Optimal Market Making and Temporal Arbitrage

Devine Group

Devine OS Whitepaper Series

October 26, 2025

Abstract

Providing immediate liquidity against assets that settle with delay is a one-sided market-making problem with asymmetric information, inventory risk, and operational constraints. We formalize a model for a liquidity provider (LP) that offers instant redemptions for Hyperliquid Provider (HLP) vault receipts that otherwise redeem after T=4.5 days via Devine OS. The objective is to maximize expected profit while tightly controlling downside risk (targeting $< 1\%$ drawdowns) and remaining market-neutral apart from temporary inventory. Using auction theory and market microstructure, we derive the economically consistent instant quote (time value + risk premium + adverse-selection discount), and we embed it in a dynamic, inventory-aware policy with strict on-chain settlement and solvency invariants.

Context

The HLP vault is the principal counterparty to exchange flow on Hyperliquid: it accrues fees but absorbs trader P&L. Users deposit USDC into HLP to earn this fee-driven yield, while withdrawals settle on a delay (approximately $T \in [4, 4.5]$ days). This delay creates a liquidity timing gap: holders who value immediacy must either wait for primary settlement or sell their receipt claims at a discount.

Devine OS introduces an LP pool that bridges this gap by purchasing users' HLP receipts for cash *now* (instant redemption) and initiating primary withdrawal to collect the terminal payout *later*. Pricing must internalize: (i) time value over T, (ii) risk of vault value changes driven by market moves and trader P&L, and (iii) adverse selection from informed sellers who redeem when they expect P_T to be low. Because no direct short on HLP exists, the LP cannot perfectly hedge receipts; it must manage exposure through conservative pricing, inventory caps, and where feasible partial hedges against systematic risk.

This paper formalizes that policy. We (a) derive the instant quote as the conditional expectation of terminal value minus explicit premia, (b) add inventory-sensitive adjustments and halt regions for safety, and (c) specify settlement, accounting, and security invariants so realized P&L matches the model.

What Devine OS Does (in this flow)

- **Primary Vault Router (on-chain):** ERC-4626–style router wired to HLP. It starts and finalizes delayed redemptions, tracks user and batch nonces, enforces NAV freshness and drift bounds, and prevents replays with per-batch key / tx-hash checks.

- **Instant Liquidity Pool (off/on-chain):** A one-sided market maker that quotes an immediate exit price based on time-to-settlement, volatility/risk, signs of informed flow, and current inventory. It can widen spreads or pause under incident or venue-health signals.
- **Settlement Orchestrator (off-chain):** Batches withdrawals; verifies exchange wallet and on-chain vault balances; posts on-chain batch funding; and executes per-user confirmations individually with replay protection, NAV freshness/drift checks, batch-liquidity caps, and expiry windows. Merkle claims are optional and not used in the current build. Idempotent by design.
- **Risk/Oracle Layer:** Streams NAV snapshots and venue/bridge health, maintains estimates of expected terminal value, volatility, informed-flow intensity, and utilization, and triggers halts when data are stale or incidents occur.
- **Policy/Limits:** Enforces inventory caps aligned to a target drawdown, drift/PnL tolerances, settlement-lock windows, and single-use batch keys / tx-hash uniqueness to block replays or double allocation.
- **Optional Pass-Through Lane (simple):** We advance part of the user's redemption today and sell the right to the final payout at unlock. The buyer of that right takes the market-price risk; we earn a transparent fee. Two modes:
  - *Consumer / non-recourse:* Conservative advance, protected by pool backstops. If the final proceeds are lower, the shortfall is the buyer's risk (not clawed back from the user).
  - *Professional / recourse:* The buyer posts margin and is responsible for any shortfall. This keeps price risk off the maker while enabling higher advance ratios.

  In both cases, our role is to deliver cash now and pass the market-price risk through to the buyer, not keep it on the pool.

Together, these components allow Devine OS to (i) quote economically sound instant liquidity, (ii) settle deterministically on-chain, and (iii) bound tail risk via inventory-aware pricing, halts, and strict invariants. Subsequent sections formalize the quote, the inventory control law, the pass-through lane, and the production invariants and TCA used to calibrate and monitor the system.

Paper Structure

We proceed as follows: In Section 1, we define the model, including the asset payoff structure, the sequence of events (trades and redemption), and the types of traders (liquidity-motivated vs. informed). Section 3 derives the optimal quoting strategy for the LP under a Bayesian framework, showing how the instant redemption price is set based on the expected future value and the probability of adverse selection. We link this to classic results where a market maker's bid equals the expected value conditional on a sell order. The analysis shows how a "discount" (bid-ask spread) emerges as a function of information asymmetry and risk.

Section 4 incorporates inventory and risk management: we formalize how the LP updates quotes after each trade, manages inventory of HLP tokens to remain near market-neutral, and imposes risk limits to achieve minimal drawdown. We adapt ideas from stochastic optimal control in market-making to our setting with periodic settlement. Section 5 discusses hedging and arbitrage considerations: what if the LP can hedge part of the exposure via correlated markets or run secondary arbitrage strategies? We also consider if the LP can offload inventory to other participants or needs to internalize all risk. Section 6 outlines practical implementation details and Transaction Cost Analysis (TCA) for evaluating the strategy's performance. Finally, Section 7 concludes with a summary and potential extensions (e.g.,

multi-asset pools, dynamic settlement periods, or integration with AMM mechanisms).

Throughout, our aim is to present a comprehensive, formal treatment akin to an academic whitepaper. We use mathematical notation to specify the LP's optimization problem and cite relevant literature to connect our approach with known optimal market making frameworks.

1 Model Setup and Assumptions

1.1 Asset and Payoff

The asset in question is a claim on the HLP vault (an HLP receipt token) that can be redeemed with a 4.5-day delay for its proportional share of the vault's underlying assets. We denote by P_0 the current net asset value (NAV) per HLP token (i.e., what one token would redeem for if one waited the full period). The future redemption value after 4.5 days is a random variable P_T (with $T = 4.5$ days).

The uncertainty in P_T comes from two main sources:

1. **Market movements and trader PnL:** The HLP vault's value can change as underlying asset prices move and as traders on the exchange win or lose against the vault. For example, if many traders are long and the market rallies, the vault pays out profits and P_T drops; if traders lose, P_T rises.
2. **Fees and yield accrual:** The vault earns trading fees which add to its value over time, providing an expected drift upward.

For modeling, we treat

$$P_T = P_0 + \Delta, \tag{1}$$

where Δ is a stochastic change over 4.5 days with $\mathbb{E}[\Delta] \geq 0$ (fees yield a small positive expected return) but with significant variance due to trader PnL and market moves. Extreme events (e.g., a "toxic flow" incident) can cause large negative Δ; indeed, incidents have caused multi-million dollar losses to HLP. We assume P_T's distribution (prior) is known to the LP or can be estimated from historical data and current market conditions.

1.2 Actors

There are two types of traders who may come to the LP requesting instant redemption (i.e., selling HLP tokens to the LP):

Uninformed (Liquidity) Traders: These users trade for idiosyncratic or liquidity reasons for example, needing immediate cash or rebalancing portfolios. They do not have private information about P_T. Their decision to redeem now versus wait is mainly based on personal preference or urgency. We model that an uninformed trader will accept the LP's instant redemption price if the discount (i.e., how much less they get now compared to the expected future value) is within their tolerance. In aggregate, the arrival of uninformed redemption requests can be treated as a random flow (e.g., Poisson process) independent of P_T. The LP earns profit from uninformed trades on average, by buying below the true expected value.

Informed Traders: These are users who may have private information or forecasts about the HLP vault's future value. For example, an informed trader might know about a large impending loss in the vault (perhaps they are aware of a risky position or a market event that will hurt HLP) such a trader has an incentive to redeem early through the LP to avoid the loss. Conversely, if a user knew P_T will

increase (e.g., they expect traders in the vault to incur losses, benefiting the vault), they would prefer to wait and redeem later at a higher value rather than sell to the LP now.

In our model, informed traders selectively interact: they will sell to the LP only if their information predicts P_T will be sufficiently below the LP's quoted price. In other words, any informed trader who chooses to use instant liquidity is likely doing so because they anticipate a drop in vault value (adverse selection against the LP). We let α denote the fraction of incoming traders who are informed (or the probability a given request is informed). Even a small α can significantly affect optimal pricing due to their selective behavior.

1.3 Liquidity Provider (LP)

Our liquidity provider is modeled as a monopolist market maker for instant redemptions (we assume no competition such that the LP can set prices optimally for profit, constrained only by users' willingness to trade). The LP has an initial capital pool used to pay users who redeem instantly, and this pool is replenished when the LP receives the delayed payouts from the HLP vault. The LP's inventory is the number of HLP tokens currently held (which have been bought from users and are awaiting redemption). We denote the LP's inventory at time t as $q(t)$ (positive q means the LP is long HLP tokens, i.e., has paid out cash and is waiting for future payouts; a negative q would mean the LP somehow short HLP, but in our case $q \geq 0$ since the LP primarily buys tokens).

The LP's *objective* is to maximize expected profit from these trades while controlling risk. Profit comes from the difference between what the LP receives at redemption (P_T per token) and what it paid users up front (Q per token), times the quantity. However, if the vault's value drops significantly or if many informed traders offload before a bad event, the LP could face losses (paying $Q >$ realized P_T). The LP is risk-averse to large drawdowns, so in the model we include risk management constraints (e.g., the LP targets a high probability that losses do not exceed 1% of capital, as mentioned). We will formalize these through either a risk penalty in the objective or explicit constraints on inventory and pricing.

1.4 Timeline and Decision Sequence

We consider a continuous-time or multi-period model over the 4.5-day horizon. For analytical tractability, it's useful to break it into discrete events:

- At $t = 0$, the current HLP NAV is P_0. The LP sets an instant redemption price quote Q_0 (per token) that it is willing to pay to anyone redeeming immediately at that time. This quote may be a function of P_0, the LP's current inventory, time remaining, etc. (We will often consider an equilibrium or steady-state where P_0 is roughly constant except for realized changes, so Q is typically somewhat below P_0).
- Traders decide whether to accept the quote. If a trader comes forward, one of two things happens:
  (a) If the trader is uninformed, they will redeem if the quote Q_0 is attractive relative to their needs (for modeling, we may assume all uninformed traders in need just trade, since any discount is the "service fee" they pay for liquidity).
  (b) If the trader is informed with signal about P_T, they will redeem only if their expected P_T (given their info) is less than Q_0 ensuring they benefit by selling at Q_0 and avoiding a lower outcome. If their info suggests P_T will be higher than Q_0, they will not trade (they prefer to wait or might even want to buy if such were offered, but this white paper focus is one-sided).

- After each trade, the LP's inventory q changes (increases by the number of tokens bought). The LP immediately initiates a primary redemption from the HLP vault for the tokens acquired (to start the 4.5-day clock for payout). We assume redemptions initiated at time t will pay the holder P_T at $t+T$ (with $T = 4.5$ days). In practice, multiple redemptions can be batched, and the vault has a pooled MPC (multi-party computation) wallet and smart contract handling these flows (as per the system design given). The settlement being $T + 4.5$ means effectively the LP will receive the funds later.
- The LP updates its beliefs and prices after observing any trade (or lack of trade). The presence of a trade, and especially its size, may carry information. No trade in a period could also carry information (in classical models, if the quote is too low, informed traders might not trade, indicating perhaps nothing is wrong). The LP can use Bayesian updating: conditioning on a sell order arriving tends to indicate a lower expected P_T (adverse selection), so the LP should adjust its estimate of the true value downward. Many market making models incorporate such belief updates. However, in our context, the LP may not need to explicitly update P_T's distribution if we assume it already prices conservatively; still, formally, after a redemption trade, the LP could revise its mean estimate of P_T given that someone was willing to sell now. We will denote the LP's belief about the expected value of a token as μ; initially $\mu = \mathbb{E}[P_T]$ under the prior, but after observing trades or other signals, μ is updated.
- The LP then sets a new quote Q for the next incoming request (this could be dynamic in continuous time or period-by-period). This process repeats for each arriving redemption request up until time T. At $t = T$ (4.5 days later), all pending redemptions initiated earlier are settled: the LP receives the actual payouts for tokens it is holding. At this point, the LP's profit/loss on each batch is realized as $P_T - Q_{\text{paid}}$ (times quantity). The LP's total profit is sum over all trades minus any hedging costs.

1.5 Pricing Decision Formalization

The core decision of the LP is setting the instant redemption price $Q(t)$ at each moment (or for each trade). We can model Q as a function or strategy $Q = f(\text{information available})$ where information includes current time, inventory, prior/posterior beliefs about P_T, etc. Because the LP is essentially bidding for the asset (HLP token), we will sometimes refer to Q as the bid price from the perspective of the LP (since LP buys at Q). There may not be an "ask price" because the LP might not normally sell HLP tokens to users (users who want to enter the HLP vault can deposit directly with the primary vault rather than buying from the LP). However, if needed, one could imagine the LP also offering an ask (selling HLP tokens at some price $> P_0$) to offload inventory if there are willing buyers this could be another mechanism to manage inventory. In the simplest case, we treat it as one-sided quoting: LP posts a bid for immediate redemption, and users either hit the bid or not.

The LP's profit from a single trade of size 1 token (for simplicity) at time 0 is:

$$\Pi = P_T - Q_0. \tag{2}$$

P_T is uncertain at time of trade, so the expected profit given the LP's information (and conditional on a trade occurring) is $\mathbb{E}[P_T \mid \text{trade}] - Q_0$. The LP will choose Q_0 to maximize this expected profit, taking into account how Q_0 influences the probability of a trade and the conditional expectation $\mathbb{E}[P_T|\text{trade}]$. We will derive Q^* (optimal quote) by solving this optimization under different information scenarios.

1.6 Trader Behavior Model

We formalize the probability of trade as a function of Q. Let V represent the true (random) final value P_T of a token. We assume for analytical clarity a binary or simplified outcome space for V in some parts of the analysis (this is a common approach in microstructure models like Glosten-Milgrom): suppose V can be either "High" or "Low". For example, $V = V_H$ (with probability p) or $V = V_L$ (with probability $1 - p$), where $V_H > V_L$. Uninformed traders do not know which state will occur; informed traders know which state will occur (or at least have a very strong signal about it). The prior expected value is

$$\mathbb{E}[V] = pV_H + (1 - p)V_L. \tag{3}$$

Without loss of generality, one can think of V_H as the case where the HLP vault performs well (no major losses, maybe gains from traders) and V_L as a case where the vault performs poorly (traders win or market downturn). The LP and uninformed traders initially share this prior p.

Now consider a sell order arriving. In classic market-making theory, a risk-neutral competitive market maker would set the bid price equal to the expected value conditional on someone wanting to sell. Intuitively, if a sale is happening, it tilts the odds that the true value is lower (because an informed trader would only sell in the low-value scenario). Formally: let $\theta \in \{\text{inf}, \text{unin}\}$ indicate trader type. Suppose:

- With probability α, the trader is informed, and with probability $1 - \alpha$ they are uninformed.
- An informed trader sells if and only if $V = V_L$ (they know it's the low state; if it were V_H, they would not sell because they'd get more by waiting).
- An uninformed trader may sell for other reasons, independent of V. For simplicity, assume an uninformed trader's decision to sell is not affected by V (they might flip a coin or have liquidity needs; in expectation, they sell with some fixed probability regardless of state). In a one-time trade setting, we can say an uninformed trader sells with probability η (and does so without knowledge of V).

If we simulate "one trader arrives" as either informed or uninformed:

- If informed (prob α): if V_L occurs, they sell (we can take this probability as 1 for V_L case); if V_H occurs, they do nothing (they would not use the service because they expect a higher payout by waiting).
- If uninformed (prob $1 - \alpha$): they sell with probability η regardless of V. Often, we consider $\eta = 0.5$ in symmetric models (uninformed equally likely to sell or not, or sell vs buy), but since here the only action is selling (redeeming), we can incorporate η into the arrival rate of uninformed sellers. Essentially, among uninformed traders who need liquidity, a certain portion will be redeeming at any given time.

Given this behavior, we can derive the conditional probabilities:

- Probability of a sell order (any) occurring is:

$$\Pr(\text{sell}) = \alpha \Pr(V = V_L) \cdot 1 + \alpha \Pr(V = V_H) \cdot 0 + (1 - \alpha)\eta[\Pr(V = V_H) + \Pr(V = V_L)]. \tag{4}$$

  If we assume at most one trade attempt in the period, this simplifies to

$$\Pr(\text{sell}) = \alpha(1 - p) + (1 - \alpha)\eta. \tag{5}$$

(Here $p = \Pr(V_H)$.)

- Conditional on seeing a sell, the probability it was an informed trader in low state is:

$$\Pr(\theta = \text{inf}, V = V_L \mid \text{sell}) = \frac{\alpha(1-p)}{\Pr(\text{sell})}. \tag{6}$$

- Conditional probability of the low-value state given a sell is higher than the prior $(1-p)$ because of this selection. Specifically:

$$\Pr(V = V_L \mid \text{sell}) = \frac{\alpha(1-p) + (1-\alpha)\eta(1-p)}{\Pr(\text{sell})}. \tag{7}$$

If η is not too large, this is $> (1-p)$. In the limit where uninformed always trade (η large or effectively $= 1$ for one trade scenario), we get

$$\Pr(V_L|\text{sell}) = \frac{\alpha(1-p) + (1-\alpha)(1-p)}{\alpha(1-p) + (1-\alpha)} = 1 - p, \tag{8}$$

meaning if almost all trades are uninformed, the conditional probability stays equal to prior (no adverse selection). But if α is significant and η is low (meaning trades are infrequent unless informed strongly wants to trade), the conditional probability of V_L given a sell can approach 1.

For a risk-neutral LP aiming for zero expected loss (in a competitive setting), the optimal bid price Q satisfies:

$$Q = \mathbb{E}[V \mid \text{sell}] = V_L \Pr(V_L \mid \text{sell}) + V_H \Pr(V_H \mid \text{sell}). \tag{9}$$

This is exactly the conditional expectation of the asset's true value given that a sell order is observed. In the Glosten–Milgrom model, market makers set bids and asks equal to these conditional expectations to avoid expected loss to informed traders. The spread emerges because

$$\mathbb{E}[V \mid \text{sell}] < \mathbb{E}[V] < \mathbb{E}[V \mid \text{buy}] \tag{10}$$

when some traders are informed. In our context, we are not necessarily assuming a perfectly competitive zero-profit LP; rather, our LP might have some monopoly power to charge a slightly larger spread (aiming for positive profit). However, if it sets Q too high, it will lose money to informed traders; if Q too low, uninformed traders might not use the service (or go to competitors, if any). So the equilibrium Q will be near this conditional expectation, possibly adjusted for risk and profit margin.

1.7 Adverse Selection Discount

We can quantify the *adverse selection component* of the discount (the difference between the fair expected value and the bid Q). Using the binary example above for intuition: say prior $\mathbb{E}[V] = pV_H + (1-p)V_L$. If $\alpha > 0$, then $\mathbb{E}[V \mid \text{sell}]$ will be tilted toward V_L. In fact, if uninformed trades are equally likely buy/sell (in a symmetric market), the classic formula for the bid price in a one-shot trade is:

$$Q_{\text{bid}} = \Pr(V_H \mid \text{sell})V_H + \Pr(V_L \mid \text{sell})V_L. \tag{11}$$

In the extreme case of $\eta \to 0$ (meaning a trade is almost surely an informed one), $\Pr(V_L|\text{sell}) \to 1$ and thus $Q \to V_L$. In the other extreme $\alpha \to 0$ (no informed traders), $\Pr(V_L|\text{sell}) = \Pr(V_L) = 1 - p$, so $Q \to \mathbb{E}[V] = pV_H + (1-p)V_L$ the LP can pay essentially the full expected value (minus perhaps a small charge for time value/profit). For intermediate cases, one can derive:

$$Q = \frac{(1-\alpha)\eta\mathbb{E}[V] + \alpha(1-p)V_L}{(1-\alpha)\eta + \alpha(1-p)}. \tag{12}$$

If η is small (traders only come when informed or very occasionally otherwise), this will be close to V_L. If η is larger (regular uninformed flow), Q moves closer to $\mathbb{E}[V]$.

In summary, the optimal quote will incorporate a discount relative to the naive expected value. This discount accounts for:

- **Time Value of Money:** The LP's capital is tied up for 4.5 days. There is an opportunity cost or interest rate r for that period. For example, if risk-free rate is r_{annual}, for 4.5 days (≈ 0.0123 of a year), the risk-free growth is negligible but not zero. The LP might factor in a tiny discount $Q \approx \mathbb{E}[P_T]/(1 + r \cdot 4.5 \text{ days})$. This is usually very small (basis points), so we focus on bigger factors.
- **Risk Premium:** Even if traders are uninformed, P_T is risky. A risk-averse LP would charge a premium for bearing this risk. If σ is the standard deviation of Δ (change in vault value over 4.5 days), the LP might discount by some fraction of σ depending on risk appetite (akin to applying a worse-case or Value-at-Risk buffer). We could incorporate this by $Q = \mathbb{E}[P_T] - \lambda\sigma$ for some λ related to a confidence level (for example, to ensure 99% of outcomes $P_T > Q$ to limit drawdown). We will formalize risk constraints later.
- **Adverse Selection (Information Asymmetry):** If $\alpha > 0$, the LP knows that a user willing to sell might have bad news. The conditional expectation $\mathbb{E}[P_T|\text{sell}]$ is lower than the unconditional $\mathbb{E}[P_T]$. We denote the adverse selection discount as

$$\delta_{\text{info}} = \mathbb{E}[P_T] - \mathbb{E}[P_T|\text{sell}] \geq 0. \tag{13}$$

  In the binary case earlier, $\delta_{\text{info}} = (pV_H + (1-p)V_L) - (\Pr(V_H|\text{sell})V_H + \Pr(V_L|\text{sell})V_L)$. This can be simplified; for instance, if $V_H - V_L = D$ is the "value range", one can show the bid-ask spread in symmetric information case is proportional to αD. For our one-sided service, effectively the "spread" is between the primary market (face value) and our bid. We ensure Q is low enough that $\mathbb{E}[P_T|\text{sell}] \approx Q$ (if we want zero expected loss). If we want positive expected profit, we set Q a bit below $\mathbb{E}[P_T|\text{sell}]$. However, setting it much below will reduce volume (uninformed might balk unless desperate), so there is an optimal point balancing volume vs. margin.

1.8 Volume vs. Price Trade-off

In a more general continuous model, we can imagine the LP sets a price schedule (demand curve) for how many tokens it is willing to buy at what price, similar to Myerson's mechanism design approach. The LP could choose a schedule $Q(q)$ giving the price as a function of total quantity q it will purchase (or equivalently, how the price slides for larger trades). If many users want to redeem at once, the LP might offer a lower price for the later units (to protect itself). However, if we assume trades arrive one by one and are typically small relative to LP capital, we can treat each trade independently for pricing. The LP's strategy effectively creates a marginal price for the next token.

According to optimal mechanism design for a monopolist with asymmetric info (like Myerson's auction theory), the profit-maximizing strategy might involve a cutoff strategy: do not trade (no quote) if the inferred value is in some range, and trade at a certain price if lower. In our setting, that could translate to: if the LP is very uncertain or risk of informed trading is too high, it might temporarily stop offering liquidity (a "no-trade" region to avoid sure loss). Otherwise, it quotes a price that includes a markup. The Myersonian result for optimal auctions implies setting a price according to "virtual value" of the trader's signal. Roughly, the LP would adjust the price to make the marginal informed trader indifferent capturing some information rent as profit while ensuring incentive-compatibility (traders reveal their need/information truthfully by the act of trading or not).

We won't delve deeply into auction math here, but note that optimal profit-maximizing liquidity provision leads to a bid-ask spread even for a monopolist, caused in part by adverse selection (and also by the desire to profit as a monopolist, which is akin to "monopoly pricing" of liquidity). Our model's results will be consistent with these general insights: a jump or gap between the price the LP is willing to buy and the expected fair value arises to protect against informed traders and to earn a return for providing the service.

1.9 Equation of Optimal Bid

Combining these considerations, we can propose a formula for the LP's instantaneous bid quote Q^*. Let μ be the current expected value of P_T (posterior mean given all info so far, prior to seeing the next trade decision). Let σ reflect risk (e.g., one standard deviation of possible change). Let α capture info asymmetry likelihood. One conceptual formula is:

$$Q^* = \mu - r_f \mu \Delta T - \gamma \sigma^2 \Delta T - \delta_{\text{info}}, \tag{14}$$

where r_f is the risk-free rate (time value adjustment, $\Delta T = 4.5/365 \approx 0.0123$ years), γ is a risk-aversion coefficient translating variance into a risk premium (this term could be $\frac{1}{2}\gamma\sigma^2$ or another function if using a CARA utility or VaR constraint), and δ_{info} is the adverse selection discount.

To be more concrete, using the binary model analysis for δ_{info}: if $V_H - V_L = D$, and $\Pr(V_H|\text{sell}) = \pi$ (lower than p), then

$$\mathbb{E}[V|\text{sell}] = \pi V_H + (1-\pi)V_L = V_L + \pi D. \tag{15}$$

Originally $\mu = pV_H + (1-p)V_L = V_L + pD$. The difference is

$$\delta_{\text{info}} = \mu - \mathbb{E}[V|\text{sell}] = (p-\pi)D. \tag{16}$$

In a scenario with continuous distribution, one can show similarly that δ_{info} relates to the hazard rate of informed trading essentially higher when the likelihood of a low outcome given trade is higher. For instance, Glosten derived that the bid-ask spread equals the probability of informed trading times the value dispersion. For our purposes, we treat δ_{info} as a parameter to be determined by calibrating to how much worse outcomes are on average when someone redeems vs overall average. In practice, the LP could estimate δ_{info} by looking at historical data: "When someone redeems early, what was the average shortfall in final payout?" If, say, historically early redeemers received on average 0.5% more than the final value (meaning the vault often dropped after early redemptions), then $\delta_{\text{info}} \approx 0.5\% \cdot P_0$. The LP would then incorporate that into quotes.

Summary. In words, the LP's quoting policy is: offer an instant redemption price that is somewhat below the expected redemption value, with the markdown reflecting (i) the financing cost for 4.5 days, (ii) a risk buffer for possible adverse vault moves, and (iii) an adverse selection buffer if informed users are likely to redeem when bad news looms. The resulting discount ensures that in expectation, the LP does not lose money on the trades in fact, ideally gains. This is analogous to how a market maker sets a bid price below the expected true value to compensate for the chance that the seller knows the value is lower.

2 Parallel Pricing Mechanism: Pass-Through Advance With True Up (Risk to Taker)

This section adds a parallel mechanism that provides instant cash while shifting price risk over the settlement window to the taker. It runs alongside the fixed price market making model developed in Sections 1–4. The key difference is that the final user cash flow references the realized terminal NAV P_T. The liquidity provider does not form a directional view and does not earn or lose on P_T in expectation. The provider earns a transparent fee net of funding and operations.

2.1 Contract Terms and Timeline

A user with S HLP receipt shares requests an instant exit at time t_0.

Advance now

$$A = S \cdot P_0 \cdot a$$

where P_0 is the current per share NAV and $a \in (0, 1)$ is the advance ratio quoted by the protocol.

Final entitlement at unlock

$$F = S(P_T - f)$$

where $f \geq 0$ is a per share fee and P_T is the realized per share payout at $t_0 + T$ with the same T as in Section 1.

True up rule

$$\text{TopUp} = \max\{0, F - A\}$$

The protocol pays TopUp from funds actually received at unlock. In a non-recourse lane no clawback is attempted if $F < A$. In a recourse lane the user posts margin and any shortfall $A - F$ is netted against that margin.

2.2 Who Bears Risk

- **Economic price risk** over $[t_0, t_0 + T]$ sits with the taker because the total user cash flow equals $P_T - f$ per share.
- **Residual shortfall risk** exists only for the non-recourse lane when $F < A$. It is removed in the recourse lane by posted margin.
- **Provider economics** are fee minus funding and operations with no directional P_T exposure in the recourse lane.

2.3 Advance Ratio as Capacity Price

The advance ratio a prices immediacy and balance sheet usage rather than a view on P_T. Write

$$a = 1 - \big(f + c_f + c_o + b\big) - \lambda_\sigma \hat{\sigma}_T - \lambda_u g(u) - \mathbf{1}_{\text{stress}} \cdot s$$

where c_f is four day funding, c_o is four day operations, b is a small safety buffer, $\hat{\sigma}_T$ is a conservative volatility proxy, u is utilization, g is increasing in u, and s is a stress add on. In practice a is clamped by a shortfall floor

$$a \le 1 - f - c_f - c_o - \mathrm{VaR}_\alpha(T)$$

so that $\Pr(F < A) \le \alpha$ in the non-recourse lane.

2.4 Two Lanes

Retail non-recourse User receives A now and a top up later if $F > A$. If $F < A$ the gap is a protocol shortfall covered by an explicit waterfall: insurance then junior capital then a capped treasury allocation. Capacity is gated by the backstop so offered advances never exceed cover.

Professional recourse User escrows shares and margin in USDC. At unlock the contract nets any shortfall $A - F$ from margin and returns the remainder. With sufficient margin the provider has no price PnL across all paths.

2.5 Settlement and Accounting

Batched workflow mirrors Section 6. The router confirms net USDC received for the batch. An off-chain job computes per user top ups and posts a Merkle root. Users claim with proofs. No loops are used on chain.

2.6 Properties

Proposition 2.1. *In the recourse lane with sufficient margin the provider's expected price PnL is zero for every path. The per share revenue equals f and the net revenue equals $f - c_f - c_o$.*

Proof. Per share user cash flow equals $P_T - f$. The provider receives P_T and pays $P_T - f$. The difference is f. Funding and operations reduce this amount deterministically. □

Proposition 2.2. *If $a \le 1 - f - c_f - c_o - \mathrm{VaR}_\alpha(T)$ then in the non-recourse lane $\Pr(F < A) \le \alpha$.*

Proof. By definition of $\mathrm{VaR}_\alpha(T)$ we have $P_T \ge P_0 - \mathrm{VaR}_\alpha(T)$ with probability at least $1 - \alpha$. The bound on a implies $A \le S(P_0 - \mathrm{VaR}_\alpha(T) - f - c_f - c_o)$. Hence $F \ge A$ with probability at least $1 - \alpha$. □

2.7 Comparison With Fixed Price Market Making

- The original model in Sections 3–4 sets a one sided bid Q and carries $P_T - Q$ price risk. Adverse selection drives the discount through $\mathbb{E}[P_T \mid \text{sell}]$ as in classical market microstructure theory.
- The pass through mechanism in this section fixes the terminal leg to P_T. Pricing focuses on capacity, utilization and risk limits. Information asymmetry affects capacity through a rather than expected price PnL.

2.8 Operational Safeguards

Advance ratio goes to zero under oracle staleness, withdrawal halts or venue incident alerts. Batches are capped by an insurance based coverage factor. Retail and professional lanes are separated at the contract boundary.

3 Optimal Quoting Strategy

Building on the model setup, we now formally derive the optimal quoting strategy. The LP's problem is to choose a bid price Q that maximizes expected profit while managing the risk of adverse selection and market movements.

3.1 Objective Function

The LP's expected profit from a single trade (normalized to one token) is:

$$\mathbb{E}[\Pi] = \Pr(\text{trade at } Q) \cdot \mathbb{E}[P_T - Q \mid \text{trade at } Q]. \tag{17}$$

The first term, $\Pr(\text{trade at } Q)$, represents the probability that a trader accepts the quote Q. This depends on the distribution of trader types and their reservation values. The second term is the expected profit conditional on a trade occurring at price Q.

For a risk-neutral LP in a competitive market, the optimal bid satisfies the zero-profit condition:

$$Q^* = \mathbb{E}[P_T \mid \text{sell}]. \tag{18}$$

However, for a monopolistic LP or one with risk aversion, we need to incorporate additional considerations.

3.2 Risk-Adjusted Pricing

For a risk-averse LP, we can formulate the objective using a mean-variance framework or utility function. Let $U(\cdot)$ denote the LP's utility function. The LP maximizes:

$$\max_Q \mathbb{E}[U(\Pi)] = \max_Q \mathbb{E}[U(P_T - Q)], \tag{19}$$

subject to the constraint that trades occur (i.e., Q must be acceptable to some traders).

For a CARA (Constant Absolute Risk Aversion) utility function with risk aversion parameter γ:

$$U(x) = -e^{-\gamma x}, \tag{20}$$

the certainty equivalent of the profit is approximately:

$$CE = \mathbb{E}[P_T - Q] - \frac{\gamma}{2}\text{Var}(P_T - Q) = \mathbb{E}[P_T] - Q - \frac{\gamma}{2}\sigma^2, \tag{21}$$

where $\sigma^2 = \text{Var}(P_T)$.

Setting the certainty equivalent to zero (for breakeven) or to some target profit π_0 gives:

$$Q = \mathbb{E}[P_T] - \frac{\gamma}{2}\sigma^2 - \pi_0. \tag{22}$$

This shows how risk aversion leads to a discount beyond the expected value.

3.3 Incorporating Information Asymmetry

When we account for the presence of informed traders, the LP must condition on the event that a trade occurs. Using Bayes' rule, the optimal quote becomes:

$$Q^* = \mathbb{E}[P_T \mid \text{sell}] - \text{margin}, \tag{23}$$

where the margin can be decomposed as:

$$\text{margin} = r_f \mu \Delta T + \frac{\gamma}{2}\sigma^2 \Delta T + \text{monopoly markup}. \tag{24}$$

The first two terms represent time value and risk premium, while the monopoly markup reflects the LP's market power.

3.4 Dynamic Belief Updating

After observing a trade (or no trade), the LP should update its beliefs about P_T using Bayesian inference. Let μ_t denote the LP's posterior mean at time t. Upon observing a sell order at time t, the updated belief is:

$$\mu_{t+1} = \mathbb{E}[P_T \mid \text{sell at } t, \mathcal{F}_t], \tag{25}$$

where $\mathcal{F}_t$ is the information set at time t.

For the binary model with $V \in \{V_H, V_L\}$, the posterior probability of the high state after observing a sell is:

$$p_{t+1} = \Pr(V = V_H \mid \text{sell}) = \frac{(1-\alpha)\eta p_t}{(1-\alpha)\eta p_t + [\alpha + (1-\alpha)\eta](1-p_t)}. \tag{26}$$

This posterior is lower than the prior p_t, reflecting the adverse selection effect. The updated quote should then be based on $\mu_{t+1} = p_{t+1}V_H + (1-p_{t+1})V_L$.

4 Inventory Management and Risk Control

While setting the right price is crucial, the LP must also manage its inventory of HLP tokens and overall exposure over time. Unlike a traditional stock market maker who can hold inventory indefinitely (subject to risk limits) or hedge continuously, our LP's inventory is tied to pending redemptions that will pay out at future times.

4.1 Inventory Risk

The LP's capital at risk at any given moment is proportional to its inventory q and the current NAV P_0. Essentially, if the LP holds q HLP tokens, a sudden drop in HLP value (e.g., due to trader wins or price crashes) would cause a mark-to-market loss of $q\Delta P$ (until redemption). The LP desires to keep this risk limited (e.g., limit potential drawdown to approximately 1% of total pool).

There is a capacity constraint: the LP cannot buy unlimited tokens; it has finite capital and risk tolerance. If too many users redeem in a short time (say a run on liquidity), the LP might have to drastically lower Q or temporarily halt trading to avoid overexposure. We may formalize a Value-at-Risk (VaR) limit: e.g., choose $q_{\max}$ such that even if the HLP vault suffers a worst-case loss (say $X\%$ drop) on the entire inventory, the loss is less than 1% of LP capital. For instance, if the LP has \$100M capital and wants $< 1\%$ loss = \$1M in worst case, and worst-case drop in HLP value in 4 days is say 20% (observing events like the \$20M incident which was approximately loss of 20% of a \$100M vault), then

$$q_{\max} \approx \frac{1M}{0.2P_0} = \frac{5M}{P_0}. \tag{27}$$

If $P_0 = 1$ (i.e., normalized per token value \$1), $q_{\max} = 5M$ tokens. This is just an example of setting a cap.

4.2 Inventory-Adjusted Quoting

Inventory also influences future quoting: A fundamental principle in market making is that if you have a long inventory, you should lower your bid (and ask) to reduce further accumulation and even try to sell some inventory off. Conversely, if you have too little inventory (or short, though here short unlikely), you might raise quotes to attract more trades. In our one-sided context, if q is large (we hold many tokens), the LP may become more cautious: lower Q substantially to deter all but the most motivated sellers, or pause quoting, until some inventory clears (either through reaching settlement or finding someone to offload to).

We can formalize a linear inventory penalty in the quote. For example, building on the Avellaneda-Stoikov style model, one might set:

$$Q(q) = Q_0^* - kq, \tag{28}$$

where Q_0^* is the base quote (as derived in previous section for zero inventory) and k is a positive constant representing how much we adjust price per token of inventory to account for risk. This essentially adds a buffer so that as inventory increases, the price offered decreases, reducing further inflow and encouraging perhaps some opportunistic buyers if any (if the LP were to offer an ask, it might come into play). The parameter k could be chosen based on risk tolerance and volatility (in Avellaneda-Stoikov, $k = \frac{\gamma\sigma^2}{2}$ for some risk aversion γ).

Because the LP cannot directly sell HLP tokens except waiting for redemption, inventory management mainly means controlling the flow of new purchases. However, if there is a possibility to arrange secondary transfers (e.g., perhaps another party wants to enter HLP instantly, the LP could match them by selling some tokens at an "ask" price), the LP could facilitate that. In absence of such external demand, the LP could also deposit additional capital into the HLP vault as needed to offset (though depositing doesn't reduce inventory; it just increases exposure). So likely the main tool is quoting and possibly hedging.

4.3 Stochastic Control Formulation

We can set up the LP's problem as a stochastic control or dynamic programming problem. The state variables are: current time t, inventory q, and perhaps current belief about P_T (mean μ). The control is the quote price Q. The objective is to maximize expected terminal profit $\mathbb{E}[\text{PNL}]$ minus a penalty for risk. We also consider that at $t = T$, all inventory is liquidated at P_T. One could write the Hamilton-Jacobi-Bellman (HJB) equation for the LP's value function $J(t, q, \mu)$, but given the complexity of adverse selection, we incorporate a simpler heuristic approach here:

- We assume the LP sets $Q(t)$ at each small interval Δt, and either a trade happens or not. The inventory update dq equals the trade size if trade happens.
- The expected infinitesimal profit $d\Pi$ in that interval is $\Pr(\text{sell at } t) \cdot (\mathbb{E}[P_T|\text{sell at } t] - Q(t)) \cdot (\text{size})$. If no trade, no immediate PNL change (though future P_T expectation might update).
- The LP's expected total profit is the sum/integral of these contributions over time, plus the final mark-to-market of any remaining inventory. If quoting continuously, the LP may receive multiple orders.
- The risk of a large loss comes if P_T turns out much lower than expected for a large inventory. This is a downside scenario we constrain.

To protect against adverse selection and inventory risk, the LP might incorporate a **no-trade region** or reservation price. If the LP's belief μ becomes very low (signaling likely losses) or inventory q is very high, the LP could even stop offering quotes (effectively Q goes to a very low level that no rational uninformed trader would accept, so only an extremely informed desperate trader would trade which in turn the LP may actually not want either). Such "no-trade intervals" have been identified in optimal market making with combining inventory and info asymmetry. In simpler terms: if uncertainty is too high, the LP widens the spread so much that trading ceases until new information arrives. In our scenario, the LP might at times say "no liquidity available right now" if conditions are too adverse (this ensures not providing liquidity to strongly informed traders at a bad price).

4.4 Minimal Drawdown Target

We now incorporate the constraint that the LP's drawdown (loss) should be very small (on the order of 1% of capital or less). We can interpret this in a few ways:

- As a chance constraint: $\Pr(\text{PNL} < -1\% \text{ capital}) < \epsilon$, with ϵ very small (say 1% or 0.1%). The LP wants a high probability of not losing more than 1%.
- As a robust worst-case: for any reasonably possible scenario (excluding extreme tail beyond some confidence), the loss is $< 1\%$.

To enforce this, the LP must be conservative in both pricing and sizing. This translates to:

- Limiting inventory q as discussed: ensure q is such that even if P_T is at its lower 1% percentile, the loss $q(Q - P_T)$ is $< 1\%$ capital.
- Dynamic hedging if possible: If any instruments exist to offset risk, use them (discussed more in hedging section).
- Real-time monitoring: If an event (like a sudden market move) suggests the vault will drop significantly, the LP might adjust instantly (e.g., reduce Q or pause) to avoid accumulating more at old prices.

- Possibly diversification: if the LP runs this service for multiple independent vaults or multiple times, diversify to not put all capital in one cycle.

In formal terms, one might add a penalty term in the objective for variance of PNL or a constraint on CVaR (Conditional Value at Risk). For example, maximize

$$\mathbb{E}[\text{PNL}] - \lambda \,\text{Var}(\text{PNL}) \tag{29}$$

for some large λ that forces low variance. Or solve for Q and acceptance criteria such that

$$\mathbb{E}[\text{PNL}] - z\sqrt{\text{Var}(\text{PNL})} \geq 0 \tag{30}$$

for z corresponding to a 2-3 standard deviation safety (if assuming normal approximations). The LP can calibrate λ or z to reflect the 1% drawdown tolerance.

4.5 Continuous Belief Updates

Each trade (or lack thereof) gives the LP information about P_T. Using Bayesian updating, the LP should adjust $\mu = \mathbb{E}[P_T]$ whenever a redemption happens. Specifically, if someone redeems, it's a signal that perhaps P_T could be on the lower side (with weight depending on how likely it is they were informed). The LP could update something like:

$$\mu_{\text{new}} = \mu_{\text{prior}} - w(Q - \mu_{\text{prior}}), \tag{31}$$

where w is related to the probability the trader was informed. In an approximate linear update (assuming normal priors etc.), one can derive formulas for how much to shift the mean. For example, if we treat the observation "a sale happened" as evidence, the updated mean might be $\lambda\mu_{\text{old}} + (1-\lambda)Q$ for some λ. Intuitively, if a sale happens, perhaps tilt the expected value a bit towards the sale price or below it. If no trades happen for a while, maybe one can raise the quote gradually (since no news could imply either nothing bad or simply no liquidity needs this part is subtle). For simplicity, the LP could maintain a running estimate of α (informed probability) and adjust μ accordingly. In practice, if the LP sees a flood of redemption requests all at once, that is a huge red flag of adverse selection likely many know something (e.g., a big position in the vault went south). In that case, μ should be slashed dramatically and Q dropped or halted to avoid being stuck with too high-priced inventory. Our model would capture that by sequentially updating μ after each trade, resulting in rapidly falling quotes.

4.6 Mathematical Example Inventory Adjusted Quote

To illustrate the combined effect, consider a simple linear model:

$$Q(t) = \mu(t) - \delta_{\text{info}}(t) - cq(t), \tag{32}$$

with $\mu(t)$ the current belief of fair value, $\delta_{\text{info}}(t)$ the adverse-selection markup needed (which could depend on updated α or variance at time t), and $cq(t)$ the inventory penalty term (with $c > 0$). If $q = 0$, this reduces to earlier Q^* formula. If q is positive, the price is set lower to slow further purchases. One can choose c such that if q hits a certain fraction of $q_{\max}$, the quote becomes very low (essentially discouraging more trades). For example, if $q_{\max} = 1000$ tokens is the limit, one might choose c such that $c \cdot 1000 =$ some large discount like 5%. So each 100 tokens adds 0.5% extra discount.

This formula is heuristic, but it captures the spirit. In well-known market making models, the optimal

bid is the base value minus an inventory term proportional to inventory. Those models assume continuous market price movements and the inventory term addresses price risk. Here, price risk is embodied in μ changing and in the possibility of lower P_T; nonetheless, we include it analogously.

4.7 No Short Hedge Available Implications

The user's note that "nowhere offers a short HLP option so we can't hedge it, only hedging is through the pool" implies the LP cannot offload the risk via derivatives or short positions directly on HLP. This is a critical constraint. If shorting HLP were possible, the LP could immediately hedge by short-selling the same amount of HLP tokens in another market (locking in the profit Q vs future P_T difference). In absence of that, the LP itself essentially is the market it must carry the position until settlement. This increases the importance of correct pricing and inventory limits. It also means the LP's only way to mitigate risk is by influencing the order flow (through pricing) and by possibly hedging indirectly in related markets:

- **Hedge underlying components:** The HLP vault's value is tied to underlying assets (e.g., if HLP holds a basket of crypto assets or stablecoins, plus PnL). If the vault has a known composition or delta exposure, the LP might hedge those. For instance, if HLP effectively is long certain coins or short certain due to open interest, the LP could take opposite positions in the open market to offset some risk. However, this is complicated: HLP's exposure to, say, BTC or ETH can change as traders enter/exit positions. If one knew HLP is currently net short ETH (because traders are long ETH perps in the exchange), then if ETH price surges, HLP loses. The LP could hedge by buying some ETH or going long an ETH future to offset that scenario. Similarly, if HLP has a lot of long exposure to an asset (through traders' shorts), then a price drop hurts HLP, which LP could hedge by shorting that asset. Essentially, the LP could try to replicate the trading book of the HLP vault externally. This requires access to real-time data on the vault's positions. If the system (divine OS mentioned) provides that, an in-house secondary arbitrage model could be to mirror trades: for example, if a whale just sold a large position to HLP (implying HLP is now long that risk), the LP might hedge by shorting that asset externally.
- **Hedge systematic risk:** Even without replicating positions, the LP knows that if overall crypto market crashes, HLP likely loses value (as traders might be net long, or even if not, some losses happen). So the LP could maintain a hedge like a put option or short futures on a crypto index to cover tail risk. This could protect against broad market moves.
- **Insurance or stop-loss:** The LP could pre-arrange a form of insurance or stop-loss trade: for example, if HLP value drops beyond a threshold, the LP might have a contract to sell HLP tokens or have someone inject funds. But these are complex to implement in decentralized setting.

We incorporate hedging in our model as an option but not a given. Formally, let H be any hedge position the LP takes (with cost and payoff that correlates with P_T). The LP's total PNL from a trade plus hedge would be $(P_T - Q) + H_{\text{payoff}} - H_{\text{cost}}$. The LP can choose H to minimize $\text{Var}(P_T - Q + H)$ ideally. A perfect hedge sets $H_{\text{payoff}} \approx -(P_T - \mathbb{E}[P_T])$ so that PNL becomes roughly $\mathbb{E}[P_T] - Q$. But without direct HLP hedge, H will be imperfect.

We formalize hedging decision as part of strategy: at time of each trade, or periodically, the LP chooses a hedge portfolio to maximize expected utility. However, to keep this paper focused, we assume limited hedging ability. Instead, the LP leans on *self-hedging through the pool*, meaning it dynamically adjusts quotes (and thus who trades with it) to manage exposures essentially the pool's behavior is the hedge.

4.8 Example Scenario

Suppose at time 0 the LP has no inventory and quotes Q_0. A few small uninformed trades happen, LP accumulates a bit of inventory (say $q = 100$ tokens). Suddenly, news breaks of a possible exploit or major win for a trader that could cost the vault money (like the whale offloading \$4M loss scenario). Informed users rush to redeem. The LP, if acting ideally, observes an increase in redemption requests (above normal flow) this triggers a belief update that P_T might end up much lower. The LP immediately lowers its quote Q significantly, perhaps by a few percent, to protect itself. Some users might still accept (if they think even worse coming), but at least the LP is now buying at a much lower price. The LP might also stop after a certain size or time. As a result, the LP might end up having bought, say, another $q = 500$ tokens but at large discount. The vault indeed loses money; when the dust settles, P_T is down say 5%. The LP's earlier inventory of 100 loses 5% (that portion is a loss). The later inventory of 500 was bought at, say, 4% discount to the original price, so maybe that portion has a small profit (bought cheap enough). Net effect, LP might roughly break even or have a small loss, instead of a huge loss if it had kept quoting high. The LP's strategy thus must be responsive to new information to minimize adverse selection impact.

In conclusion of this section, the formalism indicates that *inventory management and risk control are integral to the quoting strategy.* We have set up the necessary components: a pricing policy that depends on current beliefs and inventory, and a risk limit on inventory. Next, we discuss further the hedging possibilities and how an "in-house arbitrage model" could complement the strategy.

5 Hedging and Secondary Strategies

As mentioned, the LP's primary hedging tool is its own pricing/inventory control. However, to maximize profit and minimize risk, the LP should consider any external opportunities to hedge or arbitrage. We formalize a few ideas:

5.1 Hedging via Related Markets

Let $\mathbf{\Phi}$ denote a vector of market risk factors that influence HLP's value (e.g., BTC/ETH spot, volatility indices, aggregate open interest or funding rates). Assume the LP can trade a subset of these (BTC/ETH futures, options, etc.). Model the terminal value as $P_T \approx g(\mathbf{\Phi}T) + \varepsilon_T$, where ε_T captures idiosyncratic PnL not explained by market moves. For small changes, linearize as $dP \approx \beta^\top d\mathbf{\Phi}$ with $\beta = \nabla\mathbf{\Phi} g(\mathbf{\Phi})$. A delta hedge chooses a position H delivering payoff $-\beta^\top \mathbf{\Delta\Phi}$ to offset factor exposure. For example, if the HLP is effectively short X notional of BTC (so when BTC rises the vault loses X), the LP goes long X notional of BTC futures to neutralize that component. Residual risk arises from ε_T (e.g., trader-specific outcomes, model error, or non-linear/operational events such as oracle issues) that are not spanned by the traded factors.

We incorporate hedging by adding terms in the LP's strategy: at each time step, the LP chooses a hedge position $h(t)$ in available instruments to minimize variance. The effect is to reduce σ in our earlier equations (thus allowing a tighter Q maybe). A full formal treatment would set up a joint process for P_T and hedge asset prices and solve for optimal hedge ratio (like in mean-variance optimization). In this paper, we limit ourselves to stating that if such hedges are available, the LP should utilize them to the extent they are effective and cost-efficient. Any hedge cost (e.g., bid-ask spreads or funding for futures) should be factored into the PNL. If hedging is expensive, the LP might hedge only partially or during high-risk periods.

5.2 Running a Secondary Arbitrage Model In-House

The user's query suggests openness to running an arbitrage strategy internally if it helps. One possible arbitrage: if the LP notices mispricing between the instant liquidity price and some other market. For instance, suppose another platform or OTC market is willing to buy HLP tokens at a higher price than our Q. The LP could buy from users at Q and immediately flip to the other buyer for a profit. Currently, it's stated no other venue offers a short HLP option (likely meaning no active secondary market for HLP), but perhaps an arbitrage could be constructed via primary vault deposit. For example, if the LP's quoted discount is too high, an opportunistic arbitrageur might deposit fresh USDC into HLP, wait 4 days, then redeem at full value, effectively capturing the difference. This arbitrage is slow (it takes 4 days), but it sets a theoretical cap on how large the discount can be: if LP offers, say, a 5% discount for instant liquidity, a well-capitalized arbitrageur could take the other side by buying those HLP tokens (or equivalently depositing into HLP) to gain 5% in 4 days, which is an enormous annualized return. If such opportunities exist, competition or the LP itself could exploit them. The LP is effectively doing that by running this service, but if LP misprices, others could step in.

So, to formalize: the LP must ensure the pricing is market-efficient in the sense that it doesn't leave obvious arbitrage. The LP should monitor the relationship: Instant Price Q vs. primary market redemption value vs. any other secondary price. If Q is too low, either no one will sell (they'd just wait 4 days or others would buy and wait), or someone might buy at Q and redeem in 4 days themselves (if allowed to transfer HLP token ownership). If Q is too high, the LP loses money to informed traders or essentially gives free lunch to sellers. So Q tends toward an equilibrium reflecting fair value minus cost of liquidity and risk.

An in-house secondary arbitrage could also mean: use any slight mispricing in underlying markets to enhance PNL. For example, if a user redeems and the LP now holds an HLP token, the LP could simultaneously try to replicate a short position in the HLP by shorting a basket of underlying assets to be delta-neutral. If done perfectly, then regardless of P_T, the LP's combined position (HLP token long + underlying shorts) yields a fixed payoff (the difference ideally equal to fees earned). This is essentially hedging as described, but if there's any extra edge (like funding rate differences, etc.), the LP might earn more.

Another strategy: if the LP has superior predictive models for HLP's performance (say from on-chain data or trader behavior analytics), it might time its quoting aggressively when it expects the vault to gain value (thus willing to buy more, since likely profit) and quote more conservatively when expecting trouble. This informational edge is part of maximizing PNL, though not exactly arbitrage, more like informed market making.

To stay in formal terms: we assume the LP continuously optimizes not just the quote but also any auxiliary strategies that can enhance profit. Let Θ represent such strategies (hedges, arbitrage trades). The LP's total PNL can be viewed as:

$$\mathrm{PNL}_{\mathrm{total}} = \mathrm{PNL}_{\mathrm{quotes}}(Q) + \mathrm{PNL}_{\mathrm{hedges}}(H) + \mathrm{PNL}_{\mathrm{arbitrage}}(\Theta). \tag{33}$$

The LP chooses Q, H, Θ to maximize $\mathbb{E}[\mathrm{PNL}_{\mathrm{total}}]$ subject to risk constraints. $\mathrm{PNL}_{\mathrm{arbitrage}}$ might be zero or positive if opportunities exist. $\mathrm{PNL}_{\mathrm{hedges}}$ might be negative expectation if hedges cost money (like option premiums), but they reduce variance.

We likely set Θ aside as optional; if a clear arbitrage existed, others would likely take it too, making it

fleeting. The LP as insider to HLP flows might identify patterns though e.g., if a lot of people redeeming might mean something, maybe short the HLP's underlying assets as mentioned (which is hedging). So Θ and H overlap.

5.3 Conclusion of Hedging Section

The formal insight is that *absent a direct short HLP market, the LP's optimal strategy is largely self-contained: manage quotes and inventory.* However, to the extent external hedges exist, incorporate them. The LP effectively must act as a risk manager for the vault's liquidity: quoting in a way that any informed flow is absorbed at a price that protects the LP (like an insurance premium). If additional strategies "make sense to maximize PNL," as the user suggests, the LP should deploy them (the model allows adding any strategy with positive expected return that doesn't violate risk limits essentially the LP can invest excess capital elsewhere as long as it doesn't interfere with covering redemptions).

6 Implementation Considerations and Transaction Cost Analysis

Implementing this model in practice requires careful consideration of operational details and performance measurement. In this section, we outline how the LP would use the model and monitor its success, including conducting Transaction Cost Analysis (TCA) and other analyses to refine the strategy.

6.1 Smart Contract and Infrastructure

The LP's quoting and trading likely involves a smart contract or off-chain service integrated with the HLP vault. When a user requests instant redemption, the system should:

1. Query the LP's current quote Q (which may be computed in real-time by an off-chain algorithm referencing on-chain data, since on-chain might be too slow for dynamic pricing).
2. Execute the trade: user transfers HLP token to LP's address, LP transfers payment (in stablecoin, etc.) to the user at price $Q \times$ quantity. This can be routed via an MPC wallet as mentioned (ensuring secure custody of funds).
3. The contract then initiates a withdrawal from the primary vault for that quantity of HLP tokens. Since primary redemption is $T + 4.5$, the contract will mark that withdrawal to complete at $t + 4.5$ days.
4. When settlement occurs, the contract receives the assets and returns them to the LP's liquidity pool, replenishing capital.

This pipeline must handle partial withdrawals, track the timing of each batch, and possibly manage multiple overlapping cycles. The Devine OS mentioned likely coordinates these flows.

6.2 Parameter Estimation

Key parameters like α (informed trader probability), distribution of P_T, volatility σ, etc., should be estimated from data:

- The LP should analyze historical vault performance and redemption patterns. For example, if historical data shows that 30% of volume spiked before large vault losses, one might estimate α or at least condition probability of informed trading.

- The volatility of vault value over 4.5 days can be estimated or modeled (perhaps via a Monte Carlo simulation of market moves and trader behaviors).
- If available, analyze past cases of immediate vs delayed redemption (maybe analogues in GMX's GLP markets or other funds with lockups) to gauge typical discounts.

6.3 Calibration

Initially, the LP might start with a conservative quote (e.g., a fixed 0.5% discount plus small time value). Over time, by observing results, the LP can calibrate:

- If too few users are using the service (maybe quote is too low), the LP might narrow the discount.
- If the LP observes that every time someone redeems, the vault value tends to drop by, say, 0.3% on average relative to expectation, then that informs $\delta_{\text{info}} \sim 0.3\%$.
- The LP can compute *realized PNL per trade*: For each redemption trade, when the actual payout arrives, calculate profit = (payout − price_paid)/price_paid. Over many trades, find the distribution of this profit.
- If the LP is pricing perfectly on average, the mean profit should be positive (to cover costs) but not too high (or else maybe leaving volume on table).
- If mean profit is negative, that's bad means adverse selection or underpricing risk.
- If a significant fraction of trades are losses, perhaps α was higher or δ_{info} underestimated.

This is where TCA comes in: Typically, TCA for market making looks at metrics like realized spread (profit after some time) and impact of trades. In our context, we can do a similar analysis:

Realized Spread: For each trade (instant redemption), define mid-price as perhaps the expected value (or maybe just P_0 if no better reference). The LP's discount is $P_0 - Q$. When the vault pays out, compare the LP's revenue vs what it would have been if paid P_0. Essentially, realized profit $= P_T - Q$. We want to track average $P_T - Q$ across trades.

A positive average means LP is profitable. But we also examine variance and downside cases. The LP should track worst-case outcomes (e.g., maximum $Q - P_T$ encountered, which is a loss scenario).

If certain patterns emerge (e.g., large trades tend to be informed and result in losses), the LP might introduce size-dependent pricing (bigger sells get worse price).

If frequent small trades rarely result in loss, perhaps LP can tighten spreads for small trades to get more volume.

6.4 Transaction Cost Analysis for Users vs LP

From the user perspective, the "transaction cost" of using the instant pool is the discount they pay. If our LP is too expensive (big discount), users might only come when absolutely necessary (or not at all if they can afford to wait). If too cheap, users with minor needs will gladly use it and LP may not be compensated enough for risk. So TCA can also consider the users: e.g., measure how much value users leave on table by selling early vs what they would have gotten waiting. Ideally, for liquidity provision to be efficient, that cost should reflect fair compensation for time and risk.

The LP can measure the *average discount given*: say average Q/P_0. If the LP notices that even uninformed users are paying a high cost (like consistently, Q is 0.98 of P_0 and P_T ended up 1.0, so users lost

2%), maybe competition will eventually come or users will complain. If the LP is monopolist maybe that's fine for profit, but if it's too high, fewer trades happen. So there is a trade-off.

To optimize, the LP could simulate different pricing policies and their expected volume × margin to maximize PNL. This is like finding the optimal point on a curve of volume vs margin (akin to any monopoly pricing problem). In practice, some experimentation or iterative adjustment might be needed.

6.5 Technological and Market Risks

The model assumes all goes smoothly, but implementation must consider:

- **Settlement risk:** The 4.5-day delay has risk that something fails (smart contract bug, the vault freezes withdrawals, etc.). The LP should be aware of those. Perhaps factor a tiny probability of non-payment (if any) into pricing (a credit risk spread).
- **Regulatory/Platform risk:** If the platform is at risk (like Hyperliquid's centralization issues, regulatory crackdowns, etc.), that could drastically affect P_T or even ability to redeem. The LP in extreme cases might pause operations if such meta-risk is perceived (because if the whole platform fails, the LP could lose 100% of inventory).
- **Competition:** If others start offering similar services, pricing might become more competitive, pushing Q closer to fair value. The LP's model can adapt by including an outside option: if competitor offers discount d_{comp}, LP must offer slightly better or equal effective discount to attract trades (unless LP has other advantages). So Q could also depend on market competition, which we assume minimal here.

6.6 Monitoring Adverse Selection

The LP should continuously monitor order flow for signs of adverse selection:

- If a cluster of redemptions occur rapidly, that likely indicates an informed event. The LP's algorithm should flag this and possibly automatically widen quotes or halt until more info is known.
- The LP could integrate alerts from the vault's health: e.g., if a huge trade happens on the exchange that could hurt HLP (like the $4M loss event where a whale opened a huge position), the LP system should anticipate that HLP might drop and adjust Q proactively.
- Machine learning models could be trained on on-chain data to predict vault drawdowns; the LP can feed those into the pricing (this goes beyond our formal model but is a practical augmentation).

6.7 Empirical Backtesting

Before deployment, one would ideally backtest the strategy. Since exact data may not exist for HLP historically (if new), one can use analogies:

- Simulate a time series of HLP NAV over 5-day periods, with random jumps representing trader wins/losses, and random arrival of informed vs uninformed sellers.
- Apply the strategy (pricing and inventory rules) to the simulation and measure resulting PNL distribution.
- Adjust parameters to meet the 1% drawdown target (ensuring in, say, 99% of simulations LP loss $< 1\%$).
- Ensure profitability: The simulation should show positive average PNL.

Backtesting and TCA results can then feed back to refine the model's parameters (α, δ_{info}, k for inventory, etc.).

6.8 Transparency vs Gaming

One consideration is whether the LP reveals its pricing formula or keeps it black-box. Too much transparency might allow informed traders to game it (though essentially they just trade when beneficial anyway). But maybe the LP should commit to some algorithm to ensure users trust the fairness. Since this is an academic paper style discussion, we won't delve into strategic gaming by traders beyond adverse selection already considered.

Summary. In summary, implementing the model involves a robust system that dynamically updates quotes, executes trades atomically with transfers, and manages the queue of pending redemptions. TCA and ongoing analysis are crucial to ensure the strategy works as intended: the LP should consistently earn a small positive spread on uninformed order flow to compensate for the occasional losses to informed order flow and for the capital cost, which is the classic outcome for a market maker under asymmetric information. If the LP finds its PNL not matching theory, it must recalibrate (e.g., increase δ_{info} if losing to informed traders, or decrease it if too few trades).

7 Conclusion

We have presented a comprehensive formal model for operating as a market maker providing instant liquidity for HLP vault redemptions. By framing the problem in terms of optimal pricing under uncertainty and adverse selection, we derived how the LP should set a discounted redemption price that balances profit and risk. The key insights are:

1. **Optimal pricing under information asymmetry:** The LP's optimal quote is essentially the expected future value of the HLP token conditional on a sell order, minus a margin for profit. This creates a rational discount compared to waiting 4.5 days, compensating the LP for time value, risk, and information asymmetry. In equilibrium, this discount (or bid-ask spread analog) arises from a combination of adverse selection risk and monopolistic pricing of liquidity. If information asymmetry is high, the discount (spread) must be larger to avoid losses. If asymmetry is low, the LP can quote closer to fair value. This ensures incentive compatibility: informed traders trade only when it's optimal (which the LP prices in), and uninformed traders get liquidity at a fair cost for immediate access.

2. **Dynamic inventory and belief management:** The LP must actively manage inventory and update beliefs. After each trade, especially if sudden or large, the LP revises its outlook with the trade signal. Inventory accumulated is essentially a bet on the HLP's future value; to keep risk bounded, the LP may reduce subsequent quotes or pause to keep inventory within limits. We formalized how an inventory penalty can be added to the pricing rule to achieve near-market-neutrality aside from the necessary long HLP position the LP temporarily holds. This helps minimize the variance of outcomes and avoid large drawdowns.

3. **Stringent risk control:** Risk control is paramount: by imposing a drawdown limit (like 1% of capital), the LP ensures survival and consistency. We showed how this can translate into position limits and conservative pricing. In extreme scenarios (e.g., suspected large vault loss events), the model would dictate widening the spread or even refusing trades a phenomenon akin to a "no-trade region" when adverse selection becomes too severe. This protects the LP from being picked off by informed traders in those moments.

4. **Hedging and secondary strategies:** Although direct hedging of HLP is not possible, the LP can hedge indirectly by trading underlying assets or related derivatives to the extent feasible. While we did not provide a full hedging solution, we outlined approaches to offset market risk (like shorting/longing correlated assets) and noted that any such hedges should be utilized if available. These can reduce the effective risk of holding HLP inventory, allowing the LP to quote more competitively. Running secondary arbitrage or predictive strategies in-house can further boost profits essentially, the LP should use all information at its disposal (such as detecting when the vault is under/overvalued relative to market) to position itself advantageously.

5. **Continuous refinement through TCA:** We detailed how one would implement this strategy and continuously refine it. The LP should engage in Transaction Cost Analysis, monitoring metrics like realized profit per trade and adverse selection loss, to ensure the model's assumptions hold. If the LP consistently loses on certain trades (indicating informed trading), it must adjust the pricing formula (increase the discount) until those losses are covered by gains elsewhere. The LP also keeps an eye on user experience: providing just enough liquidity at a tolerable cost to users, without giving up excessive profit. Over time, the parameters (α, risk premium, etc.) can be tuned so that the LP's service runs optimally providing a useful function (immediate liquidity) to the market while operating at essentially zero or positive expected loss to informed traders (meaning the informed are not "free-riding" beyond what the LP has priced in).

6. **Connection to theoretical foundations:** Our formal model connects to known theoretical results in market making and mechanism design. By treating the LP as a profit-maximizing monopolist liquidity provider, we leveraged the idea of setting prices based on virtual values and ensuring incentive-compatibility. The resulting strategy mirrors an optimal auction for liquidity, where the LP "auctions" immediate liquidity to those who value it most (those willing to accept a discount) and protects itself via the pricing rule. This is a novel application of such theory to a DeFi context of vault redemptions, and it opens up further research directions: for instance, one could extend the model to multiple competing liquidity providers, or to a continuous double-auction for HLP tokens. Additionally, integrating this model on-chain (perhaps via an automated market maker smart contract that adjusts dynamically using these principles) could be an interesting future development.

In essence, the LP should behave like an optimal market maker/AMM for this delayed redemption asset always updating its beliefs with trade information, always balancing supply and demand for immediacy, and charging just enough to cover risks. By formalizing the logic in this paper, we have provided a blueprint that can be implemented and tested in a real environment. If followed, the LP can provide an efficient quoting service for instant redemptions, contributing to market efficiency (by pricing the time-value of liquidity) and maximizing its own profit with minimal risk.

References

[1] Hyperliquid platform updates confirming the need for this liquidity solution and justifying the $T + 4.5$ settlement assumption.
`https://www.sec.gov/Archives/edgar/data/1683471/000089418925005036/a21shares485a.htm`

[2] What Is Hyperliquid? A Beginner's Guide | Hyperliquid Overview | blocmates.
`https://www.blocmates.com/articles/a-complete-guide-to-hyperliquid`

[3] Milionis et al., "A Myersonian Framework for Optimal Liquidity Provision in AMMs". https://arxiv.org/pdf/2303.00208

[4] Avellaneda & Stoikov, "High-frequency trading in a limit order book" (2008) for inventory control ideas. https://arxiv.org/pdf/1210.4000

[5] Glosten & Milgrom, "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders" (1985) for market maker pricing under asymmetric information. https://www.sciencedirect.com/science/article/abs/pii/0304405X85900443

[6] Hyperliquid Loses $4M After Whale's Over $200M Ether Trade. https://finance.yahoo.com/news/hyperliquid-loses-4m-whales-over-124005325.html

[7] Kyle, A. S., "Continuous auctions and insider trading" (1985). Econometrica: Journal of the Econometric Society, 1315-1335.

[8] Back, K., & Baruch, S., "Information in securities markets: Kyle meets Glosten and Milgrom" (2004). Econometrica, 72(2), 433-465.

[9] Foucault, T., Kadan, O., & Kandel, E., "Liquidity cycles and make/take fees in electronic markets" (2013). The Journal of Finance, 68(1), 299-341.

[10] Ho, T., & Stoll, H. R., "Optimal dealer pricing under transactions and return uncertainty" (1981). Journal of Financial Economics, 9(1), 47-73.

[11] Glosten, L. R., "Insider trading, liquidity, and the role of the monopolist specialist" (1989). Journal of Business, 211-235.

[12] Myerson, R. B., "Optimal auction design" (1981). Mathematics of Operations Research, 6(1), 58-73.

[13] Maskin, E., & Riley, J., "Optimal auctions with risk averse buyers" (1984). Econometrica: Journal of the Econometric Society, 1473-1518.

[14] Milionis, J., Moallemi, C. C., Roughgarden, T., & Zhang, A. L., "Automated market making and loss-versus-rebalancing" (2022). arXiv preprint arXiv:2208.06046.

[15] Cartea, Á., Jaimungal, S., & Penalva, J., "Algorithmic and high-frequency trading" (2015). Cambridge University Press.

[16] Easley, D., & O'Hara, M., "Price, trade size, and information in securities markets" (1987). Journal of Financial Economics, 19(1), 69-90.

Addendum (Appended Content Only; Original Text Above Unchanged)

Security & Production Invariants (Non-Exhaustive)

1. **CEI discipline:** Checks-Effects-Interactions; reentrancy guards on all external entry points.
2. **NAV freshness:** Quotes and confirmations require a snapshot timestamp within $t_{\max}$; otherwise revert or auto-halt.
3. **Replay protection:** Per-batch key single-use; global tx-hash uniqueness; per-user one-time batch-key mapping.
4. **Drift bounds:** Max deviation between quote-time inputs and confirm-time state (NAV, supply, assets); exceed $\Rightarrow$ revert.
5. **Settlement lock:** Enforce batch expiry and minimum inter-batch spacing; stale batches cannot confirm.
6. **Liquidity checks:** Verify vault/exchange balances before per-user confirms; batch USDC received $\geq$ sum of obligations.
7. **Idempotency:** Confirm endpoints idempotent across retries; Merkle claims one-time per leaf.
8. **Access control:** Role-gated admin/oracle functions; emergency stop with timelocked resume.
9. **Numeric safety:** Fixed-point bounds; overflow/underflow checks; rounding floors bias to solvency.
10. **Event auditability:** Emit events on quote accept, batch confirm, per-user confirm, and parameter changes.

Stress/Halt Policy

- **Oracle staleness or venue incident:** Disable quotes; advance ratio $a \to 0$; settlement/claims remain enabled.
- **Flow spike/clustering:** Auto-increase δ_{info}, slope k, and temporarily reduce $q_{\max}$.
- **Backstop coverage:** Non-recourse capacity scales to insured backstop; auto-throttle to maintain coverage.

Parameter Defaults (Illustrative)

Parameter	Symbol	Guideline
Horizon fraction	ΔT	4.5/365 years
Risk penalty	γ	Tune to hit τ=1% drawdown at tail level ε=1%
Inventory slope	k	Choose so $Q(q_{\max})$ widens 50–100 bps vs base
Info discount	δ_{info}	Empirical conditional shortfall after sells (size/cluster aware)
Advance ratio cap	a	$1 - (f + c_f + c_o + \text{VaR}_\alpha)$
Batch expiry	–	e.g., 24h; re-quote after expiry

APY Uplift Calibration (Informal)

Let $C = 365/4.5$ cycles/year, utilization $U \in [0,1]$, deployment ratio to vault L, and per-cycle net margin (swept) g. Then

$$R_{\text{vault}} \approx L\Big[(1+g)^{UC} - 1\Big], \qquad g^\star \approx \left(1 + \frac{R^\star_{\text{vault}}}{L}\right)^{1/(UC)} - 1.$$

Discount wedge $D \approx f\Delta + g^\star + \text{risk buffer}$; raise D by Δf if funding f rises to keep g constant.

Data, Backtesting, and Validation

DGP. Simulate (P_t) with jump-diffusion (incident risk); simulate sell-arrival Cox process with intensity increasing under adverse regimes; label a fraction α as informed.
Calibration. Hit target drawdown τ and profitability via tuning $k, \delta_{\text{info}}, q_{\max}$; widen on clusters; decay with half-life after normalization.
Monitoring. Track realized $P_T - Q$, loss tails, hit-ratio by size/time, and halt efficacy.

A Binary Model Details (Supplement)

Let $V \in \{V_H, V_L\}$, prior $p = \Pr(V_H)$. With informed probability α and uninformed sell-probability η,

$$\Pr(\text{sell}) = \alpha(1-p) + (1-\alpha)\eta, \quad \mathbb{E}[V \mid \text{sell}] = \frac{\alpha(1-p)V_L + (1-\alpha)\eta(pV_H + (1-p)V_L)}{\alpha(1-p) + (1-\alpha)\eta}.$$

Thus $\delta_{\text{info}} = (p - \pi)D$ with $\pi = \Pr(V_H \mid \text{sell})$ and $D = V_H - V_L$.

B Risk Buffer from Chance Constraint (Supplement)

If PNL across n fills has mean m and variance s^2, enforce tail control via

$$m - z_\varepsilon\, s \geq 0,$$

adding a per-fill discount $\approx z_\varepsilon \sigma/\sqrt{n}$ (normal proxy) or using CVaR sizing against incident tails.

C Example Policy Pseudocode (Informal)

```
Inputs: P0, mu, sigma, alpha, q, q_max, flags, rf, gamma, delta_info
BaseQuote = mu - rf*mu*DeltaT - 0.5*gamma*sigma^2*DeltaT - delta_info - markup
Q = clamp(BaseQuote - k*q, Q_min, Q_max)
if flags.oracle_stale or flags.venue_incident or q >= q_max:
    Q = DISABLED
return Q
```