to maximize social welfare, minimize wealth transfers between accounts [in the spirit of Chitra, 2025], or maximize its own revenue. We focus on risk minimization because ADL is invoked precisely in tail scenarios, where the exchange's solvency is the dominant concern, and risk measures such as expected loss and CVaR offer a principled and economically interpretable framework for quantifying that tail exposure. Risk-based objectives also have the practical advantage of yielding tractable convex problems with structurally transparent optima, as we develop in the remainder of the paper.

- The loss $\mathcal{L}$ accounts only for the equity shortfall associated with short positions, and ignores (solvent) long accounts. This restriction is without loss of generality for the allocation problem (5). That is easiest to see in the risk-neutral case ($\rho = \mathsf{E}$), because any loss contribution from long positions is independent of the allocation variable x and therefore enters the objective only as an additive constant. The proposed allocation remains optimal even for the more general risk-based objectives considered below; see Proposition 2.18.

2.3. Minimizing Expected Loss

In this subsection, we specialize the risk-based ADL problem (5) to the case where $\rho = \mathsf{E}$, i.e., the exchange minimizes its expected loss:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & V(x) \triangleq \mathsf{E}[\mathcal{L}(x, p_T)] \\ \text{subject to} & x \in \mathcal{X}. \end{array} \tag{6}$$

We will see that this principled approach leads to a particular rule $x^\star \in \mathcal{X}$ that is straightforward to implement and has numerous desirable properties.

Under Assumption 2.3, the objective $V(x)$ is finite and convex in x (see Lemma A.1 in Appendix A). To ensure that problem (6) is well posed, we assume that the terminal price p_T is integrable. For simplicity of presentation, we further assume that the distribution of p_T admits a density with sufficiently large support:

Assumption 2.3 (Regularity of p_T). *p_T is an integrable random variable that admits a density f_T with $f_T(p) > 0$ for $p \geq p_\tau$.*

The following theorem summarizes several key insights. First, minimizing expected losses for the exchange is equivalent to minimizing the maximal post-ADL leverage across accounts; cf. (b). Solving this minimax leverage problem is further equivalent to using a threshold-rule on leverage: for a certain leverage threshold t, all accounts with initial leverage exceeding t are deleveraged down to t, whereas accounts with initial leverage below t remain untouched; cf. (c). The threshold t is set so that the total allocation across accounts equals the target buyback amount Q. In fact, this policy not only minimizes the expected loss, but even minimizes the loss in any realization of p_T;

cf. (d). Any of these formulations leads to the same solution $x^\star \in \mathcal{X}$, which we call the *water-filling rule* or *minimax leverage policy.*

Theorem 2.4 (Minimax leverage policy). *Let Assumptions 2.1–2.3 hold. For a feasible allocation $x \in \mathcal{X}$, the following are equivalent:*

(a) x minimizes the expected loss, i.e., solves (6).

(b) x minimizes the maximal post-ADL leverage, i.e., solves

$$\underset{x\in\mathcal{X}}{\text{minimize}}\ \underset{1\le i\le n}{\text{maximize}}\ \ \ell_i(x_i). \tag{7}$$

(c) x is a threshold rule on leverage, i.e., there exists a threshold $t \in [0, \max_i \ell_i(0)]$ such that

$$\begin{aligned} x_i > 0 \quad &\Longrightarrow \quad \ell_i(x_i) = t, \\ x_i = 0 \quad &\Longrightarrow \quad \ell_i(0) \le t. \end{aligned}$$

(d) x is a simultaneous pointwise minimizer of realized loss, i.e.,

$$\mathcal{L}(x,p) \le \mathcal{L}(y,p) \qquad \textit{for all} \quad y \in \mathcal{X},\ p \in \mathbb{R}.$$

The unique optimal allocation for (a)–(d) is the water-filling rule *$x^\star = (x_1^\star, \ldots, x_n^\star) \in \mathcal{X}$ defined by*

$$x_i^\star \ \triangleq\ \left(q_i - \frac{E_i}{p_\tau}\, t^\star\right)_+, \qquad i = 1, \ldots, n,$$

where the leverage threshold $t^\star > 0$ is the unique root of the equation

$$\sum_{i=1}^n \left(q_i - \frac{E_i}{p_\tau}\, t\right)_+ = Q. \tag{8}$$

As detailed in Lemma A.5 in Section A.1.2, the function $G(t) \triangleq \sum_{i=1}^n (q_i - \frac{E_i}{p_\tau}\, t)_+$ appearing in (8) is continuous and strictly decreasing on $[0, \max_i \ell_i(0)]$. Thus, it is straightforward to compute the root $t^\star$. All proofs for this section are reported in Appendix A.

Theorem 2.4 underscores that the optimal ADL allocation for minimizing expected losses induces a *leverage equalization principle*. It reallocates exposure by leveling down the highest post-ADL leverages first, until they reach the common leverage threshold. As illustrated in Figure 2, this corresponds to a water-filling (or leverage-draining) rule: Enumerate the accounts in decreasing initial leverage, so that $\ell_1(0) \ge \ell_2(0) \ge \cdots \ge \ell_n(0)$. Then as the leverage cap decreases from $\ell_1(0)$ downward, initially only the most levered account is forced to buy back. Once its leverage reaches $\ell_2(0)$, the top two accounts are deleveraged jointly so as to keep their leverages equal, and so on. The procedure stops when the target buyback amount Q is reached.

The intuition connecting minimization of expected losses with minimization of maximal leverage is as follows. A buyback at the execution price p_τ does not change the equity E_i at time τ, but it does reduce the residual short position $q_i - x_i$ and hence increases the account's "distance to default." Indeed, as derived in (38)–(39), for $x_i < q_i$ the account incurs a shortfall only when the terminal price exceeds its zero-equity (bankruptcy) level

$$p_\tau + \frac{E_i}{q_i - x_i} = p_\tau\big(1 + \ell_i(x_i)^{-1}\big),$$

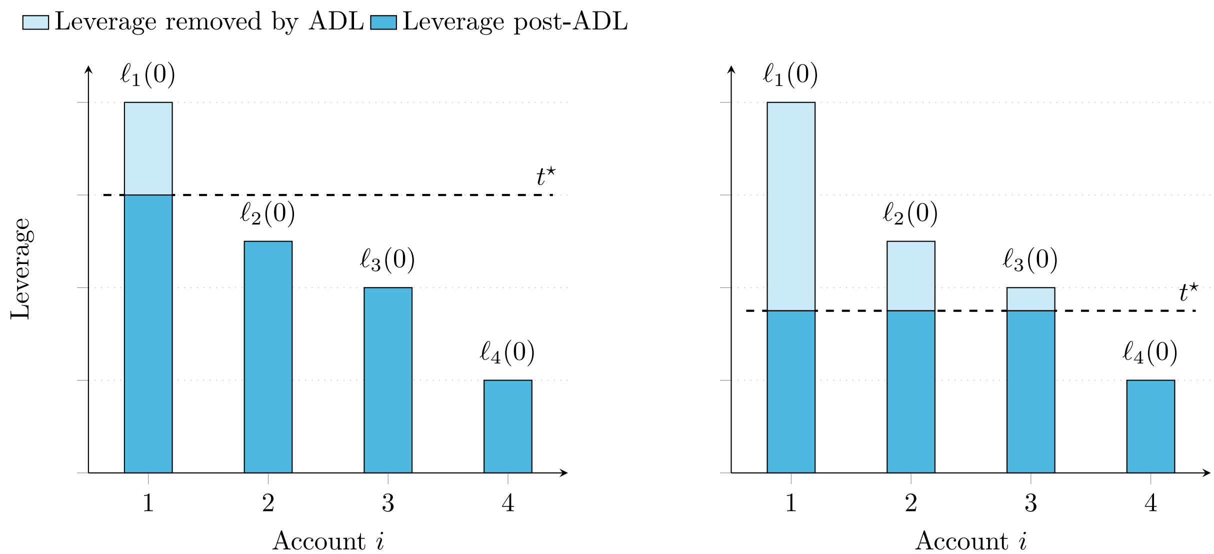

Figure 2: The minimax leverage policy equalizes leverage by water-filling ("leverage-draining") from the initial leverage $\ell_i(0)$ down to the threshold $t^\star$. A larger total buyback quantity Q leads to a lower threshold $t^\star$ and more accounts being affected (right).

so higher post-ADL leverage $\ell_i(x_i)$ corresponds to a larger region of the right tail of p_T over which the exchange bears losses from that account. Consequently, the marginal reduction in expected loss from increasing x_i is governed by how much the buyback shifts the tail threshold. This marginal benefit is strictly larger (in absolute value) for accounts with higher leverage. It follows that whenever two accounts can be adjusted at the margin, any allocation that buys back from a relatively low-leverage account while leaving a strictly higher-leverage account unaddressed can be improved by reallocating an infinitesimal unit of buyback toward the more levered account. As a result, an expected-loss minimizer must first remove exposure from the most levered accounts, which is exactly the minimax formulation (7).

2.4. Properties of the Minimax Leverage Policy

This section studies the properties of the water-filling allocation obtained in Theorem 2.4. An obvious but important feature is that the policy is *distribution-free*, i.e., does not require a model for the terminal price p_T — the allocation depends only on the state $\{(q_i, E_i)\}_{i=1}^n$ at the ADL time τ (as well as Q, p_τ), and is therefore independent of both the horizon T and the distribution of p_T.

A related property is that the *leverage ordering of accounts is invariant to the execution price.* That is, the ranking of accounts by initial leverage, which determines the water-filling priority, does not depend on the price p_τ at which ADL is executed.

Proposition 2.5 (Price-independence of leverage ordering)**.** *For any two accounts i and j, define the leverage of account i at price p (before ADL) as*

$$\ell_i(p) \triangleq \frac{p\, q_i}{E_i(p)}, \qquad E_i(p) \triangleq q_i(p_i^{(e)} - p) + m_i.$$

Then, for any prices $p, p' > 0$ at which both accounts are solvent, $\ell_i(p) \geq \ell_j(p)$ if and only if $\ell_i(p') \geq \ell_j(p')$.

As a consequence, the water-filling priority order used in Theorem 2.4 is invariant to the choice of execution price p_τ.

Additionally, unlike ADL rules that sort accounts using realized or unrealized P&L measures, the water-filling rule is *wash-trade resistant*. The reason is that the relevant state variable for minimax leverage is the pair (q_i, E_i), where E_i already includes the mark-to-market P&L at the execution price p_τ. Consequently, any trade prior to τ that leaves the position size q_i unchanged merely reallocates equity between unrealized P&L at the entry price and margin. It therefore leaves E_i unchanged, and hence does not affect the initial leverage $\ell_i(0) = p_\tau q_i / E_i$ that determines the water-filling priority.

The next two sections derive two further important properties of the minimax leverage policy. First, Sybil resistance, meaning that traders cannot evade deleveraging by operating through multiple, smaller accounts. Second, path-independence, meaning that two subsequent ADL events with quantities Q_1 and Q_2 lead to the same outcome as a single ADL event with quantity $Q_1 + Q_2$. We also show that path-independence, together with leverage-priority, characterizes the minimax leverage policy among all ADL policies.

Remark 2.6. The BitMEX-style queue-based ADL policy (see Section 1) fails to have some of these desirable properties:

- Because the BitMEX ranking is based on unrealized P&L, the priority ordering of accounts can change depending on the execution price, unlike the leverage ordering used by the minimax leverage policy (Proposition 2.5).
- Because the BitMEX policy deleverages the highest-ranked account before targeting any lower-ranked account, it is not Sybil resistant. For instance, the most levered position in the system could be split across two accounts: one with even higher leverage and one with sufficiently low leverage to appear third (or lower) in the queue. An ADL event of a given size (e.g., equal to the size of the original position) would then deleverage the first two accounts in the queue but leave the third untouched, whereas without splitting the same event could have fully deleveraged the original account.
- Prioritizing the (initially) top-ranked account also induces path dependence. Under two successive ADL events, the first event may deleverage the initially top-ranked account enough to lower its rank, so that the second event targets a different account first; if instead the same aggregate quantity were executed in a single event, the entire quantity would have been applied to the initially top-ranked account.
- Finally, because queue priority depends explicitly on unrealized P&L, an account could lower its queue priority via a wash trade.

2.4.1. Sybil Resistance

When accounts are anonymous, a participant can split their aggregate position and collateral across multiple accounts. If deleveraging allocations depend on per-account quantities, this creates a "Sybil" manipulation channel that can potentially reduce the participant's total forced buyback without changing the underlying economic exposure. Clearly, it is preferable to use ADL mechanisms that are *Sybil resistant*: splitting an account into several accounts with the same combined position and equity should not decrease the participant's total buyback. Theorem 2.9 below states that the minimax leverage policy is Sybil resistant.

To formalize this result, we define account splits and Sybil resistance mathematically. Fix a set of *non-attacker* short accounts indexed by $\mathcal{N}$, with position and equity parameters $\{(q_j, E_j)\}_{j \in \mathcal{N}}$.

We view this as the relevant account state at time τ since, up to scaling by the price, $\ell_i(0) = p_\tau q_i/E_i$ depends only on these two quantities. An *attacker* controls an aggregate pair (q^A, E^A) with $q^A \geq 0$ and $E^A > 0$. For a split into K accounts, denote the attacker index set by $\mathcal{A}_K = \{1, \ldots, K\}$ and let the full economy index set be $\mathcal{I}_K \triangleq \mathcal{N} \cup \mathcal{A}_K$. Then $\mathcal{A}_1$ represents the attacker in the *unsplit* (single-account) configuration, holding (q^A, E^A). The next definition formalizes that the attacker can split her holdings q^A and equity E^A arbitrarily over her K Sybil accounts.

Definition 2.7 (Sybil split). *A* Sybil split *of* (q^A, E^A) *is a collection* $\{(q_k, E_k)\}_{k=1}^K \in (\mathbb{R}_+ \times \mathbb{R}_{++})^K$ *of attacker-controlled accounts such that* $\sum_{k=1}^K q_k = q^A$ *and* $\sum_{k=1}^K E_k = E^A$.

Next, we formally define Sybil resistance. For a given ADL mechanism, let $x^K \in \mathbb{R}_+^{\mathcal{I}_K}$ denote the resulting allocation for the scenario with Sybil split, i.e., x_i^K is the buyback assigned to account i and $\sum_{i \in \mathcal{I}_K} x_i^K = Q$. Thus, the attacker's *total* buyback is

$$X_A^K \triangleq \sum_{k \in \mathcal{A}_K} x_k^K.$$

Definition 2.8 (Sybil resistance). *An ADL mechanism is* Sybil resistant *if splitting an account cannot reduce the attacker's total buyback. That is,* $X_A^K \geq X_A^1$ *for every attacker aggregate state* $(q^A, E^A) \in \mathbb{R}_+ \times \mathbb{R}_{++}$ *and every Sybil split* $\{(q_k, E_k)\}_{k=1}^K$ *of* (q^A, E^A).

We can now state the formal result.

Theorem 2.9. *The minimax leverage policy is Sybil resistant.*

While the proof in Appendix A.2 proceeds via the subadditivity of the function defining the leverage threshold (8), the key intuition for Theorem 2.9 is the very essence of the minimax leverage policy: it focuses on the maximal leverage, which cannot be reduced by splitting. To see this, let

$$\ell^A \triangleq \frac{p_\tau q^A}{E^A} \qquad \text{and} \qquad \ell_k \triangleq \frac{p_\tau q_k}{E_k}$$

be the pre-ADL leverages of the unsplit attacker and sub-account k, respectively. The following shows that ℓ^A is a convex combination of $\{\ell_k\}_{k=1}^K$ with weights $w_k \triangleq E_k/E^A$:

$$\ell^A = \frac{p_\tau q^A}{E^A} = \frac{p_\tau \sum_{k=1}^K q_k}{\sum_{k=1}^K E_k} = \sum_{k=1}^K \frac{E_k}{\sum_{j=1}^K E_j} \frac{p_\tau q_k}{E_k} = \sum_{k=1}^K w_k \, \ell_k.$$

In particular, $\max_{1 \leq k \leq K} \ell_k \geq \ell^A$, meaning that the *maximal leverage among the Sybil accounts is never lower than the original leverage.*

We observe that while splitting accounts is not beneficial under the minimax leverage policy, the total amount deleveraged is not necessarily invariant — it may increase.

Remark 2.10. Splitting accounts can lead to a strictly larger amount bought back from the attacker. For instance, this would occur if an account with leverage just below the threshold $t^\star$ is split so that one sub-account becomes the highest leveraged account in the system. Equivalently, merging several accounts can lead to a strict decrease in the amount deleveraged. As a result, the water-filling rule indirectly incentivizes account aggregation.

2.4.2. Path-Independence, Leverage Priority, and Axiomatic Characterization

In this section, we first observe that the minimax leverage policy has two properties that we call path-independence and leverage-priority. Then, we show that these two properties can serve as an axiomatic characterization of the minimax leverage policy among all possible ADL mechanisms.

For brevity, we limit ourselves to informal statements in this section and report the mathematical details in Section A.3. In particular, we refer to Assumptions A.10 and A.11 for precise versions of the following two definitions.

Definition 2.11 (Path-independence)**.** *An ADL mechanism is path-independent if two successive applications with buy-back quantities Q_1 and Q_2 lead to the same outcome as a single application with buy-back quantity $Q_1 + Q_2$.*

Definition 2.12 (Leverage-priority)**.** *An ADL mechanism satisfies leverage-priority if, for sufficiently small buy-back quantity $Q > 0$, it only affects the account(s) with maximal initial leverage.*

Using the water-filling representation of the minimax leverage policy, it is not hard to see that it satisfies both properties, which is one implication of Theorem 2.13 below. Intuitively, path-independence holds because water-filling to the target level $t^\star$ and then continuing to the target level $t^{\star\star}$ yields the same as directly water-filling to $t^{\star\star}$ (and one verifies that deleveraging Q_1 and then Q_2 yields the same eventual target $t^{\star\star}$ as directly deleveraging $Q_1 + Q_2$). Of course, leverage-priority is immediate from the water-filling representation.

The main result of the theorem is the reverse implication: if an ADL mechanism satisfies path-independence and leverage-priority, it must be the minimax leverage policy. To state such a result rigorously, Section A.3 formalizes a general ADL mechanism mathematically as follows. Consider the state space $\mathcal{S} \triangleq (\mathbb{R}_+ \times \mathbb{R}_{++})^n$ of all possible account states $s = (q_i, E_i)_{i=1}^n$. Each state encodes a short position size $q_i \geq 0$ and an equity level $E_i > 0$ for each account i. Then an ADL mechanism is a family

$$\{F_Q : \mathcal{S} \to \mathcal{S}\}_{Q \geq 0}$$

of maps on $\mathcal{S}$, where $F_Q(s)$ is the post-ADL state after a total buyback quantity Q has been allocated across accounts. Naturally, we only consider mechanisms that reduce (but never increase) existing positions (see Assumption A.9 for details). We then have the following axiomatic characterization of the minimax leverage policy.

Theorem 2.13 (Characterization by path-independence and leverage-priority)**.** *The minimax leverage policy satisfies path-independence and leverage-priority. Conversely, if any ADL mechanism satisfies these two properties, then it must be the minimax leverage policy.*

The intuition behind the axiomatic characterization is as follows. Split the amount Q into many (infinitesimally) small bits. By path-independence, the mechanism for Q yields the same result as consecutively buying back the small bits. By leverage-priority, the first bit is allocated to the account with the highest initial leverage, and consecutive bits will be allocated to the same account until its leverage is equalized with the second-highest initial leverage. At that point, the next small bit could be allocated to either of the two accounts (or split). However, the choice does not matter when quantities are infinitesimal. Intuitively, even if one bit was allocated to account 1 (rather than split between 1 and 2 as in the water-filling rule), the next bit would then go to account 2, realigning the leverage levels and approximately resulting in the water-filling outcome. In the limit of infinitesimally small bits, the ambiguity disappears and we recover exactly the water-filling of Theorem 2.4.

For the mathematical proof, the key is that path-independence amounts to the semigroup property $F_{Q_2} \circ F_{Q_1} = F_{Q_1+Q_2}$ where the deleveraging quantity Q plays the role of the time parameter. Moreover, the monotonicity of any ADL mechanism implies that this semigroup is Lipschitz-continuous (Lemma A.12), and once this regularity is recognized, the above intuition can be converted into a proof.

Remark 2.14. Taken individually, path-independence and leverage-priority do not imply the minimax leverage policy. For instance, consider the ADL mechanism which allocates the entire quantity

Q to the account(s) with the highest leverage at the ADL time (see also Remark 2.6). This mechanism clearly satisfies leverage-priority, but is not path-independent: After being deleveraged with Q_1, the targeted account i may no longer have the highest leverage, and then in a subsequent application of the mechanism, the quantity Q_2 would be allocated to a different account. Whereas when applied in one shot with $Q_1 + Q_2$, the entire quantity is allocated to account i. On the other hand, consider the pro-rata policy, which allocates amounts x_i proportional to positions q_i, or equivalently, decreases the leverage of all accounts by the same percentage. This policy is path-independent, but clearly does not prioritize highest leverage.

2.5. Conditional Value-at-Risk and Monotone Risk Measures

We now replace the risk-neutral objective $\mathsf{E}[\mathcal{L}(x, p_T)]$ by Conditional Value-at-Risk CVaR_β at confidence level $\beta \in (0, 1)$ as introduced by Rockafellar and Uryasev [2000], while keeping the same buyback constraints.

Definition 2.15. *For an integrable random variable Z and confidence level $\beta \in (0, 1)$, define the Value-at-Risk*

$$\mathrm{VaR}_\beta(Z) \triangleq \inf\{z \in \mathbb{R} : \mathsf{P}(Z \le z) \ge \beta\}.$$

The Conditional Value-at-Risk at level β is the tail mean[4] *beyond* $\mathrm{VaR}_\beta(Z)$*,*

$$\mathrm{CVaR}_\beta(Z) \triangleq \frac{1}{1-\beta} \int_\beta^1 \mathrm{VaR}_u(Z)\, du.$$

If Z has no atom at $\mathrm{VaR}_\beta(Z)$*, this can also be written as*

$$\mathrm{CVaR}_\beta(Z) = \mathsf{E}[Z \mid Z \ge \mathrm{VaR}_\beta(Z)].$$

Unlike the expectation, CVaR_β emphasizes tail losses and therefore changes the marginal incentives driving auto-deleveraging. In the following, we formulate the CVaR-based ADL problem and characterize its separable structure and optimality conditions. We will see that the water-filling rule remains optimal, but (in contrast to the expected loss) the optimizer need not be unique.

The CVaR-based ADL problem reads

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & \mathrm{CVaR}_\beta(\mathcal{L}(x, p_T)) \\ \text{subject to} \quad & x \in \mathcal{X}. \end{aligned} \tag{9}$$

It is straightforward to check that (9) is convex (see Appendix A.5). A key structural observation in our setting is that the individual account losses are monotone functions of a *single* scalar risk factor, the terminal price p_T. Specifically, $\sigma_i(x_i, p)$ is nondecreasing in p for each fixed x_i (see Lemma A.1 in Appendix A). Hence for any feasible x, the random variables $\{\sigma_i(x_i, p_T)\}_{i=1}^n$ are comonotone: they move in the same direction as the price p_T. Since CVaR_β is comonotone additive [cf. Propositions 2.3 & 3.4, Tasche, 2002], it follows that

$$\mathrm{CVaR}_\beta(\mathcal{L}(x, p_T)) = \mathrm{CVaR}_\beta\left(\sum_{i=1}^n \sigma_i(x_i, p_T)\right) = \sum_{i=1}^n \mathrm{CVaR}_\beta(\sigma_i(x_i, p_T)).$$

This account-by-account decomposition is a crucial simplification and yields a clean characterization of optimality. Our main result is as follows.

[4] We use the convention $\mathrm{CVaR}_0(Z) = \mathsf{E}[Z]$.

Theorem 2.16. *Let x^{WF} denote the (unique) water-filling allocation of Theorem 2.4. Then x^{WF} is an optimizer of the CVaR-ADL problem (9) for every $\beta \in (0,1)$. However, the solution may be non-unique; in general, (9) may admit a continuum of optimizers.*

Let $p_\beta \triangleq \mathrm{VaR}_\beta(p_T)$ denote the β-quantile (stress level) of the terminal price. Since each shortfall $\sigma_i(x_i, p_T)$ is nondecreasing in the single risk factor p_T, CVaR_β is determined by losses in the tail event $\{p_T \geq p_\beta\}$. This motivates the definition of the *leverage cutoff* ℓ_β via

$$p_\tau(1+\ell_\beta^{-1}) = p_\beta, \qquad \text{equivalently} \qquad \ell_\beta = \frac{p_\tau}{p_\beta - p_\tau}, \qquad \text{if} \qquad p_\beta > p_\tau,$$

whereas $\ell_\beta = +\infty$ if $p_\beta \leq p_\tau$. Thus ℓ_β is precisely the leverage level whose bankruptcy price (cf. (39)) coincides with p_β.

The cutoff separates two regimes. If an account's post-ADL leverage satisfies $\ell_i(x_i) < \ell_\beta$, then its bankruptcy threshold lies *within* the CVaR tail, and additional buyback changes tail losses in an account-specific way. By contrast, if an account remains highly levered in the sense that $\ell_i(x_i) \geq \ell_\beta$, then the tail event relevant to CVaR_β is governed by p_β rather than by the account-specific threshold. In this stressed regime the marginal CVaR_β benefit of further buyback becomes identical across all such accounts. Consequently, the water-filling rule remains optimal, but the objective can develop flat directions: redistributing buyback volume among stressed accounts (while preserving feasibility) leaves CVaR_β unchanged.

A continuum of optimal allocations can arise when the deleveraging budget Q is insufficient to bring *every* account below the cutoff ℓ_β, so that multiple accounts remain in the stressed set $\{\ell_i(x_i) \geq \ell_\beta\}$. In that case, CVaR_β is indifferent to how buybacks are redistributed among stressed accounts, producing non-uniqueness; see Appendix A.5 for a formal construction.

The optimality of water-filling can also be extended beyond the expected loss and CVaR cases, to all *monotone* risk measures.

Definition 2.17 (Monotone risk measure). *A risk measure ρ is monotone if*

$$Z_1 \leq Z_2 \ a.s. \quad \Longrightarrow \quad \rho(Z_1) \leq \rho(Z_2).$$

The CVaR criterion corresponds to focusing on a *single* tail level, while this class includes all *spectral risk measures*, which assign weights to different quantile levels of the loss distribution. Specifically, we say that $\rho(\cdot)$ is a spectral risk measure [cf. Acerbi, 2002, Shapiro, 2013] if it has the form[5]

$$\rho(Z) \ \triangleq \ \int_{[0,1)} \mathrm{CVaR}_\beta(Z)\, \mu(d\beta)$$

for all integrable random variables Z, where μ is a probability measure on $[0,1)$.

Mirroring (5), we can write the more general optimization over monotone risk measures as

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \rho(\mathcal{L}(x, p_T)) \\ \text{subject to} & x \in \mathcal{X}. \end{array} \tag{10}$$

The following is a straightforward consequence of the pointwise optimality in Theorem 2.4(d).

Proposition 2.18. *Under Assumptions 2.1–2.3 the water-filling allocation is optimal for (10) whenever $\rho(\cdot)$ is monotone.*

Note that Proposition 2.18 states optimality, not uniqueness. For general monotone risk measures the optimal allocations need not be unique, as shown in Theorem 2.16 for CVaR_β.

[5] We interpret $\rho(\cdot)$ as extended real valued on integrable random variables and tacitly assume $\mathsf{E}[Z] \geq 0$ so that $\rho(Z) \geq 0$. A sufficient condition ensuring that $\rho(Z) < \infty$ for all integrable Z is $\int_{[0,1)} (1-\beta)^{-1} d\mu(\beta) < \infty$.

2.6. Numerical Example

We illustrate the structure of CVaR-based auto-deleveraging in a one-asset setting with $n = 4$ short accounts and total short volume $\sum_{i=1}^{n} q_i = 33$. The reference price is fixed at $p_\tau \approx \$67{,}000$ (BTC spot), and deleveraging outcomes are studied as a function of the available ADL budget Q.

The terminal price p_T is modeled as a geometric Brownian motion over a horizon $\Delta t = 10/365$, with zero drift and annual volatility 60%. For the CVaR criterion we take $\beta = 0.98$. Closed-form expressions for the resulting CVaR objective under GBM are given in Appendix A.6.

Each account i is characterized by a short position q_i, entry price $p_i^{(e)}$, and margin m_i, with initial leverage

$$\ell_i(0) = \frac{p_\tau q_i}{E_i}.$$

Under the GBM model, the β-quantile p_β of the terminal price admits a closed form, yielding the stress cutoff leverage

$$\ell_\beta = \frac{p_\tau}{p_\beta - p_\tau} \approx 4.55.$$

Accounts with $\ell_i(0) > \ell_\beta$ are initially stressed.

Table 1 reports the static characteristics of four levered accounts. All four satisfy $\ell_i(0) > \ell_\beta$ and therefore lie in the stressed region at the trigger price. From these static quantities alone, there is no clear ordering of accounts in terms of which should be deleveraged first under a tail-risk criterion.

Table 1: Parameters of the four most levered accounts, sorted in increasing initial leverage.

Account	q_i	$p_i^{(e)}$ in \$1000	m_i in \$1000	$\ell_i(0)$
1	8	71	82	4.7
2	10	72	78.8	5.2
3	8	70	59.8	6.4
4	7	69.5	48.6	7.1

Figure 3 reports the CVaR-optimal deleveraging outcome as a function of the budget Q, computed using the water-filling construction that equalizes post-ADL leverage across active accounts whenever feasible. Black curves show individual post-ADL leverage trajectories $\ell_i(x_i)$, while the solid blue curve reports the mean leverage among at-risk accounts. The horizontal line indicates the stress threshold ℓ_β.

For small budgets Q, deleveraging is concentrated on the most levered accounts, which are progressively brought down toward ℓ_β. In this fully stressed regime, the CVaR objective decreases linearly in Q, reflecting identical marginal risk reduction across all stressed accounts.

Let Q_β denote the smallest budget for which all accounts can be brought below the stress threshold ℓ_β. At $Q = Q_\beta$, the system exits the stressed regime. For $Q > Q_\beta$, marginal risk reduction diminishes and the objective decreases at a strictly smaller rate, as shown by the change in slope in the inset of Figure 3.

This transition illustrates the degeneracy and non-uniqueness of CVaR-optimal allocations in the stressed regime. The water-filling rule selects a canonical representative from the set of minimizers, yielding smooth leverage paths and a transparent dependence on the available deleveraging budget.

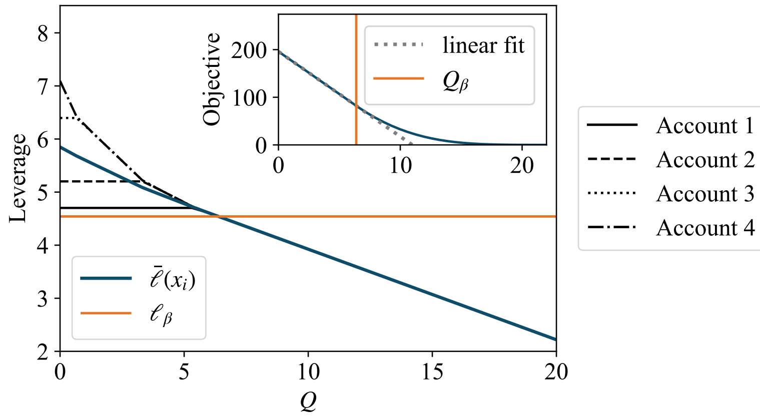

Figure 3: CVaR-optimal deleveraging under the water-filling rule. Black line-style curves correspond to individual post-ADL leverage levels $\ell_i(x_i)$, the solid blue curve shows the mean leverage among at-risk accounts, and the orange horizontal line indicates the stress threshold ℓ_β. The inset displays the CVaR objective value as a function of the total deleveraging budget Q, where the gray dotted line is a linear fit and the orange vertical line illustrates the deleveraging volume Q_β necessary to reach ℓ_β.

3. Multi-Asset Cross-Margining

This section extends the preceding single-asset analysis to a multi-asset, cross-margin setting. Under cross-margining, a single pool of collateral supports all open positions, so that gains in one position can offset losses in another and reduce the total margin required. This netting improves capital efficiency, but it also couples positions through shared collateral and correlated price moves. As a result, the ADL allocation decision is inherently multi-dimensional and must account for the joint distribution of returns across account positions. We formulate the resulting risk-based allocation problem faced by the exchange, focusing on expected loss as a risk measure. In Section 3.2 we then show that the ADL problem becomes separable *across accounts* once their coupling through the target reduction amounts Q is replaced by suitable shadow prices, opening the door to an efficient algorithm. Of course, this does not remove the possibly complicated coupling of the assets' returns. If assets are primarily driven by a common factor, this factor provides a natural dimension reduction. Section 3.4 formulates an idealized model where prices are driven by a single factor and shows that ADL then admits close parallels to the isolated-margining case of Section 2. Indeed, a generalized water-filling rule is seen to be optimal for reducing the expected loss. However, water-filling occurs in a quantity that we call factor leverage, rather than the naive (gross) leverage which does not take into account hedging effects across portfolio components.

3.1. Problem Formulation

In what follows, we consider d assets, with prices described by a vector in $\mathbb{R}^d$.[6] Each account $i \in \{1, \dots, n\}$ holds a signed position vector $q_i = (q_i^1, \dots, q_i^d) \in \mathbb{R}^d$, where $q_i^k > 0$ denotes a short position and $q_i^k < 0$ a long position in asset k. The entry-price vector is $p_i^{(e)} \in \mathbb{R}^d$, and the posted margin is $m_i \geq 0$. Auto-deleveraging is triggered at time τ with reference prices $p_\tau \in \mathbb{R}^d$ and the exchange must reduce aggregate exposure by a predetermined amount $Q \in \mathbb{R}^d$.[7] That is, each

[6] Negative prices are allowed to accommodate derivative contracts.

[7] We treat the trigger price p_τ and the aggregate reduction vector Q as exogenously given inputs to the ADL mechanism. In practice, their determination is itself a separate design problem for the exchange, involving choices

component Q^k of the vector Q specifies the aggregate position reduction required in asset k, to be allocated across the n accounts.

The ADL decision consists of vectors $x_1, \ldots, x_n \in \mathbb{R}^d$, where x_i^k is the (signed) position reduction applied to asset k in account i, so that the post-ADL position vector is $q_i - x_i$. We assume monotone deleveraging, in the sense that positions may be reduced toward zero but never increased or have their sign changed. Moreover, reductions are restricted to the direction of the target allocation Q: if $Q^k = 0$ then $x_i^k = 0$ for all i, and if $Q^k \neq 0$ then only accounts with $\operatorname{sgn}(q_i^k) = \operatorname{sgn}(Q^k)$ may be reduced in asset k. Equivalently, define the directional bounds

$$l_i^k \triangleq \begin{cases} 0, & Q^k \geq 0, \\ \min(0, q_i^k), & Q^k < 0, \end{cases} \qquad u_i^k \triangleq \begin{cases} \max(0, q_i^k), & Q^k > 0, \\ 0, & Q^k \leq 0, \end{cases} \tag{11}$$

so that, together with the aggregate reduction constraint $\sum_{i=1}^n x_i = Q$, the feasible set is

$$\mathcal{X} \triangleq \Big\{ x = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n : \sum_{i=1}^n x_i = Q, \quad l_i^k \leq x_i^k \leq u_i^k \quad \forall\, i, k \Big\}.$$

Assumption 3.1 (Feasibility). *Q satisfies*

$$Q^k \in \left[\sum_{i=1}^n \min(0, q_i^k),\ \sum_{i=1}^n \max(0, q_i^k) \right], \qquad k = 1, \ldots, d,$$

so that the feasible set $\mathcal{X}$ is nonempty.

For each account i, equity remains a scalar quantity but depends on the full price vector. If account i is reduced by x_i and the price vector at time $T > \tau$ is $p_T \in \mathbb{R}^d$, define

$$e_i(x_i, p_T) \triangleq q_i^\top (p_i^{(e)} - p_T) - x_i^\top (p_\tau - p_T) + m_i.$$

Writing $E_i \triangleq e_i(0, p_\tau) = q_i^\top (p_i^{(e)} - p_\tau) + m_i$ to be the equity at the ADL time, we equivalently have

$$e_i(x_i, p_T) = E_i + (q_i - x_i)^\top (p_\tau - p_T). \tag{12}$$

The second term captures the remaining exposure to the price move from p_τ to p_T after reducing positions by x_i. Throughout, we restrict attention to accounts that are solvent at time τ.

Assumption 3.2 (Solvency). *$E_i > 0$ for all $i = 1, \ldots, n$.*

As before, equity at time τ is unchanged by the reduction x_i, since $e_i(x_i, p_\tau) = E_i$ for all x_i. By analogy with (2) and (3), the corresponding shortfall is

$$\sigma_i(x_i, p_T) = (-e_i(x_i, p_T))_+, \tag{13}$$

and the total exchange loss is

$$\mathcal{L}(x, p_T) \triangleq \sum_{i=1}^n \sigma_i(x_i, p_T).$$

In the single-asset case, post-ADL equity is monotone in the terminal price p_T. Upward moves uniformly harm shorts and downward moves uniformly harm longs, which induces a natural ordering of accounts by "distance to default" and makes leverage a meaningful proxy for default risk. In the

of marking rules, liquidity considerations, and stress metrics. The present analysis focuses on the allocation problem conditional on these quantities.

multi-asset case, equity depends on the entire price vector and portfolios may be partially hedged across assets. There is generally no canonical stress direction and no a priori risk ordering across accounts, since insolvency depends on the joint distribution of $(p_T^1, \dots, p_T^d)$.

In multi-asset settings, exchanges often report leverage-type statistics that extend the one-dimensional notion (motivated, for instance, by the discussion of 'portfolio margin' in [Binance, 2019]). One natural extension is the *gross leverage* of account i after reduction x_i, defined by

$$\ell_i(x_i) \triangleq \frac{\sum_{k=1}^d |p_\tau^k (q_i^k - x_i^k)|}{E_i},$$

where the numerator is the gross exposure of the residual portfolio valued at the reference prices p_τ. Gross leverage is easy to compute but does not necessarily reflect default risk. Since it ignores the dependence structure of returns, it can be large even when exposures offset across assets, as in a delta-neutral portfolio. We use "gross" to emphasize that such netting is not reflected.

With this extended setup in mind, we model the exchange's ADL decision analogously to Section 2. Given a random close-out price p_T, the exchange selects reductions $x \in \mathcal{X}$ to minimize a prescribed risk functional $\rho(\cdot)$ of the induced loss:

$$\begin{array}{ll} \underset{x \in (\mathbb{R}^d)^n}{\text{minimize}} & \rho\big(\mathcal{L}(x, p_T)\big) \\ \text{subject to} & x \in \mathcal{X}. \end{array} \tag{14}$$

General risk measures such as $\rho(\cdot) = \mathrm{CVaR}_\beta(\cdot)$ couple accounts through tail scenarios of the aggregate loss. In the multi-asset setting, individual account shortfalls need not be comonotone, and feasible ADL allocations are generally not totally ordered by a single exposure variable (such as leverage). Consequently, different risk functionals can rank allocations differently and lead to different optimal ADL policies. In fact, unlike the single-asset case in which a water-filling allocation is always optimal, the sets of optimizers associated with distinct confidence levels β can be *disjoint*; see Appendix B.1 for an example.

For the subsequent analysis, we focus on the expected loss $\rho(\cdot) = \mathsf{E}[\cdot]$ and specialize (14) to

$$\begin{array}{ll} \underset{x \in (\mathbb{R}^d)^n}{\text{minimize}} & \mathsf{E}\big[\mathcal{L}(x, p_T)\big] \\ \text{subject to} & x \in \mathcal{X}. \end{array} \tag{15}$$

The following ensures that (15) is well defined.

Assumption 3.3. *The random vector p_T is integrable.*

While the expected loss criterion is more tractable thanks to additive separability across accounts, the coupling of assets remains a central complication that generally prevents reduction to single-variable problems. We emphasize that both sides of the ADL problem can be multi-dimensional in our formulation: the vector Q means that several assets undergo ADL simultaneously, and each affected account may hold several of those assets, as well as other assets not subject to ADL but still contributing to potential losses.

3.2. Separable Decomposition for the Expected Loss

As noted, the expected loss objective is additively separable across accounts:

$$\mathsf{E}\big[\mathcal{L}(x, p_T)\big] = \sum_{i=1}^n \mathsf{E}\big[\sigma_i(x_i, p_T)\big]. \tag{16}$$

Accordingly, the only coupling across accounts in (15) arises through the vector clearing constraint $\sum_{i=1}^{n} x_i = Q$. This suggests a decomposition based on convex duality, in which the clearing constraint is priced by a multiplier $\lambda \in \mathbb{R}^d$ and each account solves an independent subproblem.

Define the per-account feasible set

$$\mathcal{Y}_i \triangleq \Big\{ x_i \in \mathbb{R}^d : l_i^k \le x_i^k \le u_i^k \ \ \forall\, k \Big\}, \qquad i = 1, \ldots, n. \tag{17}$$

Let $\mathcal{Y} \triangleq \prod_{i=1}^{n} \mathcal{Y}_i$ be the joint set of constraints, and define the partial Lagrangian

$$\widehat{\mathcal{L}}(x, \lambda) \triangleq \sum_{i=1}^{n} \mathsf{E}[\sigma_i(x_i, p_T)] + \lambda^\top \Big(\sum_{i=1}^{n} x_i - Q \Big) = -\lambda^\top Q + \sum_{i=1}^{n} \Big(\mathsf{E}[\sigma_i(x_i, p_T)] + \lambda^\top x_i \Big).$$

For $\lambda \in \mathbb{R}^d$, define the primal per account value function by

$$\phi_i(\lambda) \triangleq \min_{x_i \in \mathcal{Y}_i} \Big\{ \mathsf{E}[\sigma_i(x_i, p_T)] + \lambda^\top x_i \Big\}, \qquad i = 1, \ldots, n, \tag{18}$$

and the best-response correspondence

$$X_i(\lambda) \triangleq \operatorname*{argmin}_{x_i \in \mathcal{Y}_i} \ \Big\{ \mathsf{E}[\sigma_i(x_i, p_T)] + \lambda^\top x_i \Big\}, \qquad i = 1, \ldots, n. \tag{19}$$

The associated dual function is

$$g(\lambda) \triangleq \min_{x \in \mathcal{Y}} \ \widehat{\mathcal{L}}(x, \lambda) = -\lambda^\top Q + \sum_{i=1}^{n} \phi_i(\lambda).$$

The next proposition gives a convenient characterization of optimality in terms of best responses and the dual function, which will be exploited in Section 3.3 below to develop an efficient numerical approach.

Proposition 3.4 (Dual decomposition). *Let Assumptions 3.1–3.3 hold. Then:*

(i) The program (15) is convex and admits an optimal allocation.

(ii) An allocation $x^\star = (x_1^\star, \ldots, x_n^\star) \in (\mathbb{R}^d)^n$ is optimal if and only if there exists $\lambda^\star \in \mathbb{R}^d$ such that

$$x_i^\star \in X_i(\lambda^\star) \ \ \textit{for all } i, \qquad \textit{and} \qquad \sum_{i=1}^{n} x_i^\star = Q.$$

(iii) Any such $\lambda^\star$ maximizes the dual function g, which, for any λ, has supergradient set

$$\partial g(\lambda) = \Big\{ \sum_{i=1}^{n} x_i - Q : \ x_i \in X_i(\lambda) \ \forall i \Big\}. \tag{20}$$

Proposition 3.4 admits a natural economic interpretation. The multiplier $\lambda^\star \in \mathbb{R}^d$ acts as an asset-by-asset vector of *shadow prices* for deleveraging capacity. Given λ, each account i solves the independent subproblem (18) which trades off its expected shortfall contribution against the linear charge $\lambda^\top x_i$ for consuming reductions in each asset. The clearing condition $\sum_{i=1}^{n} x_i^\star = Q$ selects $\lambda^\star$ so that the resulting best responses match the required aggregate reduction. In this sense, the optimal ADL allocation can be viewed as a competitive equilibrium in which the exchange posts $\lambda^\star$ and accounts respond optimally subject to feasibility.

3.3. Numerical Solution: ADMM with Sample-Average Approximation

Proposition 3.4 explains why the problem is structurally amenable to decomposition: the expected-loss objective is separable across accounts, while the market-clearing constraint introduces the only global coupling. At exchange scale, however, the exact expectation layer is computationally intractable, so we replace it by a sample-average approximation (SAA) and solve the resulting finite-scenario problem by the alternating direction method of multipliers (ADMM), a standard operator-splitting method for structured convex optimization [Boyd et al., 2011]. The same SAA reformulation also provides a practical numerical route for the low-dimensional examples considered below.

SAA reformulation. Draw S independent price-increment scenarios $\Delta p^{(1)}, \dots, \Delta p^{(S)}$ from the calibrated price model, where $\Delta p^{(s)} \triangleq p_\tau - p_T^{(s)}$ for scenario s. For account i, define

$$f_i(x_i) \;\triangleq\; \frac{1}{S}\sum_{s=1}^{S}\left(-E_i - (q_i - x_i)^\top \Delta p^{(s)}\right)_+ + I_{\mathcal{Y}_i}(x_i), \tag{21}$$

where $I_{\mathcal{Y}_i}$ is the indicator of the per-account box. Each f_i is closed, proper, and convex, being the sum of finitely many convex piecewise-linear terms restricted to a box. The SAA counterpart of (15) is therefore

$$\min_{x\in(\mathbb{R}^d)^n} \left\{ \sum_{i=1}^{n} f_i(x_i) \;:\; \sum_{i=1}^{n} x_i = Q \right\}.$$

Consensus splitting. Introduce an auxiliary copy $z = (z_1, \dots, z_n) \in (\mathbb{R}^d)^n$ and impose $x_i = z_i$ for all i together with the clearing condition $\sum_i z_i = Q$. The SAA problem becomes

$$\begin{aligned} \underset{x,z}{\text{minimize}} \quad & \sum_{i=1}^{n} f_i(x_i) + I_{\mathcal{C}}(z) \\ \text{subject to} \quad & x - z = 0, \end{aligned} \tag{22}$$

where $\mathcal{C} \triangleq \{z \in (\mathbb{R}^d)^n : \sum_{i=1}^{n} z_i = Q\}$. Although the first block decomposes across accounts, the optimization variables are grouped into the standard two ADMM blocks $x = (x_1, \dots, x_n)$ and $z = (z_1, \dots, z_n)$. The per-account box constraints are enforced in the x-update through the indicators $I_{\mathcal{Y}_i}(x_i)$, and the z-update is the Euclidean projection onto the market-clearing set $\mathcal{C} = \{z \in (\mathbb{R}^d)^n : \sum_i z_i = Q\}$.

ADMM updates. In scaled dual-variable form, the ADMM iteration for (22) is

$$x_i^{k+1} \in \underset{x_i\in\mathcal{Y}_i}{\operatorname{argmin}} \left\{ f_i(x_i) + \frac{\rho}{2}\|x_i - c_i^k\|_2^2 \right\}, \quad c_i^k \triangleq z_i^k - u_i^k, \quad i = 1, \dots, n, \tag{23}$$

$$z^{k+1} = \Pi_{\mathcal{C}}(x^{k+1} + u^k), \tag{24}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1}, \tag{25}$$

where $\rho > 0$ is the augmented-Lagrangian penalty parameter and $u = (u_1, \dots, u_n)$ is the scaled dual variable. Because $f(x) = \sum_i f_i(x_i)$ and the quadratic penalty is blockwise additive, the x-update (23) separates exactly across accounts. The only global coupling enters through the projection step (24), which enforces market clearing by projecting $x^{k+1} + u^k$ onto the affine set $\mathcal{C}$. The full procedure is stated in Algorithm 1 (Appendix C.1).

Convergence. For fixed penalty parameter ρ, the present formulation is an exact two-block convex consensus ADMM, so the standard convergence theory of Boyd et al. [2011, Section 3.2] applies. In particular, the primal and dual residuals converge to zero and the objective values converge to the optimal value of the SAA problem. A formal statement and proof are deferred to Appendix C.1.

Remark 3.5. The convergence result is asymptotic and does not by itself determine the iteration budgets relevant in practice. For the exchange-scale problem, the quantitatively important question is therefore how quickly the iterates recover the economically relevant part of the loss reduction. We return to that question in the exchange-scale implementation and in the empirical convergence diagnostics reported in Section C.5.

Each ADMM iteration solves n independent per-account proximal subproblems of the form (23), followed by one projection onto $\mathcal{C}$ and a dual update, both of which have cost $O(nd)$. The overall per-iteration cost depends on n, d, and the scenario count S, while the accountwise proximal updates remain fully parallelizable. The exchange-scale implementation details that make these updates tractable in the large-instance regime are described in Section 4.2.

3.4. Single Factor Model and Clipped Water-Filling

We now formulate an idealized model where asset returns are fully driven by a single factor. We will show that the solution of the ADL problem then admits a generalized water-filling structure, thus drawing a parallel with the isolated-margin case of Section 2. Importantly, water-filling needs to be applied to a quantity that we call factor leverage, not to gross leverage. Factor leverage correctly reflects the "distance to default" in this setting, whereas gross leverage ignores hedging effects. Structurally, the key simplification in the single factor model is that each account's loss depends on its post-ADL portfolio only through a *scalar* factor exposure.

Assumption 3.6 (Single factor). *There exist $v \in \mathbb{R}^d$ and a scalar random variable ϵ such that*

$$p_T = p_\tau + \epsilon v,$$

where ϵ admits a strictly positive density on $\mathbb{R}$ and $\mathsf{E}[|\epsilon|] < \infty$.

For the remainder of Section 3.4, we assume that Assumptions 3.1–3.6 hold. Next, we introduce factor leverage, taking over the role that leverage played in Section 2.

Definition 3.7 (Factor leverage). *For each account i and allocation x_i, define the* factor leverage *by*

$$\ell_i^{(v)}(x_i) \triangleq \frac{v^\top (q_i - x_i)}{E_i}.$$

Thanks to Assumptions 3.2 and 3.6, the equity (12) can be written as

$$e_i(x_i, p_T) = E_i - \epsilon\, v^\top (q_i - x_i) = E_i(1 - \epsilon\, \ell_i^{(v)}(x_i)), \tag{26}$$

so the expected shortfall contribution of account i depends on x_i only through the factor leverage $\ell_i^{(v)}(x_i)$. For each account i, recall from (17) the set $\mathcal{Y}_i$ of feasible allocations and introduce the feasible factor-leverage interval

$$[\underline{\ell}_i, \overline{\ell}_i] \triangleq \{\ell_i^{(v)}(x_i) : \; x_i \in \mathcal{Y}_i\}, \qquad i = 1, \dots, n.$$

This set is indeed a closed, bounded interval since $\ell_i^{(v)}(\cdot)$ is an affine transformation and $\mathcal{Y}_i$ is convex and compact.

We observe that the aggregate constraint $\sum_{i=1}^{n} x_i = Q$ implies a fixed equity-weighted factor exposure, denoted $L^{(v)}$ below. Indeed, for any feasible allocation $x \in \mathcal{X}$,

$$\sum_{i=1}^{n} E_i \ell_i^{(v)}(x_i) = \sum_{i=1}^{n} v^\top (q_i - x_i) = v^\top \Big(\sum_{i=1}^{n} q_i - Q \Big) \triangleq L^{(v)}. \tag{27}$$

We can then think of different allocations as redistributing this fixed total across accounts. Under certain implementability conditions, we will see that a clipped water-filling rule for the factor leverage $\ell_i^{(v)}(x_i)$ is optimal: In principle, we would like to minimize the maximal factor leverage $\ell_i^{(v)}(x_i)$ in (27). However, this water-filling is clipped to $[\underline{\ell}_i, \bar{\ell}_i]$ because an account's factor leverage can only be reduced as long as the account contains the asset(s) being deleveraged — in contrast to the isolated-margining case, the factor leverage may remain high due to other assets in the portfolio that are outside the reach of ADL. See Fig. 4 for an illustration.

Intuitively, the general idea is still to first find a target factor leverage $t^\star$ that clears the market and then allocate quantities so as to drain leverages above the target; however, the clipping means that some accounts will need to have different targets, denoted by $\ell_i^\star$ below. The following theorem formalizes this structure.

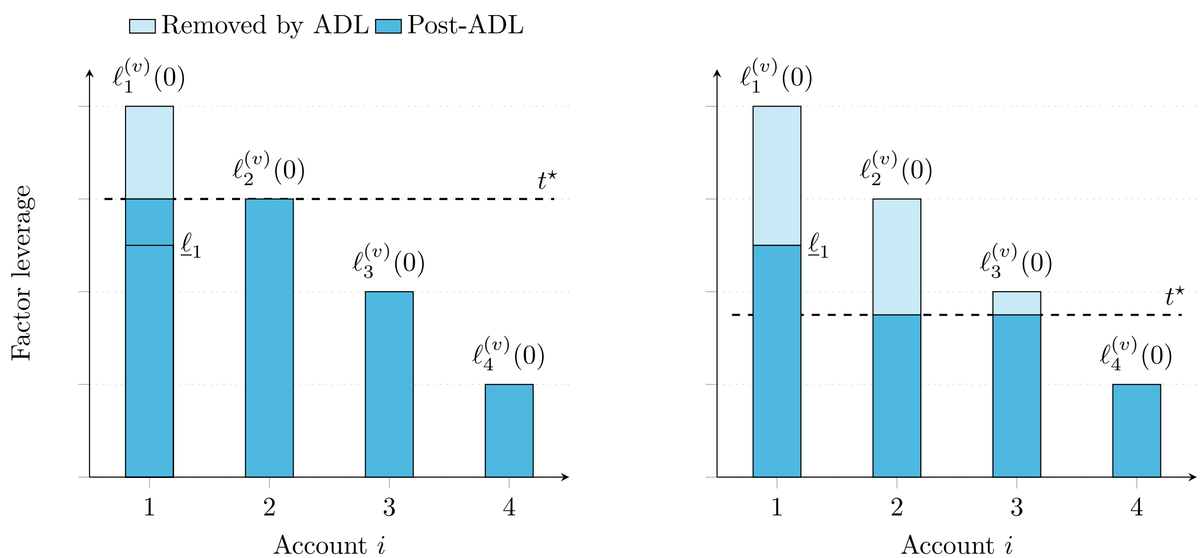

Figure 4: Water-filling with clipping. The target threshold $t^\star$ equalizes post-ADL factor leverage for interior accounts, but account-specific bounds can bind. In the right panel, account 1 cannot be reduced below $\underline{\ell}_1$, so its post-ADL factor leverage is clipped at $\underline{\ell}_1$ rather than being reduced to $t^\star$.

Theorem 3.8 (Verification theorem for factor water-filling). *For any $\eta \in \mathbb{R}$, define the* target factor leverage *for account i by*

$$\ell_i^\star(\eta) \in \operatorname*{argmin}_{z \in [\underline{\ell}_i, \bar{\ell}_i]} \{\psi(z) - \eta z\}, \quad i = 1, \dots, n, \qquad \textit{where} \quad \psi(z) \triangleq \mathsf{E}\big[(\epsilon z - 1)_+\big], \quad z \in \mathbb{R}. \tag{28}$$

(i) The minimizer $\ell_i^\star(\eta)$ is unique for any $\eta \in \mathbb{R}$, and $\eta \mapsto \ell_i^\star(\eta)$ is continuous and nondecreasing.

(ii) For any $\eta \in \mathbb{R}$, $\ell_i^\star(\eta)$ satisfies the clipped water-filling rule[8]

$$\ell_i^\star(\eta) = \begin{cases} \underline{\ell}_i, & \eta \le \psi'(\underline{\ell}_i), \\ (\psi')^{-1}(\eta), & \psi'(\underline{\ell}_i) < \eta < \psi'(\overline{\ell}_i), \\ \overline{\ell}_i, & \eta \ge \psi'(\overline{\ell}_i). \end{cases} \tag{29}$$

(iii) There exists at least one $\eta^\star \in \mathbb{R}$ such that

$$\sum_{i=1}^n E_i\, \ell_i^\star(\eta^\star) = v^\top \Big(\sum_{i=1}^n q_i - Q \Big). \tag{30}$$

Given $\eta^\star$ satisfying (30), if there exists an allocation $x^\star \in \mathcal{X}$ such that

$$\ell_i^{(v)}(x_i^\star) = \ell_i^\star(\eta^\star) \quad \textit{for all } i,$$

then $x^\star$ is an optimal allocation for (15).

By monotonicity of the map $\eta \mapsto \ell_i^\star(\eta)$, a smaller right-hand side in the budget equation (30) corresponds to a smaller value of $\eta^\star$ and therefore to smaller targets $\ell_i^\star(\eta^\star)$ for each account i. This yields the (clipped) water-filling interpretation: as the aggregate factor-exposure target is tightened, accounts are reduced along the common "water level" until they hit their individual constraints.

Mathematically, Theorem 3.8 is a *verification* result. It states that if the clipped one-dimensional minimizers (29) can be realized by some feasible allocation $x \in \mathcal{X}$, then that allocation is globally optimal. In particular, the targets $\ell_i^\star(\eta^\star)$ are always compatible with the per-account feasibility constraints $\mathcal{Y}_i$, by construction. The only remaining requirement is the clearing condition $\sum_{i=1}^n x_i = Q$. In general multi-asset ADL, such a realization need not exist because $\sum_{i=1}^n x_i = Q$ is a vector constraint, whereas (30) specifies only a single scalar target for the accounts. A degenerate example in which no feasible allocation achieves the water-filling targets is given in Appendix B.5.

However, it turns out that in many relevant situations, factor water-filling is implementable (and therefore optimal). One particular case is when only a single asset is subject to ADL, that is, only one component of Q is nonzero. Note that this situation is still quite different from the isolated-margin case in Section 2, because the accounts being deleveraged now contain further assets that influence potential future losses.

Theorem 3.9 (Factor water-filling for single-asset ADL)**.** *Suppose that only asset k_0 is subject to ADL, i.e. $Q^k = 0$ for all $k \neq k_0$, and $v^{k_0} \neq 0$. Then clipped water-filling on factor leverage is implementable and optimal for (15).*

Specifically, let $\eta^\star$ be any solution of the budget equation (30), and define the water-filling targets $\ell_i^\star(\eta^\star)$ by (28). Then there exists an allocation $x^\star \in \mathcal{X}$ such that

$$\ell_i^{(v)}(x_i^\star) = \ell_i^\star(\eta^\star) \quad \textit{for all } i,$$

and $x^\star$ is an optimal allocation for (15).

Intuitively, Theorem 3.9 holds because, when only a single asset is deleveraged, each account's factor leverage $\ell_i^{(v)}(x_i)$ is an affine function of the scalar decision $x_i^{k_0}$. As a result, the

[8] The expressions in (29) are well-defined since the convex function $\psi : \mathbb{R} \to \mathbb{R}_+$ is continuously differentiable with a strictly increasing derivative ψ'; cf. Appendix B.3.

one-dimensional water-filling targets are automatically implementable while satisfying the (scalar) clearing constraint.

The single-asset case described by Theorem 3.9 is a special case of a much more general sufficient condition for multi-asset ADL detailed in Appendix B.6. It is based on a technical *connectivity* property across the actively delevered assets (i.e., the nonzero components of Q), capturing the idea that some accounts have partial reductions in multiple assets. Mathematically, it is a weak form of an interior point condition, avoiding degenerate boundary cases like the example of Appendix B.5. Intuitively, such connectivity is most plausible in large exchanges when many large accounts maintain genuinely cross-asset portfolios, creating connections across the assets being delevered.

Theorem 3.10 (Factor water-filling for multi-asset ADL). *Suppose that $v^k \neq 0$ unless $Q^k = 0$, and that Q and the accounts being delevered satisfy the connected partial deleveraging condition*[9] *detailed in Appendix B.6. Then clipped water-filling on factor leverage is again implementable and optimal for* (15)*, as in Theorem 3.9.*

Theorem 3.10 shows that the single-asset water-filling intuition survives in the multi-asset cross-margin setting if assets are driven by a single factor. Under that structure, the residual factor leverage determines default risk, rather than gross leverage. Relative to the single-asset case, the only additional obstacle is whether the scalar water-filling targets can actually be implemented under the vector clearing constraint, and Theorem 3.10 provides a sufficient condition for this.

Remark 3.11. A similar message carries over to the CVaR_β formulation. Although CVaR_β generally couples accounts through tail scenarios and therefore does not admit the same separability as expected loss, the single-factor structure restores a one-dimensional ordering by factor leverage and the optimal ADL rule is again similar to the single-asset case — once β is fixed and assuming we further restrict to accounts with positive factor exposure, the allocation problem takes a clipped water-filling form in factor-leverage space. Thus, the extension from expected loss to CVaR preserves the basic economic insight: the exchange optimally allocates deleveraging by equalizing a marginal tail-risk criterion across accounts until individual feasibility constraints become binding. A more detailed discussion can be found in Section B.7.

3.5. Numerical Example

In this section, we illustrate ADL under a bivariate price model for BTC and ETH, and discuss the accuracy of the single-factor approximation discussed in Section 3.4. Throughout, we fix the account state at the ADL time τ and vary only the required aggregate BTC reduction. Specifically, we set

$$Q = (Q^{\mathrm{BTC}}, Q^{\mathrm{ETH}}) = (Q^{\mathrm{BTC}}, 0),$$

so that ETH positions are not deleveraged. Recall, however, that the ETH positions affect future losses and therefore the optimal BTC allocation. Terminal close-out prices are generated from a correlated bivariate geometric Brownian motion over the horizon $\Delta = T - \tau = 10/365$. Concretely,

$$p_T^k = p_\tau^k \exp\left(-\frac{1}{2}(\sigma_{\mathrm{ann}}^k)^2\Delta + \sigma_{\mathrm{ann}}^k\sqrt{\Delta}\, Z^k\right), \qquad k \in \{\mathrm{BTC}, \mathrm{ETH}\},$$

[9] See Definition B.7 and Proposition B.8. Loosely speaking, connected partial deleveraging means that each asset subject to ADL is *partially* reduced for at least one account and that the assets are linked through overlapping reductions. In other words, you can move from any delevered asset to any other by "hopping" through accounts that are partially reduced in more than one of these assets.

where $(Z^{\mathrm{BTC}}, Z^{\mathrm{ETH}})$ is standard bivariate normal with correlation ρ_{corr}. Using approximate spot prices of February 5, 2026 for p_τ and a high correlation consistent with stressed market conditions, we set

$$p_\tau^{\mathrm{BTC}} = \$67{,}000, \qquad p_\tau^{\mathrm{ETH}} = \$1{,}900, \qquad \sigma_{\mathrm{ann}}^{\mathrm{BTC}} = 60\%, \qquad \sigma_{\mathrm{ann}}^{\mathrm{ETH}} = 75\%, \qquad \rho_{\mathrm{corr}} = 0.85.$$

For each value of Q^{BTC}, we solve the SAA approximation of the expected-loss problem by the ADMM scheme of Section 3.3. The expectation is approximated with $S = 2048$ GBM scenarios, and the ADMM iteration is run with penalty $\rho = 30$ and an iteration budget of 500.

To compare the full model with the theory of Section 3.4, we also solve the same ADL problem under the following one-factor approximation. Let $\Sigma_{\Delta p}$ denote the covariance matrix of the price increment $p_T - p_\tau$ implied by the GBM calibration, and let (λ_1, u_1) be its leading eigenpair. We then set

$$v \triangleq \sqrt{\lambda_1}\, u_1, \qquad p_T = p_\tau + \epsilon v, \qquad \epsilon \sim N(0,1).$$

The calculation of v is detailed in Section B.8, where we also illustrate the approximation with a scatter plot. The one-factor model retains the dominant covariance mode and suppresses the orthogonal direction. In particular, "factor exposure" will refer to the factor leverage $\ell_i^{(v)}$ computed with this leading covariance direction v.

We consider $n = 4$ cross-margin accounts. All four are short BTC, but their ETH positions differ substantially, generating different exposures to the dominant covariance mode. Table 2 reports the account state at time τ. Positions are signed: positive entries denote shorts and negative entries denote longs. Account 4 is approximately factor-neutral despite having the largest gross leverage, account 2 is partially hedged, and accounts 1 and 3 are the most exposed to the dominant market factor.

Table 2: Account state at the ADL time τ, including gross leverage $\ell_i(0)$ and factor leverage $\ell_i^{(v)}(0)$.

Account	q_i^{BTC}	q_i^{ETH}	m_i in \$1000	E_i in \$1000	$\ell_i(0)$	$\ell_i^{(v)}(0)$
1	8.0	323.0	137.5	242.1	4.8	0.49
2	10.0	−38.7	85.3	143.0	5.2	0.41
3	8.0	326.2	75.4	180.6	6.4	0.66
4	7.0	−190.0	43.9	116.9	7.1	0.07

The left panel in Fig. 5 shows the ADL results obtained from the bivariate GBM model through the SAA–ADMM approximation, whereas the right panel shows the results under the one-factor approximation. Consistent with the theoretical results, the right panel illustrates water-filling in factor exposure: the accounts with the largest factor exposure are progressively leveled, except for the clipping that occurs when feasibility constraints become binding. The left panel shows a qualitatively similar behavior, illustrating that the overall shape of the allocation paths is governed primarily by the dominant covariance mode and that the water-filling rule captures the principal characteristics. Quantitatively, the two models are close, but not identical. In the one-factor approximation, the dominant covariance mode v completely governs the shape of the optimal allocation paths. The full GBM model retains two features that are absent from the rank-one approximation: lognormal nonlinearity and residual risk in the covariance direction orthogonal to v. These effects are economically most visible at the points where the active set changes. Once the dominant factor exposures of the currently active accounts have been brought close together, the neglected second direction becomes relatively more important, and the SAA–ADMM allocation in the full bivariate model can deviate modestly from the exact water-filling pattern.

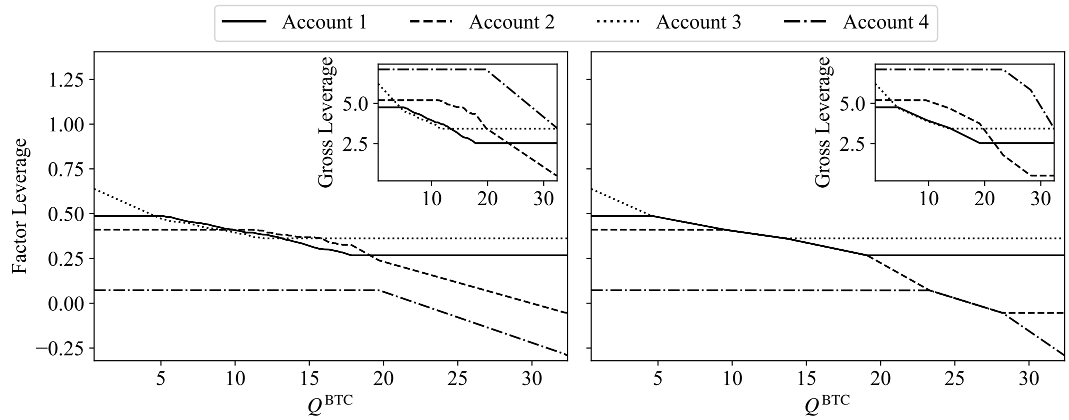

Figure 5: Risk-optimal BTC deleveraging paths under the bivariate GBM model, computed from the SAA–ADMM approximation with 2048 scenarios, $\rho = 30$, and 500 iterations (left panel), and under the one-factor approximation (right panel). The main panels plot post-ADL factor leverage $\ell_i^{(v)}(x_i^\star)$ against the BTC deleveraging budget Q^{BTC}; the insets report gross leverage $\ell_i(x_i^\star)$ for the same allocations.

Both panels highlight the distinction between factor leverage and gross leverage. The insets show that gross leverage would give a misleading ranking of risk. In particular, account 4 has the largest gross leverage throughout much of the range, yet its factor exposure is initially close to zero and it receives little or no deleveraging until BTC reductions become large. Conversely, accounts 1 and 3 are reduced first even though their gross leverage is lower. What drives the optimal ADL decision is not gross notional exposure, but exposure to the dominant risk factor after cross-asset hedges are taken into account.

4. The Exchange-Scale Problem

A perpetual futures exchange operating a cross-margin book faces a qualitatively different computational problem from the stylized instances studied in Section 3.3. When an ADL event is triggered, the exchange must solve the expected-loss program (15) across a universe of tens of thousands of accounts, each holding positions in dozens of actively traded assets, with an aggregate reduction target that can reach hundreds of millions of dollars in notional within a matter of seconds. The October 10, 2025 Hyperliquid event is a concrete instance of this problem. It is also a natural case study: it is one of the largest publicly observable ADL episodes of 2025, it unfolds in multiple distinct waves across a broad asset universe, and it generated substantial public discussion about how losses were distributed. For our purposes, the key feature is that the event yields an observed realized ADL allocation $x^{\mathrm{Hyperliquid}}$ that can be compared with model-based allocations on the reconstructed event.

This section addresses two implementation challenges and one empirical question. On the implementation side, exchange scale rules out direct expectation-based solution methods and requires an implementation of the SAA–ADMM scheme of Section 3.3 that remains tractable when both the number of accounts and the number of active assets are large. In addition, the active coins are highly heterogeneous in price level and support structure, so the implementation must normalize variable scales and exploit per-account sparsity in order to remain numerically stable and parallelizable. On the empirical side, the central question is comparative: given a reconstructed ADL event, how does the realized Hyperliquid allocation compare with the sequential one-factor Water-filling benchmark induced by Section 3.4 and with the Risk-optimal allocation computed from the same SAA–ADMM scheme?

The comparison is informative for two reasons. First, the theory above characterizes the Risk-optimal allocation and identifies a one-factor Water-filling benchmark, but it does not say how far current exchange practice is from either object once portfolios are cross-margined and assets are heterogeneous. Second, the comparison reveals not only the objective gap but also its mechanism. The empirical comparison will show that the relevant differences are not exhausted by a single expected-loss ranking. They also concern how deleveraging notional is concentrated across users, how strongly reductions align with factor leverage, and to what extent the realized event resembles partial portfolio reduction as opposed to near-full directional closeout.

Section 4.1 introduces the October 10, 2025 Hyperliquid event, the data sources, the reconstruction of pre-event account states, and the resulting wave-level optimization inputs. Section 4.2 then explains how the abstract ADMM scheme is instantiated at exchange scale. Finally, Section 4.3 focuses on the first wave τ_1 and compares Hyperliquid, Water-filling, and the Risk-optimal allocation in terms of sampled expected loss and user-level targeting. The corresponding τ_3 exercise and the robustness comparison with the directional Water-filling benchmark are deferred to the appendix.

4.1. Empirical Setup

The ADL Event. On October 10, 2025, the perpetual futures exchange Hyperliquid experienced a concentrated auto-deleveraging episode in which a large number of accounts were partially or fully closed within a five-minute window. Our empirical analysis is based on the 75-coin core universe constructed in Appendix C.2. On this universe, the total realized ADL notional on October 10 is $\$2.047 \times 10^9$. Within the event window 21:16–21:21 UTC, the ADL flow unfolds in multiple distinct bursts, identifiable by sharp increases in cumulative notional volume. We mark three reference times

$$\tau_1 = \text{21:16:05 UTC}, \qquad \tau_2 = \text{21:16:56 UTC}, \qquad \tau_3 = \text{21:17:06 UTC},$$

which correspond to the onset of the three dominant ADL waves visible in Figure 6.

The three waves differ substantially in scale and composition. The first wave τ_1 has total realized ADL notional $\$1.769 \times 10^8$; the dominant coins are HYPE, PUMP, and FARTCOIN. The second wave τ_2 is smaller at $\$9.33 \times 10^7$. The third wave τ_3 has a total realized ADL notional $\$3.866 \times 10^8$, concentrated mainly in ETH, SOL, BTC, and XRP.

Figure 7 provides complementary market context by plotting intraday price paths on October 10, 2025, normalized by each coin's average price over the day up to the ADL event. The figure shows that the exchange was not facing an isolated liquidation in a single asset, but a broad and rapid repricing across the active universe. Even the deepest markets, such as BTC, ETH, and SOL, decline by roughly 10–20%, while several thinner names lose more than half their value relative to their intraday average. In such conditions, the allocation problem must be solved under tight time constraints, and the exchange must distribute a large aggregate reduction across accounts exposed to assets with heterogeneous liquidity and common return directions.

Data sources. The empirical input combines two data sources. First, tick-level trade records are obtained from the public Hyperliquid REST API. These records contain the trade time, user address, coin, signed size, side, direction label, execution price, and liquidation mark price where applicable, allowing us to identify the realized ADL flow separately from market-driven liquidations and voluntary trades. Second, account-level state variables — position sizes, entry prices, and

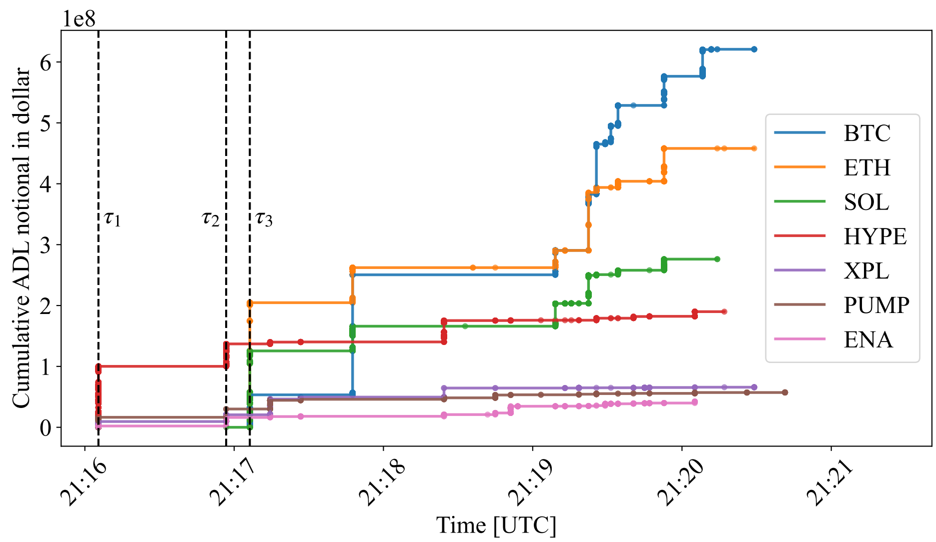

Figure 6: Cumulative notional ADL volume by coin during the October 10, 2025 event, restricted to the five-minute window 21:16–21:21 UTC. Vertical dashed lines mark τ_1, τ_2, and τ_3. The seven largest coins by total ADL notional are shown; each step corresponds to an individual ADL fill.

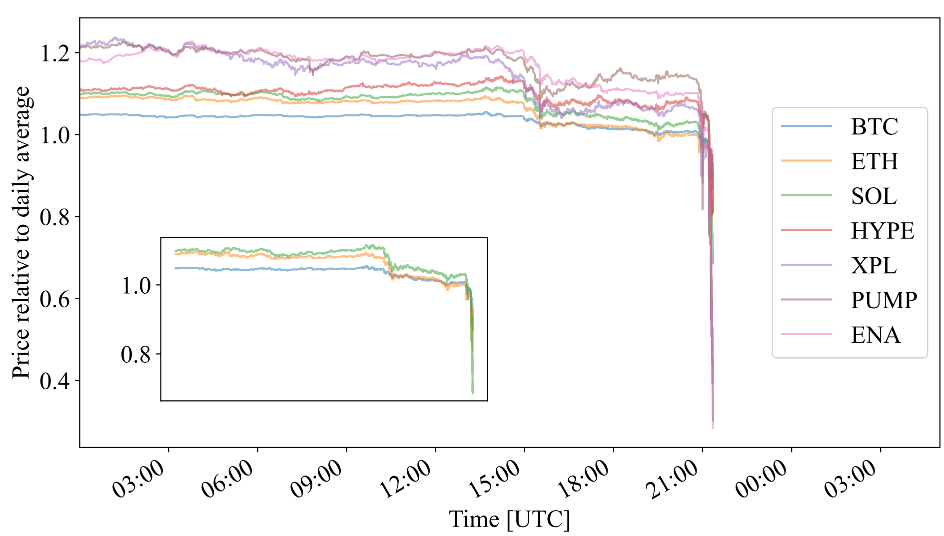

Figure 7: Intraday price paths on October 10, 2025, normalized by each coin's average price up to the ADL event. The main panel shows the most actively traded affected coins, while the inset isolates the most liquid names. The cross-section is highly heterogeneous: several thinner coins lose more than 50%, whereas BTC, ETH, and SOL still fall by roughly 10–20%.

margin-used quantities per user and coin — are obtained from Allium, which indexes on-chain Hyperliquid account data. The Hyperliquid tape therefore tells us what was actually delevered in the event, while the Allium snapshots determine which pre-event account states can be reconstructed reliably for the optimization input.

Data reconstruction and limitations. Because Allium only covers accounts with observable activity since September 22, 2025, not every user appearing in the ADL tape can be assigned a strict pre-event state. For each reconstruction time τ_j, we therefore recover the optimization input by taking the most recent available account snapshot before τ_j and rolling it forward with the Hyperliquid trade tape. The optimization is carried out on the retained reconstructed universe rather than on the full raw event stream. At τ_1, this is why the final ADL target volume used in the optimization, $\$1.539 \times 10^8$, is smaller than the raw event volume $\$1.769 \times 10^8$. Appendix C.2 contains the full reconstruction pipeline, the stage-by-stage counts, and the precise treatment of unmatched users, missing snapshots, and coin-level restrictions.

Price model inputs. The multivariate GBM serves two purposes in the empirical exercise. First, it generates the finite set of price-drop scenarios $\{\Delta p^{(s)}\}_{s=1}^{S}$, with $\Delta p^{(s)} \triangleq p_\tau - p_T^{(s)}$, used to evaluate the sampled exchange loss. Second, the same calibration supplies the one-factor direction described previously in Section 3.5. We estimate the model from hourly log-returns via EWMA-standardized covariance estimation; see Appendix C.3 for the full calibration. The key object is the conditional covariance of dollar price drops at the ADL trigger time τ,

$$\Sigma_{\tau,\Delta}^{\Delta p} \approx \mathrm{diag}(p_\tau)\,\big(\Delta\ \mathrm{diag}(\sigma_\tau)\,R_\tau\ \mathrm{diag}(\sigma_\tau)\big)\ \mathrm{diag}(p_\tau),$$

where Δ is the liquidation horizon in years, $\sigma_\tau \in \mathbb{R}_+^d$ is the vector of annualized conditional volatilities at time τ, and R_τ is the corresponding correlation matrix. Appendix C.3 explains how these objects are obtained. Let (λ_1, u_1) denote the leading eigenpair of $\Sigma_{\tau,\Delta}^{\Delta p}$, and define $v \triangleq \sqrt{\lambda_1}\,u_1$. This vector is the one-factor direction used by Water-filling in the exchange-scale implementation below.

4.2. Exchange-Scale Implementation

The Water-filling benchmark is implemented as sequential one-factor Water-filling on the active asset set. The factor direction $v \triangleq \sqrt{\lambda_1}\,u_1$ is fixed from the covariance matrix in Section 4.1, while active coins are ordered by realized ADL notional. At each step only the current coin is delevered, and account priorities are recomputed from the current one-factor leverage before moving to the next coin. Thus the benchmark preserves the one-factor prioritization logic of Section 3.4 while enforcing the multi-asset clearing constraint sequentially. To show that this one-factor sequential Water-filling is superior to the isolated-margin procedure that ranks accounts by leverage in the current coin only, we report robustness checks against the directional benchmark in Appendix C.5.

The Risk-optimal allocation instantiates the ADMM scheme of Section 3.3 after three exact reformulations: restriction to the active subset, reparameterization into dollar units, and decomposition into support-aware accountwise proximal updates. Operationally, the accountwise step is then executed in parallel, with singleton accounts handled in closed form and multi-asset support groups solved by cached warm-started Gurobi models to high accuracy. The practical calibration details are collected in Appendix C.4.

4.3. Empirical Evaluation

Loss and performance metric. We evaluate the expected loss of each allocation using the finite-scenario sample-average approximation

$$L_S(x) \triangleq \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n}\left(-E_i - (q_i - x_i)^\top \Delta p^{(s)}\right)_+, \tag{31}$$

computed over S independent GBM price-drop draws $\{\Delta p^{(s)}\}$, where $\Delta p^{(s)} \triangleq p_\tau - p_T^{(s)}$. Appendix C.5 reports the corresponding Monte Carlo robustness check for this finite-scenario evaluation. To avoid over-emphasizing a favorable single draw, the main text reports averages over 50 independent $S = 128$ scenario sets. We compare three allocations: the realized exchange allocation $x^{\text{Hyperliquid}}$, the Water-filling benchmark $x^{\text{Water-filling}}$, and the jointly calibrated Risk-optimal allocation $x^{\text{Risk-optimal}}$. To measure performance relative to the maximal amount of active-set risk that can be removed, we define the relative risk-reduction statistic

$$\mathrm{RR}(x) \triangleq \frac{L_S(0_{\mathrm{supp}(Q)}) - L_S(x)}{L_S(0_{\mathrm{supp}(Q)}) - L_S(x^{\text{Risk-optimal}})}, \tag{32}$$

where $\mathrm{supp}(Q) \triangleq \{k : Q^k \neq 0\}$ denotes the active asset set and $0_{\mathrm{supp}(Q)}$ is the zero allocation restricted to these coordinates. By construction, $\mathrm{RR}(x^{\text{Risk-optimal}}) = 1$. The realized Hyperliquid allocation and Water-filling take values strictly below 1 because they remove smaller fractions of the maximal reducible active-set risk. Thus values closer to 1 indicate better performance.

Allocation	**Mean L_{128}**	**s.d. of L_{128}**	**Mean RR**	**Runtime**
Hyperliquid	5.383×10^8	1.62×10^7	0.325	–
Water-filling	5.322×10^8	1.54×10^7	0.901	3.70 s
Risk-optimal	5.312×10^8	1.56×10^7	1.000	1.15×10^5 s

Table 3: Comparison of out-of-sample sampled-loss performance and computational efficiency for the first ADL wave τ_1. The first column lists the candidate allocations $x^{\text{Hyperliquid}}$, $x^{\text{Water-filling}}$, and $x^{\text{Risk-optimal}}$. Columns two to four evaluate these fixed allocations over 50 independent $S = 128$ GBM/SAA scenario sets and report the mean and standard deviation of sampled expected loss together with the mean risk reduction defined in Eq. (32). The last column reports the wall-clock time required to derive each allocation from the target imbalance vector and reconstructed account states.

Expected loss. To construct the Risk-optimal allocation, we solve the SAA problem once on a fixed reference scenario set of size $S = 128$ for τ_1 and use the resulting optimizer as the candidate policy. Appendix C.5 plots the associated fixed-sample convergence path. Along this run, the relative risk reduction already reaches 0.952 by iteration 500 and 0.966 by iteration 1,000, so most economically meaningful improvement arrives early; the long reported runtime mainly reflects the cost of driving the ADMM residuals to convergence.

Table 3 should therefore be read in two parts. The first four columns report an out-of-sample Monte Carlo evaluation over 50 independent $S = 128$ GBM/SAA scenario sets. Hyperliquid is observed directly, and Water-filling is computed directly from the target ADL notional and reconstructed account states, so neither allocation depends on these evaluation samples. By contrast, the Risk-optimal allocation is optimized on that fixed reference scenario set and then, like the other two allocations, evaluated on each of the 50 independent samples. The last column instead reports the wall-clock time required to derive the allocations.

Across these 50 out-of-sample evaluations, Hyperliquid attains $L_{128} = 5.383 \times 10^8$ and $\mathrm{RR} = 0.325$, Water-filling improves to $L_{128} = 5.322 \times 10^8$ and $\mathrm{RR} = 0.901$, and the Risk-optimal allocation improves further to $L_{128} = 5.312 \times 10^8$ with $\mathrm{RR} = 1.000$. The corresponding standard deviations of the loss are all of order 1.5×10^7, indicating that the Monte Carlo dispersion is of similar scale across the three candidate allocations. More importantly, Appendix C.5 shows that the ordering $L_S(x^{\text{Risk-optimal}}) < L_S(x^{\text{Water-filling}}) < L_S(x^{\text{Hyperliquid}})$ is stable across the evaluation samples, which is reassuring given the visible finite-sample variability. The table is therefore best read as a loss–runtime tradeoff: Water-filling is the fast constructive benchmark, while the Risk-optimal allocation closes the remaining gap at substantially greater computational cost.

Appendix C.5 gives the fuller robustness discussion, including the fixed-sample comparison on the reference scenario set, the corresponding τ_3 exercise, and the comparison with the directional Water-filling benchmark, which remains dominated by one-factor Water-filling. For τ_3, the reported Risk-optimal allocation is the terminal iterate at the iteration cap 20,000; the corresponding terminal residuals are reported in Appendix C.5.

Allocations. The improvement delivered by the Risk-optimal allocation is associated with structural differences in user-level targeting. The model-based allocations do not merely touch a different number of users; they also redistribute deleveraging notional across users in a systematically different way from the realized Hyperliquid event. At τ_1, one-factor Water-filling is the most concentrated rule in the upper tail, while the Risk-optimal allocation redistributes part of this mass across a broader set of economically material users. The next paragraphs make this distinction precise using the four panels of Figure 8, while Appendix C.6 reports the corresponding thresholded user-count comparisons for τ_1 and τ_3.

Before computing the diagnostics below, we apply a cleaning threshold to the model-based allocations: entries below one cent are discarded for this user-level analysis. Their total allocated volume is below \$50 for both Water-filling and the Risk-optimal allocation, so they have no economic relevance here. The realized Hyperliquid allocation $x^{\text{Hyperliquid}}$ is already aligned with this threshold; the corresponding user-count comparisons are reported in Appendix C.6.

Figure 8 collects four complementary user-level diagnostics for τ_1. The first two panels are based on the per-user deleveraging notional $\sum_{k=1}^{d} p_\tau^k \, |x_i^k|$. The upper-left and upper-right panels show distinct concentration patterns for each allocation. In the upper-left panel, the one-factor Water-filling curve decays slowest and therefore carries the heaviest upper tail. At τ_1, its largest single-user deleveraging notional is $\$6.2 \times 10^7$, compared with $\$3.1 \times 10^7$ for the Risk-optimal allocation and $\$1.6 \times 10^7$ for the realized Hyperliquid event. This top user alone accounts for 40.2% of total Water-filling deleveraging, compared with 20.1% for the Risk-optimal allocation and 10.6% for Hyperliquid. The upper-right panel confirms the same pattern cumulatively: by the top ten users, Water-filling already accounts for 76.9% of total deleveraging, compared with 57.3% for the Risk-optimal allocation and 40.9% for the realized Hyperliquid event. The Risk-optimal allocation remains concentrated, but redistributes part of this mass across a broader set of economically material accounts than Water-filling.

The lower-left panel plots the per-user deleveraging fraction

$$\mathrm{frac}_i(x) \triangleq \frac{\sum_{k=1}^{d} p_\tau^k \, \min\{|x_i^k|, c_i^k(Q)\}}{\sum_{k=1}^{d} p_\tau^k \, c_i^k(Q)}, \qquad c_i^k(Q) \triangleq u_i^k - l_i^k,$$

where $c_i^k(Q)$ is the directional capacity of account i in asset k, so that $\mathrm{frac}_i(x) \in [0, 1]$ measures the fraction of the user's deleverageable portfolio that is removed. The realized Hyperliquid event places substantially more mass near $\mathrm{frac}_i(x) = 1$, fully deleveraging a large fraction of affected accounts. By contrast, both model-based allocations usually remove only a moderate fraction of directional capacity.

The lower-right panel reinforces this picture. Restricting attention to accounts with pre-ADL equity exceeding one dollar, we order users by pre-ADL factor leverage $\ell_i^{(v)}(0)$ as defined in Definition 3.7 and plot the mean deleveraging fraction by decile. The Water-filling allocation is the lowest curve of the three and increases more smoothly with factor leverage. Risk-optimal deleverages larger fractions on average and rises more steeply, while the realized Hyperliquid event is substantially less ordered, assigning large deleveraging fractions across a broad range of factor-leverage levels.

Appendix C.6 reports the corresponding τ_3 diagnostics together with the thresholded user-count comparison and the analogous concentration patterns in the upper-left and upper-right panels. At τ_3, the lower-left panel remains qualitatively similar, and the Water-filling and Risk-optimal decile profiles in the lower-right panel remain close to their τ_1 counterparts, but the realized Hyperliquid profile is visibly more ordered. At the same time, the upper-left and upper-right panels show that the relative concentration ranking of Hyperliquid and Water-filling reverses in the upper tail, so the realized allocation is less far from the model-based rules when the active set is restricted to only four coins. Even in this case, however, the main distortion remains the same: Hyperliquid still places substantially more mass near full deleveraging of affected accounts.

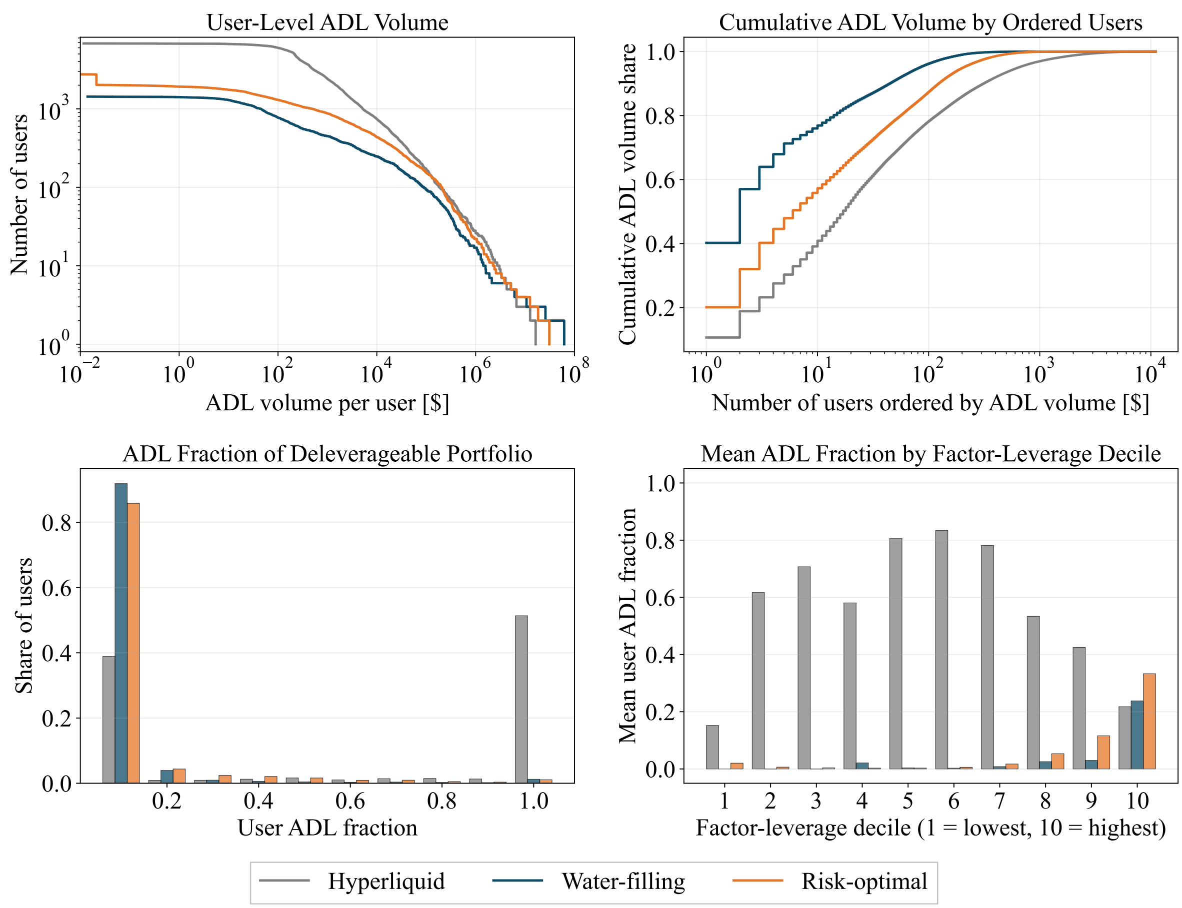

Figure 8: User-level diagnostics for τ_1: user count above deleveraging-notional threshold, cumulative top-share concentration curve, histogram of deleveraging fractions, and factor-leverage-decile targeting plot.

Code and data availability. Code and cleaned data supporting the findings are available in the accompanying GitHub repository at https://github.com/nataschahey/adl-paper-repro.

References

C. Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, 2002.

A. Armakola and J.-P. Laurent. CCP resilience and clearing membership. Technical report, 2015.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

Binance. What is auto-deleveraging (ADL) and how does it work? https://www.binance.com/en/support/faq/detail/360033525471, Sept. 2019. Accessed: 2026-02-09.

BitMEX. Auto deleveraging. https://www.bitmex.com/app/autoDeleveraging, 2017. Accessed: 2026-02-09.

S. P. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

T. Chitra. Autodeleveraging: Impossibilities and optimization. Preprint, arXiv:2512.01112, 2025.

R. Cont. The end of the waterfall: Default resources of central counterparties. *Journal of Risk Management in Financial Institutions*, 8(4):365–389, 2015.

D. Domanski, L. Gambacorta, and C. Picillo. Central clearing: Trends and current issues. Technical report, Bank for International Settlements, 2015.

D. Duffie. Resolution of failing central counterparties. *Chapter in* Making Failure Feasible: How Bankruptcy Reform Can End "Too Big to Fail", 2015.

D. Duffie and H. Zhu. Does a central clearing counterparty reduce counterparty risk? *Review of Asset Pricing Studies*, 1(1):74–95, 2011.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, July 1982.

H. Hanbali and D. Linders. Monotone tail functions: Definitions, properties, and application to risk-reducing strategies. *Journal of Computational and Applied Mathematics*, 416:114484, 2022.

J. Heilbron. Lessons from trade cancellation at the LME in March 2022. Technical Report OFR Brief 25-01, Office of Financial Research, 2025.

W. Huang. Central counterparty capitalization and misaligned incentives. Technical Report 767, Bank for International Settlements, February 2019.

W. Huang, A. J. Menkveld, and S. Yu. Central counterparty exposure in stressed markets. *Management Science*, 67(6):3596–3617, 2021.

Hyperliquid. Auto-deleveraging. https://hyperliquid.gitbook.io/hyperliquid-docs/trading/auto-deleveraging, 2024. Accessed: 2026-02-09.

International Swaps and Derivatives Association. CCP loss allocation at the end of the waterfall. Technical report, International Swaps and Derivatives Association, Aug. 2013.

R. Jia, X. Ma, C. C. Moallemi, and S. Wang. Thread on Hyperliquid's liquidation pipeline, backstop vault, and ADL outcomes. https://x.com/RuizheJia/status/2020906838653604056, Feb. 2026. X thread. Accessed 2026-03-12.

H. Lang. Crypto market volatility triggers $2.5 billion in bitcoin liquidations. https://finance.yahoo.com/news/crypto-market-volatility-triggers-2-202412904, Feb. 2026. Yahoo Finance (Reuters). Accessed 2026-02-09.

A. J. Menkveld. Crowded trades: An overlooked systemic risk for central clearing counterparties. *The Review of Asset Pricing Studies*, 7(2):209–242, 2017.

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3): 21–41, 2000.

A. Shapiro. On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, 38(1):142–152, 2013.

A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

D. Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533, 2002.

W. Thomson. Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: A survey. *Mathematical Social Sciences*, 45(3):249–297, 2003. doi: 10.1016/S0165-4896(02)00070-7.

W. Thomson. Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: An update. *Mathematical Social Sciences*, 74:41–59, 2015. doi: 10.1016/j.mathsocsci.2014.06.002.

A. Details and Proofs for Single-Asset Isolated Margining

We begin this supplementary appendix with a lemma that will be used repeatedly to prove properties of the ADL objectives.

Lemma A.1. *For each account i,*

(i) the shortfall $\sigma_i(x_i, p)$ is convex and piecewise affine in each argument, and

(ii) for every $x_i \in [0, q_i]$ the map $p \mapsto \sigma_i(x_i, p)$ is nondecreasing.

Consequently, for each p the loss $x \mapsto \mathcal{L}(x, p) = \sum_{i=1}^n \sigma_i(x_i, p)$ is convex and piecewise affine, and for each x the map $p \mapsto \mathcal{L}(x, p)$ is convex, piecewise affine, and nondecreasing.

Proof. Fix i. For $x_i \in [0, q_i]$ and $p \in \mathbb{R}$ we can write

$$-e_i(x_i, p) = (q_i - x_i)p + x_i p_\tau - q_i p_i^{(e)} - m_i,$$

which is affine in x_i for fixed p, and affine in p for fixed x_i, with slope $q_i - x_i \geq 0$ in p. Since $u \mapsto (u)_+$ is convex and nondecreasing, the composition $\sigma_i(x_i, p) = (-e_i(x_i, p))_+$ is convex and piecewise affine in x_i and in p, and is nondecreasing in p. The claims for $\mathcal{L}$ follow by summation. ■

A.1. Proof of Theorem 2.4

The proof of Theorem 2.4 will be tackled in parts. First, we will show that if x solves (6) then (c) holds (i.e., (a) $\Rightarrow$ (c)). Next, we will characterize the unique solution to (b). We will complete the equivalence of (a)–(c) by showing that any admissible $x \in \mathcal{X}$ satisfying (c) is given by the solution to (b) *and* the solution to (b) solves (a) (i.e., (c) $\Rightarrow$ (b) $\Rightarrow$ (a)). Finally, we incorporate the pointwise statement (d) by proving that the water-filling allocation is a simultaneous pointwise minimizer of realized loss (which gives (b) $\Rightarrow$ (d)) and by observing that (d) $\Rightarrow$ (a) follows from evaluating at $p = p_T$ and taking expectations.

In particular, the statement of the theorem follows from Propositions A.3 and A.6, Lemma A.7, and Proposition A.8 below.

A.1.1. Leverage Equalization

We begin with some preliminaries. Under Assumption 2.1 and the standing condition $Q > 0$, Slater's condition is trivially satisfied for the optimization problem (6) since the point

$$\bar{x}_i \triangleq \frac{Q}{\sum_{j=1}^n q_j} q_i, \qquad i = 1, \ldots, n$$

satisfies $0 < \bar{x}_i < q_i$ for all i and $\sum_i \bar{x}_i = Q$. Hence strong duality holds and the KKT conditions are necessary and sufficient for optimality.

Introduce multipliers $\lambda \in \mathbb{R}$ for $\sum_i x_i = Q$, $\nu_i \geq 0$ for $-x_i \leq 0$, and $\mu_i \geq 0$ for $x_i - q_i \leq 0$. An optimal primal–dual quadruple $(x^\star, \lambda^\star, \nu^\star, \mu^\star)$ satisfies

$$0 \in \partial_{x_i} V(x^\star) + \lambda^\star - \nu_i^\star + \mu_i^\star, \quad i = 1, \ldots, n, \tag{33}$$

$$x^\star \in \mathcal{X}, \tag{34}$$

$$\nu_i^\star \geq 0, \;\; \mu_i^\star \geq 0, \quad i = 1, \ldots, n, \tag{35}$$

$$\nu_i^\star x_i^\star = 0, \quad \mu_i^\star (x_i^\star - q_i) = 0, \quad i = 1, \ldots, n. \tag{36}$$

For each fixed p, we have $\frac{\partial}{\partial x_i}(-e_i(x_i,p)) = p_\tau - p$. Therefore, the subdifferential of the individual account shortfall satisfies

$$\partial_{x_i}\sigma_i(x_i,p) = \partial_{x_i}\big((-e_i(x_i,p))_+\big) = (p_\tau - p)\cdot \begin{cases} \{1\}, & e_i(x_i,p) < 0, \\ [0,1], & e_i(x_i,p) = 0, \\ \{0\}, & e_i(x_i,p) > 0. \end{cases}$$

Since $e_i(x_i,\cdot)$ is affine in p_T with nonzero slope for $x_i < q_i$, Assumption 2.3 implies that attaining *exactly* zero equity is a probability zero event, $\mathsf{P}(e_i(x_i,p_T) = 0) = 0$. As a result, we can obtain the pointwise partial derivative for the objective

$$\frac{\partial V}{\partial x_i}(x) = \mathsf{E}\Big[(p_\tau - p_T)\,\mathbb{I}_{\{e_i(x_i,p_T)\le 0\}}\Big], \quad x_i \in [0,q_i). \tag{37}$$

This is justified by Assumption 2.3, which implies that $\sigma_i(x_i,p_T)$ is differentiable almost surely for $x_i \in [0,q_i)$ and allows us to pass the x_i-derivative inside the expectation by dominated convergence.

Observe that the insolvency event $e_i(x_i,p_T) \le 0$ for each account admits a threshold representation in terms of p_T and the account's *bankruptcy (zero-equity) price* $p_i^{(z)}(x_i)$:

$$e_i(x_i,p_T) \le 0 \quad\Longleftrightarrow\quad p_T \ge p_i^{(z)}(x_i), \qquad p_i^{(z)}(x_i) \triangleq p_\tau + \frac{E_i}{q_i - x_i}, \tag{38}$$

with the convention $p_i^{(z)}(q_i) = +\infty$. Using (4), this can be rewritten as

$$p_i^{(z)}(x_i) = p_\tau\Big(1 + \ell_i(x_i)^{-1}\Big), \tag{39}$$

with the convention $\ell^{-1} = +\infty$ when $\ell = 0$. Combining (37) and (38) gives

$$\frac{\partial V}{\partial x_i}(x) = \mathsf{E}\left[(p_\tau - p_T)\,\mathbb{I}_{\left\{p_T \ge p_i^{(z)}(x_i)\right\}}\right]. \tag{40}$$

On $\{p_T \ge p_i^{(z)}(x_i)\}$ we have $p_T \ge p_i^{(z)}(x_i) \ge p_\tau$, hence $(p_\tau - p_T) \le 0$ and $\partial V/\partial x_i(x) \le 0$.

Using (39), the event $\{p_T \ge p_i^{(z)}(x_i)\}$ is equivalent to $\{p_T - p_\tau \ge p_\tau \ell_i(x_i)^{-1}\}$. Define next, for $\ell \in [0,\infty)$, the marginal expected shortfall exposure, $v(\ell)$, of an account with leverage ℓ:

$$v(\ell) \triangleq \mathsf{E}\Big[(p_T - p_\tau)\,\mathbb{I}_{\{\,p_T - p_\tau \ge p_\tau \ell^{-1}\}}\Big], \qquad v(0) = 0. \tag{41}$$

Equivalently, $v(\ell)$ is the "tail expectation" of $(p_T - p_\tau)$ above a leverage-indexed threshold. For $x_i \in [0,q_i)$, we arrive at the derivative representation

$$\frac{\partial V}{\partial x_i}(x) = -\,v(\ell_i(x_i)), \quad x_i \in [0,q_i), \tag{42}$$

which admits a continuous extension to $x_i = q_i$ with value 0.

Indeed, as $x_i \uparrow q_i$ we have $\ell_i(x_i) \downarrow 0$, and by definition $v(0) = 0$. Moreover, for $\ell > 0$ we have

$$v(\ell) = \mathsf{E}\Big[(p_T - p_\tau)\,\mathbb{I}_{\{\,p_T - p_\tau \ge p_\tau \ell^{-1}\}}\Big].$$

As $\ell \downarrow 0$, the threshold $p_\tau \ell^{-1} \to +\infty$, so the indicator decreases pointwise to zero and is dominated by 1, while

$$0 \le (p_T - p_\tau)\,\mathbb{I}_{\{\,p_T - p_\tau \ge p_\tau \ell^{-1}\}} \le (p_T - p_\tau)_+,$$

with $(p_T - p_\tau)_+$ integrable under our standing assumptions. By dominated convergence,

$$\lim_{\ell \downarrow 0} v(\ell) = 0 = v(0).$$

Therefore,

$$\lim_{x_i \uparrow q_i} \frac{\partial V}{\partial x_i}(x) = -\lim_{\ell \downarrow 0} v(\ell) = -v(0) = 0. \tag{43}$$

Thus the marginal effect of additional buyback vanishes as the account is fully closed. Since $V(\cdot)$ is convex in x_i, this implies that any element of the subgradient set $g_i \in \partial_{x_i} V(x)$ when $x_i = q_i$ satisfies $g_i \geq 0$.

Since $\ell_i(\cdot)$ is strictly decreasing in x_i and $v(\cdot)$ is nondecreasing in ℓ, the map $x_i \mapsto \frac{\partial V}{\partial x_i}(x)$ is nondecreasing. Under our assumptions it is straightforward to check that v is strictly increasing on $[0, \max_i \ell_i(0)]$.

Lemma A.2. *$v(\cdot)$ is strictly increasing on $[0, \max_i \ell_i(0)]$.*

Proof. A simple sufficient condition for this monotonicity is given by Assumption 2.3. Namely, that p_T admits a density f_T such that $f_T(p) > 0$ for all $p \geq p_\tau$. Indeed, let $a(\ell) \triangleq p_\tau(1 + \ell^{-1})$ (with $a(0) = +\infty$). For any $0 \leq \ell_1 < \ell_2 \leq \max_i \ell_i(0)$,

$$v(\ell_2) - v(\ell_1) = \mathsf{E}\Big[(p_T - p_\tau)\, \mathbb{I}_{\{a(\ell_2) \leq p_T < a(\ell_1)\}}\Big] = \int_{a(\ell_2)}^{a(\ell_1)} (p - p_\tau)\, f_T(p)\, dp > 0.$$

The claim follows. ■

The following proposition shows that (a) implies (c).

Proposition A.3 (Equalization of post-ADL leverage)**.** *Under Assumptions 2.1–2.3, if $x^\star$ solves (6) then $x_i^\star < q_i$ for all i and there exists a cutoff level $\bar{\ell} \in [0, \max_i \ell_i(0)]$ such that*

$$\begin{aligned} x_i^\star > 0 \quad &\Longrightarrow \quad \ell_i(x_i^\star) = \bar{\ell}, \\ x_i^\star = 0 \quad &\Longrightarrow \quad \ell_i(0) \leq \bar{\ell}. \end{aligned}$$

In particular, all accounts whose positions are reduced share the same post-ADL leverage.

Proof. We verify the claims of Proposition A.3 in turn.

First, we show that $x_i^\star < q_i$ for all i. Suppose for contradiction that $x_k^\star = q_k$ for some k. Then $\nu_k^\star = 0$ (since $x_k^\star > 0$) and $\ell_k(x_k^\star) = 0$, so using (33) we get

$$0 \in \partial_{x_k} V(x^\star) + \lambda^\star + \mu_k^\star$$

Hence there exists $g_k \in \partial_{x_k} V(x^\star)$ such that

$$0 = g_k + \lambda^\star + \mu_k^\star.$$

As noted earlier (by (43)) we have $g_k \geq 0$, so $\lambda^\star + \mu_k^\star = -g_k \leq 0$, and therefore $\lambda^\star \leq -\mu_k^\star \leq 0$. On the other hand, since $Q < \sum_i q_i$, there exists an index j with $x_j^\star < q_j$, so $\mu_j^\star = 0$. If $x_j^\star > 0$, then $\nu_j^\star = 0$ and (33) gives $\lambda^\star = v(\ell_j(x_j^\star)) > 0$, because $\ell_j(x_j^\star) > 0$ and v is strictly increasing with $v(0) = 0$. If instead $x_j^\star = 0$, then (33) gives $\lambda^\star = v(\ell_j(0)) + \nu_j^\star \geq v(\ell_j(0)) > 0$. In either case $\lambda^\star > 0$, contradicting $\lambda^\star \leq 0$. Therefore $x_i^\star < q_i$ for all i.

Since $Q > 0$ and $\sum_i x_i^\star = Q$, there exists at least one index h such that $x_h^\star > 0$. Combined with the previous step, this yields $0 < x_h^\star < q_h$. By (33)–(36) and (42),

$$-v(\ell_h(x_h^\star)) + \lambda^\star = 0,$$

so $\lambda^\star = v(\ell_h(x_h^\star))$. By Lemma A.2, v is invertible on $[0, \max_i \ell_i(0)]$, and we may therefore define

$$\bar{\ell} \triangleq v^{-1}(\lambda^\star) \in [0, \max_i \ell_i(0)].$$

Now let i be such that $0 < x_i^\star < q_i$. Then $\nu_i^\star = \mu_i^\star = 0$, and (33) together with (42) gives

$$-v(\ell_i(x_i^\star)) + \lambda^\star = 0.$$

Hence $v(\ell_i(x_i^\star)) = \lambda^\star = v(\bar{\ell})$, and strict monotonicity of v implies

$$\ell_i(x_i^\star) = \bar{\ell}.$$

Finally, if $x_i^\star = 0$, then $\mu_i^\star = 0$ and (33) gives

$$-v(\ell_i(0)) + \lambda^\star - \nu_i^\star = 0,$$

hence $v(\ell_i(0)) \leq \lambda^\star$ because $\nu_i^\star \geq 0$. Since $\lambda^\star = v(\bar{\ell})$ and v is strictly increasing, we conclude that

$$\ell_i(0) \leq \bar{\ell}.$$

■

A.1.2. The Minimax Problem

We study the minimax leverage program (7). Fix a candidate leverage cap $t \geq 0$ and ask what minimum buyback from each account is required to ensure $\ell_i(x_i) \leq t$. In view of expression (4) the constraint $\ell_i(x_i) \leq t$ is equivalent to

$$x_i \geq q_i - \frac{E_i}{p_\tau} t, \tag{44}$$

together with $x_i \geq 0$. This motivates the definition of the vector function $y(t) \triangleq (y_1(t), \ldots, y_n(t))$ where

$$y_i(t) \;\triangleq\; \left(q_i - \frac{E_i}{p_\tau}\, t\right)_+, \qquad i = 1, \ldots, n,$$

which is the minimal buyback required from account i under the leverage cap t. In particular, setting $x_i = y_i(t)$ yields

$$\ell_i(y_i(t)) = \min\{\ell_i(0),\, t\}, \qquad i = 1, \ldots, n.$$

Lemma A.4. *For any $x \in \mathcal{X}$,*

$$\max_{1 \leq i \leq n} \ell_i(x_i) \leq t \quad \Longrightarrow \quad x_i \geq y_i(t) \;\; \forall i \;\; \textit{and} \;\; Q \;\geq\; \sum_{i=1}^{n} y_i(t).$$

Proof. Fix $t \geq 0$ and $x \in \mathcal{X}$ with $\max_i \ell_i(x_i) \leq t$. Then $\ell_i(x_i) \leq t$ for each i, and hence (44) holds componentwise. Combining (44) with $x_i \geq 0$ yields $x_i \geq (q_i - \frac{E_i}{p_\tau} t)_+ = y_i(t)$ for all i. Summing over i and using $\sum_i x_i = Q$ gives $Q \geq \sum_i y_i(t) = G(t)$. ■

Summing these per-account requirements yields the total buyback needed to enforce the cap t across the system. Accordingly, define

$$G(t) \triangleq \sum_{i=1}^{n} y_i(t), \quad t \geq 0,$$

which we interpret as the aggregate deleveraging demand induced by the leverage level t. This function has several key properties and is critical to the solution of the minimax problem.

Lemma A.5.

(i) G is continuous and nonincreasing on $[0, \infty)$.

(ii) $G(0) = \sum_{i=1}^{n} q_i$, and $G(t) = 0$ for all $t \geq \max_{1 \leq i \leq n} \ell_i(0)$.

(iii) G is strictly decreasing on $\big[0, \max_i \ell_i(0)\big]$.

(iv) There exists a unique $t^\star \in \big(0, \max_i \ell_i(0)\big)$ such that $G(t^\star) = Q$.

Proof. We treat each statement in turn.

(i) is immediate.

(ii) At $t = 0$, $y_i(0) = q_i$, so $G(0) = \sum_i q_i$. For $t \geq \max_i \ell_i(0)$ we have, for each i,

$$t \geq \ell_i(0) = \frac{p_\tau q_i}{E_i} \quad \iff \quad q_i - \frac{E_i}{p_\tau} t \leq 0 \quad \iff \quad y_i(t) = 0,$$

hence $G(t) = 0$.

(iii) Fix $0 \leq t_1 < t_2 \leq \max_i \ell_i(0)$ and choose an index k such that $\ell_k(0) = \max_i \ell_i(0)$. Then, for any $t \in [0, \ell_k(0)]$ we have $t \leq \ell_k(0) = \frac{p_\tau q_k}{E_k}$. Equivalently, $q_k - \frac{E_k}{p_\tau} t \geq 0$ and thus

$$y_k(t) = q_k - \frac{E_k}{p_\tau} t \qquad \text{for all } t \in [0, \ell_k(0)].$$

In particular, $y_k(t_2) < y_k(t_1)$. On the other hand, for every j, $y_j(\cdot)$ is nonincreasing, so $y_j(t_2) \leq y_j(t_1)$. Summing across indices yields $G(t_2) < G(t_1)$. Since $t_1 < t_2$ were arbitrary in $\big[0, \max_i \ell_i(0)\big]$ the strict decrease follows.

(iv) By (ii) and the standing assumption $0 < Q < \sum_i q_i$, we have $G(0) > Q$ and $G(\max_i \ell_i(0)) = 0 < Q$. By continuity from (i), the intermediate value theorem yields existence of $t^\star \in (0, \max_i \ell_i(0))$ such that $G(t^\star) = Q$. Uniqueness follows from strict decrease in (iii). ■

Combining the preceding definitions and properties, we obtain the minimax solution.

Proposition A.6 (Minimax solution)**.** *Let $t^\star$ be the unique level from Lemma A.5(iv) such that $G(t^\star) = Q$, and define $x^\star \triangleq y(t^\star)$. Then $x^\star \in \mathcal{X}$, $x^\star$ is the unique optimizer of the minimax leverage problem (7) and $\max_{1 \leq i \leq n} \ell_i(x_i^\star) = t^\star$.*

Proof. We break up the proof into a verification of the component claims.

Feasibility. By definition, $y_i(t^\star) \in [0, q_i]$ for all i (since $(\cdot)_+ \geq 0$ and $q_i - \frac{E_i}{p_\tau} t^\star \leq q_i$), and by Lemma A.5(iv), $\sum_i y_i(t^\star) = G(t^\star) = Q$. Hence $x^\star \in \mathcal{X}$.

Optimality. Let $x \in \mathcal{X}$ be arbitrary and set $t \triangleq \max_i \ell_i(x_i) \geq 0$. Since each $\ell_i(\cdot)$ is decreasing on $[0, q_i]$ and $x_i \geq 0$, we have $\ell_i(x_i) \leq \ell_i(0)$ for all i, and hence $t \leq \max_i \ell_i(0)$. Then $\max_i \ell_i(x_i) \leq t$ holds trivially, so by Lemma A.4

$$Q = \sum_{i=1}^{n} x_i \ \geq \ \sum_{i=1}^{n} y_i(t) = G(t).$$

Because G is strictly decreasing on $[0, \max_i \ell_i(0)]$ and $G(t^\star) = Q$ (Lemma A.5), the inequality $Q \geq G(t)$ implies $t \geq t^\star$. Thus every feasible x satisfies $\max_i \ell_i(x_i) \geq t^\star$, while $x^\star$ achieves $\max_i \ell_i(x_i^\star) = t^\star$. Therefore $x^\star$ is optimal and the optimal value is $t^\star$.

Uniqueness. Now let $x \in \mathcal{X}$ be any optimal solution. Then $\max_i \ell_i(x_i) \leq t^\star$, which implies $x_i \geq y_i(t^\star)$ for all i. Summing and using $\sum_i x_i = \sum_i y_i(t^\star) = Q$ yields

$$0 = \sum_{i=1}^{n} (x_i - y_i(t^\star)) \qquad \text{with} \qquad x_i - y_i(t^\star) \geq 0 \ \ \forall i,$$

which forces $x_i = y_i(t^\star)$ for all i. Hence the optimizer is unique and equals $x^\star = y(t^\star)$.

Maximum Leverage. Fix i. If $x_i^\star > 0$, then $x_i^\star = q_i - \frac{E_i}{p_\tau} t^\star$, so

$$\ell_i(x_i^\star) = \frac{p_\tau(q_i - x_i^\star)}{E_i} = \frac{p_\tau \left(\frac{E_i}{p_\tau} t^\star\right)}{E_i} = t^\star.$$

If $x_i^\star = 0$, then $q_i - \frac{E_i}{p_\tau} t^\star \leq 0$, i.e. $t^\star \geq \frac{p_\tau q_i}{E_i} = \ell_i(0)$, so $\ell_i(x_i^\star) = \ell_i(0) \leq t^\star$. Therefore $\max_i \ell_i(x_i^\star) \leq t^\star$. Since $Q > 0$, at least one component of $x^\star$ is strictly positive, so for that index the leverage equals $t^\star$, and hence $\max_i \ell_i(x_i^\star) = t^\star$. ■

A.1.3. Pointwise Optimality

We next show that the water-filling allocation identified above is optimal not only after averaging over the terminal price distribution, but also pointwise at every price level. It is a trivial consequence of pointwise optimality that (d) ⇒ (a).

Lemma A.7 (Pointwise optimality of water-filling)**.** *Let Assumptions 2.1–2.2 hold, and let $x^\star = y(t^\star)$ be the water-filling allocation from Proposition A.6. Then*

$$\mathcal{L}(x^\star, p) \leq \mathcal{L}(x, p), \qquad x \in \mathcal{X}, \ \ p \in \mathbb{R}.$$

Proof. For $x \in \mathcal{X}$, set $r_i := q_i - x_i$ and introduce the leverage variable $u_i := \ell_i(x_i) = p_\tau r_i / E_i$. Since $x_i = q_i - E_i u_i / p_\tau$, the feasibility constraints $x \in \mathcal{X}$ are equivalent to

$$0 \leq u_i \leq \bar{u}_i := \frac{p_\tau q_i}{E_i}, \qquad i = 1, \ldots, n,$$

together with

$$\sum_{i=1}^{n} E_i u_i = M, \qquad M := p_\tau \left(\sum_{i=1}^{n} q_i - Q \right).$$

Thus the feasible set in leverage variables is

$$\mathcal{U} := \left\{ u \in \mathbb{R}_+^n : 0 \leq u_i \leq \bar{u}_i, \ \sum_{i=1}^{n} E_i u_i = M \right\}.$$

By Proposition A.6, the water-filling allocation $x^\star = y(t^\star)$ corresponds to the leverage vector $u_i^\star = \min\{\bar{u}_i, t^\star\}$.

Fix $p \in \mathbb{R}$. If $p \leq p_\tau$, then $e_i(x_i, p) = E_i + (q_i - x_i)(p_\tau - p) \geq E_i > 0$ for every i and every feasible x. Hence $\mathcal{L}(x, p) = 0$ for all $x \in \mathcal{X}$, and the desired inequality is immediate.

Suppose now that $p > p_\tau$. Then

$$\mathcal{L}(x, p) = \sum_{i=1}^{n} E_i \phi_p(u_i), \qquad \phi_p(u) := \left(\frac{p - p_\tau}{p_\tau} u - 1 \right)_+.$$

The function ϕ_p is finite-valued, convex, and nondecreasing on $[0, \infty)$.

We prove the following more general claim: for every finite-valued convex nondecreasing function $\phi : [0, \infty) \to \mathbb{R}$, the vector $u^\star$ minimizes $\sum_i E_i \phi(u_i)$ over $\mathcal{U}$. Choose $g^\star \in \partial\phi(t^\star)$. For each i with $\bar{u}_i > t^\star$, set $h_i := g^\star$. For each i with $\bar{u}_i \le t^\star$, choose $h_i \in \partial\phi(\bar{u}_i)$ such that $h_i \le g^\star$; this is possible by monotonicity of the subdifferential of a convex function, and if $\bar{u}_i = t^\star$ we simply take $h_i = g^\star$.

Let $u \in \mathcal{U}$. By the subgradient inequality,

$$\sum_{i=1}^{n} E_i\{\phi(u_i) - \phi(u_i^\star)\} \ge \sum_{i=1}^{n} E_i h_i (u_i - u_i^\star).$$

If $\bar{u}_i > t^\star$, then $h_i = g^\star$. If $\bar{u}_i \le t^\star$, then $u_i^\star = \bar{u}_i$ and $u_i - u_i^\star \le 0$, while $h_i \le g^\star$; hence $h_i(u_i - u_i^\star) \ge g^\star(u_i - u_i^\star)$. Therefore

$$\sum_{i=1}^{n} E_i h_i (u_i - u_i^\star) \ge g^\star \sum_{i=1}^{n} E_i (u_i - u_i^\star) = 0,$$

where the last equality uses $u, u^\star \in \mathcal{U}$. Thus $u^\star$ minimizes $\sum_i E_i \phi(u_i)$ over $\mathcal{U}$. Applying this claim to $\phi = \phi_p$ proves $\mathcal{L}(x^\star, p) \le \mathcal{L}(x, p)$ for all $x \in \mathcal{X}$ when $p > p_\tau$. Together with the case $p \le p_\tau$, this proves the result. ■

A.1.4. The ADL–Minimax–Equalization Equivalence

The preceding sections reported the unique solution to the minimax problem (b) and verified that (b) $\Rightarrow$ (d) $\Rightarrow$ (a). To complete the proof of Theorem 2.4 we present a final proposition that completes the equivalence with the relation (c) $\Rightarrow$ (b) $\Rightarrow$ (a).

Proposition A.8 (Equivalence). *Under Assumptions 2.1–2.3, for any $x \in \mathcal{X}$, (a), (b), (c) and (d) are equivalent.*

Proof. Since we have already shown (b) $\Rightarrow$ (d) $\Rightarrow$ (a), it suffices to show the equivalence of (a)–(c).

By Lemma A.5, there exists a unique $t^\star \in (0, \max_i \ell_i(0))$ such that $G(t^\star) = Q$. By Proposition A.6, the minimax leverage problem (7) has the unique optimizer $y(t^\star)$.

(a) $\Rightarrow$ (c). This follows directly from Proposition A.3.

(c) $\Rightarrow$ (b). If $x_i > 0$, then $\ell_i(x_i) = t$, and since $\ell_i(x) = \frac{p_\tau(q_i - x)}{E_i}$ this implies $x_i = q_i - \frac{E_i}{p_\tau} t > 0$, hence $x_i = \left(q_i - \frac{E_i}{p_\tau} t\right)_+$. If instead $x_i = 0$, then $\ell_i(0) \le t$ implies $\frac{p_\tau q_i}{E_i} \le t$, i.e. $q_i - \frac{E_i}{p_\tau} t \le 0$, hence again $x_i = \left(q_i - \frac{E_i}{p_\tau} t\right)_+$. Thus $x = y(t)$.

Using feasibility $\sum_i x_i = Q$, we have

$$Q = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \left(q_i - \frac{E_i}{p_\tau} t\right)_+ = G(t).$$

Since the equation $G(t) = Q$ has the unique solution $t^\star$, it follows that $t = t^\star$ and therefore $x = y(t^\star) = x^\star$.

(b) $\Rightarrow$ (a). A solution to (a) trivially exists by the Weierstrass Extreme Value Theorem since $\mathcal{X}$ is compact and the objective is continuous. Since any solution to (a) satisfies (c), and any feasible allocation satisfying (c) must coincide with the unique minimax optimizer from (b), every solution to (a) equals the solution to (b).

This proves the equivalence and uniqueness. ■

A.2. Proofs of Proposition 2.5 and Theorem 2.9

Proof of Proposition 2.5. Since $p > 0$ and $E_i(p), E_j(p) > 0$, the inequality $\ell_i(p) \geq \ell_j(p)$ is equivalent to $q_i\, E_j(p) - q_j\, E_i(p) \geq 0$. Expanding,

$$\begin{aligned} q_i\, E_j(p) - q_j\, E_i(p) &= q_i\big[q_j(p_j^{(e)} - p) + m_j\big] - q_j\big[q_i(p_i^{(e)} - p) + m_i\big] \\ &= q_i q_j\,(p_j^{(e)} - p_i^{(e)}) + q_i\, m_j - q_j\, m_i. \end{aligned}$$

All terms involving p cancel, so the sign of this expression is independent of p. ∎

Proof of Theorem 2.9. For any fixed $t \geq 0$, define the attacker's per-account lower bound

$$y_k(t) = \Big(q_k - \frac{E_k}{p_\tau} t\Big)_+, \qquad y_A(t) = \Big(q^A - \frac{E^A}{p_\tau} t\Big)_+.$$

Since $(\cdot)_+$ is subadditive, we have

$$\sum_{k=1}^{K} y_k(t) = \sum_{k=1}^{K} \Big(q_k - \frac{E_k}{p_\tau} t\Big)_+ \;\geq\; \Big(\sum_{k=1}^{K} q_k - \frac{\sum_{k=1}^{K} E_k}{p_\tau} t\Big)_+ = \Big(q^A - \frac{E^A}{p_\tau} t\Big)_+ = y_A(t). \tag{45}$$

Let

$$G_{\text{oth}}(t) \triangleq \sum_{j \in \mathcal{N}} y_j(t), \qquad G_{\text{att}}^{K}(t) \triangleq \sum_{k=1}^{K} y_k(t), \qquad G_{\text{att}}^{1}(t) \triangleq y_A(t),$$

and define total demand functions

$$G_{\text{tot}}^{K}(t) \triangleq G_{\text{oth}}(t) + G_{\text{att}}^{K}(t), \qquad G_{\text{tot}}^{1}(t) \triangleq G_{\text{oth}}(t) + G_{\text{att}}^{1}(t).$$

By (45), $G_{\text{tot}}^{K}(t) \geq G_{\text{tot}}^{1}(t)$ for all $t \geq 0$.

Let $t_K^\star$ and $t_1^\star$ be the unique solutions to $G_{\text{tot}}^{K}(t) = Q$ and $G_{\text{tot}}^{1}(t) = Q$, respectively (uniqueness follows from the strict decrease of the corresponding total water-level function on the relevant interval, as in Lemma A.5). Evaluating at $t = t_1^\star$ gives

$$G_{\text{tot}}^{K}(t_1^\star) \;\geq\; G_{\text{tot}}^{1}(t_1^\star) = Q = G_{\text{tot}}^{K}(t_K^\star).$$

Since G_{tot}^{K} is strictly decreasing, this implies $t_K^\star \geq t_1^\star$.

Because each $y_j(\cdot)$ is nonincreasing, $G_{\text{oth}}(\cdot)$ is nonincreasing, and hence $G_{\text{oth}}(t_K^\star) \leq G_{\text{oth}}(t_1^\star)$. Using market clearing in each economy,

$$\sum_{k=1}^{K} x_k^{\star,K} = G_{\text{att}}^{K}(t_K^\star) = Q - G_{\text{oth}}(t_K^\star) \;\geq\; Q - G_{\text{oth}}(t_1^\star) = G_{\text{att}}^{1}(t_1^\star) = x_A^{\star,1},$$

which proves the claim. ∎

A.3. Mathematical Formalization of Section 2.4.2 and Proof of Theorem 2.13

Recall that at the ADL time τ, account i is described by a short position size $q_i \geq 0$ and an equity level $E_i > 0$ evaluated at the execution price p_τ; cf. (1). We collect these into the state vector

$$s = (q_i, E_i)_{i=1}^{n} \in \mathcal{S} \triangleq (\mathbb{R}_+ \times \mathbb{R}_{++})^n.$$

To avoid overloading the notation $\ell_i(\cdot)$, which we reserve for leverage as a function of the *allocation* x_i, we introduce $\lambda_i(s)$ for leverage as a function of the *state*:

$$\lambda_i(s) \triangleq \frac{p_\tau q_i}{E_i}, \qquad \mathcal{M}(s) \triangleq \{i : \lambda_i(s) = \max_{1\le j\le n} \lambda_j(s)\}.$$

Thus $\mathcal{M}(s)$ is the set of accounts attaining the maximum leverage under the configuration s. Given an allocation $x \in \mathcal{X}$, we can define the corresponding post-ADL state

$$s(x) \triangleq (q_i - x_i,\ E_i)_{i=1}^n \in \mathcal{S},$$

so that the leverage after allocating x_i satisfies

$$\ell_i(x_i) = \frac{p_\tau(q_i - x_i)}{E_i} = \lambda_i(s(x)).$$

We emphasize that E_i is measured at the execution price p_τ, so buying back at p_τ only reshuffles equity between entry price and cash margin and does not change E_i.

Using this notation, an ADL mechanism can be formally described by a family of maps

$$\{F_Q : \mathcal{S} \to \mathcal{S}\}_{Q\ge 0},$$

where $F_Q(s)$ is the post-ADL state after a total buyback quantity Q has been allocated across accounts. The following formalizes the natural assumption that these mechanisms can only reduce existing positions by transacting at p_τ and must clear the quantity Q.

Assumption A.9. *For every $s = (q_i, E_i)_{i=1}^n \in \mathcal{S}$ and every $Q \in [0, \sum_{i=1}^n q_i]$, write*

$$F_Q(s) = \big(q_i - x_i(s,Q),\ E_i\big)_{i=1}^n. \tag{46}$$

Then the vector $x(s,Q) = (x_i(s,Q))_{i=1}^n$ satisfies

$$x_i(s,Q) \in [0, q_i] \ \text{ for all } i \text{ and } \ \sum_{i=1}^n x_i(s,Q) = Q. \tag{47}$$

The next two conditions are the mathematical versions of Definitions 2.11 and 2.12 in the main text.

Assumption A.10 (Path-independence)**.** *For any state $s \in \mathcal{S}$ and $Q_1, Q_2 \ge 0$ with $Q_1 + Q_2 \le \sum_i q_i$,*

$$F_{Q_1+Q_2}(s) = F_{Q_2} \circ F_{Q_1}(s).$$

Assumption A.11 (Leverage priority)**.** *Fix any state $s \in \mathcal{S}$. Then for sufficiently small $Q > 0$, one has $x_i(s,Q) = 0$ for all $i \notin \mathcal{M}(s)$.*

A.3.1. Proof of Theorem 2.13

Our goal is to show that an ADL mechanism satisfies Assumptions A.9–A.11 if and only if, for every initial state $s \in \mathcal{S}$ and every $Q \in (0, \sum_i q_i)$, it coincides with the minimax leverage policy.

We first provide a supporting lemma on necessary properties of any ADL mechanism.

Lemma A.12 (Absolute continuity)**.** *Let Assumptions A.9 and A.10 hold and fix $s \in \mathcal{S}$. For each i the trajectory $Q \mapsto x_i(s,Q)$ is nonnegative, nondecreasing, and 1-Lipschitz on $[0, \sum_i q_i]$. In particular, it is absolutely continuous in Q, and its a.e. derivative $\frac{\partial}{\partial Q} x_i(s,Q)$ satisfies*

$$0 \le \frac{\partial}{\partial Q} x_i(s,Q) \le 1 \quad \textit{for a.e. } Q, \qquad \sum_{i=1}^n \frac{\partial}{\partial Q} x_i(s,Q) = 1 \quad \textit{for a.e. } Q.$$

Proof. Let $s \in \mathcal{S}$ and $0 \le Q_1 \le Q_2 \le \sum_i q_i$. Set $\Delta Q \triangleq Q_2 - Q_1$. Path-independence gives

$$F_{Q_2}(s) = F_{\Delta Q}(F_{Q_1}(s)).$$

Writing both sides using (46) yields the identity

$$x_i(s, Q_2) = x_i(s, Q_1) + x_i(F_{Q_1}(s), \Delta Q), \qquad i = 1, \ldots, n. \tag{48}$$

By (47) applied at the intermediate state $F_{Q_1}(s)$, we have $x_i(F_{Q_1}(s), \Delta Q) \ge 0$. Moreover, since $\sum_i x_i(F_{Q_1}(s), \Delta Q) = \Delta Q$, we also have $x_i(F_{Q_1}(s), \Delta Q) \le \Delta Q$. Substituting into (48) gives

$$0 \le x_i(s, Q_2) - x_i(s, Q_1) \le \Delta Q = Q_2 - Q_1.$$

Thus $Q \mapsto x_i(s, Q)$ is nondecreasing and 1-Lipschitz. The derivative properties follow from standard facts for absolutely continuous functions and by differentiating $\sum_i x_i(s, Q) = Q$ a.e. ■

We can now show the main result.

Proof of Theorem 2.13. We prove the two implications separately.

The water-filling rule satisfies the axioms. Fix a state $s = (q_i, E_i)_{i=1}^n \in \mathcal{S}$ and define

$$G_s(t) \triangleq \sum_{i=1}^n \Big(q_i - \frac{E_i}{p_\tau} t\Big)_+, \qquad t \ge 0.$$

The function G_s is continuous and strictly decreasing on $[0, \lambda_{\max}(s)]$, where

$$\lambda_{\max}(s) \triangleq \max_{1 \le i \le n} \frac{p_\tau q_i}{E_i},$$

with

$$G_s(0) = \sum_{i=1}^n q_i, \qquad G_s(\lambda_{\max}(s)) = 0.$$

For $Q \in (0, \sum_i q_i)$, let $t_s(Q)$ be the unique threshold satisfying

$$G_s(t_s(Q)) = Q.$$

At the boundary points, set

$$t_s(0) \triangleq \lambda_{\max}(s), \qquad t_s\Big(\sum_i q_i\Big) \triangleq 0.$$

Define

$$x_i^{\mathrm{WF}}(s, Q) \triangleq \Big(q_i - \frac{E_i}{p_\tau} t_s(Q)\Big)_+$$

and

$$F_Q^{\mathrm{WF}}(s) \triangleq (q_i - x_i^{\mathrm{WF}}(s, Q),\ E_i)_{i=1}^n.$$

Market clearing holds immediately, since

$$0 \le x_i^{\mathrm{WF}}(s, Q) \le q_i$$

and

$$\sum_{i=1}^n x_i^{\mathrm{WF}}(s, Q) = G_s(t_s(Q)) = Q.$$

We next verify leverage priority. Let

$$\mathcal{M}(s) \triangleq \{i : \lambda_i(s) = \lambda_{\max}(s)\}.$$

If $\mathcal{M}(s) = \{1, \ldots, n\}$, the claim is immediate. Otherwise define

$$\lambda_{\text{sub}}(s) \triangleq \max_{i \notin \mathcal{M}(s)} \lambda_i(s) < \lambda_{\max}(s)$$

and set

$$\delta \triangleq G_s(\lambda_{\text{sub}}(s)) > 0.$$

If $0 < Q < \delta$, then

$$G_s(t_s(Q)) = Q < G_s(\lambda_{\text{sub}}(s)).$$

Since G_s is strictly decreasing,

$$t_s(Q) > \lambda_{\text{sub}}(s).$$

Hence, for every $i \notin \mathcal{M}(s)$,

$$\lambda_i(s) \le \lambda_{\text{sub}}(s) < t_s(Q),$$

and therefore

$$x_i^{\text{WF}}(s, Q) = 0.$$

Thus, for sufficiently small quantities, ADL is imposed only on maximally leveraged accounts.

Finally, we verify path independence. Fix $Q_1, Q_2 \ge 0$ such that

$$Q_1 + Q_2 \le \sum_{i=1}^{n} q_i,$$

and write

$$t_1 \triangleq t_s(Q_1), \qquad t_{12} \triangleq t_s(Q_1 + Q_2).$$

Since G_s is decreasing and $Q_1 + Q_2 \ge Q_1$, we have

$$t_{12} \le t_1.$$

After the first step, the residual position of account i is

$$q_i^{(1)} = q_i - x_i^{\text{WF}}(s, Q_1) = \min\Big\{q_i, \frac{E_i}{p_\tau} t_1\Big\}.$$

Since $t_{12} \le t_1$, for each i we have

$$\Big(q_i^{(1)} - \frac{E_i}{p_\tau} t_{12}\Big)_+ = \Big(q_i - \frac{E_i}{p_\tau} t_{12}\Big)_+ - \Big(q_i - \frac{E_i}{p_\tau} t_1\Big)_+.$$

The restarted demand function therefore satisfies

$$\begin{aligned}
G_{F_{Q_1}^{\text{WF}}(s)}(t_{12}) &= \sum_{i=1}^{n} \Big(q_i^{(1)} - \frac{E_i}{p_\tau} t_{12}\Big)_+ \\
&= \sum_{i=1}^{n} \Big[\Big(q_i - \frac{E_i}{p_\tau} t_{12}\Big)_+ - \Big(q_i - \frac{E_i}{p_\tau} t_1\Big)_+\Big] \\
&= G_s(t_{12}) - G_s(t_1) \\
&= (Q_1 + Q_2) - Q_1 \\
&= Q_2.
\end{aligned}$$

By uniqueness of the restarted threshold, together with the boundary conventions above, the second-step threshold is t_{12}. The residual position after both steps is therefore

$$\begin{aligned}\min\Big\{q_i^{(1)}, \frac{E_i}{p_\tau}t_{12}\Big\} &= \min\Big\{\min\Big\{q_i, \frac{E_i}{p_\tau}t_1\Big\}, \frac{E_i}{p_\tau}t_{12}\Big\} \\ &= \min\Big\{q_i, \frac{E_i}{p_\tau}t_{12}\Big\},\end{aligned}$$

where the last equality follows from $t_{12} \leq t_1$. This is exactly the residual position obtained by applying the water-filling rule directly with quantity $Q_1 + Q_2$. Therefore,

$$F_{Q_2}^{\mathrm{WF}}(F_{Q_1}^{\mathrm{WF}}(s)) = F_{Q_1+Q_2}^{\mathrm{WF}}(s).$$

Thus the water-filling mechanism satisfies Assumptions A.9–A.11.

Any mechanism satisfying the axioms is the water-filling rule. Conversely, suppose that an ADL mechanism satisfies Assumptions A.9–A.11.

Fix an initial state $s = (q_i, E_i)_{i=1}^n \in \mathcal{S}$ and a target quantity $Q_1 \in (0, \sum_i q_i)$. Define the maximum leverage at quantity $Q \in [0, Q_1]$ by

$$t(Q) \triangleq \max_{1\leq i\leq n} \lambda_i(F_Q(s)) = \max_{1\leq i\leq n} \frac{p_\tau\,(q_i - x_i(s,Q))}{E_i}.$$

By Lemma A.12, each $x_i(s,\cdot)$ is continuous and nondecreasing. Hence $Q \mapsto \lambda_i(F_Q(s))$ is continuous and nonincreasing, and therefore $t(\cdot)$ is continuous and nonincreasing.

Claim 1. If $\lambda_i(s) < t(Q_1)$, then $x_i(s, Q_1) = 0$.

Assume for contradiction that $x_i(s, Q_1) > 0$ and let

$$Q_0 \triangleq \inf\{Q \in [0, Q_1] : x_i(s,Q) > 0\}.$$

Continuity gives $x_i(s, Q_0) = 0$, hence $\lambda_i(F_{Q_0}(s)) = \lambda_i(s) < t(Q_1) \leq t(Q_0)$, so $i \notin \mathcal{M}(F_{Q_0}(s))$. By leverage-priority at $F_{Q_0}(s)$ there exists $\varepsilon > 0$ such that

$$x_i(F_{Q_0}(s), h) = 0 \qquad \forall h \in [0, \varepsilon].$$

By path-independence, for all $h \in [0, \varepsilon]$,

$$x_i(s, Q_0 + h) = x_i(s, Q_0) + x_i(F_{Q_0}(s), h) = 0,$$

contradicting the definition of Q_0. Hence $x_i(s, Q_1) = 0$.

Claim 2. If $\lambda_i(s) \geq t(Q_1)$, then $\lambda_i(F_{Q_1}(s)) = t(Q_1)$.

We always have $\lambda_i(F_{Q_1}(s)) \leq t(Q_1)$. Suppose for contradiction that $\lambda_i(F_{Q_1}(s)) < t(Q_1)$. Set

$$\varepsilon \triangleq t(Q_1) - \lambda_i(F_{Q_1}(s)) > 0.$$

Since $\lambda_i(s) \geq t(Q_1) > t(Q_1) - \varepsilon/2$ and $Q \mapsto \lambda_i(F_Q(s))$ is continuous and nonincreasing, the intermediate value theorem yields some $Q_* < Q_1$ such that

$$\lambda_i(F_{Q_*}(s)) = t(Q_1) - \frac{\varepsilon}{2}.$$

By path-independence, $F_{Q_1}(s) = F_{Q_1 - Q_*}(F_{Q_*}(s))$. Define the restarted maximum leverage

$$\tilde{t}(h) \triangleq \max_{1 \le j \le n} \lambda_j(F_h(F_{Q_*}(s))), \qquad h \ge 0,$$

so that $\tilde{t}(Q_1 - Q_*) = t(Q_1)$. Since $\lambda_i(F_{Q_*}(s)) < \tilde{t}(Q_1 - Q_*)$, Claim 1 applied to the initial state $F_{Q_*}(s)$ and terminal quantity $Q_1 - Q_*$ implies $x_i(F_{Q_*}(s), Q_1 - Q_*) = 0$. Since the proof of Claim 1 applies to any initial state, this invocation is valid. Using path-independence in the form $x_i(s, Q_1) = x_i(s, Q_*) + x_i(F_{Q_*}(s), Q_1 - Q_*)$ (see (48) in Lemma A.12), we obtain $x_i(s, Q_1) = x_i(s, Q_*)$, hence $\lambda_i(F_{Q_1}(s)) = \lambda_i(F_{Q_*}(s)) = t(Q_1) - \varepsilon/2 > \lambda_i(F_{Q_1}(s))$, a contradiction. We conclude $\lambda_i(F_{Q_1}(s)) = t(Q_1)$.

Combining Claims 1–2 yields, for every i,

$$\lambda_i(F_{Q_1}(s)) = \min\{\lambda_i(s),\, t(Q_1)\}.$$

Since $\lambda_i(F_{Q_1}(s)) = \frac{p_\tau(q_i - x_i(s, Q_1))}{E_i}$ and $\lambda_i(s) = \frac{p_\tau q_i}{E_i}$, this is equivalent to

$$x_i(s, Q_1) = \Big(q_i - \frac{E_i}{p_\tau}\, t(Q_1)\Big)_+.$$

Summing over i and using market clearing $\sum_i x_i(s, Q_1) = Q_1$ gives

$$Q_1 = \sum_{i=1}^{n} \Big(q_i - \frac{E_i}{p_\tau}\, t(Q_1)\Big)_+.$$

By the strict decrease of $t \mapsto \sum_i \big(q_i - \frac{E_i}{p_\tau} t\big)_+$ on $[0, \max_i \lambda_i(s)]$ (cf. Lemma A.5), $t(Q_1)$ is the unique solution to the equation

$$Q_1 = \sum_{i=1}^{n} \Big(q_i - \frac{E_i}{p_\tau}\, t\Big)_+.$$

Hence the induced allocation $x_i(s, Q_1) = \big(q_i - \frac{E_i}{p_\tau} t(Q_1)\big)_+$ coincides with the water-filling allocation of Theorem 2.4. ■

A.4. Connection to the Claims Literature

There is an exact connection between the claims framework of Thomson [2003, 2015] and the water-filling rule in the single-asset isolated-margin model. If x_i is the buyback imposed on account i and $r_i \triangleq q_i - x_i$ is the residual position left open after ADL, then the ADL feasibility constraints become

$$0 \le r_i \le q_i \qquad \text{and} \qquad \sum_i r_i = \sum_i q_i - Q.$$

Thus the residual positions form a claims problem with claims q_i and endowment $\sum_i q_i - Q$. The water-filling allocation of Theorem 2.4 can be written in terms of residual positions as

$$r_i^\star = \min\Big\{q_i,\, \frac{E_i}{p_\tau}\, t^\star\Big\},$$

which is precisely the *weighted constrained equal awards rule* for the residual positions, with weights proportional to E_i/p_τ. Equivalently, in the buyback variables $x_i = q_i - r_i$, the rule is the dual weighted constrained equal losses rule. In the risk-neutral case, our expected-loss objective selects this allocation uniquely; for CVaR objectives, the same water-filling allocation remains a canonical optimizer, although the optimizer need not be unique (cf. Theorem 2.16). For spectral and more general monotone risk objectives, water-filling optimality is established in Proposition 2.18.

The interpretation, however, is different from the classical claims model. In the claims literature, weights are typically primitive priority or entitlement parameters. In our setting, the weights E_i/p_τ arise endogenously from account equity and from the bankruptcy thresholds induced by future price moves. Thus our contribution in the isolated-margin case is not the abstract water-filling formula itself, but the derivation of the particular weighted claims rule from an exchange shortfall-risk minimization problem.

Several properties of the minimax leverage policy also have counterparts in the claims literature. The path-independence property of Section 2.4.2 is analogous in spirit to composition properties studied in the claims framework, and our Sybil resistance requirement is related to the axiom of no advantageous splitting. However, the ADL axioms operate on account states (q_i, E_i) and are tied to leverage and future default risk, rather than to claims alone.

A.5. Proof of Theorem 2.16

We break the verification of the theorem claims into two parts. The first part shows that the water-filling strategy x^{WF} is optimal for (9). The second shows that the optimizer need not be unique in general. The proof of the theorem follows directly from Lemma A.17 and Proposition A.18.

A.5.1. Optimality of x^{WF}

Before we begin, we provide a useful representation for the account shortfall that will be used repeatedly.

Lemma A.13. *Fix i and $x_i \in [0, q_i]$. For any price $p \in \mathbb{R}$,*

$$\sigma_i(x_i, p) = (q_i - x_i)\big(p - p_i^{(z)}(x_i)\big)_+.$$

Proof. Fix $x_i \in [0, q_i]$ and $p \in \mathbb{R}$. Starting from

$$e_i(x_i, p) = q_i(p_i^{(e)} - p) - x_i(p_\tau - p) + m_i,$$

we expand and regroup terms:

$$\begin{aligned} e_i(x_i, p) &= q_i p_i^{(e)} - q_i p - x_i p_\tau + x_i p + m_i \\ &= (q_i p_i^{(e)} + m_i - x_i p_\tau) - (q_i - x_i)p. \end{aligned}$$

Hence

$$-e_i(x_i, p) = (q_i - x_i)p - \big(q_i p_i^{(e)} + m_i - x_i p_\tau\big).$$

If $x_i < q_i$, define

$$p_i^{(z)}(x_i) \triangleq \frac{q_i p_i^{(e)} + m_i - x_i p_\tau}{q_i - x_i}.$$

Then

$$-e_i(x_i, p) = (q_i - x_i)\big(p - p_i^{(z)}(x_i)\big),$$

and therefore

$$\sigma_i(x_i, p) = \big(-e_i(x_i, p)\big)_+ = (q_i - x_i)\big(p - p_i^{(z)}(x_i)\big)_+.$$

If $x_i = q_i$, then $e_i(q_i, p) = q_i(p_i^{(e)} - p_\tau) + m_i = E_i > 0$ is constant in p, so $\sigma_i(q_i, p) = 0$. With the convention $p_i^{(z)}(q_i) = +\infty$, the formula still holds since $(p - \infty)_+ = 0$. ∎

We can now derive necessary and sufficient conditions for optimality.

Lemma A.14. *Define, for $x_i \in [0, q_i]$,*

$$\bar{p}_i(x_i) \triangleq \max\{p_\beta,\ p_i^{(z)}(x_i)\}, \qquad h_\beta(x_i) \triangleq \mathsf{E}\Big[(p_T - p_\tau)\,\mathbb{I}_{\{p_T \geq \bar{p}_i(x_i)\}}\Big].$$

Under Assumptions 2.1–2.3, a point $x^\star \in \mathcal{X}$ solves (9) if and only if there exists $\theta^\star \in \mathbb{R}$ such that for each i,

$$\begin{aligned} 0 < x_i^\star < q_i \implies h_\beta(x_i^\star) = \theta^\star, \\ x_i^\star = 0 \implies h_\beta(0) \leq \theta^\star, \\ x_i^\star = q_i \implies h_\beta(q_i) \geq \theta^\star. \end{aligned}$$

Moreover, feasibility and the assumption $0 < Q < \sum_{i=1}^n q_i$ imply that $\theta^\star \geq 0$. Indeed, there exists an index j such that $x_j^\star < q_j$. If $0 < x_j^\star < q_j$, then

$$\theta^\star = h_\beta(x_j^\star) \geq 0.$$

If $x_j^\star = 0$, then

$$h_\beta(0) \leq \theta^\star,$$

and hence $\theta^\star \geq 0$ since $h_\beta(0) \geq 0$. Consequently, if $x_i^\star = q_i$ for some i, then

$$0 = h_\beta(q_i) \geq \theta^\star \geq 0,$$

so $\theta^\star = 0$.

Proof. In what follows we will exploit the separable form of the objective that was derived from the comonotonicity of the individual account shortfalls. The Rockafellar–Uryasev representation [Rockafellar and Uryasev, 2000] for CVaR gives that

$$\sum_{i=1}^n \mathrm{CVaR}_\beta(\sigma_i(x_i, p_T)) = \sum_{i=1}^n f_i(x_i),$$

where

$$f_i(x_i) \triangleq \inf_{\alpha_i \in \mathbb{R}} \left\{ \alpha_i + \frac{1}{1-\beta}\,\mathsf{E}\Big[\big(\sigma_i(x_i, p_T) - \alpha_i\big)_+\Big] \right\}.$$

Letting $\alpha = (\alpha_1, \ldots, \alpha_n)$ we can write the CVaR optimization equivalently as:

$$\begin{array}{ll} \underset{(x,\alpha)\in\mathbb{R}^{2n}}{\text{minimize}} & \tilde{V}(x,\alpha) \triangleq \sum_{i=1}^n \left\{ \alpha_i + \frac{1}{1-\beta}\,\mathsf{E}\Big[\big(\sigma_i(x_i,p_T) - \alpha_i\big)_+\Big] \right\} \\ \text{subject to} & x \in \mathcal{X}. \end{array}$$

For each i, the map $(x_i, \alpha_i) \mapsto \alpha_i + \frac{1}{1-\beta}\mathsf{E}[(\sigma_i(x_i, p_T) - \alpha_i)_+]$ is (jointly) convex in (x_i, α_i). So, this "lifted" representation is a convex problem. Moreover, partial minimization over α_i preserves convexity, so f_i (and the original CVaR optimization problem) is also convex. Finally, as in Appendix A.1, under Assumption 2.1 Slater's condition is trivially satisfied and so the KKT conditions are both necessary and sufficient for optimality.

As before, introduce multipliers $\lambda \in \mathbb{R}$ for $\sum_i x_i = Q$, $\nu_i \geq 0$ for $-x_i \leq 0$, and $\mu_i \geq 0$ for $x_i - q_i \leq 0$. An optimal primal–dual quintuple $(x^\star, \alpha^\star, \lambda^\star, \nu^\star, \mu^\star)$ satisfies

$$\begin{aligned} & 0 \in \partial_{x_i}\tilde{V}(x^\star, \alpha^\star) + \lambda^\star - \nu_i^\star + \mu_i^\star, \quad i = 1, \ldots, n, \\ & 0 \in \partial_{\alpha_i}\tilde{V}(x^\star, \alpha^\star), \quad i = 1, \ldots, n, \\ & x^\star \in \mathcal{X}, \\ & \nu_i^\star \geq 0,\ \mu_i^\star \geq 0, \quad i = 1, \ldots, n, \\ & \nu_i^\star x_i^\star = 0, \quad \mu_i^\star(x_i^\star - q_i) = 0, \quad i = 1, \ldots, n. \end{aligned} \tag{49}$$

The α_i-stationarity condition. Fix i and abbreviate $Z_i \triangleq \sigma_i(x_i, p_T)$. For each realization of p_T, the map $\alpha_i \mapsto (Z_i - \alpha_i)_+$ is convex and its subdifferential is

$$\partial_{\alpha_i}(Z_i - \alpha_i)_+ = \begin{cases} \{-1\}, & Z_i > \alpha_i, \\ [-1, 0], & Z_i = \alpha_i, \\ \{0\}, & Z_i < \alpha_i. \end{cases}$$

Since $\tilde{V}$ is a sum across i, one can compute $\partial_{\alpha_i}\tilde{V}$ by a direct calculation of the one-sided difference quotients as

$$\partial_{\alpha_i}\tilde{V}(x, \alpha) = \left[1 - \frac{1}{1-\beta}\mathsf{P}(Z_i \geq \alpha_i), \quad 1 - \frac{1}{1-\beta}\mathsf{P}(Z_i > \alpha_i)\right].$$

where $[a, b]$ denotes the closed interval.

Thus using the stationarity condition (49) and selecting the extreme points of the subgradient interval yields the pair of inequalities

$$1 - \frac{1}{1-\beta}\,\mathsf{P}(Z_i > \alpha_i) \;\geq\; 0 \;\geq\; 1 - \frac{1}{1-\beta}\,\mathsf{P}(Z_i \geq \alpha_i).$$

Rearranging gives the quantile condition

$$\mathsf{P}(Z_i \leq \alpha_i) \;\geq\; \beta \;\geq\; \mathsf{P}(Z_i < \alpha_i).$$

That is, any optimal $\alpha_i^\star$ is a β-quantile of $\sigma_i(x_i^\star, p_T)$. But, for each fixed x_i, the map $p \mapsto \sigma_i(x_i, p)$ is continuous and nondecreasing, and (under Assumption 2.3) p_T is atomless. Then, under the standing assumptions, the β-quantile of $\sigma_i(x_i, p_T)$ satisfies (see [Hanbali and Linders, 2022, Theorem 3.1])

$$\mathrm{VaR}_\beta(\sigma_i(x_i, p_T)) = \sigma_i(x_i, p_\beta). \tag{50}$$

Hence, we may take $\alpha_i^\star = \sigma_i(x_i^\star, p_\beta)$.

The x_i-stationarity condition. Fix i and notice that since $\sigma_i(x_i, p) \geq 0$ for all (x_i, p), the minimization in α_i can be restricted to $\alpha_i \geq 0$ without loss of generality (since $\alpha_i^\star = \sigma_i(x_i^\star, p_\beta)$).

For $\alpha_i \geq 0$ we have the identity

$$(\sigma_i(x_i, p_T) - \alpha_i)_+ = (-e_i(x_i, p_T) - \alpha_i)_+,$$

because $\sigma_i = 0$ on $\{e_i > 0\}$ and $\sigma_i = -e_i$ on $\{e_i \leq 0\}$.

For each fixed p, the map $x_i \mapsto -e_i(x_i, p) - \alpha_i$ is affine with slope $p_\tau - p$ for $x_i < q_i$. Hence, for each (x_i, α_i), a subgradient of $x_i \mapsto (-e_i(x_i, p) - \alpha_i)_+$ is given by

$$(p_\tau - p) \cdot \mathbb{I}_{\{-e_i(x_i, p) > \alpha_i\}},$$

with the usual interval of subgradients at the kink $\{-e_i(x_i, p) = \alpha_i\}$. Under Assumption 2.3, the kink event has probability zero for $x_i < q_i$, hence for $x_i < q_i$

$$\partial_{x_i}\tilde{V}(x^\star, \alpha^\star) = \left\{\frac{1}{1-\beta}\mathsf{E}\Big[(p_\tau - p_T)\,\mathbb{I}_{\{-e_i(x_i^\star, p_T) > \alpha_i^\star\}}\Big]\right\}. \tag{51}$$

Define the threshold price

$$\bar{p}_i(x_i, \alpha_i) \triangleq \frac{q_i p_i^{(e)} + m_i + \alpha_i - x_i p_\tau}{q_i - x_i}, \qquad x_i < q_i.$$

A rearrangement yields

$$-e_i(x_i, p) - \alpha_i = (q_i - x_i)(p - \bar{p}_i(x_i, \alpha_i)),$$

and since $q_i - x_i > 0$,

$$\{-e_i(x_i, p) > \alpha_i\} = \{p > \bar{p}_i(x_i, \alpha_i)\}.$$

Substituting into (51) gives

$$\partial_{x_i} \tilde{V}(x^\star, \alpha^\star) = \left\{ \frac{1}{1-\beta} \mathsf{E}\Big[(p_\tau - p_T)\, \mathbb{I}_{\{p_T > \bar{p}_i(x_i^\star, \alpha_i^\star)\}}\Big] \right\}.$$

Finally, note that

$$p_i^{(z)}(x_i) = \frac{q_i p_i^{(e)} + m_i - x_i p_\tau}{q_i - x_i} \qquad \Rightarrow \qquad \bar{p}_i(x_i, \alpha_i) = p_i^{(z)}(x_i) + \frac{\alpha_i}{q_i - x_i}.$$

Moreover, since $p \mapsto \sigma_i(x_i, p) = (q_i - x_i)\,(p - p_i^{(z)}(x_i))_+$ is continuous and nondecreasing, the β-quantile satisfies (see (50) and Lemma A.13)

$$\alpha_i^\star = \mathrm{VaR}_\beta(\sigma_i(x_i^\star, p_T)) = \sigma_i(x_i^\star, p_\beta) = (q_i - x_i^\star)\big(p_\beta - p_i^{(z)}(x_i^\star)\big)_+.$$

Therefore

$$\bar{p}_i(x_i^\star, \alpha_i^\star) = p_i^{(z)}(x_i^\star) + \big(p_\beta - p_i^{(z)}(x_i^\star)\big)_+ = \max\{p_\beta,\ p_i^{(z)}(x_i^\star)\}.$$

As a result, by comparing with the definition in the statement of the Lemma, we see that $\bar{p}_i(x_i^\star) = \bar{p}_i(x_i^\star, \alpha_i^\star)$ by the calculation above. We obtain, for $x_i^\star < q_i$,

$$\partial_{x_i} \tilde{V}(x^\star, \alpha^\star) = \left\{ -\frac{1}{1-\beta}\, h_\beta(x_i^\star) \right\}.$$

Remark A.15. In what follows we will make use of the following standard facts for convex functions. Let $g : [a, b] \to \mathbb{R}$ be convex.

1. For every $x \in (a, b)$ the one-sided derivatives $g'_-(x)$ and $g'_+(x)$ exist, are finite, and satisfy $g'_-(x) \le g'_+(x)$. Moreover g'_- is left-continuous and g'_+ is right-continuous on (a, b).

2. At the endpoints,

$$\partial g(a) = (-\infty,\ g'_+(a)], \qquad \partial g(b) = [\,g'_-(b),\ \infty),$$

(with the convention that $(-\infty, -\infty] = \varnothing$ and $[+\infty, \infty) = \varnothing$) so that $\sup \partial g(a) = g'_+(a)$ and $\inf \partial g(b) = g'_-(b)$. If g is differentiable on (a, b), then $g'_+(a) = \lim_{x \downarrow a} g'(x)$ and $g'_-(b) = \lim_{x \uparrow b} g'(x)$.

The endpoint $x_i = q_i$. Suppose that $x_i^\star = q_i$. Then

$$\alpha_i^\star = \sigma_i(q_i, p_\beta) = 0.$$

Since

$$p_i^{(z)}(x_i) = \frac{q_i p_i^{(e)} + m_i - x_i p_\tau}{q_i - x_i} = p_\tau + \frac{E_i}{q_i - x_i},$$

we have

$$p_i^{(z)}(x_i) \longrightarrow +\infty \qquad \text{as } x_i \uparrow q_i.$$

Hence, for $x_i < q_i$ sufficiently close to q_i,

$$p_i^{(z)}(x_i) \geq p_\beta,$$

and therefore

$$\bar{p}_i(x_i, \alpha_i^\star) = p_i^{(z)}(x_i) + \frac{\alpha_i^\star}{q_i - x_i} = p_i^{(z)}(x_i).$$

Using the derivative formula derived above and the atomlessness of p_T, we obtain, for $x_i < q_i$ sufficiently close to q_i,

$$\frac{\partial}{\partial x_i} \tilde{V}(x_1^\star, \dots, x_{i-1}^\star, x_i, x_{i+1}^\star, \dots, x_n^\star, \alpha^\star) = -\frac{1}{1-\beta} h_\beta(x_i).$$

Moreover,

$$h_\beta(x_i) = \mathsf{E}\left[(p_T - p_\tau)\mathbb{I}_{\left\{p_T \geq p_i^{(z)}(x_i)\right\}}\right] \longrightarrow 0 \qquad \text{as } x_i \uparrow q_i,$$

by dominated convergence, since $p_i^{(z)}(x_i) \to +\infty$ and the integrand is eventually dominated by $(p_T - p_\tau)_+$. It follows that the left derivative of the coordinate objective at q_i equals 0. By convexity,

$$\partial_{x_i} \tilde{V}(x^\star, \alpha^\star) \subset [0, \infty)$$

and

$$0 \in \partial_{x_i} \tilde{V}(x^\star, \alpha^\star)$$

at $x_i^\star = q_i$.

Conclusion. Define $\theta^\star \triangleq (1-\beta)\lambda^\star$.

(Only if). Assume $x^\star$ is optimal. Consider any index i.

<u>Interior: $0 < x_i^\star < q_i$.</u> Then $\nu_i^\star = \mu_i^\star = 0$ and stationarity gives

$$0 = -\frac{1}{1-\beta} h_\beta(x_i^\star) + \lambda^\star \quad \iff \quad h_\beta(x_i^\star) = \theta^\star.$$

<u>Lower bound: $x_i^\star = 0$.</u> Then $\mu_i^\star = 0$ and stationarity yields

$$0 \in \partial_{x_i} \tilde{V}(x^\star, \alpha^\star) + \lambda^\star - \nu_i^\star.$$

Since $\partial_{x_i} \tilde{V}(x^\star, \alpha^\star) \subset (-\infty, -\frac{1}{1-\beta} h_\beta(0)]$ at $x_i^\star = 0$ (the maximal subgradient is the right derivative), existence of $\nu_i^\star \geq 0$ forces

$$-\frac{1}{1-\beta} h_\beta(0) + \lambda^\star \geq 0 \quad \iff \quad h_\beta(0) \leq \theta^\star.$$

<u>Upper bound: $x_i^\star = q_i$.</u> Then $\nu_i^\star = 0$ and stationarity gives

$$0 \in \partial_{x_i} \tilde{V}(x^\star, \alpha^\star) + \lambda^\star + \mu_i^\star, \qquad \mu_i^\star \geq 0.$$

As shown above,

$$\partial_{x_i} \tilde{V}(x^\star, \alpha^\star) \subset [0, \infty)$$

at $x_i^\star = q_i$. Hence stationarity requires

$$\lambda^\star \leq 0,$$

or equivalently

$$\theta^\star \le 0 = h_\beta(q_i).$$

Thus

$$h_\beta(q_i) \ge \theta^\star.$$

(If). Conversely, suppose $x^\star \in \mathcal{X}$ and there exists $\theta^\star$ satisfying the three displayed implications. Set $\lambda^\star \triangleq \theta^\star/(1-\beta)$ and define $\alpha_i^\star \triangleq \sigma_i(x_i^\star, p_\beta)$.

By the preceding argument, $\theta^\star \ge 0$.

We construct $\nu^\star, \mu^\star$ coordinatewise to satisfy KKT.

$\underline{\text{If } 0 < x_i^\star < q_i}$: set $\nu_i^\star = \mu_i^\star = 0$. Then stationarity holds because $-\frac{1}{1-\beta} h_\beta(x_i^\star) + \lambda^\star = 0$ due to $h_\beta(x_i^\star) = \theta^\star$.

$\underline{\text{If } x_i^\star = 0}$: set $\mu_i^\star = 0$ and choose the maximal subgradient $g_i^\star \triangleq -\frac{1}{1-\beta} h_\beta(0) \in \partial_{x_i} \tilde{V}(x^\star, \alpha^\star)$. Define

$$\nu_i^\star \triangleq g_i^\star + \lambda^\star = \frac{\theta^\star - h_\beta(0)}{1-\beta} \ge 0,$$

where the inequality uses $h_\beta(0) \le \theta^\star$. Then stationarity holds: $0 = g_i^\star + \lambda^\star - \nu_i^\star$.

$\underline{\text{If } x_i^\star = q_i}$: the condition $h_\beta(q_i) \ge \theta^\star$ gives

$$0 = h_\beta(q_i) \ge \theta^\star.$$

Together with $\theta^\star \ge 0$, this implies $\theta^\star = 0$, and hence $\lambda^\star = 0$. Choose $g_i^\star = 0 \in \partial_{x_i} \tilde{V}(x^\star, \alpha^\star)$ and set $\nu_i^\star = 0$, $\mu_i^\star = 0$. Then stationarity holds.

Finally, primal feasibility holds by assumption ($x^\star \in \mathcal{X}$), dual feasibility holds by construction, and complementary slackness holds by how we set $\nu_i^\star, \mu_i^\star$ in each case. Therefore the KKT conditions are satisfied, and since the problem is convex with Slater's condition, $x^\star$ is optimal. ∎

By interpreting these conditions in terms of the account leverage, we can get a result that is reminiscent of the original equalization characterization from Proposition A.3. Observe that

$$\{p_T > \bar{p}_i(x_i^\star)\} = \{p_T > \max\{p_\beta,\, p_i^{(z)}(x_i^\star)\}\} = \{p_T - p_\tau > \max\{p_\beta - p_\tau,\, p_\tau\, \ell_i(x_i^\star)^{-1}\}\}.$$

As in Section 2.5, write $\ell_\beta = \left(\frac{p_\beta - p_\tau}{p_\tau}\right)^{-1}$ with the convention that $\ell_\beta = +\infty$ if $p_\beta \le p_\tau$. Then, we can define the function $v_\beta(\cdot)$

$$v_\beta(\ell) = \mathsf{E}\left[(p_T - p_\tau)\,\mathbb{I}_{\left\{p_T - p_\tau > p_\tau \max\{\ell_\beta^{-1},\, \ell^{-1}\}\right\}}\right].$$

We also use the conventions

$$(+\infty)^{-1} = 0, \qquad 0^{-1} = +\infty.$$

Since p_T is atomless under Assumption 2.3, the strict and weak inequalities are interchangeable in these tail events. Since the threshold $\max(p_\beta, p_\tau(1+\ell^{-1}))$ is nonincreasing in ℓ, the map $\ell \mapsto v_\beta(\ell)$ is nondecreasing. Moreover, $v_\beta(\ell) \ge 0$ for all ℓ, because the integrand is nonnegative on the indicator event. We may immediately restate the preceding lemma as follows.

Corollary A.16. *Under Assumptions 2.1–2.3, a point $x^\star \in \mathcal{X}$ solves (9) if and only if there exists $\theta^\star \in \mathbb{R}$ such that for each i,*

$$\begin{aligned}
&0 < x_i^\star < q_i \implies v_\beta(\ell_i(x_i^\star)) = \theta^\star,\\
&x_i^\star = 0 \implies v_\beta(\ell_i(0)) \le \theta^\star,\\
&x_i^\star = q_i \implies v_\beta(\ell_i(q_i)) \ge \theta^\star.
\end{aligned}$$

Moreover, $\theta^\star \ge 0$. If $x_i^\star = q_i$ for some i, then

$$0 = v_\beta(\ell_i(q_i)) \ge \theta^\star \ge 0,$$

so $\theta^\star = 0$.

We can now verify that the water-filling solution is optimal.

Lemma A.17. *The water-filling solution is optimal for (9).*

Proof. Let $x^{\mathrm{WF}} = y(t^\star)$ be the water-filling allocation of Theorem 2.4. Then

$$\ell_i(x_i^{\mathrm{WF}}) = \min\{\ell_i(0), t^\star\} \qquad \text{for all } i.$$

Set $\theta^\star \triangleq v_\beta(t^\star)$.

Since $0 < Q < \sum_i q_i$, we have $t^\star > 0$. Hence

$$q_i - x_i^{\mathrm{WF}} = \min\left\{q_i, \frac{E_i}{p_\tau} t^\star\right\} > 0,$$

so $x_i^{\mathrm{WF}} < q_i$ for every i.

If $x_i^{\mathrm{WF}} > 0$, then

$$\ell_i(x_i^{\mathrm{WF}}) = t^\star,$$

and therefore

$$v_\beta(\ell_i(x_i^{\mathrm{WF}})) = v_\beta(t^\star) = \theta^\star.$$

If $x_i^{\mathrm{WF}} = 0$, then

$$\ell_i(0) \le t^\star,$$

and monotonicity of v_β yields

$$v_\beta(\ell_i(0)) \le v_\beta(t^\star) = \theta^\star.$$

Thus x^{WF} satisfies the conditions of Corollary A.16, and is therefore optimal. ■

A.5.2. Non-Uniqueness

Proposition A.18. *Let $x^\star$ be an optimizer of (9). Define the stressed set for a feasible allocation x by*

$$\mathcal{D}(x) \triangleq \{\, i : \ \bar{p}_i(x_i) = p_\beta \,\} = \{\, i : \ \ell_i(x_i) \ge \ell_\beta \,\}.$$

Suppose there exist $i \ne j$ in $\mathcal{D}(x^\star)$ and $\varepsilon_0 > 0$ such that for all $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$ the perturbed point $x^\star(\varepsilon)$ defined by

$$x_i^\star(\varepsilon) = x_i^\star + \varepsilon, \qquad x_j^\star(\varepsilon) = x_j^\star - \varepsilon, \qquad x_k^\star(\varepsilon) = x_k^\star \ (k \notin \{i,j\})$$

remains feasible and satisfies $i, j \in \mathcal{D}(x^\star(\varepsilon))$. Then $x^\star(\varepsilon)$ is optimal for all $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$, and the set of optimizers contains a continuum.

Proof. Fix $i \in \mathcal{D}(x^\star)$. Since $\bar{p}_i(x_i) = \max\{p_\beta, p_i^{(z)}(x_i)\}$, $\bar{p}_i(x_i) = p_\beta$ means $p_i^{(z)}(x_i) \le p_\beta$. In this regime, the β-quantile of $\sigma_i(x_i, p_T)$ is $\mathrm{VaR}_\beta(\sigma_i) = \sigma_i(x_i, p_\beta)$ and the excess loss above this quantile is

$$\left(\sigma_i(x_i, p_T) - \mathrm{VaR}_\beta(\sigma_i)\right)_+ = (q_i - x_i)\,(p_T - p_\beta)_+,$$

so the Rockafellar–Uryasev formula gives

$$f_i(x_i) = \mathrm{CVaR}_\beta(\sigma_i(x_i, p_T)) = \sigma_i(x_i, p_\beta) + \frac{1}{1-\beta}\mathsf{E}\big[(q_i - x_i)(p_T - p_\beta)_+\big].$$

Let $\mu_{\text{tail}} \triangleq \mathsf{E}[p_T \mid p_T \geq p_\beta]$. Since p_T is atomless, $\mathsf{E}[(p_T - p_\beta)_+] = (1-\beta)\,(\mu_{\text{tail}} - p_\beta)$, and $\sigma_i(x_i, p_\beta) = (q_i - x_i)(p_\beta - p_i^{(z)}(x_i))$ on $p_i^{(z)}(x_i) \leq p_\beta$. Hence

$$f_i(x_i) = (q_i - x_i)\big(\mu_{\text{tail}} - p_i^{(z)}(x_i)\big).$$

Using $(q_i - x_i)p_i^{(z)}(x_i) = q_i p_i^{(e)} + m_i - x_i p_\tau$ (by definition of $p_i^{(z)}$), we obtain the affine representation

$$f_i(x_i) = q_i(\mu_{\text{tail}} - p_i^{(e)}) + x_i(p_\tau - \mu_{\text{tail}}) - m_i, \qquad \text{whenever } \bar{p}_i(x_i) = p_\beta,$$

so on the stressed regime every such f_i has the same constant slope $p_\tau - \mu_{\text{tail}}$.

Now fix $i \neq j$ as in the statement and consider $x^\star(\varepsilon)$. Feasibility holds since $\sum_k x_k^\star(\varepsilon) = Q$ and the box constraints are preserved by assumption. Moreover, for all $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$ both i and j remain in the stressed regime, and the objective changes by

$$f_i(x_i^\star + \varepsilon) + f_j(x_j^\star - \varepsilon) - f_i(x_i^\star) - f_j(x_j^\star) = (p_\tau - \mu_{\text{tail}})\varepsilon + (p_\tau - \mu_{\text{tail}})(-\varepsilon) = 0,$$

while all other coordinates are unchanged. Therefore $\sum_k f_k(x_k^\star(\varepsilon)) = \sum_k f_k(x_k^\star)$, so $x^\star(\varepsilon)$ is also optimal. This yields a continuum of optima. ■

The following example illustrates that the Proposition is not vacuously true (i.e., it provides a simple example where the stated conditions hold and lead to a continuum of CVaR-optima).

Example A.19. Fix $p_\tau = 1$ and take the terminal price to be

$$p_T \;=\; 1 + Y, \qquad Y \sim \mathrm{Exp}(1),$$

so that p_T admits the density $f_T(p) = e^{-(p-1)}\mathbb{I}_{\{p \geq 1\}}$, which is strictly positive on $[p_\tau, \infty)$.

Let $\beta = 1/2$, so that

$$p_\beta = \mathrm{VaR}_\beta(p_T) = 1 + \log 2, \qquad \mu_{\text{tail}} = \mathsf{E}[p_T \mid p_T \geq p_\beta] = p_\beta + 1 = 2 + \log 2,$$

where the last identity follows from the memoryless property of the exponential distribution.

Consider $n = 2$ short accounts with identical parameters

$$q_1 = q_2 = 4, \qquad p_1^{(e)} = p_2^{(e)} = 1, \qquad m_1 = m_2 = 1,$$

so that $E_1 = E_2 = q_i(p_i^{(e)} - p_\tau) + m_i = 1 > 0$. Choose a buyback quantity $Q = 1$. Then the feasible set is

$$\mathcal{X} = \{(x_1, x_2) \in \mathbb{R}^2 :\; x_1 + x_2 = 1,\; 0 \leq x_1, x_2 \leq 4\}.$$

For any $x \in \mathcal{X}$ we have $x_i \leq 1$, hence

$$p_i^{(z)}(x_i) = p_\tau + \frac{E_i}{q_i - x_i} = 1 + \frac{1}{4 - x_i} \;\leq\; 1 + \frac{1}{3} = \frac{4}{3} \;<\; 1 + \log 2 = p_\beta.$$

Therefore $\bar{p}_i(x_i) = \max\{p_\beta, p_i^{(z)}(x_i)\} = p_\beta$ for both $i = 1, 2$ and for every feasible x, i.e. $\mathcal{D}(x) = \{1, 2\}$ for all $x \in \mathcal{X}$. Arguing as in Proposition A.18, every feasible allocation $x \in \mathcal{X}$ is optimal for (9), so the optimizer set contains a continuum. In particular,

$$(1/2,\, 1/2) \quad \text{and} \quad (1,\, 0) \quad \text{and} \quad (0,\, 1)$$

are distinct CVaR-optimal solutions with $x^{\mathrm{WF}} = (Q/2, Q/2) = (1/2, 1/2)$ being optimal but not unique.

A.6. CVaR Closed-Form under Geometric Brownian Motion

Assume that the terminal price follows a geometric Brownian motion,

$$p_T = p_\tau \exp\Big((\mu - \tfrac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}\, Z\Big), \qquad Z \sim N(0,1), \qquad \Delta \triangleq T - \tau, \tag{52}$$

with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$. Then $\log p_T$ is Gaussian with

$$\mathsf{E}[p_T] = p_\tau e^{\mu\Delta}, \qquad \operatorname{Var}(p_T) = p_\tau^2 e^{2\mu\Delta}\big(e^{\sigma^2\Delta} - 1\big).$$

Let $z_\beta \triangleq \Phi^{-1}(\beta)$ so that (by monotonicity) the β-quantile of p_T is

$$p_\beta = p_\tau \exp\Big((\mu - \tfrac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}\, z_\beta\Big). \tag{53}$$

We will also use the tail conditional mean

$$\mu_{\text{tail}} \triangleq \mathsf{E}\big[p_T \,\big|\, p_T \geq p_\beta\big] = \frac{\mathsf{E}\big[p_T \mathbf{1}_{\{p_T \geq p_\beta\}}\big]}{1-\beta}.$$

A direct computation yields a closed form for μ_{tail}.

Lemma A.20 (Tail mean of GBM)**.** *Under (52), one has*

$$\mu_{\text{tail}} = \frac{p_\tau e^{\mu\Delta}\, \Phi(\sigma\sqrt{\Delta} - z_\beta)}{1-\beta}. \tag{54}$$

Proof. Write $p_T = p_\tau e^{(\mu - \frac{1}{2}\sigma^2)\Delta + \sigma\sqrt{\Delta}Z}$ and set $s \triangleq \sigma\sqrt{\Delta}$. Then

$$\mathsf{E}\big[p_T \mathbf{1}_{\{p_T \geq p_\beta\}}\big] = p_\tau e^{(\mu - \frac{1}{2}\sigma^2)\Delta}\, \mathsf{E}\big[e^{sZ}\mathbf{1}_{\{Z \geq z_\beta\}}\big].$$

Using the identity $\mathsf{E}[e^{sZ}\mathbf{1}_{\{Z \geq a\}}] = e^{\frac{1}{2}s^2}\Phi(s-a)$ for $Z \sim N(0,1)$, we obtain

$$\mathsf{E}\big[p_T \mathbf{1}_{\{p_T \geq p_\beta\}}\big] = p_\tau e^{(\mu - \frac{1}{2}\sigma^2)\Delta}\, e^{\frac{1}{2}s^2}\Phi(s - z_\beta) = p_\tau e^{\mu\Delta}\Phi(s - z_\beta).$$

Dividing by $1-\beta$ yields (54). ■

Since $p \mapsto \sigma_i(x,p)$ is nondecreasing and continuous, the β-quantile of $\sigma_i(x,p_T)$ is realized by the price quantile p_β, and thus

$$\operatorname{VaR}_\beta(\sigma_i(x,p_T)) = \sigma_i(x,p_\beta) = (q_i - x)\,\big(p_\beta - p_i^{(z)}(x)\big)_+. \tag{55}$$

The CVaR admits a piecewise closed form depending on the relative position of $p_i^{(z)}(x)$ and p_β.

Lemma A.21 (Per-account CVaR under GBM)**.** *Fix i and $x \in [0, q_i)$. Let p_β be given by (53). Then*

$$\operatorname{CVaR}_\beta(\sigma_i(x,p_T)) = \mathsf{E}\big[\sigma_i(x,p_T) \,\big|\, p_T \geq p_\beta\big]$$

and is given explicitly as follows.

(i) If $p_i^{(z)}(x) \leq p_\beta$, then

$$\operatorname{CVaR}_\beta(\sigma_i(x,p_T)) = (q_i - x)\big(\mu_{\text{tail}} - p_i^{(z)}(x)\big), \tag{56}$$

where μ_{tail} is given by (54). Equivalently, using $(q_i - x)p_i^{(z)}(x) = q_i p_i^{(e)} + m_i - x p_\tau$,

$$\mathrm{CVaR}_\beta(\sigma_i(x, p_T)) = q_i(\mu_{\text{tail}} - p_i^{(e)}) + x(p_\tau - \mu_{\text{tail}}) - m_i. \tag{57}$$

(ii) If $p_i^{(z)}(x) \geq p_\beta$, then

$$\mathrm{CVaR}_\beta(\sigma_i(x, p_T)) = \frac{q_i - x}{1-\beta}\,\mathsf{E}\big[(p_T - p_i^{(z)}(x))_+\big]. \tag{58}$$

Moreover, $\mathsf{E}[(p_T - K)_+]$ for $K > 0$ admits the closed form

$$\mathsf{E}\big[(p_T - K)_+\big] = p_\tau e^{\mu\Delta}\Phi(d_1(K)) - K\,\Phi(d_2(K)), \tag{59}$$

where

$$d_1(K) \triangleq \frac{\log(p_\tau/K) + (\mu + \frac{1}{2}\sigma^2)\Delta}{\sigma\sqrt{\Delta}}, \qquad d_2(K) \triangleq d_1(K) - \sigma\sqrt{\Delta}. \tag{60}$$

In particular, taking $K = p_i^{(z)}(x)$ in (59)–(60) yields an explicit expression for (58).

Proof. We first note from Lemma A.13 that $p \mapsto \sigma_i(x,p)$ is nondecreasing and continuous. Since p_T is atomless under (52), the β-quantile of $\sigma_i(x,p_T)$ is $\sigma_i(x,p_\beta)$, giving (55).

Although $\sigma_i(x,p_T)$ can have an atom at 0 (since $\sigma_i(x,p_T) = 0$ on $\{p_T \leq p_i^{(z)}(x)\}$), we can still express CVaR_β as a conditional expectation on the *price* tail $\{p_T \geq p_\beta\}$. Let $F(p) \triangleq \mathsf{P}(p_T \leq p)$ denote the cdf of p_T; under (52), F is continuous. For $u \in (0,1)$ define the (left) quantile

$$p_u \triangleq \mathrm{VaR}_u(p_T) = \inf\{p \in \mathbb{R} : \; F(p) \geq u\}.$$

Set $U \triangleq F(p_T)$. Since F is continuous, U is uniformly distributed on $(0,1)$ and $p_T = p_U$ with probability one.

Now write $K \triangleq p_i^{(z)}(x)$ and recall $\sigma_i(x,p_T) = (q_i - x)\,(p_T - K)_+$. For $y \geq 0$ we have

$$\mathsf{P}\big((p_T - K)_+ \leq y\big) = \mathsf{P}(p_T \leq K + y) = F(K+y),$$

hence, for all $u \in (0,1)$,

$$\mathrm{VaR}_u((p_T - K)_+) = (p_u - K)_+, \qquad \text{and therefore} \qquad \mathrm{VaR}_u(\sigma_i(x,p_T)) = (q_i - x)\,(p_u - K)_+ = \sigma_i(x,p_u).$$

Using the quantile-integral definition of CVaR_β,

$$\mathrm{CVaR}_\beta(\sigma_i(x,p_T)) = \frac{1}{1-\beta}\int_\beta^1 \mathrm{VaR}_u(\sigma_i(x,p_T))\,du = \frac{1}{1-\beta}\int_\beta^1 \sigma_i(x,p_u)\,du.$$

Since U is uniform on $(0,1)$, for any integrable function g one has $\mathsf{E}[g(U) \mid U \geq \beta] = \frac{1}{1-\beta}\int_\beta^1 g(u)\,du$. Applying this with $g(u) = \sigma_i(x,p_u)$ yields

$$\frac{1}{1-\beta}\int_\beta^1 \sigma_i(x,p_u)\,du = \mathsf{E}\big[\sigma_i(x,p_U)\,\big|\,U \geq \beta\big] = \mathsf{E}\big[\sigma_i(x,p_T)\,\big|\,p_T \geq p_\beta\big],$$

where the last equality uses $p_T = p_U$ with probability one and $\{U \geq \beta\} = \{p_T \geq p_\beta\}$ (by continuity and monotonicity of F).

If $p_i^{(z)}(x) \le p_\beta$, then $(p_T - p_i^{(z)}(x))_+ = p_T - p_i^{(z)}(x)$ on $\{p_T \ge p_\beta\}$, so using Lemma A.13 we obtain

$$\mathrm{CVaR}_\beta(\sigma_i(x, p_T)) = (q_i - x)\mathsf{E}\big[p_T - p_i^{(z)}(x) \,|\, p_T \ge p_\beta\big] = (q_i - x)\big(\mu_{\text{tail}} - p_i^{(z)}(x)\big),$$

which is (56). The affine form (57) follows by substituting $(q_i - x)p_i^{(z)}(x) = q_i p_i^{(e)} + m_i - x p_\tau$.

If $p_i^{(z)}(x) \ge p_\beta$, then on the event $\{p_T < p_\beta\}$ we also have $p_T < p_i^{(z)}(x)$, and therefore

$$\sigma_i(x, p_T) = (q_i - x)\big(p_T - p_i^{(z)}(x)\big)_+ = 0.$$

Hence, using the conditional-tail representation established above,

$$\mathrm{CVaR}_\beta(\sigma_i(x, p_T)) = \mathsf{E}\big[\sigma_i(x, p_T) \mid p_T \ge p_\beta\big] = \frac{1}{1-\beta}\mathsf{E}\big[\sigma_i(x, p_T)\big].$$

Substituting $\sigma_i(x, p_T) = (q_i - x)\big(p_T - p_i^{(z)}(x)\big)_+$ yields

$$\mathrm{CVaR}_\beta(\sigma_i(x, p_T)) = \frac{q_i - x}{1-\beta}\mathsf{E}\big[(p_T - p_i^{(z)}(x))_+\big],$$

which is (58). Finally, (59)–(60) are obtained by a standard lognormal calculation (equivalently, the Black–Scholes call expectation with drift μ rather than the risk-free rate). ■

B. Details and Proofs for Multi-Asset Cross-Margining

B.1. Example: Distinct Optimizers for Different Risk Measures

In this section, we construct an example where two risk measures CVaR_{β_j} with $\beta_1 \ne \beta_2$ lead to different (and unique) optimal allocations.

Consider $d = 2$ assets and $n = 2$ accounts. Let the ADL time be τ with reference prices

$$p_\tau = (1, 1), \qquad p_i^{(e)} = p_\tau \ \text{ for } i = 1, 2,$$

and write the price changes as $r = p_T - p_\tau$. Then for any ADL allocation $x_i \in \mathbb{R}^2$,

$$e_i(x_i, p_T) = m_i - (q_i - x_i)^\top r, \qquad \sigma_i(x_i, p_T) = \big((q_i - x_i)^\top r - m_i\big)_+, \qquad \mathcal{L}(x, p_T) = \sum_{i=1}^{2} \sigma_i(x_i, p_T).$$

Take the signed position vectors (all shorts) $q_1 = (10, 0)$, $q_2 = (10, 10)$, and margins $m_1 = 18$, $m_2 = 40$. Assume the exchange must reduce aggregate exposure by $Q = (10, 0)$. By monotone deleveraging, $0 \le x_i^k \le q_i^k$ for all i, k and $x_1 + x_2 = Q$. Moreover, feasibility allows no changes in positions for the second asset. As a result, we can parameterize the feasible set by

$$a = x_1^1 \in [0, 10], \qquad x_1 = (a, 0), \qquad x_2 = (10 - a, 0).$$

Then the post-ADL positions are

$$q_1 - x_1 = (10 - a, 0), \qquad q_2 - x_2 = (a, 10).$$

Let p_T take three values (expressed via $r = p_T - p_\tau$) with the following probabilities:

$$r_0 = (0, 0) \ \text{ w.p. } 0.90, \qquad r_1 = (3, 0) \ \text{ w.p. } 0.05, \qquad r_2 = (1, 4) \ \text{ w.p. } 0.05.$$

Example B.1. In the stated model, the CVaR_β-optimal ADL allocation depends on β. Specifically, for $\beta \in \{0.90, 0.95\}$, the optimizers are unique and given by

$$a^\star_{0.90} = 4 \neq 3 = a^\star_{0.95}.$$

Equivalently, the optimal allocations are, respectively,

$$x_1(a^\star_{0.90}) = (4,0), \quad x_2(a^\star_{0.90}) = (6,0) \qquad \text{and} \qquad x_1(a^\star_{0.95}) = (3,0), \quad x_2(a^\star_{0.95}) = (7,0).$$

Proof. For $a \in [0,10]$, write $x(a)$ for the induced allocation and denote the losses in each scenario by $\mathcal{L}_j(a) = \mathcal{L}(x(a),\, p_\tau + r_j)$, $j = 0,1,2$. In scenario $r_1 = (3,0)$,

$$\mathcal{L}_1(a) = \big(3(10-a) - 18\big)_+ + \big(3a - 40\big)_+ = (12-3a)_+, \qquad a \in [0,10],$$

since $3a \le 30 < 40$ on $[0,10]$. In scenario $r_2 = (1,4)$,

$$\mathcal{L}_2(a) = \big((10-a) - 18\big)_+ + \big(a + 10\cdot 4 - 40\big)_+ = 0 + (a)_+ = a, \qquad a \in [0,10],$$

and in scenario $r_0 = (0,0)$ we have $\mathcal{L}_0(a) = 0$ for all a. Because $\mathbb{P}(r = r_0) = 0.90$ and $\mathcal{L}_0(a) = 0$ for all a, while $\mathbb{P}(r \in \{r_1, r_2\}) = 0.10$ and

$$\mathcal{L}_1(a) = (12-3a)_+ \ge 0, \qquad \mathcal{L}_2(a) = a \ge 0, \qquad a \in [0,10],$$

the loss is identically zero in scenario 0 and can be positive only in scenarios 1 and 2 (which together have total probability 0.10). In particular, $\mathrm{VaR}_{0.90}(\mathcal{L}(a)) = 0$, so the 10% upper tail at level $\beta = 0.90$ can be identified, for the purpose of computing $\mathrm{CVaR}_{0.90}$, with scenarios 1 and 2.[10]

Hence, for $\beta = 0.90$,

$$\mathrm{CVaR}_{0.90}(\mathcal{L}(a)) = \frac{0.05\,\mathcal{L}_1(a) + 0.05\,\mathcal{L}_2(a)}{0.10} = \frac{\mathcal{L}_1(a) + \mathcal{L}_2(a)}{2} = \frac{(12-3a)_+ + a}{2}.$$

If $a \le 4$, then $(12-3a)_+ = 12 - 3a$ and $\mathrm{CVaR}_{0.90}(\mathcal{L}(a)) = 6 - a$; if $a \ge 4$, then $(12-3a)_+ = 0$ and $\mathrm{CVaR}_{0.90}(\mathcal{L}(a)) = a/2$. Thus the unique minimizer is $a^\star_{0.90} = 4$.

For $\beta = 0.95$, the tail mass is 5%; by the quantile-integral definition of CVaR, since each nonzero scenario has probability 0.05, we have

$$\mathrm{CVaR}_{0.95}(\mathcal{L}(a)) = \max\{\mathcal{L}_1(a), \mathcal{L}_2(a)\} = \max\{(12-3a)_+,\ a\}.$$

For $a \le 4$, this equals $\max\{12 - 3a,\ a\}$, minimized when $12 - 3a = a$, i.e. at $a = 3$, yielding value 3. For $a \ge 4$, it equals a, which is minimized at $a = 4$ with value 4. Therefore the unique minimizer is $a^\star_{0.95} = 3$. ■

B.2. Proof of Proposition 3.4

We consider (15) and write

$$f_i(x_i) \triangleq \mathsf{E}[\sigma_i(x_i, p_T)]\,, \qquad f(x) \triangleq \mathsf{E}[\mathcal{L}(x, p_T)] = \sum_{i=1}^n f_i(x_i).$$

[10] Note that there is an atom at 0, so the 10% tail is not always uniquely determined: it consists of all scenarios with loss strictly greater than 0, together with an arbitrary selection of additional probability mass at loss 0 to reach total mass 0.10. In particular, we may choose a representative tail by taking scenarios 1 and 2 (and, when either loss is 0 for a given a, interpreting the scenario as part of the mass at the VaR level).

For each fixed $p_T \in \mathbb{R}^d$, equity $e_i(x_i, p_T)$ is affine in x_i, hence $\sigma_i(x_i, p_T) = (-e_i(x_i, p_T))_+$ is convex and piecewise affine in x_i. Therefore $f_i(x_i) = \mathsf{E}[\sigma_i(x_i, p_T)]$ is convex as an expectation of convex functions, and it is finite under Assumption 3.3.

Recall the definition of $\mathcal{Y}_i$ from (17) and define $\mathcal{Y} \triangleq \prod_{i=1}^n \mathcal{Y}_i$, so that the feasible set is equivalently represented as $\mathcal{X} = \{x \in \mathcal{Y} : \sum_{i=1}^n x_i = Q\}$. Each $\mathcal{Y}_i$ is nonempty, closed, and bounded, hence compact, and Assumption 3.1 implies $\mathcal{X} \neq \emptyset$. Since f is convex and finite on $\mathcal{Y}$ and $\mathcal{X}$ is compact, an optimizer of (15) exists.

We now turn to the characterization of the solution. Rewrite the primal problem as

$$\min_{x \in \mathcal{Y}} \sum_{i=1}^n f_i(x_i) \qquad \text{s.t.} \qquad \sum_{i=1}^n x_i = Q. \tag{61}$$

Introduce a multiplier $\lambda \in \mathbb{R}^d$ for the coupling constraint and form the partial Lagrangian

$$\widehat{\mathcal{L}}(x, \lambda) \triangleq \sum_{i=1}^n f_i(x_i) + \lambda^\top \Big(\sum_{i=1}^n x_i - Q \Big) = -\lambda^\top Q + \sum_{i=1}^n \big(f_i(x_i) + \lambda^\top x_i \big).$$

Define the dual function $g(\lambda) \triangleq \min_{x \in \mathcal{Y}} \widehat{\mathcal{L}}(x, \lambda)$ and the per-account value functions

$$\phi_i(\lambda) \triangleq \min_{x_i \in \mathcal{Y}_i} \{ f_i(x_i) + \lambda^\top x_i \}, \qquad i = 1, \dots, n,$$

so that

$$g(\lambda) = -\lambda^\top Q + \sum_{i=1}^n \phi_i(\lambda), \qquad \max_{\lambda \in \mathbb{R}^d} g(\lambda) = \max_{\lambda \in \mathbb{R}^d} \min_{x \in \mathcal{Y}} \widehat{\mathcal{L}}(x, \lambda).$$

Note that $\mathcal{Y}$ is compact so a minimizer always exists. Moreover, each $\phi_i(\cdot)$ is the pointwise minimum over $x_i \in \mathcal{Y}_i$ of affine functions of λ. As a result, the functions $\phi_i(\lambda)$ and $g(\lambda)$ are concave.

If for some k we have $Q^k = \sum_i l_i^k$ or $Q^k = \sum_i u_i^k$, then feasibility forces $x_i^k = l_i^k$ or $x_i^k = u_i^k$ for all i, so these coordinates are effectively equality constraints and may be treated as pinned. On the remaining coordinates where $Q^k \in (\sum_i l_i^k, \sum_i u_i^k)$, one can construct a feasible $\bar{x} \in \mathcal{X}$ with $l_i^k < \bar{x}_i^k < u_i^k$ whenever $u_i^k > l_i^k$ as follows. Since $\sum_{i=1}^n (u_i^k - l_i^k) > 0$ we can define

$$\alpha^k \triangleq \frac{Q^k - \sum_{i=1}^n l_i^k}{\sum_{i=1}^n (u_i^k - l_i^k)} \in (0, 1), \qquad \bar{x}_i^k \triangleq l_i^k + \alpha^k (u_i^k - l_i^k).$$

Hence $\mathcal{X}$ has nonempty relative interior in its affine hull, which yields a Slater-type constraint qualification for the polyhedral constraints. Consequently strong duality holds for (61), and there exists a primal–dual optimal pair $(x^\star, \lambda^\star)$ forming a saddle point of $\widehat{\mathcal{L}}$.

Recall the best-response correspondence $X_i(\lambda)$ from (19). Since $\mathcal{Y}$ is a product set and $\widehat{\mathcal{L}}$ separates across i, we have

$$\operatorname*{argmin}_{x \in \mathcal{Y}} \ \widehat{\mathcal{L}}(x, \lambda) = \prod_{i=1}^n X_i(\lambda).$$

Moreover, $\widehat{\mathcal{L}}$ is differentiable in λ with

$$\nabla_\lambda \widehat{\mathcal{L}}(x, \lambda) = \sum_{i=1}^n x_i - Q.$$

Because $g(\lambda) = \min_{x \in \mathcal{Y}} \widehat{\mathcal{L}}(x, \lambda)$ is concave, its superdifferential satisfies an extension of Danskin's theorem (cf. [Shapiro et al., 2021, Thm. 9.27]),

$$\partial g(\lambda) = \operatorname{conv} \Big\{ \sum_{i=1}^n x_i - Q : \ x_i \in X_i(\lambda) \ \forall i \Big\} \tag{62}$$

where $\mathrm{conv}\{\cdot\}$ denotes the convex hull. Since $\mathcal{Y}_i$ is compact and $x_i \mapsto f_i(x_i) + \lambda^\top x_i$ is convex and continuous, $X_i(\lambda)$ is nonempty and convex for each i. It is then straightforward to verify that the convex hull in (62) is redundant. Thus

$$\partial g(\lambda) = \Big\{ \sum_{i=1}^{n} x_i - Q : \ x_i \in X_i(\lambda) \ \forall i \Big\}.$$

Since $g(\cdot)$ is a finite concave function, $\lambda^\star$ is an unconstrained maximizer of $g(\cdot)$ if and only if it satisfies $0 \in \partial g(\lambda^\star)$. This is, in turn, equivalent to the existence of selections $x_i^\star \in X_i(\lambda^\star)$ such that $\sum_{i=1}^{n} x_i^\star = Q$.

Finally, by strong duality, $(x^\star, \lambda^\star)$ is primal–dual optimal if and only if it is a saddle point of $\widehat{\mathcal{L}}$, which holds if and only if $x^\star \in \mathrm{argmin}_{x\in\mathcal{Y}} \ \widehat{\mathcal{L}}(x, \lambda^\star)$ and $\sum_i x_i^\star = Q$. Using the product characterization of the minimizers yields exactly the stated condition

$$x_i^\star \in X_i(\lambda^\star) \ \text{ for all } i, \qquad \text{and} \qquad \sum_{i=1}^{n} x_i^\star = Q.$$

This proves Proposition 3.4. ∎

B.3. Properties of $\psi(\cdot)$

Proposition B.2. *Under Assumption 3.6, $\psi(z) \triangleq \mathsf{E}\big[(\epsilon z - 1)_+\big]$ is convex and continuously differentiable on $\mathbb{R}$, with derivative*

$$\psi'(z) = \mathsf{E}\big[\epsilon\, \mathbf{1}_{\{\epsilon z > 1\}}\big], \qquad z \in \mathbb{R}. \tag{63}$$

Moreover, ψ' is strictly increasing *on $\mathbb{R}$.*

Proof. For each fixed $x \in \mathbb{R}$, the map $z \mapsto (xz - 1)_+$ is convex. Therefore

$$\psi(z) = \mathsf{E}\big[(\epsilon z - 1)_+\big]$$

is convex as an expectation of convex functions.

Since under Assumption 3.6 ϵ has a strictly positive density φ_ϵ and $\mathsf{E}[|\epsilon|] < \infty$, we can write

$$\psi(z) = \int_{\mathbb{R}} (\ xz - 1\ )_+ \, \varphi_\epsilon(x)\, dx.$$

For $z > 0$, the condition $xz - 1 > 0$ is equivalent to $x > 1/z$, hence

$$\psi(z) = \int_{1/z}^{\infty} (xz - 1)\, \varphi_\epsilon(x)\, dx, \qquad z > 0. \tag{64}$$

For $z < 0$, the condition $xz - 1 > 0$ is equivalent to $x < 1/z$, hence

$$\psi(z) = \int_{-\infty}^{1/z} (xz - 1)\, \varphi_\epsilon(x)\, dx, \qquad z < 0. \tag{65}$$

Finally, $\psi(0) = \mathsf{E}[(-1)_+] = 0$.

We differentiate (64) for $z > 0$ using Leibniz' rule. The boundary term vanishes and we obtain

$$\psi'(z) = \int_{1/z}^{\infty} x\, \varphi_\epsilon(x)\, dx, \qquad z > 0.$$

Similarly, differentiating (65) for $z < 0$ yields

$$\psi'(z) = \int_{-\infty}^{1/z} x\, \varphi_\epsilon(x)\, dx, \qquad z < 0.$$

At $z = 0$, we compute the one-sided limits:

$$\lim_{z \downarrow 0} \psi'(z) = \lim_{z \downarrow 0} \int_{1/z}^{\infty} x\varphi_\epsilon(x)\, dx = 0, \qquad \lim_{z \uparrow 0} \psi'(z) = \lim_{z \uparrow 0} \int_{-\infty}^{1/z} x\varphi_\epsilon(x)\, dx = 0,$$

using $\mathsf{E}[|\epsilon|] < \infty$ and dominated convergence (as $1/z \to +\infty$ for $z \downarrow 0$ and $1/z \to -\infty$ for $z \uparrow 0$). By the right (resp. left) continuity of right (resp. left) derivatives of convex functions, we conclude that ψ' exists at 0 with $\psi'(0) = 0$. Altogether, we see that ψ' is continuous on $\mathbb{R}$.

To verify (63), note that for any $z > 0$,

$$\int_{1/z}^{\infty} x\, \varphi_\epsilon(x)\, dx = \int_{\mathbb{R}} x\, \mathbf{1}_{\{xz > 1\}} \varphi_\epsilon(x)\, dx,$$

and similarly for $z < 0$, so the integral expressions above are exactly $\mathsf{E}[\epsilon\, \mathbf{1}_{\{\epsilon z > 1\}}]$. For $z = 0$ the indicator is identically 0, so (63) holds for all $z \in \mathbb{R}$.

We finally show strict monotonicity of ψ'. Let $z < z'$.

Case 1: $0 < z < z'$. Then $1/z' < 1/z$ and

$$\psi'(z') - \psi'(z) = \int_{1/z'}^{\infty} x\varphi_\epsilon(x)\, dx - \int_{1/z}^{\infty} x\varphi_\epsilon(x)\, dx = \int_{1/z'}^{1/z} x\varphi_\epsilon(x)\, dx.$$

On $(1/z', 1/z)$ we have $x > 0$ and $\varphi_\epsilon(x) > 0$, hence the integral is strictly positive.

Case 2: $z < z' < 0$. Then $1/z' < 1/z < 0$ and

$$\psi'(z') - \psi'(z) = \int_{-\infty}^{1/z'} x\varphi_\epsilon(x)\, dx - \int_{-\infty}^{1/z} x\varphi_\epsilon(x)\, dx = -\int_{1/z'}^{1/z} x\varphi_\epsilon(x)\, dx.$$

On $(1/z', 1/z) \subset (-\infty, 0)$ we have $x < 0$ and $\varphi_\epsilon(x) > 0$, so $-\int_{1/z'}^{1/z} x\varphi_\epsilon(x)\, dx > 0$.

Case 3: $z < 0 < z'$. Then $\psi'(z) = \int_{-\infty}^{1/z} x\varphi_\epsilon(x)\, dx < 0$ (as the integrand is strictly negative), while $\psi'(z') = \int_{1/z'}^{\infty} x\varphi_\epsilon(x)\, dx > 0$, hence $\psi'(z') > \psi'(z)$.

If $z = 0 < z'$, then $\psi'(0) = 0 < \psi'(z')$. If $z < 0 = z'$, then $\psi'(z) < 0 = \psi'(0)$.

In all cases, $z < z'$ implies $\psi'(z') > \psi'(z)$, hence ψ' is strictly increasing on $\mathbb{R}$, by continuity. ■

B.4. Proof of Theorem 3.8

By Proposition B.2 ψ is strictly convex. Fix $\eta \in \mathbb{R}$ and define $h_\eta(z) \triangleq \psi(z) - \eta z$. Since ψ is strictly convex, h_η is strictly convex, so the minimizer of h_η over the interval $[\underline{\ell}_i, \overline{\ell}_i]$ is unique; this proves the first part of (i). The standard first-order/KKT conditions for minimizing a differentiable convex function over an interval lead to the cases

$$\begin{cases} \psi'(\ell_i^\star(\eta)) = \eta, & \underline{\ell}_i < \ell_i^\star(\eta) < \overline{\ell}_i, \\ \psi'(\underline{\ell}_i) \geq \eta, & \ell_i^\star(\eta) = \underline{\ell}_i, \\ \psi'(\overline{\ell}_i) \leq \eta, & \ell_i^\star(\eta) = \overline{\ell}_i. \end{cases}$$

Since ψ' is continuous and strictly increasing, this is equivalent to the clipped rule (29), proving (ii). By using that $(\psi')^{-1}$ must also be strictly increasing on the range of ψ', the same characterization implies that $\eta \mapsto \ell_i^\star(\eta)$ is continuous and nondecreasing, proving the second part of (i).

Next, define

$$H(\eta) \triangleq \sum_{i=1}^{n} E_i\, \ell_i^\star(\eta), \qquad R \triangleq v^\top \Big(\sum_{i=1}^{n} q_i - Q \Big).$$

By (i), H is continuous and nondecreasing. Moreover, by (29),

$$\lim_{\eta \to -\infty} H(\eta) = \sum_{i=1}^{n} E_i\, \underline{\ell}_i, \qquad \lim_{\eta \to +\infty} H(\eta) = \sum_{i=1}^{n} E_i\, \overline{\ell}_i.$$

In fact, these endpoint values are attained. If

$$\eta \le \min_{1 \le i \le n} \psi'(\underline{\ell}_i),$$

then (29) implies

$$\ell_i^\star(\eta) = \underline{\ell}_i \qquad \text{for all } i,$$

so that

$$H(\eta) = \sum_{i=1}^{n} E_i\, \underline{\ell}_i.$$

Similarly, if

$$\eta \ge \max_{1 \le i \le n} \psi'(\overline{\ell}_i),$$

then

$$\ell_i^\star(\eta) = \overline{\ell}_i \qquad \text{for all } i,$$

and hence

$$H(\eta) = \sum_{i=1}^{n} E_i\, \overline{\ell}_i.$$

Since $\mathcal{X} \neq \emptyset$ by Assumption 3.1, pick any feasible $x \in \mathcal{X}$. Then for each i, $\ell_i^{(v)}(x_i) \in [\underline{\ell}_i, \overline{\ell}_i]$, so

$$\sum_{i=1}^{n} E_i\, \underline{\ell}_i \ \le\ \sum_{i=1}^{n} E_i\, \ell_i^{(v)}(x_i) \ \le\ \sum_{i=1}^{n} E_i\, \overline{\ell}_i.$$

Using the identity $\sum_i E_i \ell_i^{(v)}(x_i) = \sum_i v^\top (q_i - x_i) = v^\top (\sum_i q_i - Q) = R$, we obtain

$$\sum_{i=1}^{n} E_i\, \underline{\ell}_i \ \le\ R \ \le\ \sum_{i=1}^{n} E_i\, \overline{\ell}_i.$$

By the intermediate value theorem, there exists $\eta^\star$ such that $H(\eta^\star) = R$, proving (iii).

Fix $\eta^\star$ satisfying (30) and suppose there exists $x^\star \in \mathcal{X}$ with $\ell_i^{(v)}(x_i^\star) = \ell_i^\star(\eta^\star)$ for all i. Set the shadow-price vector $\lambda^\star \triangleq \eta^\star v$ and recall (26). For each i and any $x_i \in \mathcal{Y}_i$ we have

$$\mathsf{E}[\sigma_i(x_i, p_T)] + (\lambda^\star)^\top x_i = E_i \psi(\ell_i^{(v)}(x_i)) + \eta^\star v^\top x_i = \eta^\star v^\top q_i + E_i\big(\psi(\ell_i^{(v)}(x_i)) - \eta^\star \ell_i^{(v)}(x_i)\big),$$

so minimizing $f_i(x_i) + (\lambda^\star)^\top x_i$ over $x_i \in \mathcal{Y}_i$ is equivalent to minimizing $\psi(z) - \eta^\star z$ over $z \in [\underline{\ell}_i, \overline{\ell}_i]$. Hence any $x_i \in \mathcal{Y}_i$ satisfying $\ell_i^{(v)}(x_i) = \ell_i^\star(\eta^\star)$ is a best response to $\lambda^\star$. In particular, $x_i^\star$ is a best response for each i, and the clearing constraint $\sum_i x_i^\star = Q$ holds because $x^\star \in \mathcal{X}$. Therefore, by Proposition 3.4, $x^\star$ is optimal for (15).

B.5. Example Where Water-Filling Fails

In general, the factor water-filling construction with clipping can fail to produce a solution because it enforces only the scalar constraint from (27),

$$\sum_i E_i \ell_i^{(v)}(x_i) = v^\top \left(\sum_{i=1}^n q_i - Q \right),$$

and cannot generally satisfy the *vector* clearing constraint $\sum_i x_i = Q$. We provide a simple illustrative counterexample here.

Example B.3. Suppose there are $d = 2$ assets and $n = 2$ accounts. Assume the single-factor model

$$p_T = p_\tau + \epsilon v, \qquad v = (1, 1),$$

where ϵ satisfies Assumption 3.6. We impose that the exchange has the ADL requirement

$$Q = (Q^1, Q^2) = (0.2, 0.8).$$

Take equal initial equities $E_1 = E_2 = 1$ and assume that account 1 holds only asset 1 and account 2 holds only asset 2:

$$q_1 = (1, 0), \qquad q_2 = (0, 1).$$

As a result, any feasible allocation $x = (x_1, x_2)$ satisfies

$$0 \le x_1^1 \le 1, \qquad 0 \le x_2^2 \le 1, \qquad x_1^2 = x_2^1 = 0.$$

Since we require $x_1 + x_2 = Q$, we conclude that the *unique* feasible allocation is

$$x_1^1 = Q^1 = 0.2, \qquad x_2^2 = Q^2 = 0.8.$$

In terms of factor leverage, this unique feasible allocation leads to

$$\ell_1^{(v)}(x_1) = 1 - x_1^1 = 0.8, \qquad \ell_2^{(v)}(x_2) = 1 - x_2^2 = 0.2, \qquad \ell_1^{(v)}(x_1) + \ell_2^{(v)}(x_2) = 1.$$

If this problem were instead approached via 1D water filling on the factor leverage, each account solves an identical 1D problem (since their equities and factor leverage constraints are the same),

$$\ell_i^\star(\eta) \in \operatorname*{argmin}_{z \in [0,1]} \left\{ \mathsf{E}\big[(\epsilon z - 1)_+\big] - \eta z \right\}, \quad i = 1, 2.$$

Therefore, by symmetry, the solution produces equal factor leverages

$$\ell_1^\star(\eta) = \ell_2^\star(\eta) = \ell^\star(\eta)$$

for every η. At the same time, the necessary condition (27) at an optimal $\eta^\star$ reads

$$\ell_1^\star(\eta^\star) + \ell_2^\star(\eta^\star) = 1.$$

Taken together, this yields $\ell^\star(\eta^\star) = 0.5$. But $\ell_1^{(v)}(x_1) = 0.5$ would imply $x_1^1 = 1 - \ell_1^{(v)}(x_1) = 0.5 \neq Q^1 = 0.2$, and similarly $\ell_2^{(v)}(x_2) = 0.5$ would imply $x_2^2 = 0.5 \neq Q^2 = 0.8$. Hence there is *no* $\eta^\star$ for which the water-filling exposures can be realized by an allocation satisfying $\sum_i x_i = Q$.

B.6. Sufficient Conditions for Water-Filling on Factor Leverage

This section describes a regime in which the expected-loss ADL problem (15) admits a generalized water-filling structure, but applied to *factor leverage* rather than gross leverage. The key simplification is that, under a one-factor price model, each account's expected shortfall depends on its post-ADL portfolio only through a scalar factor exposure.

Theorem 3.8 is stated as a verification result because, in general multi-asset ADL, the vector constraint $\sum_i x_i = Q$ need not allow one to realize an arbitrary collection of scalar targets $\{\ell_i^\star(\eta)\}_{i=1}^n$. A complementary observation is that, *at an optimum*, the dual multiplier $\lambda^\star$ can inherit a strong structure under a natural "overlap" condition across assets.

Recall the definitions

$$f_i(x_i) \triangleq \mathsf{E}[\sigma_i(x_i, p_T)]\,, \qquad f(x) \triangleq \mathsf{E}[\mathcal{L}(x, p_T)] = \sum_{i=1}^n f_i(x_i).$$

Under Assumptions 3.1–3.3, the expected-loss problem is convex and a relative-interior/polyhedral constraint qualification holds; see Appendix B.2. Consequently, the KKT conditions are necessary and sufficient for optimality.

Introduce multipliers $\lambda \in \mathbb{R}^d$ for the clearing constraint $\sum_i x_i = Q$, and for each (i,k) introduce $\mu_i^k \geq 0$ for $l_i^k - x_i^k \leq 0$ and $\nu_i^k \geq 0$ for $x_i^k - u_i^k \leq 0$. The Lagrangian is

$$\mathcal{L}(x, \lambda, \mu, \nu) = f(x) + \sum_{k=1}^d \lambda^k \Big(\sum_{i=1}^n x_i^k - Q^k \Big) + \sum_{i=1}^n \sum_{k=1}^d \mu_i^k (l_i^k - x_i^k) + \sum_{i=1}^n \sum_{k=1}^d \nu_i^k (x_i^k - u_i^k).$$

A primal–dual quadruple $(x^\star, \lambda^\star, \mu^\star, \nu^\star)$ is optimal if and only if:

$$\sum_{i=1}^n x_i^\star = Q, \qquad l_i^k \leq x_i^{k,\star} \leq u_i^k \quad \forall i, k, \tag{66}$$

$$\mu_i^{k,\star} \geq 0, \qquad \nu_i^{k,\star} \geq 0 \quad \forall i, k, \tag{67}$$

$$\mu_i^{k,\star}\,(l_i^k - x_i^{k,\star}) = 0, \qquad \nu_i^{k,\star}\,(x_i^{k,\star} - u_i^k) = 0 \quad \forall i, k, \tag{68}$$

$$0 \in \partial_{x_i} f_i(x_i^\star) + \lambda^\star - \mu_i^\star + \nu_i^\star \quad \forall i. \tag{69}$$

Assume now the single-factor model of Assumption 3.6. As in (12), under $p_T = p_\tau + \epsilon v$ one has

$$\sigma_i(x_i, p_T) = E_i(\epsilon\, \ell_i^{(v)}(x_i) - 1)_+,$$

so we may write

$$f_i(x_i) = E_i\, \psi(\ell_i^{(v)}(x_i)). \tag{70}$$

Moreover, ψ is continuously differentiable with

$$\psi'(z) = \mathsf{E}\Big[\epsilon\, \mathbf{1}_{\{\epsilon z > 1\}}\Big]\,, \qquad z \in \mathbb{R},$$

and ψ' is strictly increasing (see Proposition B.2). Using $\frac{\partial}{\partial x_i^k} \ell_i^{(v)}(x_i) = -v^k / E_i$ and (70), we obtain

$$\frac{\partial}{\partial x_i^k} f(x) = \frac{\partial}{\partial x_i^k} f_i(x_i) = -\, v^k\, \psi'(\ell_i^{(v)}(x_i)), \qquad \forall i, k. \tag{71}$$

Substituting (71) into (69) yields the single-factor stationarity condition

$$0 = -v^k \psi'(\ell_i^{(v)}(x_i^\star)) + \lambda^{k,\star} - \mu_i^{k,\star} + \nu_i^{k,\star}, \qquad \forall i, k. \tag{72}$$

In particular, whenever $v^k \neq 0$, by complementary slackness

$$l_i^k < x_i^{k,\star} < u_i^k \quad \Longrightarrow \quad \psi'(\ell_i^{(v)}(x_i^\star)) = \frac{\lambda^{k,\star}}{v^k}. \tag{73}$$

Let

$$K \triangleq \{k \in \{1, \dots, d\} : Q^k \neq 0,\ v^k \neq 0\}$$

denote the set of *active* assets whose reductions affect factor exposure. Given an optimal solution $x^\star$, define the interior coordinates of account i by

$$F_i \triangleq \{k \in K :\ l_i^k < x_i^{k,\star} < u_i^k\}, \qquad i = 1, \dots, n,$$

and let $K_{\text{int}}(x^\star) \triangleq \cup_{i=1}^n F_i$ be the set of active coordinates that are strictly interior for at least one account.

Definition B.4 (Interior-coverage graph). *The* interior-coverage graph *is the simple undirected graph*

$$G(x^\star) \triangleq (K_{\text{int}}(x^\star), E),$$

where

$$E \triangleq \Big\{\{k, k'\} \subseteq K_{\text{int}}(x^\star) : k \neq k' \text{ and } \exists\, i \text{ s.t. } \{k, k'\} \subseteq F_i\Big\}.$$

Equivalently, there is an edge between k and k' if and only if there exists an account whose reduction is strictly interior in both coordinates.

Proposition B.5. *Let $(x^\star, \lambda^\star, \mu^\star, \nu^\star)$ satisfy the KKT system (66)–(69). If $K_{\text{int}}(x^\star) \neq \emptyset$ and the interior-coverage graph $G(x^\star)$ is connected*[11]*, then there exists $\eta^\star \in \mathbb{R}$ such that*

$$\lambda^{k,\star} = \eta^\star v^k, \qquad \forall k \in K_{\text{int}}(x^\star).$$

Moreover, for any i and any $k \in F_i$, one has $\eta^\star = \psi'(\ell_i^{(v)}(x_i^\star))$.

Proof. Fix any i and any $k \in F_i$. Since $k \in K$, we have $v^k \neq 0$, and (73) gives

$$\frac{\lambda^{k,\star}}{v^k} = \psi'(\ell_i^{(v)}(x_i^\star)).$$

If $k, k' \in F_i$ for the same account i, then the same identity holds for k' and hence

$$\frac{\lambda^{k,\star}}{v^k} = \frac{\lambda^{k',\star}}{v^{k'}}. \tag{74}$$

Thus, along any edge $\{k, k'\} \in E$, the ratios $\lambda^{k,\star}/v^k$ and $\lambda^{k',\star}/v^{k'}$ agree.

If $|K_{\text{int}}(x^\star)| = 1$, the conclusion follows immediately by setting $\eta^\star \triangleq \lambda^{k,\star}/v^k$ on the unique vertex k. Otherwise, connectivity implies that for any $k \in K_{\text{int}}(x^\star)$ there exists a path from a fixed reference vertex k_0 to k, and the edge-by-edge identity (74) propagates along the path to yield $\lambda^{k,\star}/v^k = \lambda^{k_0,\star}/v^{k_0}$ for all k. Setting $\eta^\star \triangleq \lambda^{k_0,\star}/v^{k_0}$ gives $\lambda^{k,\star} = \eta^\star v^k$ on $K_{\text{int}}(x^\star)$, and the identification $\eta^\star = \psi'(\ell_i^{(v)}(x_i^\star))$ for $k \in F_i$ follows from the interior stationarity equality above. ■

Remark B.6. Proposition B.5 is an *ex post* structural statement: if the optimal allocation features enough strictly-interior overlap across active assets so $K = K_{\text{int}}(x^\star)$ and $G(x^\star)$ is connected, then the optimal shadow prices for the active assets must align with the factor loading. In large venues, such overlap is more plausible when many large accounts hold genuinely cross-asset portfolios, creating interior reductions in multiple assets simultaneously.

[11] If $K_{\text{int}}(x^\star)$ is a singleton it is trivially connected.

The alignment $\lambda^\star = \eta^\star v$ on interior active coordinates is the structural condition under which the expected-loss ADL problem reduces to a one-dimensional (clipped) water-filling rule in factor-leverage space.

Definition B.7 (Connected Partial Deleveraging). *We say an optimal allocation $x^\star$ satisfies the* connected partial deleveraging condition *if $K_{\rm int}(x^\star) = K$ and the interior-coverage graph $G(x^\star)$ is connected.*

Proposition B.8. *Let $(x^\star, \lambda^\star, \mu^\star, \nu^\star)$ satisfy the KKT conditions for (15) and assume $x^\star$ satisfies the connected partial deleveraging condition of Definition B.7. Suppose further that if $Q^k \neq 0$ then $v^k \neq 0$. Then there exists $\eta^\star \in \mathbb{R}$ such that for every account i,*

$$\ell_i^{(v)}(x_i^\star) \in \operatorname*{argmin}_{z \in [\underline{\ell}_i, \bar{\ell}_i]} \{\psi(z) - \eta^\star z\},$$

equivalently,

$$\ell_i^{(v)}(x_i^\star) = \ell_i^\star(\eta^\star),$$

where $\ell_i^\star(\eta)$ is the clipped factor water-filling target from Theorem 3.8.

Proof. By Proposition B.5 and the assumptions $K_{\rm int}(x^\star) = K$ and connectivity of $G(x^\star)$, there exists $\eta^\star \in \mathbb{R}$ such that

$$\lambda^{k,\star} = \eta^\star v^k, \qquad \forall k \in K. \tag{75}$$

Fix an account i. By Proposition 3.4, $x_i^\star$ minimizes the per-account Lagrangian

$$\min_{x_i \in \mathcal{Y}_i} \Big\{ f_i(x_i) + (\lambda^\star)^\top x_i \Big\}. \tag{76}$$

Under the single-factor model,

$$f_i(x_i) = E_i\, \psi(\ell_i^{(v)}(x_i)), \qquad \ell_i^{(v)}(x_i) = \frac{v^\top (q_i - x_i)}{E_i}.$$

We partition coordinates into two disjoint sets:

$$\mathcal{Z} \triangleq \{k : \ Q^k = 0\}, \qquad K = \{k : \ Q^k \neq 0\}.$$

Note that under the standing assumptions K takes this simplified form. By the directional bounds (11), $k \in \mathcal{Z}$ implies $x_i^k = 0$ for all $x_i \in \mathcal{Y}_i$, so $\sum_{k \in \mathcal{Z}} \lambda^{k,\star} x_i^k \equiv 0$ and can be dropped from (76). Thus, up to additive constants independent of x_i, the per-account objective in (76) reduces to

$$E_i\, \psi(\ell_i^{(v)}(x_i)) + \sum_{k \in K} \lambda^{k,\star} x_i^k. \tag{77}$$

Using (75), we have

$$\sum_{k \in K} \lambda^{k,\star} x_i^k = \eta^\star \sum_{k \in K} v^k x_i^k.$$

Moreover, since $x_i^k = 0$ for all $k \in \mathcal{Z}$ we obtain

$$\sum_{k \in K} v^k x_i^k = \sum_{k=1}^{d} v^k x_i^k = v^\top x_i.$$

Finally, by definition of $\ell_i^{(v)}$,

$$v^\top x_i = v^\top q_i - E_i\, \ell_i^{(v)}(x_i).$$

Substituting these identities into (77) yields, up to the additive constant $\eta^\star v^\top q_i$,

$$E_i\Big(\psi(\ell_i^{(v)}(x_i)) - \eta^\star \ell_i^{(v)}(x_i)\Big).$$

Therefore minimizing (76) over $x_i \in \mathcal{Y}_i$ is equivalent to minimizing

$$\psi(z) - \eta^\star z \qquad \text{over} \qquad z = \ell_i^{(v)}(x_i) \in [\underline{\ell}_i, \overline{\ell}_i].$$

Hence

$$\ell_i^{(v)}(x_i^\star) \in \operatorname*{argmin}_{z\in[\underline{\ell}_i,\overline{\ell}_i]} \{\psi(z) - \eta^\star z\}.$$

By uniqueness of $\ell_i^\star(\eta^\star)$ (Theorem 3.8), we conclude that $\ell_i^{(v)}(x_i^\star) = \ell_i^\star(\eta^\star)$ for all i. ■

The need for verification disappears when ADL is required in only one asset, because the vector clearing constraint reduces to a single scalar equation.

Lemma B.9. *Assume $Q^k = 0$ for $k \neq k_0$ and $v^{k_0} \neq 0$, so $x_i^k = 0$ for all $k \neq k_0$. Then for each account i,*

$$\ell_i^{(v)}(x_i) = \frac{1}{E_i}\Big(\sum_{k\neq k_0} v^k q_i^k\Big) + \frac{v^{k_0}}{E_i}(q_i^{k_0} - x_i^{k_0}),$$

and the map $x_i^{k_0} \mapsto \ell_i^{(v)}(x_i)$ is an affine bijection from $[l_i^{k_0}, u_i^{k_0}]$ onto $[\underline{\ell}_i, \overline{\ell}_i]$. Its inverse is

$$x_i^{k_0} = q_i^{k_0} - \frac{E_i\,\ell - \sum_{k\neq k_0} v^k q_i^k}{v^{k_0}}, \qquad \ell \in [\underline{\ell}_i, \overline{\ell}_i]. \tag{78}$$

Proof. With $x_i^k = 0$ for $k \neq k_0$, the displayed affine form follows directly from the definition $\ell_i^{(v)}(x_i) = v^\top(q_i - x_i)/E_i$. Since $v^{k_0} \neq 0$, the coefficient of $x_i^{k_0}$ is nonzero, so the map is injective and maps the compact interval $[l_i^{k_0}, u_i^{k_0}]$ onto a compact interval, which must equal $[\underline{\ell}_i, \overline{\ell}_i]$ by definition. Solving the affine relation for $x_i^{k_0}$ yields (78). ■

The following special case corresponds to Theorem 3.9 in the main text.

Corollary B.10. *Assume $Q^k = 0$ for $k \neq k_0$ and $v^{k_0} \neq 0$, so $x_i^k = 0$ for all $k \neq k_0$. Let $\eta^\star$ satisfy the budget equation (30), and define targets $\ell_i^\star(\eta^\star)$ by (28). Define an allocation $x^\star$ by setting $x_i^{\star,k} = 0$ for $k \neq k_0$ and*

$$x_i^{\star,k_0} \;\triangleq\; q_i^{k_0} - \frac{E_i\,\ell_i^\star(\eta^\star) - \sum_{k\neq k_0} v^k q_i^k}{v^{k_0}}, \qquad i = 1, \dots, n.$$

Then $x^\star \in \mathcal{X}$ and $x^\star$ is optimal for (15).

Proof. By Lemma B.9, each target $\ell_i^\star(\eta^\star) \in [\underline{\ell}_i, \overline{\ell}_i]$ corresponds to a unique $x_i^{\star,k_0} \in [l_i^{k_0}, u_i^{k_0}]$, so the constructed $x^\star$ satisfies the per-account constraints.

It remains to verify the clearing constraint $\sum_i x_i^{\star,k_0} = Q^{k_0}$. Using the definition of $\ell_i^{(v)}$ and that $x_i^{\star,k} = 0$ for $k \neq k_0$, we compute

$$\sum_{i=1}^n E_i\,\ell_i^{(v)}(x_i^\star) = \sum_{i=1}^n \sum_{k\neq k_0} v^k q_i^k + v^{k_0}\Big(\sum_{i=1}^n q_i^{k_0} - \sum_{i=1}^n x_i^{\star,k_0}\Big).$$

On the other hand, since $Q^k = 0$ for $k \neq k_0$,

$$v^\top\Big(\sum_{i=1}^{n} q_i - Q\Big) = \sum_{i=1}^{n}\sum_{k\neq k_0} v^k q_i^k + v^{k_0}\Big(\sum_{i=1}^{n} q_i^{k_0} - Q^{k_0}\Big).$$

Subtracting these equalities and using (30) gives $v^{k_0}\big(\sum_i x_i^{\star,k_0} - Q^{k_0}\big) = 0$, hence $\sum_i x_i^{\star,k_0} = Q^{k_0}$ because $v^{k_0} \neq 0$. Thus $x^\star \in \mathcal{X}$.

Finally, by construction $\ell_i^{(v)}(x_i^\star) = \ell_i^\star(\eta^\star)$ for all i, so Theorem 3.8 implies that $x^\star$ is globally optimal. ■

Remark B.11. Corollary B.10 formalizes the sense in which the single-asset case admits a true water-filling solution in factor-leverage space: the scalar budget equation (30) is then equivalent to the (scalar) clearing constraint, so the water-filling targets are automatically implementable. In contrast, when $|K| > 1$, implementability can fail because a multi-dimensional clearing constraint cannot generally be enforced by matching only one scalar target.

B.7. Single-Factor Analysis for More General Risk Measures

This section briefly sketches how the single-factor analysis for the expected loss extends to more general risk measures. The crucial observation is that, in the single-factor case, comonotonicity leads to separability across accounts, even for risk measures such as CVaR_β.

Assume the single-factor model $p_T = p_\tau + \epsilon v$ and recall that

$$e_i(x_i, p_T) = E_i(1 - \epsilon\,\ell_i^{(v)}(x_i)), \qquad \ell_i^{(v)}(x_i) \triangleq \frac{v^\top(q_i - x_i)}{E_i}.$$

We restrict attention to accounts with nonnegative factor exposure. Namely, we impose $\ell_i^{(v)}(x_i) \geq 0$ as an additional feasibility constraint for allocations x and restrict to accounts with nonnegative factor exposure before the ADL event. We assume that, within this subset of accounts, the modified feasible set

$$\mathcal{X} \cap \{x:\ \ell_i^{(v)}(x_i) \geq 0,\ \forall i\}$$

is nonempty. Then $\sigma_i(x_i, p_T)$ is a nondecreasing function of the common scalar factor ϵ for each i, and hence the collection $\{\sigma_i(x_i, p_T)\}_{i=1}^n$ is *comonotone*.

For any comonotone-additive risk functional ρ, including spectral risk measures such as $\rho(\cdot) = \mathsf{E}[\cdot]$ or $\rho(\cdot) = \mathrm{CVaR}_\beta(\cdot)$, we then have

$$\rho\left(\sum_{i=1}^{n} \sigma_i(x_i, p_T)\right) = \sum_{i=1}^{n} \rho(\sigma_i(x_i, p_T))\,.$$

Defining the one-dimensional factor-risk function

$$\psi_{i,\rho}(\ell) \triangleq \rho(E_i(\epsilon\ell - 1)_+)\,, \qquad \ell \geq 0,$$

the exchange objective becomes separable across accounts,

$$\rho(\mathcal{L}(x, p_T)) = \sum_{i=1}^{n} \psi_{i,\rho}(\ell_i^{(v)}(x_i)).$$

This is analogous to (16), but now driven by comonotonicity rather than linearity of the expectation.

After this observation, the analysis becomes analogous to the expected loss in the single-factor model. As in (27), the clearing constraint implies a fixed equity-weighted aggregate factor exposure,

$$\sum_{i=1}^{n} E_i\, \ell_i^{(v)}(x_i) = \sum_{i=1}^{n} v^\top (q_i - x_i) = v^\top \Big(\sum_{i=1}^{n} q_i - Q \Big), \qquad \forall x \in \mathcal{X}.$$

If the KKT multiplier satisfies $\lambda^\star = \eta^\star v$ (as implied, for instance, by an analogous interior-coverage[12] connectivity condition; cf. Proposition B.5), the corresponding per-account Lagrangian subproblem (cf. (18) for the expected loss) depends on x_i only through $\ell_i^{(v)}(x_i)$ and reduces to the one-dimensional program

$$\min_{\ell \in [\underline{\ell}_i, \overline{\ell}_i] \cap \mathbb{R}_+} \{\psi_{i,\rho}(\ell) - \eta^\star E_i\, \ell\},$$

where $[\underline{\ell}_i, \overline{\ell}_i] = \{\ell_i^{(v)}(x_i) : x_i \in \mathcal{Y}_i\}$ is the feasible factor-leverage interval. If $\psi_{i,\rho}$ is differentiable and strictly convex on $\mathbb{R}_+$, the induced optimizer satisfies the clipped water-filling rule

$$\ell_i^\star(\eta^\star) = \begin{cases} \underline{\ell}_i^+, & \eta^\star \le \psi'_{i,\rho}(\underline{\ell}_i^+)/E_i, \\ (\psi'_{i,\rho})^{-1}(\eta^\star E_i), & \psi'_{i,\rho}(\underline{\ell}_i^+)/E_i < \eta^\star < \psi'_{i,\rho}(\overline{\ell}_i)/E_i, \\ \overline{\ell}_i, & \eta^\star \ge \psi'_{i,\rho}(\overline{\ell}_i)/E_i, \end{cases}$$

where $\underline{\ell}_i^+ \triangleq \max\{\underline{\ell}_i, 0\}$. As was the case for the expected loss, even when the targets $\{\ell_i^\star(\eta)\}$ satisfy the scalar budget

$$\sum_{i=1}^{n} E_i\, \ell_i^\star(\eta) = v^\top \Big(\sum_{i=1}^{n} q_i - Q \Big),$$

a joint selection of allocations $x_i \in \mathcal{Y}_i$ achieving these targets and satisfying $\sum_{i=1}^n x_i = Q$ need not exist in general because the clearing constraint $\sum_{i=1}^n x_i = Q$ is vector-valued. Nonetheless, as in Theorem 3.9, when only a single asset is deleveraged, the same argument shows that the "clipped water-filling" solution in factor-leverage space is attainable by admissible reductions satisfying the clearing constraint and is therefore optimal.

B.8. Numerical Example

This appendix details the one-factor approximation of the bivariate GBM model in Section 3.5. Note that we take v from the price increment $\Delta p \triangleq p_T - p_\tau$ of the GBM model, rather than from log returns, so that v has the correct dollar units in the additive model $p_T = p_\tau + \epsilon v$. Under the model stated in Section 3.5, the covariance matrix of Δp is

$$\begin{aligned}\Sigma_{\Delta p} &= \begin{pmatrix} (p_\tau^{\mathrm{BTC}})^2 \Big(e^{(\sigma_{\mathrm{ann}}^{\mathrm{BTC}})^2 \Delta} - 1 \Big) & p_\tau^{\mathrm{BTC}} p_\tau^{\mathrm{ETH}} \Big(e^{\rho\, \sigma_{\mathrm{ann}}^{\mathrm{BTC}} \sigma_{\mathrm{ann}}^{\mathrm{ETH}} \Delta} - 1 \Big) \\ p_\tau^{\mathrm{BTC}} p_\tau^{\mathrm{ETH}} \Big(e^{\rho\, \sigma_{\mathrm{ann}}^{\mathrm{BTC}} \sigma_{\mathrm{ann}}^{\mathrm{ETH}} \Delta} - 1 \Big) & (p_\tau^{\mathrm{ETH}})^2 \Big(e^{(\sigma_{\mathrm{ann}}^{\mathrm{ETH}})^2 \Delta} - 1 \Big) \end{pmatrix} \\ &\approx \begin{pmatrix} 44{,}494{,}130.91 & 1{,}341{,}048.70 \\ 1{,}341{,}048.70 & 56{,}064.46 \end{pmatrix}.\end{aligned}$$

We then compute its principal eigenpair numerically via the symmetric eigendecomposition of $\Sigma_{\Delta p}$, normalizing the eigenvector to unit Euclidean norm and choosing the sign so that its BTC loading is positive. This gives

$$\lambda_1 \approx 44{,}534{,}564.19, \qquad u_1 \approx \begin{pmatrix} 0.99954578 \\ 0.03013680 \end{pmatrix}, \qquad v \triangleq \sqrt{\lambda_1}\, u_1 \approx \begin{pmatrix} 6670.3910 \\ 201.1156 \end{pmatrix}.$$

[12] In the present setting, an "interior" condition must also account for the additional constraint $\ell_i^{(v)}(x_i) \ge 0$.

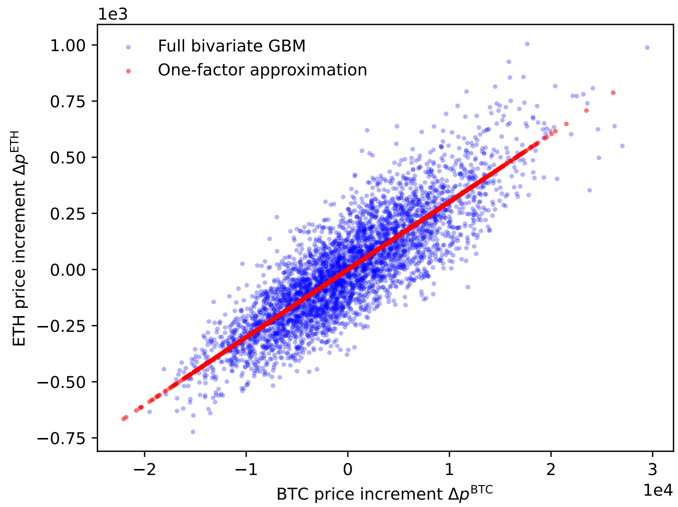

Figure 9: Samples of price increments under the bivariate GBM model (blue) and the one-factor approximation collapsing the bivariate distribution onto the dominant covariance direction (red).

C. Details for the Exchange-Scale Problem

C.1. ADMM: Algorithm and Convergence

Algorithm 1 summarizes the abstract two-block ADMM iteration. Section 4.2 explains how this abstract scheme is instantiated on the exchange-scale instance. The box is intentionally stated at the theory level: $\varepsilon_{\text{stop}}$ denotes a generic residual tolerance, while ρ and the iteration cap K are algorithmic inputs whose concrete values depend on the application. In the exchange-scale implementation we use more detailed empirical calibration for the penalty parameter and stopping rule; see Appendix C.4.

Require: aggregate target $Q \in \mathbb{R}^d$; per-account boxes $\{\mathcal{Y}_i\}_{i=1}^n$; SAA scenarios $\{\Delta p^{(s)}\}_{s=1}^S$; penalty $\rho > 0$; tolerance $\varepsilon_{\text{stop}} > 0$; iteration cap K.
Ensure: accountwise iterate $x^K \in (\mathbb{R}^d)^n$, projected market-clearing iterate $z^K \in (\mathbb{R}^d)^n$, and primal/dual residuals (r^K, s^K).

Initialize z^0 (zero initialization or warm start), and set $u^0 = 0$.
for $k = 0, 1, \ldots, K-1$ **do**
  for $i = 1, \ldots, n$ **do in parallel**
    $c_i^k \leftarrow z_i^k - u_i^k$;
    Solve per-account proximal subproblem (23).
  end for
  $z^{k+1} \leftarrow \Pi_{\mathcal{C}}(x^{k+1} + u^k)$ ▷ project onto the clearing set
  $u^{k+1} \leftarrow u^k + x^{k+1} - z^{k+1}$
  $r^{k+1} \leftarrow x^{k+1} - z^{k+1}$ ▷ primal residual
  $s^{k+1} \leftarrow \rho(z^{k+1} - z^k)$ ▷ dual residual
  if $\|r^{k+1}\| \le \varepsilon_{\text{stop}}$ **and** $\|s^{k+1}\| \le \varepsilon_{\text{stop}}$ **then**
    break.
  end if
end for
return $x^\star \leftarrow x^{k+1}$, $z^\star \leftarrow z^{k+1}$
($x^\star$ satisfies the per-account box constraints, $z^\star$ satisfies the market-clearing constraint $\sum_i z_i^\star = Q$ by construction, and primal feasibility is recovered up to the residual $\|x^\star - z^\star\|$.)

Algorithm 1: Abstract two-block ADMM with Sample-Average Approximation

Proposition C.1. *Fix $\rho > 0$ and suppose the feasible set is nonempty. Then the iterates generated by (23)–(25) satisfy*

$$\|x^k - z^k\| \to 0, \qquad \|\rho(z^k - z^{k-1})\| \to 0,$$

and

$$\sum_{i=1}^n f_i(x_i^k) + I_{\mathcal{C}}(z^k) \to L_S^\star,$$

where $L_S^\star$ denotes the optimal value of the SAA problem. Moreover, ρu^k converges to a dual optimal point.

Proof. Writing

$$F(x) \triangleq \sum_{i=1}^n f_i(x_i), \qquad G(z) \triangleq I_{\mathcal{C}}(z),$$

the problem takes the standard two-block ADMM form

$$\min_{x,z} \; F(x) + G(z) \qquad \text{subject to} \qquad x - z = 0.$$

Each f_i is closed, proper, and convex, since it is the sum of a finite convex piecewise-linear loss term and the indicator of the corresponding box. Hence F is closed, proper, and convex. The set $\mathcal{C}$ is a nonempty closed affine set, so $G = I_{\mathcal{C}}$ is closed, proper, and convex as well.

The feasible set is contained in the product of the per-account boxes and is therefore compact. Since the objective is lower semicontinuous, the primal optimum is attained. Because the problem is a finite-dimensional convex program with polyhedral objective and affine constraints, a saddle point of the unaugmented Lagrangian exists. Therefore the assumptions of the standard two-block ADMM convergence theorem in Boyd et al. [2011, Section 3.2] are satisfied, and the stated residual, objective, and dual convergence follow. ∎

C.2. Data reconstruction

For each wave time τ_j, the empirical input is constructed in four steps. The cleaned reconstructed wave-level states used in the numerical analysis are available in the accompanying GitHub repository.

First, we define the raw ADL wave directly from the Hyperliquid event tape by restricting to trades whose direction is Auto-Deleveraging in the corresponding event window. Throughout the empirical analysis, we work on a 75-coin core universe: all ADL coins are retained, and non-ADL coins are included only if their aggregate gross notional exceeds USD 100,000, so that the covariance calibration and scenario generation remain supported on names with sufficient trading activity. This determines the raw realized ADL users, the raw set of affected coins, and the raw wave notional.

Second, we reconstruct the pre-event account state strictly before τ_j from Allium account snapshots. For each retained user–coin pair, we take the most recent snapshot strictly before τ_j as the baseline and roll it forward using the full trade tape. Realized PnL accrued during this roll-forward is added back to the account margin variable. User–coin pairs without a strict pre-event baseline are excluded from the reconstructed state.

Third, we restrict the raw ADL flow in the wave window to the retained reconstructed user–coin pairs and aggregate the retained fills. This yields the realized Hyperliquid allocation $x^{\text{Hyperliquid},(j)}$ and the corresponding target vector

$$Q^{(j)} \triangleq \sum_i x_i^{\text{Hyperliquid},(j)}.$$

This is the step at which the optimization target becomes smaller than the raw event total.

Fourth, we restrict the optimization problem to the active asset set $\operatorname{supp}(Q^{(j)})$ and apply the directional-box reduction. Accounts whose feasible interval has zero width on every active asset are removed before solving. The resulting active-account instance is the final optimization problem used in the empirical comparison. Concretely, the solver is run on a $11{,}026 \times 37$ instance at τ_1 and on a $7{,}185 \times 4$ instance at τ_3, rather than on the full reconstructed ambient state.

Table 4 summarizes these reductions for the first wave τ_1 and for the third wave τ_3.

Two distinctions are worth emphasizing. First, the row "ADL users with reconstructable pre-event state" is a user-level coverage statistic, whereas the row "Users with positive retained ADL in $x^{\text{Hyperliquid},(j)}$" reflects the stricter pair-level filtering applied when constructing the retained realized allocation. Second, the row "Active optimization accounts" refers to the final directional-box reduction used before solving. It contains all accounts with positive retained ADL in $x^{\text{Hyperliquid},(j)}$, together with additional reconstructable accounts that were not delevered in the realized event but still hold positions on $\operatorname{supp}(Q^{(j)})$ that can be delevered under the directional box constraints. This is why the active optimization set is larger than the support of $x^{\text{Hyperliquid},(j)}$, while still being smaller than the full reconstructed pre-event state.

C.3. Covariance Estimation

ADL events occur in precisely the market conditions where volatility clustering is most severe: sharp directional moves concentrated in short windows. Crypto assets are known to exhibit pronounced

Table 4: Data reconstruction summary for the first wave τ_1 and the third wave τ_3. The reconstructed pre-event state is built on the full 75-asset modeling universe, while the wave-specific retained target $Q^{(j)}$ is supported only on the coins that remain active after matching the raw ADL flow to reconstructable user–coin pairs.

Quantity	τ_1	τ_3
Raw ADL users in wave	8,048	2,384
Raw ADL coins in wave	37	4
Raw ADL notional	1.7686×10^8	3.8661×10^8
ADL users with reconstructable pre-event state	6,956	1,961
Accounts in reconstructed pre-event state	60,324	59,978
Users with positive retained ADL in $x^{\text{Hyperliquid},(j)}$	6,820	1,814
Retained target coins $\lvert\text{supp}(Q^{(j)})\rvert$	37	4
Retained target notional $\sum_k p_{\tau,k}\lvert Q_k^{(j)}\rvert$	1.5390×10^8	3.4606×10^8
Active optimization accounts	11,026	7,185

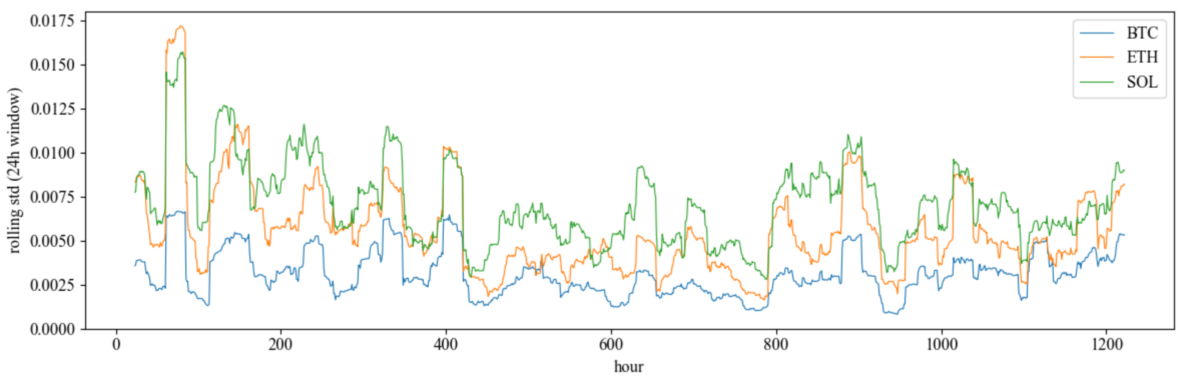

Figure 10: Rolling-window volatility estimates for selected assets. Pronounced clustering motivates conditioning on the volatility level at the ADL trigger time τ rather than using an unconditional estimate.

time-varying volatility, and conditioning on the current volatility level at the trigger time τ is therefore essential for a meaningful risk assessment. An unconditional covariance estimate would understate risk in stressed regimes and overstate it in calm ones. Figure 10 illustrates this clustering for the assets in the ADL universe: rolling volatility estimates vary substantially over time, with spikes that coincide with periods of market stress. This motivates the Exponentially Weighted Moving Average (EWMA)-standardized decomposition below.

EWMA-standardized covariance model. We decompose hourly log-returns as $r_t = D_t \varepsilon_t$, where D_t is a diagonal matrix of time-varying asset volatilities and ε_t is the standardized residual vector, whose covariance $\Sigma_\varepsilon = \text{Cov}(\varepsilon_t)$ is approximately stationary over time. The conditional 1-hour log-return covariance at the ADL trigger time τ is then

$$\Sigma_{\tau,1\text{h}}^{\log} = D_\tau \Sigma_\varepsilon D_\tau, \tag{79}$$

where D_τ is the EWMA volatility diagonal at time τ [Engle, 1982]. This separates the two empirical ingredients relevant for ADL risk: the slowly varying cross-sectional dependence structure Σ_ε, and the rapidly changing volatility level D_τ that must be captured at the trigger time. The matrix

Σ_ε is approximated by a diagonal-plus-rank-one structure, as described in the empirical support paragraph below and used in the price model inputs of Section 4.1.

Since hourly return autocorrelation is empirically weak, we scale the hourly covariance linearly across horizons. Expressed in annualized GBM units, this gives

$$\Sigma^{\log}_{\tau,\Delta} = \Delta\,(24 \times 365)\,\Sigma^{\log}_{\tau,1\mathrm{h}},$$

which is consistent with the empirical absence of meaningful serial correlation in hourly returns once the covariance matrix has been annualized. Throughout the exchange-scale experiments, we set $\Delta = 10/365$, corresponding to a ten-day liquidation horizon under annualized inputs.

Price simulation models. To evaluate the SAA objective (21), we sample S terminal price vectors $p_T^{(s)}$ from a multivariate geometric Brownian motion calibrated to the horizon-Δ log-return covariance matrix $\Sigma^{\log}_{\tau,\Delta}$. Writing

$$X \triangleq \log(p_T/p_\tau),$$

we model the horizon-Δ log-return vector as

$$X \sim \mathcal{N}\Big(-\tfrac{1}{2}\,\mathrm{diag}(\Sigma^{\log}_{\tau,\Delta}),\ \Sigma^{\log}_{\tau,\Delta}\Big),$$

where $\mathrm{diag}(\Sigma^{\log}_{\tau,\Delta}) \in \mathbb{R}^d$ denotes the vector of diagonal entries of $\Sigma^{\log}_{\tau,\Delta}$. Equivalently, if L_τ is any matrix square root satisfying

$$L_\tau L_\tau^\top = \Sigma^{\log}_{\tau,\Delta},$$

and $Z \sim \mathcal{N}(0, I_d)$, then

$$X = -\tfrac{1}{2}\,\mathrm{diag}(\Sigma^{\log}_{\tau,\Delta}) + L_\tau Z, \qquad p_T = p_\tau \odot \exp(X), \tag{80}$$

where $\odot$ denotes componentwise multiplication. This representation makes the cross-asset covariance structure explicit.

For implementation, the same covariance matrix may equivalently be written in volatility-correlation form as

$$\Sigma^{\log}_{\tau,\Delta} = \Delta\ \mathrm{diag}(\sigma_\tau)\,R_\tau\ \mathrm{diag}(\sigma_\tau),$$

where $\sigma_\tau \in \mathbb{R}^d_+$ is the vector of annualized conditional volatilities at time τ and R_τ is the corresponding correlation matrix.

The water-filling algorithm additionally requires a factor direction $v \in \mathbb{R}^d$ defining the dominant covariance mode. This is extracted from the same GBM calibration. Defining the horizon-Δ dollar price drop by $\Delta p \triangleq p_\tau - p_T$, we have the first-order approximation

$$\Delta p \approx -\,\mathrm{diag}(p_\tau)\,X.$$

This yields the dollar price-drop covariance

$$\Sigma^{\Delta p}_{\tau,\Delta} \approx \mathrm{diag}(p_\tau)\,\Sigma^{\log}_{\tau,\Delta}\ \mathrm{diag}(p_\tau), \tag{81}$$

and the leading eigenpair (λ_1, u_1) of $\Sigma^{\Delta p}_{\tau,\Delta}$ gives $v \triangleq \sqrt{\lambda_1}\,u_1$. This vector enters the water-filling solver as the one-factor direction defining factor leverage (cf. Definition 3.7).

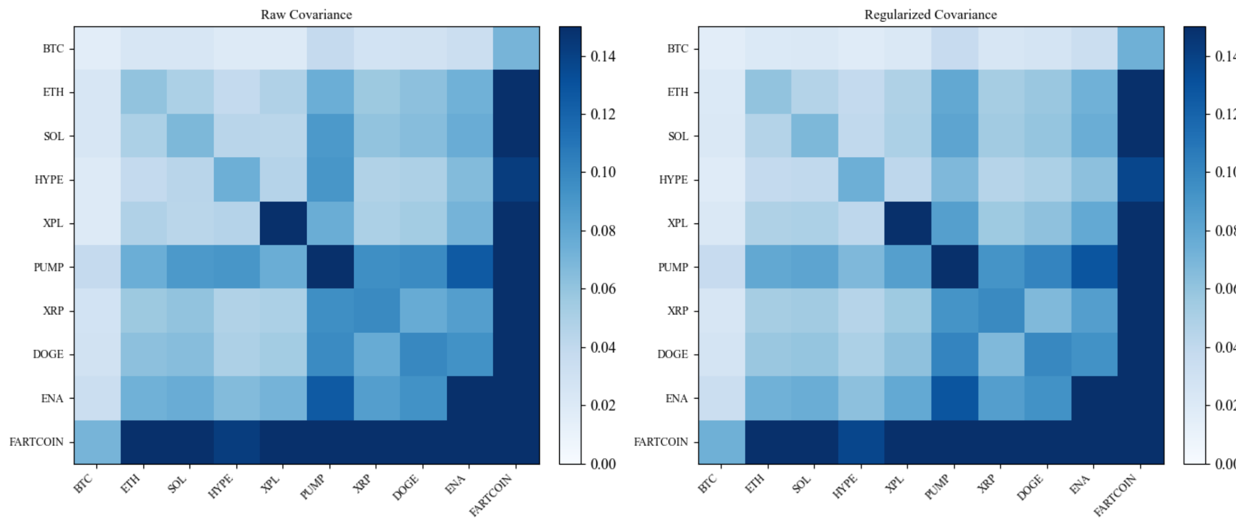

Figure 11: Left: stationary covariance matrix Σ_ε of EWMA-standardized residuals. Right: diagonal-plus-rank-one approximation $\Sigma_\varepsilon^{(1)}$, preserving the original diagonal and retaining only the leading principal component in the off-diagonal structure. The relative Frobenius error is approximately 5%, with residual variation predominantly idiosyncratic.

Empirical support for the one-factor approximation. Figure 11 shows the stationary covariance matrix Σ_ε and its diagonal-plus-rank-one approximation

$$\Sigma_\varepsilon^{(1)} = \lambda_1^\varepsilon u_1^\varepsilon (u_1^\varepsilon)^\top + \operatorname{diag}\big(\Sigma_\varepsilon - \lambda_1^\varepsilon u_1^\varepsilon (u_1^\varepsilon)^\top\big),$$

where $(\lambda_1^\varepsilon, u_1^\varepsilon)$ is the leading eigenpair of Σ_ε. The leading principal component explains approximately 56% of the total variance of Σ_ε. Despite this moderate explained-variance share, the rank-one approximation reproduces the main cross-asset dependence patterns accurately: the relative Frobenius error is approximately 5% in covariance space and 4% in correlation space. The residual variation is predominantly idiosyncratic, concentrated on the diagonal rather than in additional common factors. This supports the use of the one-factor model as a reduced-form benchmark and justifies the rank-one plus idiosyncratic structure used for Σ_ε.

C.4. Exchange-Scale ADMM Implementation Details

Active-set restriction. The ADMM iteration is never run on the full reconstructed ambient array. For a given wave, we first restrict the optimization to the active asset set $\operatorname{supp}(Q)$ and then remove every account whose directional box has zero width on that set. This reduction is exact at the level of the decision variables: accounts that cannot contribute in any active asset are fixed at zero in every feasible allocation. The scenario-dependent shortfall intercept in (82), however, still incorporates the PnL of the retained non-active positions through the term $-q_i^\top \Delta p^{(s)}$, so the optimization is carried out on the active subset while preserving the contribution of the remaining retained positions to the shortfall threshold.

Dollar reparameterization. Different assets have price scales that differ by several orders of magnitude, which causes numerical ill-conditioning in the raw position-unit formulation. We therefore reparameterize the decision variable componentwise as

$$y_i^k \triangleq p_\tau^k x_i^k,$$

so that y_i^k is the dollar notional removed from account i in asset k. Defining

$$E_i \triangleq m_i + q_i^\top (p_i^{(e)} - p_\tau), \qquad a^{(s)} \triangleq \Delta p^{(s)} \oslash p_\tau,$$

where $\oslash$ denotes componentwise division, the SAA program is written equivalently in dollar units as

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n}\Big(-E_i - q_i^\top \Delta p^{(s)} + (a^{(s)})^\top y_i\Big)_+ \\ \text{subject to} \quad & \sum_{i=1}^{n} y_i = \operatorname{diag}(p_\tau)\, Q, \\ & y_i^k \in [\, p_\tau^k\, l_i^k,\ p_\tau^k\, u_i^k\,], \quad i = 1, \dots, n,\ k \in \operatorname{supp}(Q). \end{aligned} \tag{82}$$

This transformation is an exact equivalence: problem (82) has the same optimal value as (22), and its optimal $y^\star$ recovers $x^\star$ by componentwise division. In dollar units, the decision variables, clearing constraints, and loss coefficients all lie on a common numerical scale, which substantially improves numerical conditioning.

Support-aware per-account decomposition. Because the x-update (23) separates exactly across accounts, the dollar-variable y-update does as well, and each per-account subproblem is solved on its own active support — the set of assets in which account i has nonzero feasible movement width. The majority of accounts are movable in exactly one asset and reduce to a scalar proximal problem. The remaining multi-asset accounts are grouped by their support pattern; within each group the subproblems share the same constraint structure and differ only in the proximal center and the scenario right-hand sides. This turns the accountwise ADMM step from one large heterogeneous optimization problem into a collection of many small low-dimensional subproblems.

Parallel implementation of the accountwise proximal step. In each ADMM iteration, the computational core is the per-account proximal x-update (23), or equivalently the dollar-variable y-update described above. Because this step still separates fully across accounts after the active-set reduction, the implementation distinguishes only between two solver types: accounts movable in a single asset are updated by the fast scalar routine, whereas accounts movable in multiple assets are updated by small exact per-account quadratic programs. For the multi-asset accounts, constructing a fresh optimization model for every solve would create substantial overhead, so accounts are grouped by their active support pattern and one cached quadratic-program model is reused for each pattern. For a given pattern, the quadratic part of the proximal objective is fixed once, so across ADMM iterations and across accounts sharing that pattern only the account-specific linear objective coefficients, the scenario right-hand side, and the box bounds are updated before re-solving. These multi-asset solves are then dispatched concurrently across worker tasks and warm-started from the previous iterate, while the only global synchronization point in each ADMM iteration is the projection step that restores market clearing. The long exchange-scale runs themselves are initialized from the saved directional sequential Water-filling allocation, whereas the one-factor Water-filling allocation is retained as the main constructive benchmark in the loss comparisons.

Penalty parameter calibration. The augmented-Lagrangian penalty ρ controls the relative weight of the market-clearing residual against the per-account objectives. Because the SAA objective is measured in dollars and the clearing constraint involves dollar notionals of order $\sim 10^8$, the numerically effective range of ρ lies well below standard defaults. We determine ρ by a broad logarithmic search; the exchange-scale runs reported in Section 4.3 use $\rho = 5 \times 10^{-6}$ together with absolute and relative stopping tolerances $\varepsilon_{\text{stop}} = 10^{-6}$, and we evaluate the final ADMM iterate produced at termination.

C.5. Risk-Reduction Robustness and Convergence

The following tables and figures collect the risk-reduction diagnostics deferred from the main text. They serve three purposes. First, they report the third wave τ_3 using the same definitions as the main τ_1 comparison. Second, they compare one-factor Water-filling with the directional coin-by-coin benchmark inspired by the isolated-margin ranking. Third, they show both Monte Carlo robustness and the evolution of the relative risk-reduction statistic along the fixed-sample ADMM trajectories.

Wave	**Allocation**	**Sampled Loss** L_S	**Relative RR**
τ_1	No allocation	5.218×10^8	0.000
	Hyperliquid	5.175×10^8	0.354
	Isolated Water-Filling	5.130×10^8	0.732
	One-factor Water-filling	5.114×10^8	0.871
	Risk-optimal	5.098×10^8	1.000
τ_3	No allocation	5.291×10^8	0.000
	Hyperliquid	5.178×10^8	0.568
	Isolated Water-Filling	5.115×10^8	0.886
	One-factor Water-filling	5.106×10^8	0.931
	Risk-optimal	5.092×10^8	1.000

Table 5: Sampled-loss comparison on the fixed reference scenario set for τ_1 and for the third wave τ_3. The first row in each panel reports the zero-allocation baseline $L_S(\mathbf{0})$. The directional coin-by-coin benchmark is reported only as a robustness check. In both waves, one-factor Water-filling dominates the directional heuristic under the updated relative-risk-reduction metric.

Table 5 records the fixed-sample comparison on the reference scenario set used to compute the risk-optimal allocation. It confirms that the one-factor sequential benchmark is the stronger constructive comparison. At τ_1, replacing the isolated margin/directional ranking by one-factor Water-filling raises the relative risk reduction from 0.732 to 0.871. At τ_3, the same refinement raises the relative risk reduction from 0.886 to 0.931. The directional heuristic is therefore retained only as a robustness benchmark.

To gauge the Monte Carlo sensitivity of the reported τ_1 losses, we keep the candidate allocations fixed and re-evaluate them over 50 independent $S = 128$ scenario sets under the same calibrated model, using evaluation seeds $400, \ldots, 419, 421, \ldots, 450$ so that the optimization seed 420 is excluded. The corresponding means and standard deviations of the out-of-sample losses are reported in Section 4.3; relative to those averages, the fixed reference scenario set in Table 5 yields losses that are uniformly lower by about 2.1×10^7 for all candidate allocations, while leaving the ordering in risk reduction of the allocations unchanged.

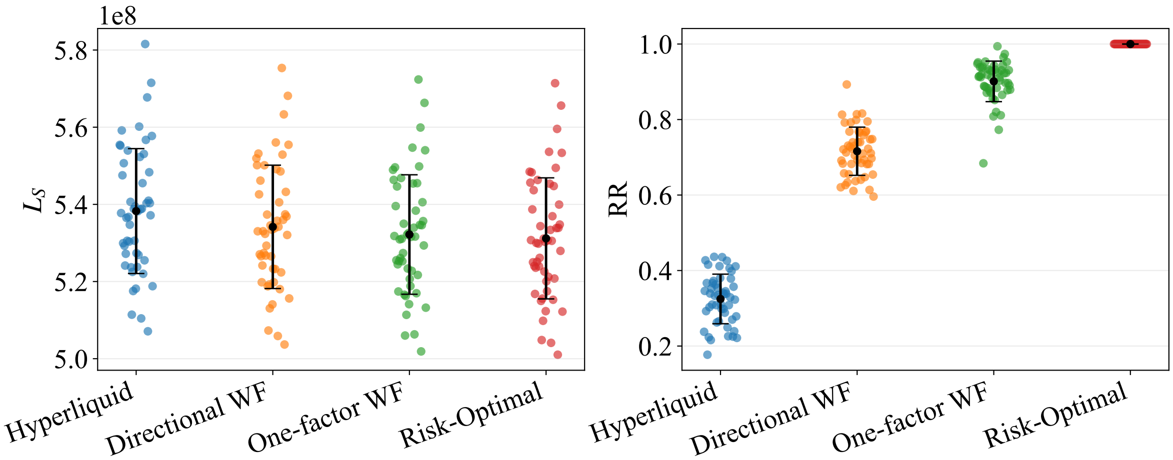

Figure 12: Monte Carlo seed-robustness check for the fixed τ_1 allocations at $S = 128$. Each point is a full-state re-evaluation of the sampled loss or the relative risk-reduction statistic on an independent scenario set under the same calibrated model; the black markers report the seed mean with one standard-deviation error bars. The spread is economically visible at this small scenario count, but the ordering Risk-optimal < One-factor WF < Directional WF < Hyperliquid is stable across seeds.

Convergence. Figure 13 plots the relative risk-reduction statistic along the ADMM iterate sequence. The initial point in each curve is the Water-filling warm start. At each iteration, we evaluate the induced sampled loss of the projected feasible allocation on the same fixed reference scenario set as in Table 5, and then map it into the statistic RR defined in Eq. (32). The horizontal axis is the ADMM iteration count. The resulting trajectories show that the relative risk-reduction statistic approaches its terminal level quickly: at iteration 500 it already equals 0.952 at τ_1 and 0.887 at τ_3, and at iteration 1,000 it reaches 0.966 and 0.917, respectively. The τ_1 run reaches the stopping rule before the iteration cap. The reported τ_3 trajectory ends at the cap $K = 20{,}000$ rather than at the formal tolerance, with terminal residuals $\|r^K\|_2 \approx 5.19$ and $\|s^K\|_2 \approx 8.0 \times 10^{-3}$. Hence most of the reduction in sampled loss is attained well before termination; the remaining iterations primarily improve the ADMM residuals rather than materially changing the induced allocation.

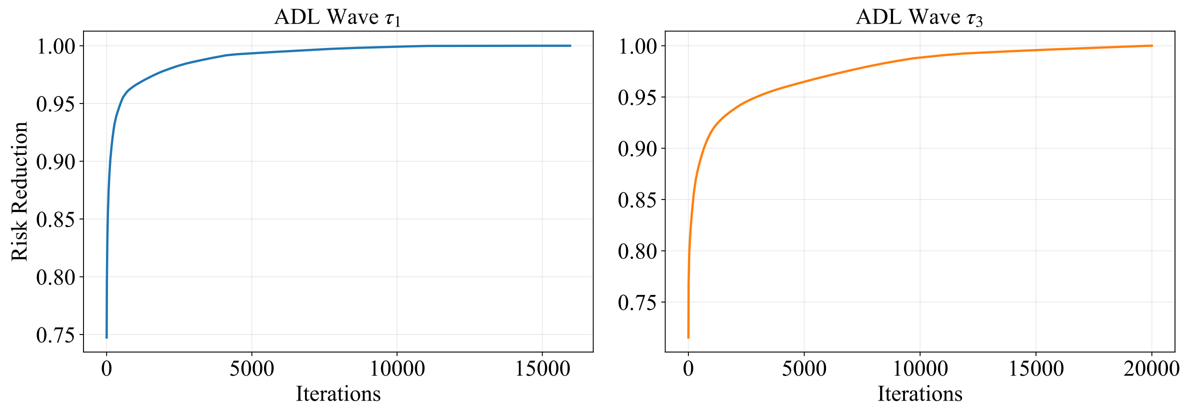

Figure 13: Relative-risk-reduction trajectories along the Risk-optimal runs on the fixed reference scenario sets for τ_1 and τ_3. The curves show that most of the improvement arrives in the first few hundred iterations, while the remaining runtime is largely spent reducing ADMM residuals. Each curve starts from the isolated water-filling warm start and evaluates the projected feasible allocation on the fixed reference scenario set for that wave.

C.6. User-Level Characteristics

The following tables and figures provide the additional user-level diagnostics deferred from the analysis in Section 4.3. They report thresholded user counts for τ_1 and τ_3 and then summarize the analogous τ_3 four-panel diagnostics using the same definitions as the main τ_1 comparison.

Threshold	Hyperliquid	One-factor WF	Risk-optimal
> 0	6,820	9,480	2,794
$>$ \$1	6,780	1,409	1,920
$>$ \$10	6,706	1,292	1,749
$>$ \$100	5,988	773	1,303
$>$ \$1k	2,416	448	874
$>$ \$10k	739	246	436
$>$ \$100k	170	92	154

Table 6: Thresholded user-count comparison at τ_1. At low thresholds, one-factor Water-filling and the Risk-optimal allocation touch many more users than the realized Hyperliquid event, but much of this mass disappears once even a one-dollar cutoff is imposed. Above $>$ \$100k, one-factor Water-filling concentrates more notional on fewer users than either Hyperliquid or the Risk-optimal allocation.

Threshold	Hyperliquid	One-factor WF	Risk-optimal
> 0	1,814	3,913	2,805
$>$ \$1	1,811	847	1,044
$>$ \$10	1,765	755	917
$>$ \$100	1,344	579	695
$>$ \$1k	650	376	437
$>$ \$10k	237	203	246
$>$ \$100k	88	91	98

Table 7: Thresholded user-count comparison at τ_3. The same dust pattern is visible at low thresholds: Hyperliquid changes almost not at all between > 0 and $>$ \$1, whereas one-factor Water-filling and the Risk-optimal allocation lose most of their touched users once a one-dollar cutoff is imposed. In the far upper tail, the ordering now reverses relative to τ_1: above $>$ \$100k, Hyperliquid touches slightly fewer users than one-factor Water-filling.

Tables 6 and 7 make the dust-allocation pattern explicit. At both τ_1 and τ_3, the realized Hyperliquid allocation changes very little between the > 0 and $>$ \$1 thresholds, whereas one-factor Water-filling and the Risk-optimal allocation touch many more users in total but much more lightly. At τ_1, one-factor Water-filling also carries more dollar notional on fewer large users, as seen most clearly in the $>$ \$100k row. At τ_3, this upper-tail ordering no longer holds: the roles of Hyperliquid and one-factor Water-filling are essentially reversed.

Figure 14 reports the analogous four-panel diagnostics for τ_3. The upper-left and upper-right panels recover the same concentration facts in graphical form. At τ_3, Hyperliquid touches only 1,814 users overall, compared with 3,913 under one-factor Water-filling and 2,805 under the Risk-optimal allocation, but its largest-user deleveraging notional rises to \$$1.7 \times 10^8$, compared with \$$1.0 \times 10^8$ for the Risk-optimal allocation and \$$6.9 \times 10^7$ for one-factor Water-filling. Its top ten users already account for 87.5% of total deleveraging, compared with 72.3% for the Risk-optimal allocation and 76.0% for one-factor Water-filling. Thus the upper-left and upper-right panels show that, at τ_3, the realized Hyperliquid allocation is the most concentrated of the three in the upper tail.

The lower-left panel looks qualitatively similar to τ_1: Hyperliquid places more mass near full closeout, whereas one-factor Water-filling and the Risk-optimal allocation usually remove more moderate fractions of the deleverageable portfolio. The lower-right panel is more ordered for Hyperliquid than at τ_1, while one-factor Water-filling and the Risk-optimal allocation retain approximately the same monotone factor-leverage targeting pattern.

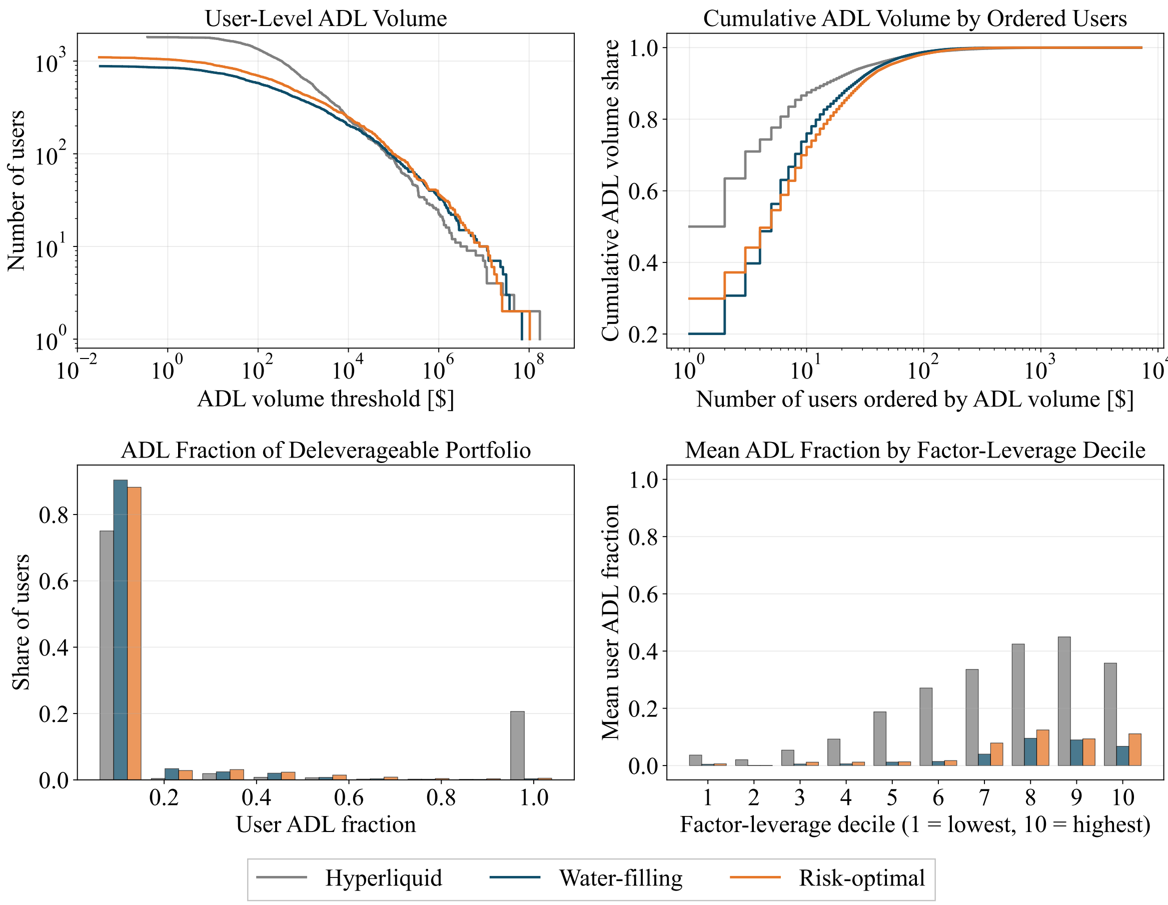

Figure 14: User-level diagnostics for τ_3: user count above deleveraging-notional threshold, cumulative top-share concentration curve, histogram of deleveraging fractions, and factor-leverage-decile targeting plot.